Complex and Chaotic Nonlinear Dynamics

Thierry Vialar

Complex and Chaotic Nonlinear Dynamics

Advances in Economics and Finance,
Mathematics and Statistics

🕮 Springer

Dr. Thierry Vialar
University of Paris X
Dept. of Economics
200, ave de la République
92001 Nanterre Cedex
France
www.isnld.com
t.vialar@isnld.com

Pr. Alain Goergen
École normale supérieure de Cachan
Dept. of Economics
61, ave du Président Wilson
94235 Cachan Cedex
France

Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper

springer.com

To
Mr. Professor Alain GOERGEN
with my deep gratitude
and admiration.

Preface

Since the 1970s, Complex and chaotic nonlinear dynamics (in short, Complex Dynamics) constitute a growing and increasingly important area that comprises advanced research activities and strongly interdisciplinary approaches. This area is of a fundamental interest in many sciences, including Economics.

Let us start with a comment about the interest of Complex Dynamics in Economics and in so doing the necessity of such a book and its interdisciplinarity: Mathematics in Economics have a very strong didactic role. Mathematics state theoretical models and paradigms that must conform to measurements. However, in Economics the measurements are rare, more often of a small number of points and of a very low density, except for stock markets. This chasm that separates Economics from the "dense" measurable reality has maintained economists in a kind of "isolation" (compared with other sciences): (1) that of the qualitative approach that can be conducted, for instance, in a literary way, very often of great relevance but which is not quantitative, (2) and that of the construction of models ("mathematical idealities") with a strong didactic or mechanistic vocation often far from the richness of "the living". This is above all a problem of *measurement*. Beyond the epistemological revolution of nonlinear theory, today there is that of the information systems and networks, which will provide the exceptional opportunity to capture dense measurements in numerous fields of Economics (e.g. from consumer behaviors up to national accounts). Thus, the economists will be in an opposite situation than before. The measure flows for the economists will have densities increasingly similar to those of other sciences whose measures come from "the living" for example. Economics will have to treat these measure flows with relevant tools, which are necessary to master. This is a turning point for Economics. Thus, Mathematics and its analytic tools are more relevant than ever for economists, in particular to study Complex Dynamics and to bring closer theoretical models and information coming from signal or time series measurements. Calculation capabilities, networks, measurements and information treatment also make the existence of such a book legitimate and necessary.

In the same vein, Economic Policy needs tools going beyond simple observation (shifted in time) offered by statistical series in order to recognize the exact and not

only the apparent state of conjuncture. This book describes these tools, with a wealth of details and precisions, and not only the tools but also many concrete applications to economic series in general. When Clive W.J. Granger published his work about the time series analysis 40 years ago, he exposed the means available at that time, of which Fourier series decomposition. These means had been refined and improved with the aim of applying them, for example, to telecommunications. The series on which telecommunications analysts work contain a great number of points, the edge effects are very often negligible and, especially, the series are almost always stationary. The use of more elaborated tools than the Fourier series decomposition is a necessity. In Economics, this had been different for a long time; everything created a problem in time series analysis, a weak extent, edge effects, a non-stationariness which cannot be reduced to the existence of a tendency, as the basic handbooks could make it believe, but is expressed by a high volatility, relative to an average value which does not have great sense, and variable according to the selected sample.

In such a context, the content of this book shows how much the recent contributions of signal theory in relation with nonlinear dynamics are powerful means of analysis and have a so important potential. Information systems and networks will contribute to this goal. In this regard, let us point out that the third part of the book is a quite essential contribution. It covers signal theory, not only in a didactic way (Fourier, Wiener, Gabor, etc.) but also by presenting highly advanced contents (polyspectra already used by economists, best basis, multi-resolution analysis, hybrid waveform dictionaries, matching pursuit algorithms with time-frequency atoms, etc.). The applications are numerous and demonstrative: stock market indexes, standard signals, signals of coupled oscillators or turbulent phenomena highlighting coherent structures. Signal theory certainly has to be promoted in Economics; this book contributes to this aim.

More than in the past, Economics calls for nonlinear formalizations which provide complex formal solutions. The increasingly frequent necessity to carry out digital simulations after still largely heuristic "calibrations" leads to thorough analyses of simulated series and reference-series (often reconstructed) for which this book offers particularly adapted tools.

What appears most clearly is the innovation and originality of many parts of this book, the diversity of the applications and the richness of the theoretical exploration possibilities. This is what makes this book a document from now on impossible to disregard for economists as for econometricians, and potentially for practitioners of other disciplines.

To end this preface, may I wish that the readers have as much pleasure as I to peruse this work that numerous illustrations make less austere without ceasing to be rigorous, and then, convinced by the diversity of the applications, that the readers implement themselves the tools.

Ecole Normale Supérieure of Cachan,   *Professor Alain Goergen*
France

Summary

Contents

Part II Statistics of Complex and Chaotic Nonlinear Dynamics: Invariants and Rare Events

Introduction

The aim of this work is to try to offer a stimulating environment for the study of complex or chaotic nonlinear Dynamics. The topicality of this type of dynamics results from widely different scientific disciplines. And although keeping an economic or financial prevalence, the assigned objective can only be approached by an opening to the other disciplines related to the subject.

Economic models have long been elaborated from constructions whose algebraic nature was of a linear order. This factor, coupled with the fact that still a few decades ago, constraints linked with calculation possibilities were strong, weighed heavily on the way of apprehending and understanding economic and scientific phenomena in general. The discovery or rediscovery, more than 30 years ago, of the different types of behavior that a simple equation of a nonlinear nature can offer, opened considerable possibilities in the formalization of economic and financial phenomena. "The great discovery of the nineteenth century was that the equations of Nature are linear, and the great discovery of the twentieth century is that they are not" (Körner 1988). This assertion which consists in saying that *the world is nonlinear* penetrates economic realities which do not escape from this observation. Even if the writing of nonlinear models precedes this rediscovery, the possibilities of simulation and experimentation are immense today. It is around this concept of *nonlinearity*, adapted to the formalization of natural phenomena and around *chaotic dynamics* that this whole book is organized. They constitute the vital leads of the four Parts of this book:

- *Part I.* The first part presents investigation methods of complex and chaotic nonlinear dynamics, among which the concepts of *nonlinear theory* (often called chaos theory) and also *nonlinear signal processing*.
- *Part II.* The second part reviews the evolution of *statistical analysis* towards nonlinear and chaotic dynamics.
- *Part III.* The third part, dedicated to *spectral and time-frequency analyses*, underlines the contributions of waveforms and atomic decompositions to the study of nonlinear phenomena.
- *Part IV.* The last part aims to depict the evolution of linear *economic growth models* towards nonlinear models, and the growing importance of nonlinearities in the construction of models.

Part I

Since the 1970s the irruption of the "nonlinear" led to a profound transformation of numerous scientific and technical fields. Economics does not escape this revolution. The taking into account of *nonlinearities is an infinite source of behavior diversity*, that makes it possible to better understand *natural phenomena* and phenomena considered as complex which were adverse to any modeling before. It is indeed a true *epistemological rupture*, that occurred in 1971,[1] with the introduction of the concepts of "deterministic chaos", *sensitive dependence on initial conditions* and dissipative systems. Henceforth, we understand how *an apparent disorder can dissimulate a subjacent order*. Thus, it appears important to assemble the *theory core* and the *tools* to *investigate* these *complex* and *nonlinear dynamics*, which escaped any analytic effort before. These investigation methods are not necessarily recent. Indeed, they can have distant origins in time, but today it is possible to speak of a "modern conceptual unification" through the notions of sensitive dependence on initial conditions, bifurcations,[2] subharmonic cascades, attractors, Lyapunov exponent, saddle-connection, transitions to chaos, dissipative systems, conservative systems, hyperbolicity, hyperbolic systems, etc. Although we are faced with a coherent set, this set is only at an early stage.

Inside this set appear techniques developed within the last 20 years, gathered together under the name "nonlinear signal processing" based mainly on the Takens theorem. It is a theorem of time series reconstruction, based on the concept of *topological equivalence*, which enables to identify the nonlinear nature of an original time series, for example, periodic, quasiperiodic, aperiodic or chaotic, while saving a huge amount of calculation. The stake here is extremely high. Indeed, without necessarily knowing the equations of the dynamical system which generated a series and by working in a *reconstructed phase space* of a very low dimension, it is possible to reproduce the essential features of a system or an original trajectory. The study of the *geometrical objects of low dimension can provide all the information* which we need. In Economics very often we have to face this type of problem where we do not necessarily know the number of variables involved in dynamics nor the dimension of the system which can be infinitely large. However, it is known that large attractors or infinite systems can have low dimensions. Thus, with largely reduced series, the *study of low-dimension objects can reveal the information* that we need to identify dynamics. It is possible to consider this as a diagnosis method, like the Poincaré map. This approach leads directly to a major concept which is that of the capacity dimension,[3] which is a non-integer dimension. This concept is an instrument quite as important as the Lyapunov exponent. On one hand, it makes it possible

---

[1] But the origin dates back Lorenz in the 1960s, where new mathematics were born.

[2] In connection with these tools, it seemed important to highlight the symptomatic phenomenon which links the *speed* of transition between two states of a dynamical system with the characteristics of the periodic or chaotic regime of the final state of the system. The section is entitled the *bifurcation paradox* exhibiting the title of the authors of the study.

[3] Also called Kolmogorov dimension, or "box counting", also named Hausdorff dimension; the whole being regrouped under the name of fractal dimension.

to characterize the attractor that we have to face and, on the other hand, to make *the difference between deterministic chaos* and *random walk*. Generally, it is said for example that a "Brownian motion" has a capacity dimension equal to two, which is not necessarily the case of an apparent deterministic chaos.[4] It is the concept of stability of a dynamics, approached in particular through the *Floquet theory*, which leads us to define the notion of *topological invariant set* which then leads to define the *attractor* concept.[5] An attractor can be known as a fixed-point, limit-cycle, toric or strange attractor. It is characterized by its capacity dimension, as well as the Euclidean dimension of the system in which the attractor appears, knowing that the latter can have only a non-integer dimension per definition. A system which does not have an attraction area in the phase space is known as *conservative*, in the reverse case it is known as *dissipative*. Thus, the existence of an attractor type in a dynamics characterizes *dissipative systems*. It will be noticed however that *chaotic or very complex behaviors can singularly exist in models considered conservative*. This is indeed the case in systems of the *Hamiltonian type*, as it is possible in Economics.[6]

Like the methods mentioned above, which are assembled under the name of "nonlinear signal processing", there is a major tool which will be also largely developed from another aspect in the part III, which is *spectral analysis* or power spectrum of a dynamics. Spectral analysis results from two basic concepts which are the Fourier transform and the autocorrelation function of a signal, which is an extraction of the Wiener–Khintchine theorem. *The temporal autocorrelation function* measures the resemblance or the similarity of values of a variable in time. In fact *this temporal correlation function corresponds to the Fourier transform of the power spectrum*. For chaotic regimes, for example, the similarity decreases with time. It is said that they are unpredictable due to the loss of internal similarity to their processes. When we have a time series, an essential task is to determine the type of dynamics which engendered it, be it a more or less complex oscillation but of a defined period, or a superposition of several different oscillations, or other types of dynamics. The periodicities are identified with the spectral analysis method, whether the subjacent model is known or not. This already evoked phenomenon, highlights the fact that certain dynamics result from a superposition of oscillations with different amplitudes and oscillations but also from harmonics of these oscillations. In the last case, the regime which is described as quasiperiodic has an associated attractor which is of a higher order than the limit-cycle, i.e. a torus for example. For chaotic but deterministic regimes, i.e. for dynamics represented by a limited number of nonlinear differential equations, the attractor in the phase space is a strange attractor. Thus, this method of spectral analysis makes it possible to identify the nature of a dynamics, periodical, aperiodic or quasiperiodic. Besides, we experiment this technique on various types of signals resulting from the

---

[4] See section about "(non-fractional) Brownian motions".

[5] We sometimes find an unsuitable terminology to characterize an attractor and in particular the term of "chaotic attractor", which in fact is sometimes used to indicate an attractor which exists in a chaotic regime.

[6] Refer for cyclic growth models in Economics to Goodwin (1967).

logistic equation, from ARMA processes, from stock market courses as the French stock-market index (Cac40), or from the Van der Pol oscillator.

The logistic model symbolizes the paradigm of a nonlinear model. Thus, it is used transversely to highlight numerous notions quoted previously concerning nonlinearities. It is also used to experiment the delay-model applied to the logistic equation which introduces by convolution a discrete delay into the construction of an economic model. The lengths of the delays are distributed in a random way in the population. The delay is in fact modelled by means of a random variable which is characterized by its probability distribution. It will be noted that such a system topples more tardily in the chaos, i.e. there is a shift of the bifurcation points but also an unhooking in the trajectories.

The *singular spectrum analysis* (SSA) method is the last investigation method concerning the complex dynamics presented in this part. The method associates the Takens reconstruction technique and the technique called the *singular value decomposition* in matrix algebra.[7] If we simplify, the method consists in the projection of a time series on a basis of eigenvectors extracted from this same time-series. In fact, a *matrix trajectory* is projected on the space described by the eigenvectors of the time series covariance matrix. The eigenvalues obtained can be ordered and be filtered for the extraction of the *deterministic* part of the signal cleaned of its background noise.

Part II

This part aims to depict the evolution of statistical analysis towards nonlinear stochastic processes and chaotic processes. The purpose is to state the main developments concerning this subject: (a) End of the domination of the ARMA model, (b) Nonlinearity Tests (BDS), (c) Stakes of the non-parametric analysis, (d) A statistical theory of chaotic dynamics is to be built, (e) Long-memory process and self-similarity, (f) Construction of ARFIMA models, (g) FIGARCH models and volatility of variances, (h) Lo and MacKinlay tests about the rejection of the random-walk hypothesis for stock-exchange markets, (i) Estimations of the Hurst exponent (in particular by means of the Abry–Veitch wavelet method), (j) Estimators of density, (k) Invariant measurement of a dynamical system and (limit) invariant density,[8] (l) Ergodic theory, etc.

ARMA modeling is representative of linear modeling, but the linearity Hypothesis is unsuitable to represent real phenomena in many fields. The nonlinear economic models introduced in 1950 and 1951 by Hicks and Goodwin, because of the absence of statistical tools, did not have the deserved resonance that they should have had at that time. This explains the long domination of the ARMA models until the nonlinearity Hypothesis found its own consistency, in particular in statistics.

---

[7] Karhunen-Loève method.

[8] *Invariant density:* Invariant density is also called "natural" invariant or "natural" density.

Thus in 1976, the first linearity Test has been developed by Granger and Newbold. More recently, the non-parametric Test Statistic (BDS test) was taken as a nonlinearity test. It helps to provide information about nonlinearity, however it is not a measurement of the nature of the time series chaos, but in this part it is used as an introduction to non-parametric statistical analysis. Parametric analysis aims to rebuild a deterministic model subjacent to a time series. The rebuilding is done by means of a stochastic model with the implicit idea of the existence of a subjacent structure. The stochastic parametric modeling of nonlinear processes does unfortunately not make it possible for a great number of processes to produce robust estimates. In front of this lack of specification about the complex or chaotic processes, our attention is drawn to *non-parametric analysis*, which does not seek to specify models, but generally aims to rebuild trajectories without a deterministic or stochastic model. The method by means of the extraction of the time-series properties and its estimators allows the reconstruction of the dynamic relation which links the time-series terms. It is said that the dynamic relation is estimated in a non-parametric way.

Faced with these complex or chaotic dynamics, Chaos theory provides chaos detection tests, as the Lyapunov test and the correlation dimension, but they are not statistical tests in a strict sense. Thus, the statisticians have been led to elaborate statistical validation tests of these detection tests; in particular the *random mixture* test. The method is surprising and the idea is that the mixture destroys the possible deterministic structure of a time series. If the mixed series loses its structure, the correlation dimension and the Lyapunov exponent of the mixture must make it possible to distinguish it. In spite of encouraging results, contradictions between the results of Lyapunov and the correlation dimension prevent convincing conclusions about the deterministic or non-deterministic nature of the observed chaotic dynamics. In the light of this type of example, it seems fundamental to build a statistical theory core of chaotic dynamics, which still remains to be worked out in spite of numerous current contributions.

A way of characterizing the nature of chaotic dynamics is to study the structure of long-memory processes. These processes are observed in numerous fields, for example in the telecommunications sector in connection with the information flow on the Internet, but also in connection with financial markets. They are detected by the observation of their autocorrelation functions, which decrease hyperbolically towards zero, whereas they decrease exponentially for short-memory processes. (The hyperbolic decrease can also express a nonstationariness.) The long memory is also detected by spectral concentrations which increase when we approach the central frequency centered at zero, or by a persistent or anti-persistent behavior. It is commonly said that the more a process is persisting, the more the convergence is slow and the more the sum of the autocorrelations is high. For a process with short memory, the sum of the autocorrelations is weak. In short, we are interested in the speed of the hyperbolic and geometrical convergence towards zero. Weak lags and strong correlations rather characterize models with short memory of the ARMA type. The long memory processes lead us to outline the ARFIMA processes which integrate the long memory phenomena (Long Range Dependence: LRD).

The Hurst exponent allows to introduce long memory into an artificially generated process. It is the purpose of the numerical generators of fractional Brownian motions. The parameter of fractional integration, i.e. unit root, is used in ARIMA processes to test stationarity. Moreover, a functional relation was highlighted between this parameter and the Hurst exponent, consequently assigning a new role to it. This role is to introduce long-term dependence (or long memory) into new models, i.e. the ARFIMA processes.[9] Recently, the estimate methods of the Hurst exponent have been developed in empirical series.[10] And the effectiveness of these methods can be tested with a good safeness, since it is possible to estimate a parameter that we have fixed before (a priori) to construct an experimental series. We can proceed in a similar way with an ARFIMA process. Among the estimation methods of the Hurst exponent, the technique developed in 1998 by Abry and Veitch using the wavelets properties appears to be important, to which we will give a particular place.

The ARCH processes supplanted the ARMA processes, unsuited to financial series which have asymmetrical structures and strong volatility of variance. ARCH processes integrate the parameters of the conditional variance in an endogenous way and have often been used in the optimization of (financial) portfolio choices. The study of the conditional variance makes it possible to highlight the persistence of shocks, by using an extension of IGARCH processes (integrated GARCH), i.e. FIGARCH processes (Fractionally Integrated GARCH).

Closely related to the birth of probability theory, the random walk hypothesis has a famous history, whose actors are Bachelier, Lévy, Kolmogorov and Wiener. More recently, one of the first applications of the random-walk hypothesis to the financial markets dates back to Paul Samuelson in 1965, whose contribution has been developed in an article entitled "Proof that Properly Anticipated Prices Fluctuate Randomly". He explains why in an efficient market, concerning information, the price changes are unpredictable if they are properly anticipated, i.e. if they fully incorporate the expectations, information and forecasts of all the market participants. In 1970, Fama summarizes what precedes in a rather explicit formula: "the prices fully reflect all available information". Contrary to numerous applications of the random walk hypothesis in natural phenomena, for which the randomness is supposed almost by default due to the absence of any natural alternative, Samuelson argues that the randomness is achieved through the active participation of many investors who seek increase of their wealth. They attempt to take advantage of the smallest information at their disposal. And while doing so, they incorporate their information into the market prices and quickly eliminate the capital-gain and profit opportunities. If we imagine an "ideal" market without friction and trading-cost, then the prices must always reflect all information available and no profits can be garnered from the trading based on information *because "such profits have already been captured"*. Thus in a contradictory way, the more the market is efficient, the more the price time-series generated by such a market is random, and "the most

---

[9] Hosking in 1981 and Granger, Joyeux in 1980.

[10] Some of them use the concept of Embedding space resulting from the Takens theorem.

efficient of markets is one in which the price changes are completely random and unpredictable". Thus the random walk hypothesis and the efficient markets hypothesis became emblematic in Economics and Finance, although more recently, in 1980, Grossman and Stiglitz considered that the efficient market assumption is an economically unrealizable idealization. Moreover, some recent works done during the last 15 years initiated an approach aiming to reject the random walk hypothesis. Econometric studies conducted by Lo and MacKinlay (since 1988) relating to the US stock-exchange market rejected the random walk hypothesis for weekly values of the courses of the New York Stock Exchange, using a simple test based on the volatility of the courses. They specified however that the rejection of the random walk assumption does not necessarily imply the inefficiency of the stock-price formation. We will outline this test to show how the academic assumption of random walk for the financial markets is subjected to critiques from statisticians today.

A way of approaching the study of statistical properties of *chaotic processes* results from the *Birkhoff and Von Neumann works* about the *invariant distributions*[11] (distributions which have a positive Lebesgue measure). Dynamics of *aperiodic nature* sometimes have variables with distributions of this type, *which indicate the frequency with which they take values in a given interval*. The most analyzed invariant distributions are those which can be represented by a density function. Techniques have been developed in order to build such functions for chaotic processes.

Part III

One of the characteristics of behaviors belonging to nonlinear models is to highlight transitory phenomena, intermittencies, turbulences or chaotic dynamics. They correspond to a tool which allows the representation of phenomena of the stationary types already depicted by linear models, and, also at the same time, phenomena of the periodic and turbulent or chaotic type. This faculty of representation is particularly useful. The complexity of the phenomena observed empirically thus finds an algebraic tool which makes it possible to depict this complexity. The writing of dynamical systems must be able to *depict the coexistence of simple and complex solutions*, by the elementary play of parameter setting. Apparently, deterministic chaos resulting from nonlinear models does not seem different from a random walk measured on a natural phenomenon, a stock-exchange or economic phenomena. However, in the first case, we are the "holders" of the equations of the system, whereas in the second case, we do not know if these equations exist.

This evolution of the "linear" towards the "nonlinear" can find a symbolic illustration in the overtaking (which proved to be necessary) of the Fourier analysis of certain natural phenomena. In his study about Heat in the nineteenth century, Fourier showed how his works help to understand the Natural phenomena by helping to

---

[11] Absolutely continuous invariant distributions which have a positive Lebesgue measure.

numerically solve the equations which hitherto were refractory. For a number of differential equations, the Fourier transformation replaces a complicated equation by a series of simple equations. Each one of these equations indicates the temporal evolution of the coefficient of one of the sinusoids composing the initial function. Each coefficient has its own equation. Thus the Fourier transform had an important success, insomuch it was often used for problems to which it was unsuited (Meyer 1992, p. 86). The Fourier analysis is not appropriate for all signals nor for all problems. In fact, "the Fourier analysis helps to solve *linear problems, for which the effect is proportional to the cause*" (Hubbard 1998). Nonlinear problems are more difficult to solve. Generally, the interacting variables form systems whose behaviors are unstable. And faced with this type of problem we treat it "as if it was linear". Indeed, certain complex natural phenomena that would require the use of nonlinear partial differential equations are not solved in this manner because they cannot be solved. It is by reducing the difficulty to a linear equation, which can be solved by the Fourier analysis, that answers were brought. In addition, this is a reality that the economists had to face in the past. The elements of the Fourier analysis are sines and cosines which oscillate infinitely with a fixed frequency. And "in this context of infinite time, the following expression, *changing frequency,* becomes a contradiction".[12] "The Fourier analysis is thus unsuited to the signals which change abruptly and in an unpredictable way" (Hubbard 1998), and however these events contain the most information. Now, "wavelet analysis is a manner of expressing our sensitivity to the variations".[13] Thus, the abrupt and unforeseeable variations are "read" by the wavelets. The study of turbulence phenomena, i.e. sudden unpredictable chaotic variations which are representative of nonlinear problems, are in this respect symbolic of the efficiency lack of the previous investigation methods.

The Fourier analysis is not appropriate to the nonlinear problems that we meet in turbulence phenomena, the nature of the wavelets is more adapted. The turbulence appears on scales of very diverse frequencies and the wavelets are also adapted to the analysis of the interactions between scales. In spite of what precedes, there is no ideal tool to resolve nonlinear behaviors. All of these observations about waveform analysis is one of the objects of the third part of this work.

The time-frequency theory offers transformation methods of time-series. A complete statistical theory about time-frequency analysis does not exist yet, in spite of certain recent work developed from the starting point of the Wiener deterministic spectral theory. Without covering this vast and very difficult topic here, a very particular interest however is granted to the statistical and econometric properties of the wavelet transform of stock-exchange and economic time series. We will argue in favor of wavelets and their particular properties, their adaptation to nonstationarities and to abrupt variations of signals, as it was evoked previously. Most of all, we will highlight the role of *atomic decompositions* of signals which use *time-frequency atoms* (i.e. *waveforms dictionaries*, which contain different types of waveforms to decompose a signal). *Numerical imaging* in time-frequency planes

---

[12] D. Gabor quotation, Nobel Prize of Physics 1971 for the invention of holography. Recovery in the work of Hubbard (1998).

[13] Yves Meyer, one of the founders of the wavelet theory.

(or in three-dimensional time-frequency-amplitude spaces) are also a considerable contribution to the comprehension of the subjacent (or hidden) structures for signals whose origin is natural, financial or economic. A contribution for which it is probably necessary, in economic and financial matters, to develop reference frameworks, i.e. for example data banks of *imagings* and *forms* of *structures* calibrated on academic signals. This could be the case for signals whose origin is the stock-exchange that we commonly consider as following a "random walk".

Before all new analysis methods briefly mentioned previously, to which it is advisable to add *polyspectral analysis*, the classical "spectral analysis" *has been already experimented in* Economics in the past. Indeed, the first work dates back to Kendall in 1922, spectral analysis has been experimented on the Beveridge corn price index. More recently, it is possible to quote the fundamental contribution of C.W. Granger in 1969, those of M.W. Watson in 1993 (Watson 1993), but also the work of Wen Y. in 1998 (Wen 1998), applied to the Real Business Cycle theory, as those of Perli and Sakellaris (1998) about the same topic. We previously evoked the fact that, apparently, the deterministic chaos resulting from nonlinear models does not seem different from a random walk measured on a stock-exchange market for example. However for the first, we know the equations, whereas for the second we do not know if they exist.

The recent Lo and MacKinlay results mentioned in the preceding part, which reject the random walk hypothesis about stock-exchange markets, can be related to the conclusions of quite recent work (although of very different nature) presented by J.B. Ramsey and Z. Zhang about the Standard & Poor's stock 500 index;[14] Thus, we are interested in describing the "Matching Pursuit Accord" (Mallat and Zhang 1993) algorithm with its dictionaries of "waveforms" (i.e. Gabor or time-frequency atoms) conceived by S. Mallat and Z. Zhang, and that we furthermore applied to the French stock-exchange index (cac40). Meeting the statistical Lo and MacKinlay analysis, the decomposition (into "time-frequency atoms") of a signal could make it possible to discriminate its random or non-random characteristic. The argument is based on the fact that the number of waveform structures necessary for the decomposition of a "natural" random series is higher than the number of waveform structures used to decompose a stock-exchange series. Such an analysis would indeed plead in favor of the idea that stock-exchange markets do not follow a random-walk, contrary to the "traditional" conclusions of statistical analysis. Moreover, within this same framework, a detailed attention was given to the treatment of *internal or external shocks of a signal*, in particular concerning the Dirac (delta) function. Indeed, this same work attempts to show that the high-energy bursts of a time series which involve *almost all* frequencies of a spectrum would make it possible *to discriminate* the internal or external origin of shocks of the aforementioned time series.

---

[14] The S&P500 is decomposed through the "Matching Pursuit Accord" by means of dictionaries of time-frequency atoms.

Part IV

The last part aims to depict the growing place of nonlinearities in the conception of economic growth models. The models used during the twentieth century were primarily of linear nature, including those aiming to represent business cycles; in particular we think about the Frisch–Slutsky linear model (1930). The cycles resulted then from the propagation of external shocks in our economies and were not generated in an endogenous way by the model. All the models deriving from these principles quickly showed their limits. The use of nonlinear models proved to be necessary, Hicks in 1950 and Goodwin in 1951 had already introduced them, but they had at that time a weak resonance, amongst other reasons due to the lack of statistical and computational tools. Note that the first application of "catastrophe theory" to economics, and in particular to Kaldor model (1940) seems to date back to H.R. Varian (1979). Nevertheless, the linearity hypothesis is so simple to pose and exploit, that is why historically it has long been used.

The *AK* models for example concern this hypothesis, where the output is constantly proportional to the input and *A* is taken as a parameter. Endogenous growth is generated because of a linearity in the differential equation of the model. However, this hypothesis is subject to critiques and is not necessarily relevant. Indeed, on the one hand, the basic rules of the economic theory, which state the decrease of marginal productivity or marginal utility, implicitly imply that the relation between input and output is of rather nonlinear nature. In this respect, the Von Neumann model, based on relations of linear nature, avoids this pitfall, since it is not built from the neo-classical production function.

Moreover, even within the framework of models whose construction is linear, the nature of the domain of definition in particular at its boundaries, can be the source of nonlinearities because of "edge-effects". In this respect, we can evoke the variables which are rates and ratios, like the per capita variables for example. The "nonlinearity" hypothesis is more relevant than its opposite which, beyond its *didactic qualities*, must give way to more realistic concepts concerning the notion of nonlinearity. As said hereinbefore, certain endogenous growth models are built from linearities, but the endogenous growth phenomenon can appear without linearity, for example from a model with two differential equations whose respective convexities of trajectories are different. Thus, in the absence of linearity, certain configurations allow the observation of endogenous growth. *The economic models gradually endogenize the mechanisms of growth*. Exogenous in the various Solow models, the factors of Growth integrated (directly or indirectly) the heart of the models, whether they result from internal mechanisms of choice optimization between consumption and savings, or from positive externalities as in the Romer model.

The conception of optimal growth models highlights a system of differential equations of order two, which determines the optimal trajectory of consumption and investment. The solutions of this type of model generally appear as a saddle-point (i.e. saddle path) which is a kind of instability basin, in which the trajectory of steady balance is narrow and depends on the initial conditions. In economic models, this type of instability basin represented by the saddle-point is symptomatic of the

prevalence of unstable trajectories where equilibrium and stability are understood as *singularities*. The optimal growth model, applied to the strategy of portfolio choice, leads to the same observation and suggests a way to explain the instability of stock markets. In the Boldrin–Woodford model (1990), the optimal growth exhibits cyclic or chaotic endogenous fluctuations. The axis of the saddle, which represents the stable trajectory, converges towards an equilibrium point, which becomes (under some conditions) a limit-cycle at the equilibrium in the Benhabib–Nishimura optimal growth model. This analysis is developed through the "Rational Expectations" school and explains the cycles (and fluctuations) of economy under a new light. The model replaces the academic notion of shock by internal behaviors of optimization which become the source of the Equilibrium cycles. An explanation of an endogenous type thus leads to a better understanding of these phenomena, but this requires to accept, as a preliminary, the probable preeminence of nonlinearities in the construction of the economic models. The hypothesis of nonlinearities, more plausible for the comprehension of natural and economic phenomena, allows to explain equilibrium cycles but also of chaotic behaviors. The models based on nonlinear algebraic structures, such as those of Day and Day–Lin, also depict these periodic or chaotic behaviors. The structure of overlapping generations models developed for instance by Benhabib–Day (1982) and Grandmont (1985) also allows to produce cyclic or chaotic dynamics.

The economic growth models attempt to depict the evolution of national income. However, stock markets, which are known as perfect markets and also known as advanced indicators of economic activity, exhibit trajectories very different from those of the national product. The characterization of these growth disparities (i.e. gaps) is not an easy task, but can be approached from various angles through the concepts of Market and Economic *Value-Creation*,[15] through the rational expectation concept and that of rational bubbles, but also through the contribution of the sunspots models.

---

[15] e.g., indicators such as *Market Value Added* (MVA) and *Economic Value Added* (EVA).

Chapter 1
Nonlinear Theory

In the general introduction we observed that the irruption of the "nonlinear" led to a profound transformation of a great number of scientific fields. The behaviors resulting from the "nonlinear" make it possible to better understand the natural phenomena considered as complex. The "nonlinear" introduced a set of concepts and tools, i.e. analysis and investigation instruments of dynamics generated by the "nonlinear". We try to gather these investigation tools, knowing that today it is possible to say that there exists a kind of "conceptual unification" through the notions of attractors, period-doublings, subharmonic cascades, bifurcations, Lyapunov exponent, sensitive dependence on initial conditions, etc. Belonging to this set, there are techniques gathered under the name of *nonlinear signal processing* based mainly on the *Takens theorem*. The purpose is the reconstruction of the time series based on the topological equivalence concept. While saving an important amount of calculation, it allows the identification of the nonlinear nature[1] of the studied time-series. The stakes are important: while working in a reconstructed phase-space of very low dimension, the objective is to reproduce the essential features of the original dynamics, without necessarily knowing the equations of the dynamical system which generated the studied series.

The *topological equivalence* enables us to study geometrical objects of low dimension which will provide the desired information about the original dynamics. In Economics, we can face this type of problems. Indeed, the number of variables implied in a dynamics is not necessarily known, just as the dimension of the system which can be infinitely large. Moreover, it is known that the *attractor of large dynamical systems (even infinite) can have low dimensions*. Consequently, with considerably reduced series, the study of low dimension objects can make it possible to extract information which we need to identify dynamics. These concepts of topological equivalence and attractor of low dimensions lead to a major concept which is that of the *capacity dimension* (non-integer dimension). Indeed, it makes it possible, on the one hand *to characterize the attractor* which we have to face and, on the other hand, *to highlight the difference between deterministic chaos* and *random*

---

[1] Periodic, quasi-periodic and aperiodic or chaotic.

walk. It is said for example that a "Brownian motion"[2] has a capacity dimension equal to two, which is not necessarily the case of an apparent deterministic chaos. In connection with these (investigation) tools of nonlinear dynamics, it is interesting to highlight the results of recent work completed by Butkovskii, Kravtsov and Brush concerning the predictibility of the final state of a nonlinear model. The model used here is that of the logistic equation with its cascade of subharmonic bifurcations. The observed phenomenon is rather singular and paradoxical in relation to an a priori knowledge that we could have before the experimentation. It results from this work that the predictibility of the final state of the model, after the first bifurcation point, depends on the speed of change of its control parameter and on the background noise. A relation is established between the probability of the final state, the transition speed and the noise level. A critical value of the speed is highlighted by the experimentation. Indeed, when the speed is strictly higher than its critical value, for a given noise level, the probability is close to 1, whereas if the speed is lower than its critical value, the probability of the state is close to 1/2. Such works, which introduce the control speed in nonlinear dynamic models, could find interesting transpositions in economic models. Thus, this type of experimental formalization introduces a further dimension for the effectiveness and the control of economic policies.

The "nonlinear signal processing" methods evoked above make it possible, in particular, to diagnose the nature of a dynamics from the time series itself. *Spectral analysis* is a tool which concerns the same objective. Spectral analysis results from two mechanisms which are the Fourier transform and the autocorrelation function. However, the autocorrelation function, which measures the resemblance of values of a variable in time, corresponds to the Fourier transform of the power spectrum. In the case of chaotic trajectories, the resemblance decreases with time. There is an *internal loss of similarity* to their processes which is the cause of our incapacity to make forecasts, because these anticipations are established from the resemblances in relation to the past. Thus, the forecasts rest on resemblances, recurrences and periodicities. The periodicities of a dynamics are identified by spectral analysis, whether the model which underlies it is known or not. Certain dynamics can result from simple periodicities, however other dynamics (less simple) can be the result of a superposition of oscillations of different amplitudes. *Finally, certain complex dynamics can result from harmonics of primary oscillations. And in this case, the regime of the dynamics, characterized as quasiperiodic, has an associated attractor, which is of a higher order than the limit-cycle, it is a torus. And for chaotic but deterministic regimes, i.e. for dynamics represented by a limited number of nonlinear differential equations, the attractor in the phase space is a "strange attractor".*

The spectral analysis is thus a tool with which one can identify the nature of dynamics, periodical, aperiodic or quasiperiodic. This tool was tested on different types of time-series, such as: standard ARMA processes, stock-exchange time series (e.g. the French stock-exchange index: *cac40*), behaviors of the *Van der Pol*

---

[2] See section about the (non-fractional) Brownian motions.

oscillator, and also various trajectories resulting from the logistic equation, for many values of the fertility parameter. In this last case, we will particularly observe the way in which spectral analysis, at the threshold of the chaotic regime occurrence, highlights the rise of spectral background noise, which comes to "enfold" the *harmonic frequencies* which existed before this threshold. A particular place is given to the "Takens theorem" and to the principle of time-series reconstruction, with various applications. Moreover in the second chapter, one of its possible derived applications will be exposed, the technique of the "singular spectrum analysis".

Main subjects treated in this first chapter: (a) *Dynamical systems* (Solutions, flows, maps, vector fields, existence and uniqueness of the solutions, stability and Floquet theory, Center manifold theorem, Poincaré–Bendixson theorem, fixed-points, saddle-connection,...). (b) *Invariant sets* (Simple and strange attractors, simulations,...). *Dissipative and conservative systems* (Hamiltonian system and optimal growth model,...). (c) *Instruments and concepts* of *Chaos theory* (Deterministic chaos, Sensitivite dependence on initial conditions, Poincaré map, Lyapunov exponent, period-doubling and cascades of bifurcations, stable-superstable and unstable cycles, Kolmogorov capacity dimension or box counting, Entropy or the measure of disorder, KAM theory, Invariant torus and chaos, rotation number, coupling of frequencies,...). (d) *Topological equivalence* (From the topological equivalence of flow towards the Poincaré map, orientation-preserving diffeomorphisms and topological equivalence of flows, rotation number, suspension,...). (e) *Reconstruction theorems* (Topological equivalence, embedding spaces, Whitney embedding theorem, Broomhead and King theorem, Takens reconstruction theorem, window concept,...). (f) *Navier–Stokes dynamics and turbulent phenomena* (Invariant ergodic measures, Floquet theory, characteristic exponents and Hilbert spaces, Dirac delta functions, Invariant measure of probability, nonlinear operators,...). (g) *Oscillators and Pendulums* (Approach of dynamical systems by means of oscillators, approach of turbulence via the concept of "Mode"[3] and the Fourier intuition, approach of turbulence via the "nonlinear" analysis). (h) *Nonlinearity and Hyperbolicity* (Cantor sets, contradiction between dissipative systems and chaos solved by the capacity dimension, structural stability, Smale horseshoe maps, hyperbolic sets, Anosov diffeomorphisms, symbolic dynamics, Smale–Birkhoff homoclinic theorem, Hartman–Grobman theorem, hyperbolic nonlinear fixed-points, hyperbolic structures, homoclinic orbits and perturbations, Melnikov method, Shilnikov phenomenon,...). (i) *Transitions to chaos* (Stability Boundaries, Saddle-connection, blue sky catastrophe, Intermittency, Period-doubling, Quasiperiodic route to chaos [Ruelle–Takens],...). (j) *Spectral analysis* (Temporal correlation, power spectrum density, spectral analysis of various academic signals, spectra of van der Pol oscillator,...).

---

[3] *Mode*: Collection of independent oscillators.

1.1 Dynamical Systems

1.1.1 Differential Equation and Difference Equation

If we pose $x \in U \subset \mathbb{R}^n$, $t \in I \in \mathbb{R}^1$, and $\alpha \in V \subset \mathbb{R}^p$, where U et V are open sets respectively in $\mathbb{R}^n$ et $\mathbb{R}^p$, we use the two following definitions to indicate a *dynamical system*:

The *differential equations,* or "vector fields",

$$\dot{x} = f(x,t;\alpha), \quad with \quad \dot{x} \equiv \frac{dx}{dt}. \tag{1.1}$$

or, in a developed form, with n differential equations:

$$\frac{dx_1}{dt} = f_1(x_1,\ldots,x_n;t;\alpha)$$
$$\vdots \tag{1.2}$$
$$\frac{dx_n}{dt} = f_n(x_1,\ldots,x_n;t;\alpha).$$

The *difference equations,* or the "maps",

$$x \to g(x;\alpha)$$
$$x_{n+1} = g(x_n;\alpha). \tag{1.3}$$

The vector x indicates the studied physical or economic variables (or one of their transformations), t indicates the "time" and α symbolizes the parameter(s) of the system. It is also said that in a dynamical system "time" is the independent variable. The dependent variables x, spread in $\mathbb{R}^n$ which is called *the phase space* or *the state space*. Whereas $\mathbb{R}^n \times \mathbb{R}$ is called the *motion space* of the system.

1.1.2 Solution-Trajectory of a Dynamical System

(a) The term *solution* of the system $\dot{x} = f(x,t;\alpha)$ means a *map x*, from an interval $I \subset \mathbb{R}^1$ to $\mathbb{R}^n$, which is written as follows:

$$x: \quad I \longrightarrow \mathbb{R}^n,$$
$$t \longmapsto x(t), \tag{1.4}$$

such that $x(t)$ satisfies $\dot{x} = f(x,t;\alpha)$, i.e.

$$\dot{x}(t) = f(x(t),t;\alpha). \tag{1.5}$$

The geometrical interpretation of the map x is a *curve* in $\mathbb{R}^n$, and $\dot{x} = f(x,t;\alpha)$ is the *tangent vector* at each point of the curve. Thus, we refer to $\dot{x} = f(x,t;\alpha)$ as being a *vector field*.[4] In addition, we refer to the space $\mathbb{R}^n$ of independent variables as the phase space of $\dot{x} = f(x,t;\alpha)$. The goal is to understand the *geometry of "solution-curves" in the phase-space*. It will be noticed that in many cases, the structure of the phase space can be more general than $\mathbb{R}^n$, we can mention for example, phase-spaces such as: *Spherical*, *Cylindrical*, or *Toric*.[5] We enunciated that the equation $\dot{x} = f(x,t;\alpha)$ is called *vector field*. Indeed, this term is often used because its solution at each point x is a curve in $\mathbb{R}^n$, for which the *speed* is given by $f(x)$.

(b) We will prefer the following notation to indicate the solution of the differential equation $\dot{x} = f(x,t;\alpha)$:

$$\phi : I \to \mathbb{R}^n. \tag{1.6}$$

Generally, in Economics $I = [0,+\infty)$, such that ϕ is differentiable on I, $[\phi(t)] \in U$ for all $t \in I$, and $\dot{\phi}(t) = f[\phi(t)]$, for all $t \in I$. The set $\{\phi(t) : t \in I\}$ is the *orbit* of ϕ: it is contained in the phase space. The set $\{t, \phi(t) : t \in I\}$ is the *trajectory* of ϕ: it is said that the trajectory is contained in the *space of motions*. In practice the terms *trajectories and orbits* are often synonyms.

(c) In order to write the solution of the system, if we wish to mention the initial conditions, i.e. the point x_0 at time t_0, then we write $\phi(t,t_0,x_0)$. Or if $t_0 = 0$, we write $\phi(t,x_0)$. In order that a solution $\phi(t,x_0)$ exists, it suffices that f is continuous. In order that the solution is unique, it suffices that f is of C^1 class in U^3.

(d) The writing of solutions, most largely used, is as follows for an autonomous system:

$$\phi_t(x) : U \to \mathbb{R}^n; \quad \text{with} \quad \phi(t,x) \equiv \phi_t(x). \tag{1.7}$$

Whereas we have written in (a) for x:

$$\begin{aligned} x: \ I &\longrightarrow \mathbb{R}^n, \\ t &\longmapsto x(t). \end{aligned} \tag{1.8}$$

Indeed, for this system, *the "translated" solutions*, *remain solutions*, i.e. if $\phi(t)$ is a solution, $\phi(t+\tau)$ is also a solution for all $\tau \in \mathbb{R}$. If we consider t as a fixed parameter, we can write $\phi(t,x) \equiv \phi_t(x)$, or $\phi_t(x) : U \to \mathbb{R}^n$. Consequently, all the solutions can be regarded as a *family of maps* with one parameter from the phase space in itself. The map ϕ_t is called the *flow* generated by the vector field f. This terminology refers to the evolution of a *fluid flow* in the course of time, e.g. a natural stream. In *a discrete-time dynamical system*, where the evolution of the variable will be represented by sequences of discrete values, we will write (where $g = \phi$ and τ is the value of the parameter t):

$$\begin{aligned} x &\longrightarrow g(x), \\ x_{n+1} &= g(x_n). \end{aligned} \tag{1.9}$$

---

[4] Term used in particular in physics to indicate the orientation of the solutions of a system.

[5] Introduction to applied non linear dynamical systems and chaos. S. Wiggins. Springer.

1.2 Autonomous and Non-Autonomous Flows, Fixed-Point

1.2.1 Definition of a Flow

The flow of a dynamical system is the expression of its trajectory or beam of its trajectories in *the phase space*, i.e. the movement of the variable(s) in time. If we consider the following autonomous dynamical system $\dot{x} = dx/dt = f(x)$, we can give the following definition of a flow:

Definition 1.1 (Flow). Given $x_0 \in U$, let $x(x_0,t)$ be a solution of $\dot{x} = dx/dt = f(x)$, with as initial conditions $x(0) = x_0$. We call flow of $\dot{x} = f(x)$, or of the vector field f, the map $\phi_t : U \to \mathbb{R}^n$ defined by

$$\phi_t(x_0) = x(x_0,t), \tag{1.10}$$

where ϕ_t has the following properties:

$$
\begin{array}{ll}
(1) & \phi_t(x_0) \text{ is of class } C^r, \\
(2) & \phi_0(x_0) = x, \\
(3) & \phi_{t+s}(x_0) = \phi_t(\phi_s(x_0)).
\end{array}
\tag{1.11}
$$

1.2.1.1 Remark About the Flow

We note that ϕ_t can be expressed in two manners:

- Either as the following map:

$$\phi_t : I \to \mathbb{R}^n, \quad \text{with} \ \ I \in \mathbb{R}^1, \tag{1.12}$$

  where I expresses the "time" of the system.
- Or as the following map:

$$\phi_t : U \to \mathbb{R}^n, \quad \text{with} \ \ U \in \mathbb{R}^n, \tag{1.13}$$

  U expresses the "phase space" of the system. Thus, as seen previously, the set of solutions ϕ_t becomes a map of the phase space in itself.

1.2.1.2 Autonomous and Non-Autonomous Flows

The flow can be a solution of a dynamical system, non-autonomous: $\dot{x} = f(x,t;\alpha)$ and autonomous: $\dot{x} = f(x;\alpha)$. It will be also noticed that if the system is non-parametric, the flows are written respectively non-autonomous flow and autonomous flow: $\dot{x} = f(x,t)$, or $\dot{x} = f(x)$.

1.2.2 Continuous and Discrete System

The expression of such systems of differential equations in the *continuous* case is written

$$\frac{dx}{dt} = \dot{x} = f(x,t;\alpha). \tag{1.14}$$

On the other hand, in the *discrete case*:

$$x_{t+1} = f(x_t;\alpha). \tag{1.15}$$

The values of x spread themselves in the $\mathbb{R}^n$ space, which is called *the phase space*, where n corresponds to the number of variables of the system, x which can obviously be a set of different variables, i.e. a vector field.

1.2.3 Definition of a Fixed Point and a Stable Fixed Point

Definition 1.2 (Fixed point). A *fixed point* (also called a stationary point, or equilibrium point) of $dx/dt = f(x)$, is the point $\bar{x}$ of the phase space obtained by canceling the second member of $dx/dt = f(x)$, i.e. $f(\bar{x}) = 0$.

(An usual approach is to bring back the fixed point $\bar{x}$ to the origin by the "change of variables" $y = x - \bar{x}$).

Definition 1.3 (Stable fixed point). A fixed point $\bar{x} \in \mathbb{R}^n$ is stable if $\forall \varepsilon > 0, \exists \delta > 0$, such that $\|x(0) - \bar{x}\| < \delta \Rightarrow \|x(t) - \bar{x}\| < \varepsilon$, and if $\exists \delta_0$ with $0 < \delta_0 < \delta$, such that $\|x(0) - \bar{x}\| < \delta_0 \Rightarrow \lim_{t \to \infty} x(t) = \bar{x}$, $\bar{x}$ is asymptotically stable (Fig. 1.1).[6]

We will present the "stability" concept in one of the following sections.
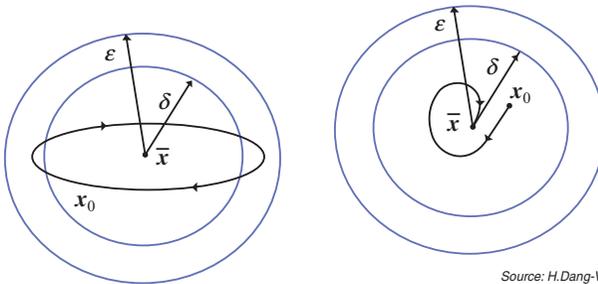
Source: H.Dang-Vu. C. Delcarte

Fig. 1.1 Stable fixed point (*left*). Asymptotically stable (*right*)

---

[6] Where $\|.\|$ is the norm in $\mathbb{R}^n$.

1.3 Introduction to the Resolution of Nonlinear Dynamical Systems

1.3.1 Linearization of Nonlinear Models

Many models are nonlinear, and we are led to linearize in order to analyze their stability. It is posed the following non parametric and autonomous model:

$$\frac{dx}{dt} = \dot{x} = f(x). \tag{1.16}$$

$f(x)$ can be approximated by Taylor series truncated around the stable point $\bar{x}_s$ (corresponding to a solution flow $\phi(t)$):

$$f(x) = f(\bar{x}_s) + \frac{\partial f}{\partial x}(x - \bar{x}_s) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x - \bar{x}_s)^2 + o(\cdot). \tag{1.17}$$

Neglecting $o(\cdot)$ the terms of high degrees, we write

$$f(x) \approx f(\bar{x}_s) + \frac{\partial f}{\partial x}(x - \bar{x}_s). \tag{1.18}$$

Being an equilibrium-point, we know that: $f(\bar{x}_s) = d\bar{x}_s/dt = \dot{\bar{x}}_s = 0$ (and for the map $\bar{x} = g(\bar{x})$). Thus we can write

$$\frac{dx}{dt} = f(x) \approx \frac{\partial f}{\partial x}(x - \bar{x}_s). \tag{1.19}$$

Since the derivative of a constant $\bar{x}_s$ is zero, we can write $\dfrac{d(x - \bar{x}_s)}{dt} = \dfrac{dx}{dt}$. Consequently:

$$\frac{d(x - \bar{x}_s)}{dt} \approx \frac{\partial f}{\partial x}(x - \bar{x}_s). \tag{1.20}$$

$\bar{x}_s$ being stable, we can note y as being the difference between $\bar{x}_s$ and x, to express the *deviation* in relation to the values of the stable state

$$y = x - \bar{x}_s. \tag{1.21}$$

(It is sometimes also written $x' = x - \bar{x}_s$).

$$\frac{dy}{dt} \approx \frac{\partial f}{\partial x}y. \tag{1.22}$$

Thus, we write: $\dfrac{dy}{dt} = ay$, where $a = \dfrac{\partial f}{\partial x}$. Thus:

$$\dot{y} = ay, \tag{1.23}$$

consequently, a general solution of the model is written

$$y = e^{at} y_0. \tag{1.24}$$

We have just seen how to linearize a simple system, then we can generalize the approach.

1.3.2 Linearization Generalized to All Nonlinear Models

Beforehand, we can give some details about the vocabulary generally used when we describe dynamical systems to differentiate the different variables of a system. It is said that a *state variable* is a measurable quantity which indicates the state of a system; this must be connected to the concept of *dependent variable,* or *output* variable. In addition, we speak of the *"input" variable* as a variable which must be specified before seeking the solution of a system, i.e. to some extent, it is a process on which we can operate. By way of example, we can quote the speed of the flow in Fluid Mechanics. The input variables "are often handled" to control processes in order to obtain "desired results". We can relate them to the group of independent variables of a system. This type of variable is defined, in contrast, to the *parameters* which also must be specified before the resolution of the model, but correspond to an *element fixed by nature* in the dynamics that we study. In Economics, it is true that the distinction between parameters and "input variables" is weak, knowing that the economist has rarely the opportunity to become a "process controller", because economist cannot do experiments in laboratory.

Let us consider the general nonlinear model in which $\mathbf{x}$ is a vector of n state-variables, and $\mathbf{u}$ is a vector of "input-variables" of dimension m, and $\mathbf{y}$ is a vector of r "output variables":

$$\dot{\mathbf{x}}_1 = f_1(x_1, \ldots, x_n; u_1, \ldots, u_m),$$

$$\vdots$$

$$\dot{\mathbf{x}}_n = f_n(x_1, \ldots, x_n; u_1, \ldots, u_m), \tag{1.25}$$

$$\mathbf{y}_1 = g_1(x_1, \ldots, x_n; u_1, \ldots, u_m),$$

$$\vdots$$

$$\mathbf{y}_r = g_r(x_1, \ldots, x_n; u_1, \ldots, u_m),$$

or, in a simplified form:

$$\dot{\mathbf{x}} = \mathbf{f(x,u)}, \tag{1.26}$$

$$\mathbf{y} = \mathbf{g(x,u)}. \tag{1.27}$$

Elements of linearization matrices:

$$A_{ij} = \partial f_i / \partial x_j,$$
$$B_{ij} = \partial f_i / \partial u_j,$$
$$C_{ij} = \partial g_i / \partial x_j,$$
$$D_{ij} = \partial g_i / \partial_j u. \tag{1.28}$$

After linearization, ("/" expresses the linearized vectors), we obtain

$$\dot{\mathbf{x}}' = \mathbf{A}\mathbf{x}' + \mathbf{B}\mathbf{u}', \tag{1.29}$$
$$\mathbf{y}' = \mathbf{C}\mathbf{x}' + \mathbf{D}\mathbf{u}'. \tag{1.30}$$

We often omit "/" and the system is written

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \tag{1.31}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}. \tag{1.32}$$

The measured (output) variable is not a direct function of the input variable; thus it is more usual to write the state-space system:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}, \tag{1.33}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x}. \tag{1.34}$$

1.4 Resolution of the Zero-Input Form

In (1.31) $\mathbf{x}$, $\mathbf{u}$ are (deviation) variable vectors for the states and inputs. The inputs are now supposed constant at their steady-state values and the states can be perturbed from steady-state. The model is

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \tag{1.35}$$

where $\mathbf{x}$ the state-variable can be a linearized variable. From the model above, we analyze the stability and the behavior of the system. Remember: $\dot{x} = ax$ has as solution $x(t) = e^{at}x(0)$, which is stable for $a < 0$.[7] In the same way, for the system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, the *solution* is written

$$x(t) = e^{At}x(0), \tag{1.36}$$

(e^{At} could be considered as a series:

$$e^{At} = \sum_{n=0}^{\infty} \frac{1}{n!}(At)^n = I + At + \frac{1}{2!}A^2t^2 + \cdots). \tag{1.37}$$

---

[7] The solution for a model $\dot{x} = Ax$ can be written $x(t) = e^{At}c$.

The solution of our *system is "stable"* if *all eigenvalues of A* are lower than zero. The solution is *oscillatory* if the eigenvalues are *complex*. e^{At} could be considered as a flow ϕ_t, (i.e. $\phi_t = e^{At}$). At this stage, there are several manners of calculating the exponential of the matrix. Here the following method will be used

$$\mathbf{AV} = \mathbf{V\Lambda}, \tag{1.38}$$

where $\mathbf{V}$ is the eigenvector of $\mathbf{A}$ and $\mathbf{\Lambda}$ is the matrix of eigenvalues of $\mathbf{A}$. For a matrix $\mathbf{A}$ of dimension 2×2, we write the matrix of eigenvectors of $\mathbf{A}$ as follows:

$$\mathbf{V} = [\xi_1, \xi_2] = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}, \tag{1.39}$$

where $\xi_1 = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}$ is the first eigenvector associated with the eigenvalue λ_1, and

where $\xi_2 = \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix}$ is the second eigenvector associated with the eigenvalue λ_2.

Thus, the matrix of eigenvalues is written

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \tag{1.40}$$

Moreover, $\mathbf{AV} = \mathbf{V\Lambda}$ by multiplying by $\mathbf{V}^{-1}$ can be written

$$\mathbf{A} = \mathbf{V\Lambda V}^{-1}. \tag{1.41}$$

Multiplying by the scalar t and by taking exponential matrices:

$$e^{At} = \mathbf{V}e^{\Lambda t}\mathbf{V}^{-1}, \tag{1.42}$$

where

$$e^{\Lambda t} = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix}. \tag{1.43}$$

Here, we understand the reason why if $\lambda < 0$ the solution is stable. The solution of the system is

$$\mathbf{x}(t) = \mathbf{V}e^{\Lambda t}\mathbf{V}^{-1}x(0). \tag{1.44}$$

It will be noticed that if the vector of initial conditions $x(0)$ is equal to ξ_i, then the response of the model will be similar to ξ_i.

1.4.1 Solution of the General State-Space Form

We point out that a solution of a simple differential equation $\dot{x} = ax + bu$ is

$$x(t) = e^{at}x(0) + (e^{at} - 1)(b/a)u(0), \quad \text{when} \quad u(t) = constant = u(0). \tag{1.45}$$

In a same way, consider now the general form

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}, \tag{1.46}$$

for all $t > 0$ and for $\mathbf{u}(t) = constant = \mathbf{u}(0)$, the *solution* is

$$\mathbf{x}(t) = \mathbf{Px}(0) + \mathbf{Qu}(0), \tag{a}$$

where $\mathbf{Q} = (\mathbf{P} - \mathbf{I})\mathbf{A}^{-1}\mathbf{B}$ with $\mathbf{P} = e^{At}$. This equation can be used to solve a system in which the input changes from time step to time step by means of:

$$\mathbf{x}(t + \Delta t) = \mathbf{Px}(t) + \mathbf{Qu}(t) \tag{b}$$

or in another form:

$$\mathbf{x}(k + 1) = \mathbf{Px}(k) + \mathbf{Qu}(k), \tag{1.47}$$

where k is the k-*th* time step (or the k-*th* iteration). The initial form of the system is often used for the *numerical integration* calculations of systems aiming to their resolution (Ode). The evoked methods are of the type: Euler, Runge–Kutta integrations, (Newton–Rapson). They are in particular approximation methods of differential equations solutions, using the slope concept and the Taylor expansion.

1.5 Existence and Uniqueness of Differential System Solutions

1.5.1 Lipschitz Condition

If we consider the function $f(x,t)$, the differential equation $\dot{x} = f(x,t)$ must satisfy some conditions, among which the Lipschitz condition is most important.

Definition 1.4 (Lipschitz condition). Let us consider the function $f(x,t)$, where

$$f : \mathbb{R}^{n+1} \to \mathbb{R}^n, \tag{1.48}$$

$$|t - t_0| \leq a, x \in D \subset \mathbb{R}^n, \tag{1.49}$$

$f(x,t)$ satisfies the Lipschitz condition, i.e. if x in $[t_0 - a, t_0 + a] \times D$, where:

$$\|f(x_1,t) - f(x_2,t)\| \leq L\|x_1 - x_2\|, \tag{1.50}$$

with x_1 and $x_2 \in D$ and L a constant. L is called the Lipschitz constant.

Theorem 1.1. *If we consider the differential system $\dot{x} = f(x,t)$ with the initial conditions $x(t_0) = x_0$, with $x \in D \subset \mathbb{R}^n, |t - t_0| \leq a; D = \{x : \|x - x_0\| \leq d\}$, a and b are positive constants. The function $f(x,t)$ satisfies the following conditions: (1) $f(x,t)$ is continuous in $g = [t_0 - a, t_0 + a] \times D$, (2) $f(x,t)$ continuous in the Lipschitz sense on x, then the system has a unique solution for $|t - t_0| \leq \min(a, d/M)$, where $M = \sup_g \|f\|$.*

Here is an elementary example: $\dot{x} = x$, with $x(0) = 1$, $t \geq 0$, the solution exists for $0 \leq t \leq a$ with a as arbitrary positive constant. The solution, $x(t) = e^t$, can be continuous for any positive t and becomes unlimited for $t \to \infty$.

1.6 Stability of a Dynamical System

Literature about the subject is vast. We evoke the few essential definitions to pose the necessary tools in order to study the orbit stability of a system. Let us pose again the model $\dot{x} = f(x)$, with

$$f : U \to \mathbb{R}^n, \qquad (1.51)$$

U is an open set of $\mathbb{R}^n$ and f is of class C^r.

Definition 1.5 (Stability in the Lyapunov sense). A solution ϕ_t of the system is *stable in the Lyapunov sense* if for $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ such that for any other given solution $\psi(t)$ of system, for which $|\phi(t_0) - \psi(t_0)| < \delta$, we have

$$|\phi(t) - \psi(t)| < \varepsilon, \quad \text{for any } t \geq 0. \qquad (1.52)$$

Definition 1.6 (Asymptotic stability). Let us suppose that an orbit $\phi(t)$ is stable in accordance with the preceding definition and moreover suppose that there exists $\bar{\delta} > 0$, such that if $|\phi(t_0) - \psi(t_0)| < \bar{\delta}$, then:

$$lim_{t \to +\infty} |\phi(t) - \psi(t)| = 0; \qquad (1.53)$$

then $\phi(t)$ is said *asymptotically stable*.

Expressed in a different way, it will be said that an orbit is stable, if, when it is slightly *perturbated*, *the motion of the system does not diverge more than the non-perturbated orbit*. Moreover, if the orbit is asymptotically stable, the impact of the perturbation *will be gradually eliminated in time*. We exposed the definitions of the Stability through the flow concept ϕ_t, we can also present them through the map concept.

Definition 1.7 (Stability in Lyapunov sense for the solutions of a map). A solution $x \in \mathbb{R}^n$ of the map:

$$x \longmapsto g(x), \qquad (1.54)$$

where $x = \{x_0, x_1 = g(x_0), x_2 = g^2(x_0), \dots, x_n = g^n(x_0)\}$ is *stable in the Lyapunov sense,* if for all $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$, such that for any other solution y with $|x_0 - y_0| < \delta$, we have

$$|g^n(x_0) - g^n(x_0)| < \varepsilon \quad \text{for} \quad n \in [0, +\infty). \qquad (1.55)$$

Definition 1.8 (Asymptotic stability for the solution of a map). The solution x is *asymptotically stable* if it is stable and if, moreover, it exists $\delta > 0$ such as if $|x_0 - y_0| < \bar{\delta}$,

$$lim_{n \to \infty} |g^n(x_0) - g^n(y_0)| = 0. \qquad (1.56)$$

Remark 1.1 (Phase-space). In a dynamical system or transformation group, the topological space whose points are being moved about by the given transformations. A phase space can also be defined as a mathematical space with one point for every possible state of the system, which has as many dimensions as there are degrees of freedom in the system. (In Quantum Mechanics, the phase space is replaced by the state space formalism, usually a Hilbert space.) A phase space is the space in which all possible states of a system are represented. For mechanical systems, the phase space usually consists of all possible values of position and momentum variables. In a phase space, every degree of freedom or parameter of the system is represented as an axis of a multidimensional space. A phase space may contain very numerous dimensions.

Remark 1.2 (Phase plane). For a map with two degrees of freedom, the two-dimensional *phase space* which is possible for the map is called its phase plane.

Remark 1.3 (Phase portrait). A phase portrait can be understood as a plot of the different phase curves corresponding to the different initial conditions in the same phase plane.

Remark 1.4 (Phase diagram). A plot of position and momentum variables as a function of time is sometimes called a phase diagram.

1.7 Floquet Theory

1.7.1 Stability, Floquet Matrix and Eigenvalues

We introduce the floquet theory, which is connected with the stability concept of solutions for dynamical systems. To approach the subject, we focus on the *periodic solution of a nonlinear dynamical system* and on *dissipative systems*.

1.7.1.1 Stability Problem of a Periodic Solution

Hereafter, we describe the same problem by using different ways, in order to clarify the subject and to present it from three slightly different angles.

1st Presentation

Remember at this stage that we can write the periodicity T of a system in the following way: for a point P_0 for which we do not know yet if it is a stable solution (not fixed) of the system $P_0 = T(P_0) = T^2(P_0) = \cdots$
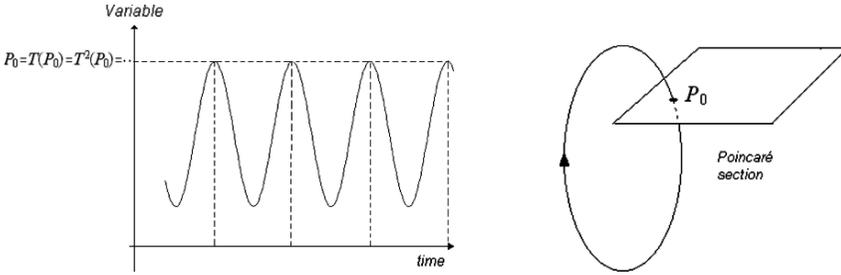
On the other hand, it is possible to say that P_0 *is a fixed point of the map T above.* This observation is very interesting and is in connection with the Poincaré section concept. This *property is so important that, starting from it, we can study the stability of the periodic solution of a system.* The *map T is studied in the neighborhood of P_0 by the Floquet matrix. Floquet matrix*[8] is

$$M = \left[\frac{\partial T}{\partial x_i} \right]_{x_i^0} \quad \text{with} \quad i = 1, 2. \tag{1.57}$$

We know that *after the action of the map T* (i.e. after one period T), the initial point P_0 plus a small variation (spacing, gap) denoted δ, i.e. $P_0 + \delta$, is near P_0. The spacing between the two points is written

$$T(P_0 + \delta) - P_0 \simeq M\delta \tag{1.58}$$

with $\|\delta\| \to 0$. *The characteristics of eigenvalues of M define the stability of solutions.* If the map T acts successively a number of times equal to p (i.e. after p time one period T), then it comes

$$T^p(P_0 + \delta) - P_0 \simeq M^p \delta. \tag{1.59}$$

Thus, the initial gap δ is multiplied by M^p. ($|\lambda_M| < 1$). Therefore: (1) If all eigenvalues of M have *a modulus lower than 1, then the gap decreases exponentially towards zero.*[9] *Then the periodic solution is linearly stable* because any deviation compared to the fixed point P_0 tends to decline. (2) Otherwise, i.e. if at least one eigenvalue of M has a modulus higher than 1, then the spacing exponentially increases with time, the limit cycle is (or periodic solution) is unstable. Thus the stability loss of the limit cycle corresponds to unit circle crossing by one (or several) eigenvalue(s) of the Floquet matrix.

---

[8] Stability is studied for very small perturbations, and a linear analysis limited to the terms of order 1 is considered sufficient.

[9] The modulus of a possible *complex eigenvalue* $\lambda_c = a \mp ib$ *is written:* $|\lambda_c| = \sqrt{a^2 + b^2}$ $= ((\text{real}(\lambda_c))^2 + (\text{imaginary}(\lambda_c))^2)^{1/2}$.

2nd Presentation

Let $\dot{x} = f(x)$ be a dynamical system where $x \in \mathbb{R}^n$, its flow is written $\phi(t,x)$. If we assume a periodic orbit of period T, we can write $\phi(t,x_0) = \phi(t+T,x_0)$, i.e. there is a return after one period to the starting point. We assume the existence of a "hyperplane", or a "hyper-surface" denoted Σ of dimension $n-1$, in $\mathbb{R}^{n-1}$, which is intersected by the trajectory of the flow of dimension n. Then, we can define a "map" associated with this plane of a dimension $n-1$, which exists through the periodicity T of the flow. We write

$$U \to \Sigma,$$
$$x \to \phi(\tau(x),x), \tag{1.60}$$

where τ is the "step" of the first return of the point x in the section Σ. If we linearize the initial system $\dot{x} = f(x)$ around the periodic solution as we have done previously, with $y = x - \bar{x}$, where $\bar{x}$ are the solutions of the system. Then, we write

$\dot{y} = \dfrac{\partial \dot{\bar{x}}}{\partial t}(x - \bar{x})$. At this stage, it is necessary to study the *stability* of solutions of the system linearized by *analyzing the eigenvalues of the matrix* e^{TR} which is called the *Floquet characteristic Multiplier*, because we remember that the solutions of the new linearized system are written

$$y = e^{TR}y_0, \tag{1.61}$$

where R is the matrix which determines the system of which it is necessary to study the eigenvalues to analyze the stability. Here we are thus brought back to the set of remarks about the behavior of eigenvalues on the unit circle.

1.7.1.2 Floquet Theory in Dissipative Systems

Given a *nonlinear dynamical system:* $\dot{x} = f(x)$ or its solution map $x_{n+1} = g(x_n)$. Suppose that *the solution of the system is periodic of period T*. This means that the point x_t after one period T *will be unchanged,* we can write: $x_{t+T} = x_t$. To *study* the *stability of this solution,* we can briefly say that it is necessary to observe, around the solution, the *behavior of a very small initial gap (spacing, variation)* that we call δx in relation to the solution of the system, i.e. with the limit-cycle in the phase space. The observation of δx *at the end of one period gives us indications about the stability:*

(1) If the variation *decreases* in all the directions of the space, we conclude to the *stability* of solutions.
(2) If *the variation increases in at least one direction*, we conclude to the *instability* of solutions.

If *we linearize the flow around the periodic solution in the phase space (to bring back to the solution),* we can write that the initial distance $(x_0 + \delta x)$ becomes after

one period T: $(x_0 + M\delta x)$, where M is the Floquet matrix. Thus, the study of the stability of the periodic solution will be done through eigenvalues of the matrix M. The eigenvalue of the Floquet matrix in our case is equal to 1 and this value corresponds to δx. This observation does not give us information about stability. The only information that we obtain is that at the end of one period, we logically come back to the same point. On "the Poincaré section" of the trajectory, one can find the variation δx at the same place in the plane. It is essential to study the stability to observe what occurs in the plane perpendicular to the trajectory. The eigenvalues of M depend of course on the solutions of the flow, i.e. the shape of the limit-cycle. The point or the set of points $(x_0 + \delta x)$ becomes after one period $(x_0 + M\delta x)$: (1) If the set of eigenvalues of M is located inside the unit circle of the complex plane, presented in Fig. 1.2, then the solution is linearly stable, and all the components of (the vector) δx perpendicular to the limit-cycle decreases regularly at each period. (2) If at least one of the eigenvalues of M is outside of the unit circle and δx increases regularly in at least one direction, this means that the trajectory moves away from the limit-cycle, there are divergence and thus instability. It is the case presented in Fig. 1.3, which depicts instability. M and its eigenvalues change while time evolves. It is when the eigenvalues leave the unit circle that there is instability and a *bifurcation* of the trajectory occurs. There are three ways for eigenvalues to leave the unit circle: $a \pm ib, -1, +1$. The consequence of these crossing types is the stability loss and the ulterior system behavior which depends on the nonlinearity type and bifurcation type.
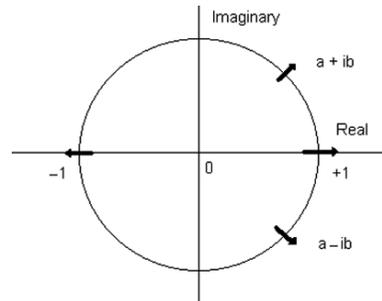
Fig. 1.2 Unit circle crossing by eigenvalues of M
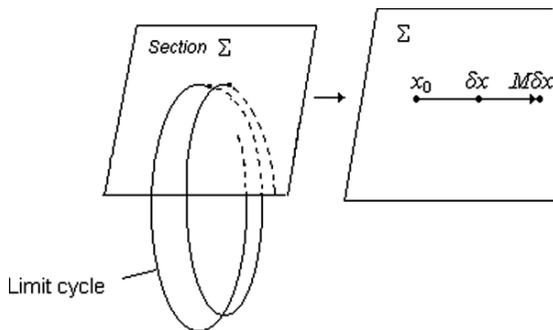
Fig. 1.3 Instability

1.7.2 Transitions Stemming from the Linear Stability Loss in Dissipative Systems

The study of eigenvalues of the matrix M $(Df(\bar{x}))$ allows a first approach of the bifurcation concept. For a fixed point $\bar{x}$, there are three ways of losing its stability:

(1) If the unit circle is crossed by $(+1)$ then occurs a *saddle-node* bifurcation. The solution is not only unstable but disappears completely, then just beyond the bifurcation threshold, the system generates a specific regime called type-1 inter-mittency, characterized by a mixing of *laminar phases* interrupted by apparently anarchical behaviors (*turbulence bursts*).

(2) If the unit circle is crossed by (-1), we have to analyze the normal or inverse character of the bifurcation that then occurs. (a) If we face a *subharmonic normal bifurcation*, a new stable periodic solution (whose period is twice long) appears. The repetition of this period-doubling for each periodic solutions gives rise to an infinite sequence of bifurcations, i.e. a *subharmonic cascade of bifurcations*, to finally produce chaos. (b) If we face an inverse bifurcation, this leads to a type-3 intermittency whose usual manifestations qualitatively look like those of the type-1 intermittency, i.e. long phases of almost periodic behaviors interrupted sporadically by chaotic bursts. However, the type-3 gives rise to a gradual increasing of the amplitude of the subharmonic during the almost periodic phase because the nonlinear effects amplify the subharmonic instability of the limit cycle. Consequently, the amplitude difference between successive oscillations will grow; Beyond a critical value, the laminar phase is interrupted.

(3) If the unit circle is simultaneously crossed by two complex conjugate eigen-values $a \pm ib$. (a) If this Hopf bifurcation is *supercritical* (normal), this leads to a stable attractor close to the limit cycle from now on unstable. This attrac-tor is a torus T^2 whose surface contains the new solution corresponding to a *quasiperiodic regime*.[10] A second supercritical Hopf bifurcation can generate then a transition $T^2 \rightarrow T^3$, and a strange attractor occurs after a third bifurcation (possibly confused with the second one). (b) If this is a *subharmonic* bifurca-tion, it is possible to meet another phenomenon: the type-2 intermittency. This intermittency has the same overall qualitative characteristics as those of type-1 and type-3 intermittencies, except that the instabilities that arise during the laminar phases have a frequency without a priori relationship with the basic frequency of the initial limit cycle. By contrast, this new frequency is linked to the ratio of the real and imaginary parts a/b of the eigenvalues $(a \pm ib)$ which crosses the unit circle at the bifurcation. Furthermore, the dynamical process which allows to restart a new laminar period after a turbulent burst is complex to describe in this case, because this requires to consider a flow in a

---

[10] **Definition (Quasiperiodic function).** A function $\psi: \mathbb{R} \rightarrow \mathbb{R}$ is quasiperiodic if it can be written $\psi(t) = \Phi(\omega_1 t, \dots, \omega_n t)$ (ω_i: frequencies) where Φ is periodic of period 2π in each of its arguments, and two or more of the n frequencies are incommensurable, i.e. with irrational ratios.

phase space of dimension 6 or a Poincaré section in $\mathbb{R}^5$. Here is a summary of scenarios:

Crossing	Bifurcation		Phenomena
$(+1)$	Node-saddle		Type-1 Intermittency
(-1)	Subharmonic	Normal	Subharmonic cascade
		Inverse	Type-3 Intermittency
$(a \pm ib)$	Hopf	Normal	Quasiperiodicity
		Inverse	Type-2 Intermittency

1.8 The Bifurcation Concept

The subject corresponds to a vast theory. Although it is difficult to give a general definition of a bifurcation, it is possible to say that bifurcations are commonly associated with the *topology change* notion *of the trajectory of dynamical systems,* when a parameter or several parameters are modified. Mathematically the bifurcation point is defined as opposed to the concept of "topological equivalence" under the action of a homeomorphism. We can trivially define a bifurcation by saying that it is a behavior change of a system, while when modifying the parameter which predetermines it, we reach a critical point during which we pass from stable solutions to unstable solutions. The value of the parameter which separates the two types of solution is called the critical value, critical point, or bifurcation point. Hereafter, we restrict the analysis schematically to different types of one parameter bifurcations.

1.8.1 Codimension-1 Bifurcations of Fixed Points

The normal forms correspond to *supercritical* situations:

> Normal form: Bifurcation type:
> *(a):* $\dot{x} = \mu - x^2$ Node-Saddle,
> *(b):* $\dot{x} = \mu x - x^2$ Transcritical (stability exchange),
> *(c):* $\dot{x} = \mu x - x^3$ Pitchfork,
> *(d):* $\dot{z} = (\mu + i\gamma)z - z|z|^2$ Hopf,

where x is a real variable, z is a complex variable, μ is a parameter measuring the threshold gap, i is a pure imaginary, γ is an arbitrary constant.

(a) For the *node-saddle* bifurcation, the stationary solution is $x = \pm\sqrt{\mu}$ (defined only for $\mu > 0$) and appears thus at $\mu = 0$. There is no stable or unstable solution for $\mu = 0$. (b) For the *transcritical* bifurcation (stability exchange), two stationary solutions $x = 0$ and $x = \mu$ coexit. $x = 0$ is stable if $\mu < 0$ and unstable if $\mu > 0$, whereas it is the opposite for $\mu = 0$. Thus, there is stability exchange at the

bifurcation point. (c) For the *pitchfork bifurcation*, the stationary solutions are $x = 0$ and $x = \pm\sqrt{\mu}$ (defined only for $\mu > 0$). This normal form is invariant by the transformation $x \to (-x)$. Each time we meet this problem of insensitivity to a reflection symmetry (e.g. first convective instability of the Benard experiment), we face this normal form. (d) For the *Hopf bifurcation* (i.e. complex equivalent of the pitchfork bifurcation). It is convenient for example to pose $z = x + iy$, the normal formal is written: $\dot{x} = [\mu + (x^2 + y^2)]x - \gamma y$, $\dot{y} = \gamma x + [\mu - (x^2 + y^2)]y$. The stationary solutions are $z = 0$ ($x = y = 0$) and also $|\dot{z}|^2 = x^2 + y^2 = \mu$. This condition defines the equation of a circle in the plane (x, y) whose radius is $\sqrt{\mu}$ (Fig. 1.4).

In the Hopf bifurcation diagram, we observe that the bifurcation point precedes a limit-cycle. But this bifurcation is characterized by an essential property, i.e. the amplitude of the cycle is proportional to $|\mu - \mu_{Critical}|^{1/2}$. Such a characteristic is useful to identify this type of bifurcation. The Hopf bifurcation was studied very early by Poincaré. A way of studying this type of bifurcation is to pose the dynamical system of Van der Pol (self-sustained oscillator) written $\ddot{x} + x = \mu(1 - x^2)\dot{x}$ which is typical of such phenomena. (Remark: at $\mu = 0$ we deal with supercritical pitchfork bifurcation.)

stable ⎤
           ⎬ fixed points
unstable ⎦
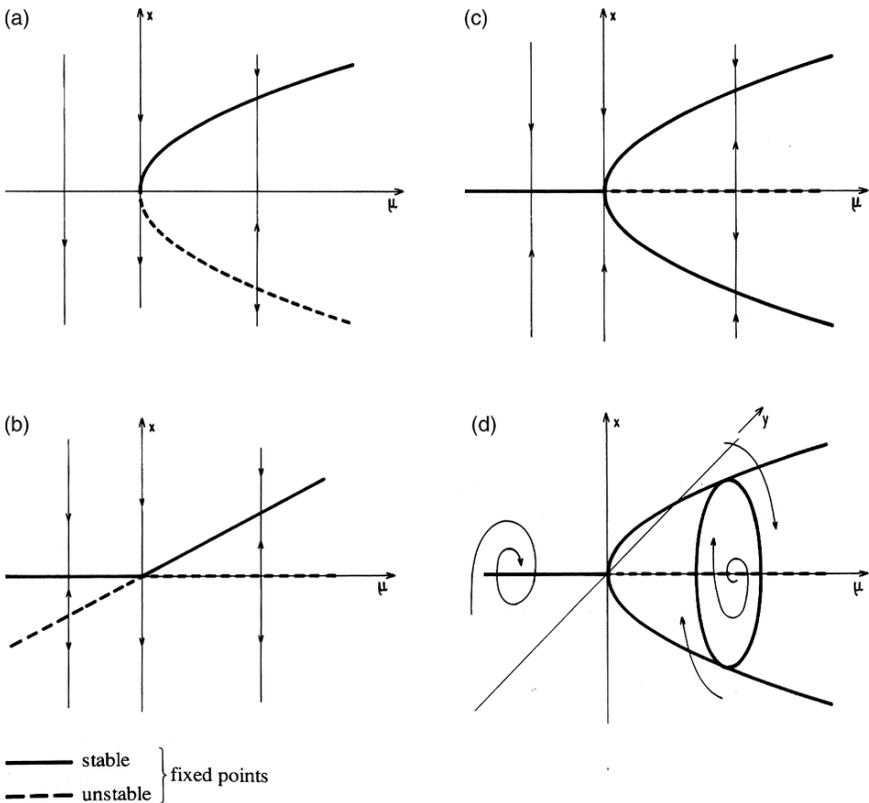
Fig. 1.4 Diagram of codimension-1 bifurcations (normal form)
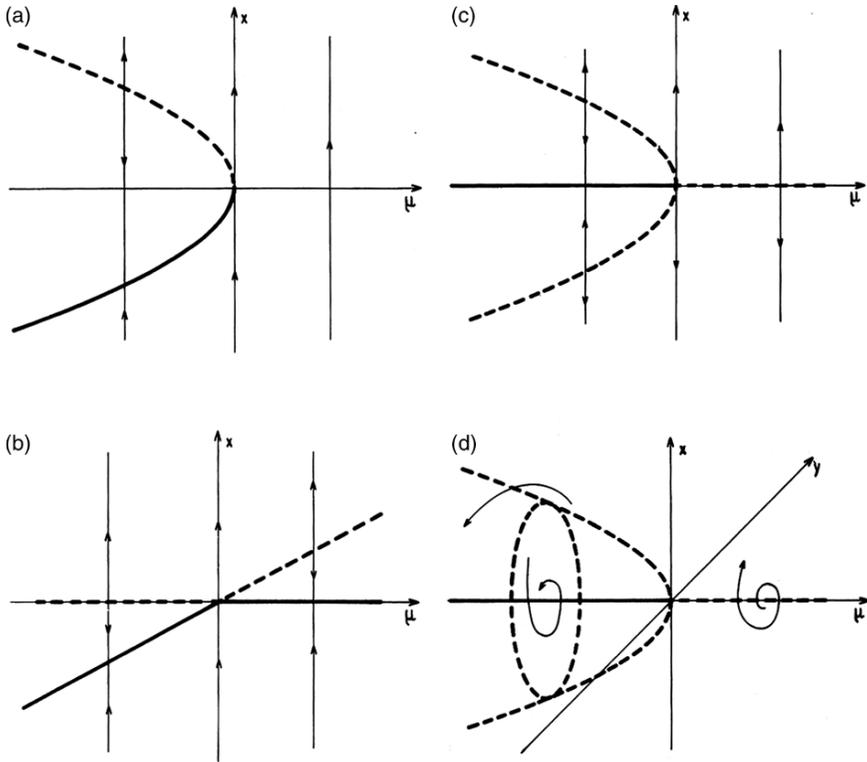
Fig. 1.5 Diagram of codimension-1 subcritical (or inverse) bifurcations. **(a)** $\dot{x} = \mu + x^2$, **(b)** $\dot{x} = \mu x + x^2$, **(c)** $\dot{x} = \mu x + x^3$, **(d)** $\dot{z} = (\mu + i\gamma)z + z|z|^2$

1.8.2 Subcritical Bifurcations of Fixed Points

As explained in the previous section, the normal forms correspond to *supercritical* situations (also named normal). That means that the nonlinear terms x^2 or x^3 have an effect opposite to that of the instability generated by the term of lower order (e.g. $\dot{x} = \mu x - x^3$). By writing this equation as follows $\dot{x} = \mu x(1 - (x^2/\mu))$, we observe easily that the nonlinear term "saturates" the effect of the linear instability for $x^2 = \mu$ (see Fig. 1.5).

1.8.3 Codimension-1 Bifurcations of Periodic Orbits

At least locally, it is possible to wonder if the bifurcation notion of fixed points can be extended to periodic orbits. In particular because by using a Poincaré section, a periodic orbit is observed as a fixed point. We understand that mathematically the

problem is difficult to treat (because to return to the cut plane, it requires the integration of equations along the orbit). Fortunately, the Floquet theory allows to avoid this difficulty and to study the linear stability of the periodic orbit without being obliged to follow the flow step by step. In such a case, there are three possibilities according the manner whose eigenvalue of the Floquet matrix crosses the unit circle of the complex plane. (1) If the crossing is carried out by $+1$, the theory is not radically different from that of fixed point bifurcations. The normal form, topologically equivalent to a node-saddle bifurcation is now written: $\dot{x} = x + \mu - x^2$, the term x is introduced by the eigenvalue $+1$. Similarly, there are normal forms which lead to pitchfork bifurcations or stability exchanges according to the considered problem. (2) If the loss of linear stability results from an eigenvalue -1, there is no equivalence in fixed point bifurcations of the flow. In such a case, the bifurcation generates the occurrence of a double periodic orbit by the *subharmonic instability*. The bifurcation diagram is similar to that of a pitchfork bifurcation (see Fig. 1.5) but the dynamics the radically different. Indeed, after the *subharmonic bifurcation*, the two branches are continuously visited one after another by the (periodic) solution. Whereas in a pitchfork bifurcation, one and only one stationary solution is established, after the adoption of one or other of the two branches of a pitchfork bifurcation. (3) If the loss of stability results from the crossing of the unit circle by two complex conjugate eigenvalues, something similar to a Hopf bifurcation occurs, and is often called like this in spite of its specific characteristics. It follows that in the Poincaré section, the fixed point is replaced by a set of points on a curve and conjures up a limit cycle (in a certain way). However, the analysis needs to be refined because this curve (that is not an orbit) is covered in a very specific way. This is here that the notion of rotation number appears with its rational or irrational characteristics, in connection with the quasiperiodicity notion and phenomena of frequency coupling.

1.9 Hopf Bifurcation

1.9.1 Codimension-1 Hopf Bifurcation

Let us describe more precisely the Hopf bifurcation by using vector fields. Beforehand, consider the following notations and definitions: Dynamical models usually consist of a *family* of *differential equations* (or also *diffeomorphisms*). Given t a continuous time variable, dynamical systems are written by a differential equation as follows: $d\mathbf{x}/dt = \dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$. $\mathbf{X}$ is a m-parameter, C^r family, of vector fields and $\mathbf{x}$ represents the state of the system and takes values in the state or phase space. The family $\mathbf{X}$ is said to have a bifurcation point at $\mu = \mu^*$ if, in every neighborhood of μ^*, there exist values of μ such that the corresponding vector fields $\mathbf{X}(\mu, \cdot) = \mathbf{X}_\mu(\cdot)$ exhibit topologically distinct behavior.[11] In addition, the partial derivatives are denoted $\mathbf{D}$.

---

[11] For a family of diffeomorphisms we could equivalently write $\mathbf{f}(\mu, \cdot) = \mathbf{f}_\mu(\cdot)$.

Definition 1.9 (Local family). A local family $\mathbf{X} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ is induced by the family $\mathbf{Y} : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^n$ by means of a continuous map $\varphi : \mathbb{R}^m \to \mathbb{R}^s$, $\varphi(0) = 0$, if $\mathbf{X}(\mu, \mathbf{x}) = \mathbf{Y}(\varphi(\mu), \mathbf{x})$.

Definition 1.10 (Unfolding). Any local family, $\mathbf{X}(\mu, \mathbf{x})$, at $(\mathbf{0}, \mathbf{0})$ is said to be an unfolding of the vector field $\mathbf{X}(\mathbf{0}, \mathbf{x}) = \mathbf{X}_0(\mathbf{x})$. When $\mathbf{X}_0(\mathbf{x})$ has a singularity at $\mathbf{x} = \mathbf{0}$, $\mathbf{X}(\mu, \mathbf{x})$ is referred to as an unfolding of the singularity.[12]

Definition 1.11 (Versal unfolding). A given m-parameter local family $\mathbf{X}(\gamma, \mathbf{x})$ on $\mathbb{R}^n$ is said to be a versal unfolding of the vector field $\mathbf{X}(\mathbf{0}, \mathbf{x}) = \mathbf{X}_0(\mathbf{x})$ if every other unfolding is equivalent to one induced by the given family.

Hopf bifurcation[13] (In the case of codimension-1 singularity). Let $\mathbf{X}_0$ be a smooth vector field with a Hopf singularity at the origin. Then $tr\, \mathbf{DX}_0(\mathbf{0}) = 0$ and the normal form[14] of $\mathbf{X}_0$ can be given by

$$\begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1^2 + x_2^2) \left\{ a_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b_1 \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \right\} + O(|\mathbf{x}|^5), \qquad (1.62)$$

where $\alpha = (\text{Det}\, \mathbf{DX}_0(\mathbf{0}))^{1/2} > 0$ and $a_1 \neq 0$. Consider now the local family

$$\mathbf{X}(\mu, \mathbf{x}) = \begin{pmatrix} \mu & -\alpha \\ \alpha & \mu \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1^2 + x_2^2) \left\{ a_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b_1 \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \right\} + O(|\mathbf{x}|^5) \quad (1.63)$$

is a versal unfolding of the Hopf singularity. Suppose now that the terms of order $|\mathbf{x}|^5$ are absent (or removed) then the family above is more malleable in the plane polar coordinates (r, θ). It comes: $\dot{r} = r(\mu + a_1 r^2)$, $\dot{\theta} = \alpha + b_1 r^2$.

- Suppose $a_1 < 0$; the phase portrait of $\dot{r}$ and $\dot{\theta}$ consists of a hyperbolic, stable focus at the origin. If $\mu = 0$, $\dot{r} = a_1 r^3$ and the origin is still asymptotically stable, even if it is no longer hyperbolic. For $\mu > 0$, $\dot{r} = 0$ for $r = (\mu / |a_1|)^{1/2}$ or for $r = 0$. Then we have for $\mu > 0$ a stable limit cycle, of radius proportional to $\mu^{1/2}$, which surrounds a hyperbolic, unstable focus at the origin. This is a *supercritical* Hopf bifurcation (see Fig. 1.6a).
- Suppose $a_1 > 0$, the limit cycle occurs for $\mu < 0$, it is unstable and surrounds a stable fixed point. When μ increases, the radius of the cycle decreases to zero at $\mu = 0$, where the fixed point at the origin becomes a (weakly) unstable focus. If $\mu > 0$, $(x_1, x_2)^T = \mathbf{0}$ is unstable and hyperbolic. This is a *subcritical* Hopf (see Fig. 1.6b). The presence or absence of the terms $O(|\mathbf{x}|^5)$ does not affect the characteristics of diagrams.

---

[12] See Arnold (1983).

[13] Hassard et al. (1981); Marsden and MacCracken (1976).

[14] **Normal Form Theorem:** Given a smooth vector field $\mathbf{X}(\mathbf{x})$ on $\mathbb{R}^n$ with $\mathbf{X}(\mathbf{0}) = \mathbf{0}$, there is a polynomial transformation to new coordinates $\mathbf{y}$ such that the differential equation $\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$ takes the form $\dot{\mathbf{y}} = \mathbf{J}\mathbf{y} + \sum_{r=2}^{N} \mathbf{w}_r(\mathbf{y}) + O(|\mathbf{y}|^{N+1})$, where $\mathbf{J}$ is the real Jordan form of $\mathbf{A} = \mathbf{DX}(\mathbf{0})$ and $\mathbf{w}_r$ a complementary subspace.
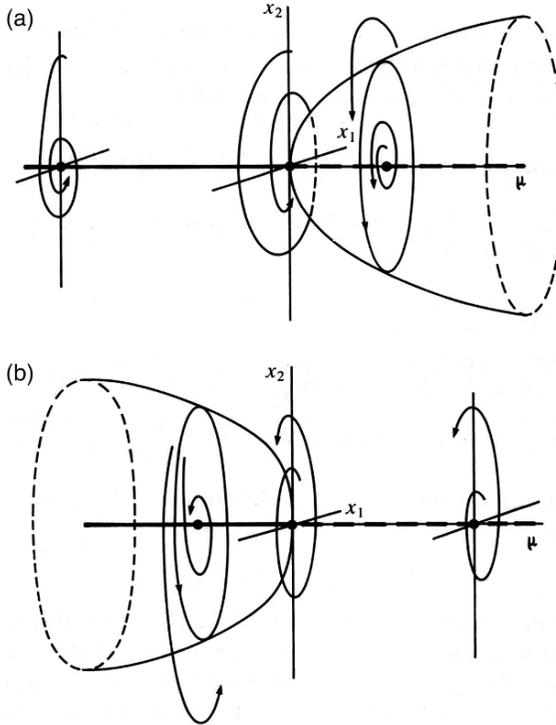
Fig. 1.6 Hopf bifurcations: (**a**) supercritical ($a_1 < 0$); (**b**) subcritical ($a_1 > 0$)

Given $\mathbf{Y}(\upsilon,\mathbf{x})$, $\upsilon \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^2$ a family of vector fields. Then a Hopf bifurcation occurs at $(\upsilon^*,\mathbf{x}^*)$ when $tr\,\mathbf{DY}_\upsilon(\mathbf{x}^*)$ passes through zero at $\upsilon = \upsilon^*$, which is thus associated with a change of stability of the fixed point $\mathbf{x}^*$. The following theorem provides the conditions for which the family $\mathbf{Y}(\upsilon,\mathbf{x})$ generates a complete Hopf bifurcation $\big($with $\upsilon \in \upsilon$, i.e. υ is one of υ_j, $j = 1,\ldots,m$ and $\mathbf{Y}(\upsilon,\mathbf{x}) = \mathbf{Y}(\upsilon,\mathbf{x})\big|_{\upsilon_k = \upsilon_k^*\ k \neq j}\big)$:

Theorem 1.2 (Hopf bifurcation theorem). *Given $\dot{\mathbf{x}} = \mathbf{Y}(\upsilon,\mathbf{x})$ a parameterized dynamical system, with $\mathbf{x} \in \mathbb{R}^2$ and $\upsilon \in \mathbb{R}$, which has a fixed point at the origin for all values of the real parameter υ. In addition, suppose the eigenvalues, $\lambda_1(\upsilon)$ and $\lambda_2(\upsilon)$, of $\mathbf{DY}(\upsilon,\mathbf{0})$ are pure imaginary for $\upsilon = \upsilon^*$. If the real part of the eigenvalues, $\mathrm{Re}\,\lambda_1(\upsilon)(=\mathrm{Re}\,\lambda_2(\upsilon))$, verifies*

$$\frac{d}{d\upsilon}\,\{\mathrm{Re}\,\lambda_1(\upsilon)\}\big|_{\upsilon=\upsilon^*} > 0 \tag{1.64}$$

and the origin is an asymptotically stable fixed point when $\upsilon = \upsilon^*$ then: (1) $\upsilon = \upsilon^*$ is a bifurcation point of the system; (2) for $\upsilon \in (\upsilon_1,\upsilon^*)$, some $\upsilon_1 < \upsilon^*$, the origin is a stable focus; (3) for $\upsilon \in (\upsilon^*,\upsilon_2)$, some $\upsilon_2 > \upsilon^*$, the origin is an unstable focus surrounded by a stable limit cycle, whose size increases with υ.

1.9.2 Cusp and Generalized Hopf Bifurcations

Here the bifurcations concern singularities with codimension greater than 1. These are only unfolded by local families that depend on more than one parameter.

1.9.2.1 Cusp Bifurcation

Given $\mathbf{X_0}$ a smooth vector field with a *cusp singularity* at the origin. Then $\mathbf{DX_0}(\mathbf{0}) \neq 0$ but $tr\, \mathbf{DX_0}(\mathbf{0}) = \det \mathbf{DX_0}(\mathbf{0}) = 0$ and $\mathbf{X_0}$ has normal form (with $a_2 \neq 0$ and $b_2 \neq 0$).

$$\begin{pmatrix} x_2 + a_2 x_1^2 \\ b_2 x_1^2 \end{pmatrix} + O(|\mathbf{x}|^3). \tag{1.65}$$

The local family

$$\mathbf{X}(\mu, \mathbf{x}) = \begin{pmatrix} x_2 + \mu_2 x_1 + a_2 x_1^2 \\ \mu_1 + b_2 x_1^2 \end{pmatrix} + O(|\mathbf{x}|^3) \tag{1.66}$$

is a versal unfolding of the cusp singularity above. (Note that the cusp singularity requires at least two parameters to completely unfold it.) Considering this local family $\mathbf{X}(\mu, \mathbf{x})$, under some conditions (not shown here) in particular about parameters, it is possible to generate three bifurcation types: (1) *Saddle node*, (2) *Hopf bifurcation*, (3) *Saddle connection*.

1.9.2.2 Generalized Hopf Bifurcations

Generalized Hopf bifurcations. (Takens 1973b, Guckenheimer and Holmes 1983, Arrowsmith and Place 1990). Here, we describe the *generalized Hopf singularity* of type p. The vector field in the previous codimension-1 Hopf bifurcation was obtained by posing $a_1 \neq 0$ in the normal form

$$\begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sum_{k=1}^{\infty} (x_1^2 + x_2^2)^k \left[a_k \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b_k \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \right]. \tag{1.67}$$

If a degeneracy is introduced by posing $a_1 = a_2 = \cdots = a_{p-1} = 0$ with $a_p \neq 0$, then the system above is said to have a *generalized Hopf singularity* of type p at $(x_1, x_2)^T = \mathbf{0}$. (Note that a codimension-1 Hopf singularity is said to be of type 1.) *The framework of the generalized Hopf singularity highlights the relationship between the degeneracy of a singularity and the number of parameters contained in its versal unfolding.*

   If a Hopf singularity of type p occurs, it is possible to reduce the system above to a simpler system. Given

$$\phi(x_1, x_2) = \left(\alpha + \sum_{k=1}^{\infty} b_k (x_1^2 + x_2^2)^k \right) \tag{1.68}$$

that can also be written

$$\phi(x_1,x_2)\left[\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sum_{k=1}^{\infty} \gamma_k(x_1^2+x_2^2)^k\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right], \tag{1.69}$$

with $\gamma_1 = a_1/\alpha$. Because $\alpha > 0$, we have $\phi(x_1,x_2) > 0$ for all (x_1,x_2) enough close to the origin. Thus there exists a neighborhood of $(x_1,x_2)^T = \mathbf{0}$ on which the system is topologically equivalent to the (simpler) *vector field:*

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sum_{k=1}^{\infty} \gamma_k(x_1^2+x_2^2)^k\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{1.70}$$

While preserving the topological equivalence, we can replace γ_p by $sgn(\gamma_p) = sgn(a_p)$ in the vector field above. Two cases are then possible according to a singularity of the type: $(p,+)$ or $(p,-)$. A versal unfolding for the vector field above can be obtained through the following Takens theorem (1973b):

Theorem 1.3. *The versal unfolding of a generalized Hopf singularity of type p given in the simpler vector field above is written (where $s_i \in R$, $i = 0,1,\ldots,p-1$):*

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sum_{k=1}^{\infty} \gamma_k(x_1^2+x_2^2)^k\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sum_{i=0}^{p-1} s_i(x_1^2+x_2^2)^i\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{1.71}$$

According to the sign of γ_p, the unfoldings of in this theorem gives two cases $(p,+)$ and $(p,-)$. Here, we only describe the first case $(p,+)$, and the latter case can be related to the other by a time reversal (so the stabilities of fixed points and limit cycles are inverted). The formula in this theorem, through the *polar form* $((\dot{\theta},\dot{r})$ with here $\dot{\theta} = 1$), gives for $(p,+)$ the bifurcational behavior by using the equivalent equation: $\dot{r} = \sum_{i=0}^{p-1} s_i r^{2i+1} + r^{2p+1} + O(r^{2p+3})$. The behavior is independent of the terms of order r^{2p+3} (as described for the codimension-1 Hopf bifurcation), so it suffices to consider $\dot{r} = \sum_{i=0}^{p-1} s_i r^{2i+1} + r^{2p+1}$. This equation in particular provides limit cycles. (1) For $p = 1$, by posing $r^2 = \lambda$, we obtain $\lambda^2 + s_0 = 0$, then a limit cycle occurs whose radius is $(-s_0)^{1/2}$ only when $s_0 < 0$. (2) For $p = 2$, by posing $r^2 = \lambda$, we obtain $\lambda^2 + s_1\lambda + s_0 = 0$ and non-trivial zeros result from $\lambda = (\pm(s_1^2 - 4s_0)^{1/2} - s_1)/2$. Then we obtain (a) If $s_0 > 0$ and $s_1 > 0$ (or $s_1^2 - 4s_0 < 0$): no limit cycle but a repelling spiral. (b) If $s_1 < 0$ and $s_1^2 - 4s_0 > 0$: two limit cycles. (c) If $s_0 < 0$: one limit cycle (see Fig. 1.7).

Figure 1.7 shows the bifurcation diagram for the type-$(2,+)$ Hopf bifurcation. It shows that a supercritical (subcritical) Hopf bifurcation occurs when s_0 increases through zero with $s_1 < 0$ (subcritical with $s_1 > 0$), and a double limit cycle bifurcation occurs on the semi-parabola. If s_0 decreases through the semi-parabola, a non-hyperbolic limit cycle occurs and then divides into two hyperbolic cycles.
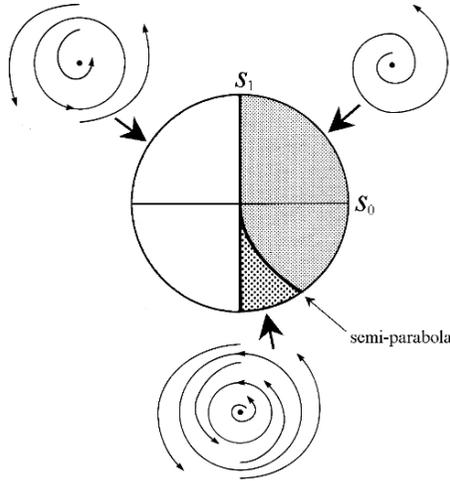
Fig. 1.7 Bifurcation diagram for the type-(2, +) Hopf bifurcation

1.10 Examples of Dynamical System Resolution

1.10.1 A Stable System

$$\dot{x}_1 = -0,5x_1 + x_2, \tag{1.72}$$
$$\dot{x}_2 = -2x_2. \tag{1.73}$$

The *Jacobian matrix* of the system $\dot{x} = Ax$ is $A = \begin{bmatrix} -0.5 & 1 \\ 0 & -2 \end{bmatrix}$. The eigenvalues are solutions of $\det(\lambda I - A) = 0$:

$$\det \begin{bmatrix} \lambda + 0.5 & -1 \\ 0 & \lambda + 2 \end{bmatrix} = (\lambda + 0.5)(\lambda + 2) = 0 \tag{1.74}$$

thus: $\lambda_1 = -0.5$, $\lambda_2 = -2$. Then, the eigenvectors are written: $\xi_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\xi_2 = \begin{bmatrix} -0.5547 \\ 0.8321 \end{bmatrix}$. Note that ξ_1 is the *"slow"* subspace, since it corresponds to an eigenvalue $\lambda_1 = -0.5$, and ξ_2 is *the "fast"* subspace since it corresponds to $\lambda_2 = -2$. Consequently, we can write: $x(t) = Ve^{\Lambda t}V^{-1}x(0)$ as follows:

$$x(t) = \begin{bmatrix} 1 & -0.5547 \\ 0 & 0.8321 \end{bmatrix} \begin{bmatrix} e^{-0.5t} & 0 \\ 0 & e^{-2t} \end{bmatrix} \begin{bmatrix} 1 & 0.6667 \\ 0 & 1.2019 \end{bmatrix} x(0) \tag{1.75}$$

(a) If the initial condition is $x(0) = \xi_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, then the solutions are $x(t) = \begin{bmatrix} e^{-0.5t} \\ 0 \end{bmatrix}$.

(b) If the initial condition is $x(0) = \xi_2 = \begin{bmatrix} -0.5547 \\ 0.8321 \end{bmatrix}$, then the solutions are $x(t) = \begin{bmatrix} -0.5547e^{-2t} \\ 0.8321e^{-2t} \end{bmatrix}$. Note that $x(0) = \xi_1$ is the slow initial condition and $x(0) = \xi_2$ is the fast initial condition. In other words, in this system, the response of the model for an initial condition, $x(0) = \xi_1$, will first be a transitory motion before an asymptotic stabilization.

1.10.2 An Unstable System with a Saddle Point

$$\dot{x}_1 = 2x_1 + x_2, \tag{1.76}$$
$$\dot{x}_2 = 2x_1 - x_2. \tag{1.77}$$

The Jacobian matrix of the system

$$\dot{x} = Ax, \tag{1.78}$$

can be written

$$A = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}. \tag{1.79}$$

The eigenvalues are the solutions of

$$\det(\lambda I - A) = 0, \tag{1.80}$$

that give

$$\det \begin{bmatrix} \lambda - 2 & -1 \\ -2 & \lambda + 1 \end{bmatrix} = (\lambda + 1.5616)(\lambda - 2.5616) = 0. \tag{1.81}$$

Thus:
$$\lambda_1 = -1.5616 \text{ and } \lambda_2 = 2.5616. \tag{1.82}$$

Then, the eigenvectors are written

$$\xi_1 = \begin{bmatrix} 0.2703 \\ -0.9628 \end{bmatrix}, \qquad \xi_2 = \begin{bmatrix} 0.8719 \\ 0.4896 \end{bmatrix}. \tag{1.83}$$

Since $\lambda_1 < 0$, ξ_1 is a stable subspace; since it corresponds to an eigenvalue $\lambda_2 > 0$, ξ_2 is an unstable subspace. Consequently, we can write the following equation:

$$x(t) = Ve^{\Lambda t}V^{-1}x(0) \tag{1.84}$$

as follows:

$$x(t) = \begin{bmatrix} 0.2703 & 0.8719 \\ -0.9628 & 0.4896 \end{bmatrix} \begin{bmatrix} e^{-1.5616t} & 0 \\ 0 & e^{-2.5616t} \end{bmatrix} \begin{bmatrix} 0.5038 & 0.8972 \\ 0.9907 & 0.2782 \end{bmatrix} x(0)$$

(a) If the initial condition is $x(0) = \xi_1 = \begin{bmatrix} 0.2703 \\ -0.9628 \end{bmatrix}$, then the solutions are

$$x(t) = \begin{bmatrix} 0.2703e^{-1.5616t} \\ -0.9628e^{-1.5616t} \end{bmatrix}.$$

(b) If the initial condition are $x(0) = \xi_2 = \begin{bmatrix} 0.8719 \\ 0.4896 \end{bmatrix}$, then the solutions are

$$x(t) = \begin{bmatrix} 0.8719e^{2.5616t} \\ 0.4896e^{-2.5616t} \end{bmatrix}.$$

It will be also noted that $x(0) = \xi_1$ is the slow initial condition, and $x(0) = \xi_2$ is the fast initial condition. In other words, in this system the response of the model for an initial condition $x(0) = \xi_1$ will first be a transitory motion before an asymptotic stabilization.

1.11 Typology of Second-Order Linear Systems

We have to solve the generic system $\dot{x} = Ax$, where the *Jacobian matrix* is $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. The *eigenvalues* are obtained by solving: $det(\lambda I - A) = (\lambda - a_{11})$
$(\lambda - a_{22}) - a_{12}a_{21} = 0$, also written $det(\lambda I - A) = \lambda^2 - tr(A)\lambda + det(A) = 0$. The quadratic formula is used to find the eigenvalues $\lambda = \dfrac{tr(A) \pm \sqrt{(tr(A)^2 - 4\det(A))}}{2}$, or expressed separately:

$$\lambda_1 = \frac{tr(A) - \sqrt{(tr(A)^2 - 4\det(A))}}{2}, \quad \lambda_2 = \frac{tr(A) + \sqrt{(tr(A)^2 - 4\det(A))}}{2}.$$

If the trace $tr(A)$ is negative, then we will have at least one negative eigenvalue. It will be also noted that the eigenvalue will be complex if $4\,det(A) > tr(A)^2$.
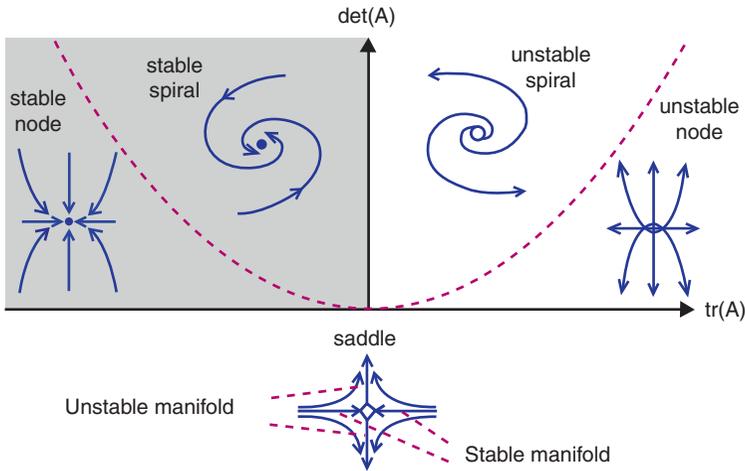
Fig. 1.8 Behaviors of dynamical systems

1.11.1 Eigenvalues Interpretation

The behavior of the system results from its eigenvalues λ_1 and λ_2 (Fig. 1.8):

Sinks (Stable nodes)   : $Real(\lambda_1) < 0$ and $Real(\lambda_2) < 0$
Saddles (Unstable point) : $Real(\lambda_1) < 0$ and $Real(\lambda_2) > 0$
Sources (Unstable nodes) : $Real(\lambda_1) > 0$ and $Real(\lambda_2) > 0$
Spirals   : λ_1 and λ_2 are complex conjugates.
                                              If $Real(\lambda_1) < 0$, then there is stability,
                                              if $Real(\lambda_1) > 0$, there is instability.

1.11.2 Some Representations in the Phase-Plane

1. *A stable equilibrium point: The node-sink (Fig. 1.9).* Given the system:

$$\dot{x}_1 = -x_1, \quad \dot{x}_2 = -4x_2. \tag{1.85}$$

Solutions and representation in the phase-plane:

$$x_1(t) = x_1(0)e^{-t}, \quad x_2(t) = x_2(0)e^{-4t}. \tag{1.86}$$

2. *An unstable equilibrium: the saddle-point (Fig. 1.10).* Given the system:

$$\dot{x}_1 = -x_1, \quad \dot{x}_2 = 4x_2. \tag{1.87}$$

Fig. 1.9 Node attractor, node-sink

Node-sink
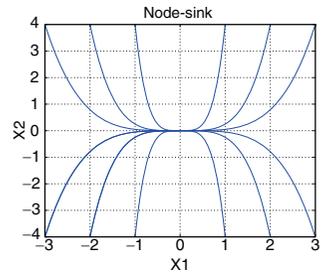
Fig. 1.10 Saddle-point with the flow orientation
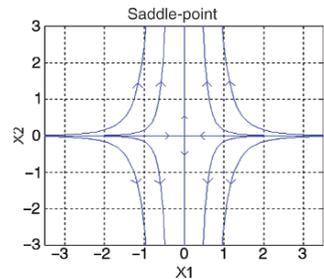
Saddle-point

Fig. 1.11 Another saddle point with flow orientation
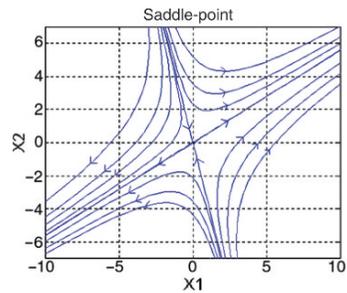
Saddle-point

Solutions and representation in the phase-plane:

$$x_1(t) = x_1(0)e^{-t}, \quad x_2(t) = x_2(0)e^{4t}. \tag{1.88}$$

3. *Another unstable solution: another saddle-point (Fig. 1.11).* Given the system: $\dot{x}_1 = 2x_1 + x_2, \dot{x}_2 = 2x_1 - x_2$. The Jacobian matrix is written $A = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}$. See the representation in the phase plane.

4. An *unstable zone: the spiral source (Fig. 1.12).* Given the system: $\dot{x}_1 = x_1 + 2x_2$, $\dot{x}_2 = -2x_1 + x_2$. Jacobian matrix is written $A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$. See phase plane.

5. *The center (Fig. 1.13):* Given the system: $\dot{x}_1 = -x_1 - x_2, \dot{x}_2 = 4x_1 + x_2$. The Jacobian matrix is written $A = \begin{bmatrix} -1 & -1 \\ 4 & 1 \end{bmatrix}$. See phase plane.

Fig. 1.12 Spiral source
(repelling spiral) with the
flow orientation
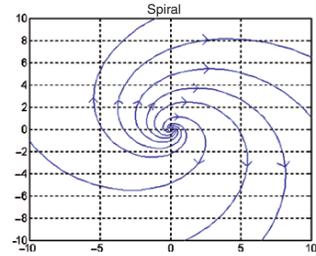
Fig. 1.13 Node center with
the flow orientation
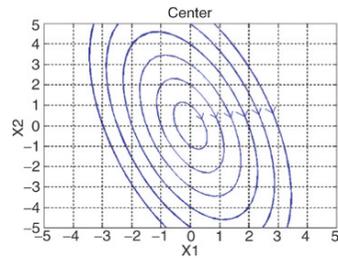
1.11.3 Behavior Summary of Second-Order Linear Systems

In order to identify the behavior of a system of second-order ordinary differential
equations as a function of the trace and the determinant of the Jacobian matrix of
the aforesaid system, we can refer to the following central graph (Fig. 1.14). In the
central diagram, in abscissa $tr(A)$ is replaced by p and in ordinate $\det(A)$ is replaced
by q. Thus, if $\triangle = p^2 - 4q = 0$; $p \neq 0$: the fixed point is a node, it is asymptotically
stable for $p < 0$, and unstable for $p > 0$. Note the Jacobian matrix of the node center
computed above $A = \begin{bmatrix} -1 & -1 \\ 4 & 1 \end{bmatrix}$, we observe that the Jacobian Matrix has a trace
equal to 0 and a determinant equal to 3, which we localize easily in the graph.

1.11.3.1 Example of Eigenvalue Positions on the Unit Circle

See Figs. 1.15–1.20.
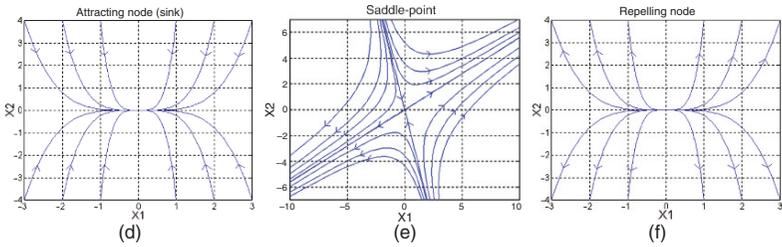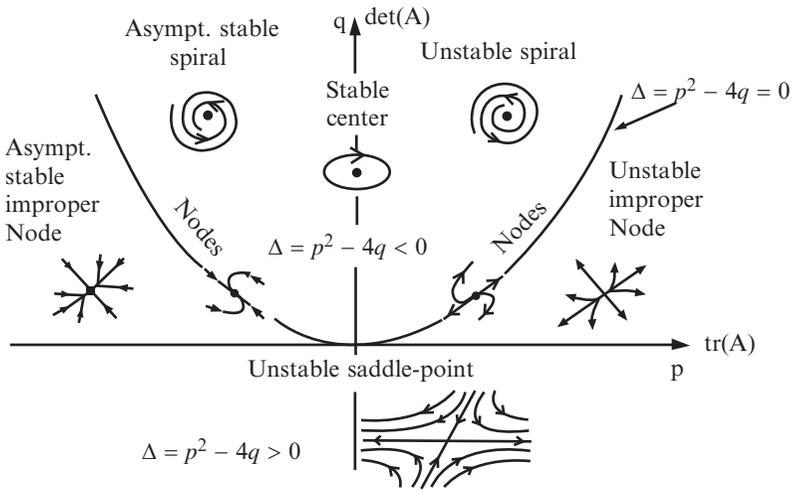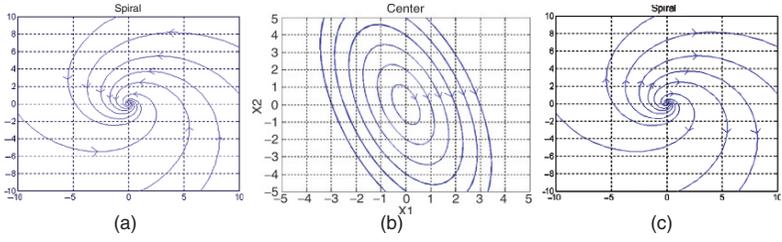
Fig. 1.14 (**a**) Spiral sink, (**b**) node center, (**c**) repelling spiral, (**d**) attracting node sink, (**e**) saddle point, and (**f**) repelling node
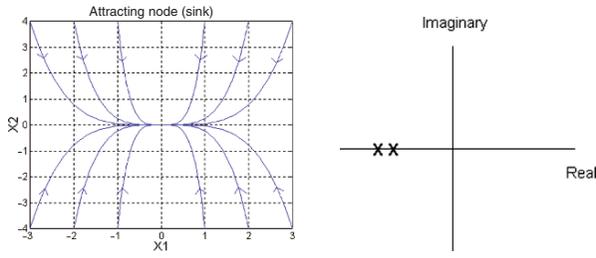
Fig. 1.15 Attracting node sink (*left*), eigenvalues (*right*)
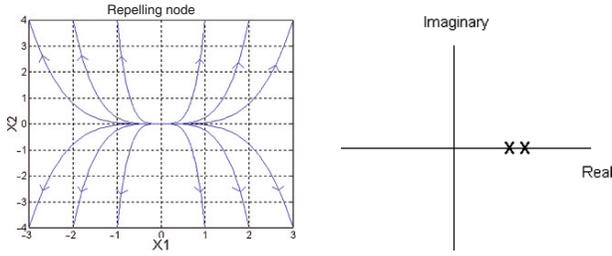
Fig. 1.16 Repelling node
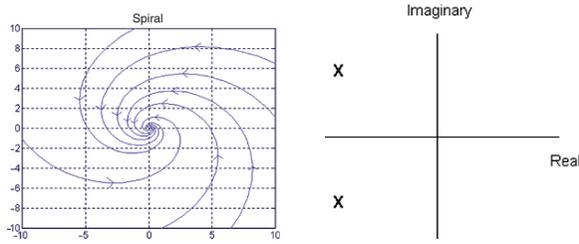
Fig. 1.17 Attracting spiral sink
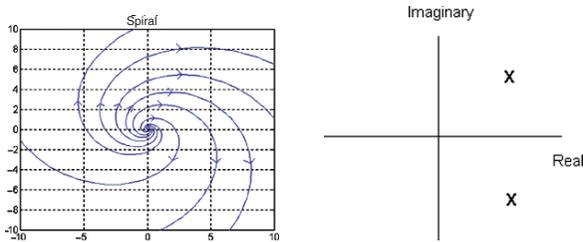
Fig. 1.18 Repelling spiral
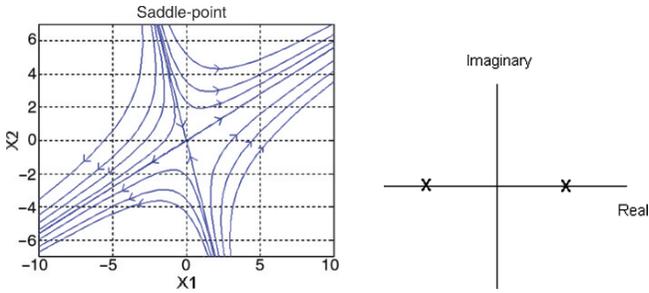
Fig. 1.19 Saddle point with flow orientation
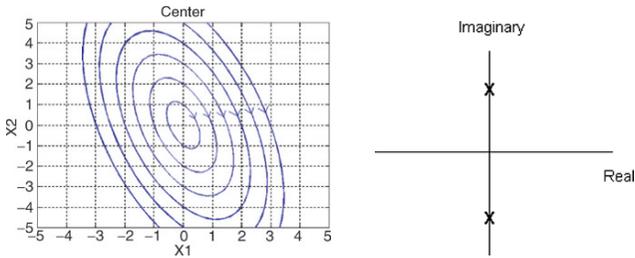
Fig. 1.20 Node center

1.12 Examples of Nonlinear System Resolution

1.12.1 A (Bilinear) Nonlinear System and a Saddle-Point

Given the system

$$\frac{dz_1}{dt} = z_2(z_1 + 1), \tag{1.89}$$

$$\frac{dz_2}{dt} = z_1(z_2 + 3). \tag{1.90}$$

This system has two equilibrium solutions:

$$\text{Trivial equilibrium:} \quad z_{1,s} = 0, \qquad z_{2,s} = 0, \tag{1.91}$$
$$\text{Non-trivial equilibrium:} \quad z_{1,s} = -1, \qquad z_{2,s} = -3. \tag{1.92}$$

If we linearize the system, we find the following Jacobian matrix:

$$A = \begin{bmatrix} z_{2,s} & z_{1,s} + 1 \\ z_{2,s} + 3 & z_{1,s} \end{bmatrix}. \tag{1.93}$$

Then, we *analyze the stability* of each equilibrium point:

(a) Equilibrium (1): The Jacobian matrix is

$$A = \begin{bmatrix} 0 & 1 \\ 3 & 0 \end{bmatrix} \tag{1.94}$$

and the eigenvalues are

$$\lambda_1 = -\sqrt{3}, \quad \lambda_2 = \sqrt{3}. \tag{1.95}$$

The analysis of the *linear system* enables us to say that the first equilibrium point is a *saddle point*, since its *first eigenvalue is stable* and the second is *unstable*.
The stable eigenvector is: $\xi_1 = \begin{bmatrix} -0.5 \\ 0.866 \end{bmatrix}$.

The unstable eigenvector is: $\xi_2 = \begin{bmatrix} 0.5 \\ 0.866 \end{bmatrix}$.

The phase space or the phase-plane of the model linearized around the equilibrium point (1) is a "saddle". The *linearized model* is: $\dot{x} = \begin{bmatrix} 0 & 1 \\ 3 & 0 \end{bmatrix} x$, where

$x = z - z_s$.

(b) Equilibrium (2): The Jacobian matrix is

$$A = \begin{bmatrix} -3 & 0 \\ 0 & -1 \end{bmatrix} \tag{1.96}$$

and the eigenvalues are $\lambda_1 = -3$, $\lambda_2 = -1$. The analysis of the *linear system* enables us to say that the second equilibrium point is a *stable node*, since its *eigenvalues are stable*.
The stable "fast" eigenvector is $\xi_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

The unstable "slow" eigenvector is $\xi_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

1.12.2 Pitchfork Bifurcation

The differential equation is written

$$\dot{x} = \mu x + x^3. \tag{1.97}$$

This system has three equilibrium solutions for $\dot{x} = 0$:

$$\bar{x} = 0, \quad \bar{x} = \sqrt{\mu}, \quad \bar{x} = -\sqrt{\mu}. \tag{1.98}$$
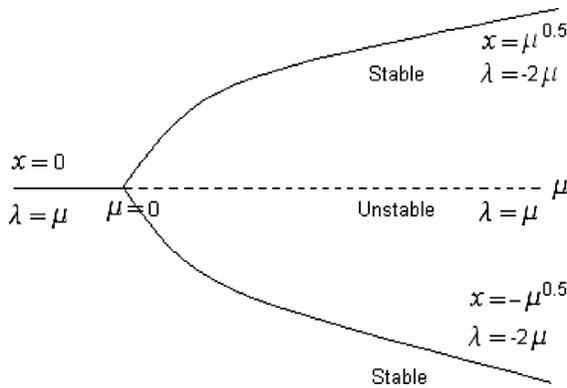
Fig. 1.21 Pitchfork bifurcation diagram, abscissa μ, ordinate x

We have the *Jacobian matrix* corresponding to the eigenvalues:

$$\lambda = -3x^2 + \mu. \tag{1.99}$$

For $\lambda < 0$ the solution is stable, for $\lambda > 0$ the solution is unstable. As we can suppose, it is the parameter μ which will determine the behavior of the system. In fact, we can write:

(a) For $\mu < 0$, we have the following eigenvalue for $\bar{x} = 0 : \lambda = \mu$.
(b) For $\mu > 0$, we have the following eigenvalue: (1) with $\bar{x} = 0$ then $\lambda = \mu =$ unstable, (2) with $\bar{x} = \sqrt{\mu}$ then $\lambda = -2\mu$, the behavior is stable, (3) with $\bar{x} = -\sqrt{\mu}$ then $\lambda = -2\mu$, the behavior is stable. Consequently, we obtain the bifurcation diagram (Fig. 1.21).

1.12.3 Supercritical Hopf Bifurcation

Given the initial system:

$$\dot{x}_1 = x_2 + x_1(\mu - x_1^2 - x_2^2), \tag{1.100}$$
$$\dot{x}_2 = -x_1 + x_2(\mu - x_1^2 - x_2^2), \tag{1.101}$$

we can also write this model by means of polar coordinates:

$$\dot{r} = r(\mu - r^2), \tag{1.102}$$
$$\dot{\theta} = -1. \tag{1.103}$$

If we indicate f such that: $\dot{r} = f(r) = r(\mu - r^2)$ and Jacobian is written: $\partial f / \partial r = r(\mu - r^2)$. The equilibrium (steady-state) point is $f(r) = 0$, which yields $r(\mu - r^2) = 0$, which has three solutions:

$$r = 0 \text{ (trivial solution)}, \tag{1.104}$$

$$r = \sqrt{\mu}, \tag{1.105}$$

$$r = -\sqrt{\mu} \text{ (not physically realizable).} \tag{1.106}$$

(1) For $\mu < 0$, only the single possible solution is $r = 0$, i.e. $\partial f / \partial r = \mu$. This solution is stable because $\mu < 0$. (2) For $\mu = 0$, all of the steady-state solution are $r = 0$ and the Jacobian is $\partial f / \partial r = -3r^2$ which is stable, but has slow convergence to $r = 0$. (3) For $\mu > 0$, the trivial solution ($r = 0$) is unstable, because $\partial f / \partial r = \mu$.

The nontrivial solution ($r = \sqrt{\mu}$) is stable is a stable solution because $\partial f / \partial r = \mu - 3r^2 = -2\mu$ and we find the following phase-plane plots (Fig. 1.22).

Their trajectories for identical initial-values are shown in Fig. 1.23.

Hereafter, the bifurcation diagram indicates that the origin ($r = 0$) is stable when $u < 0$. When $u > 0$ the origin becomes unstable, but a *stable limit cycle* (with radius $r = \sqrt{u}$) *emerges* (Fig. 1.24).
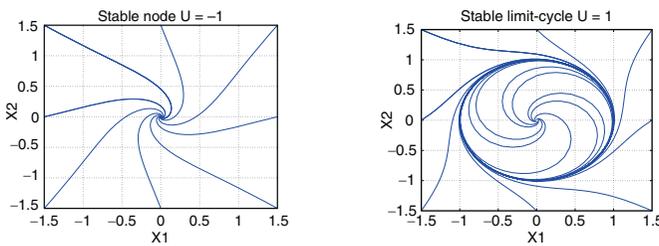
Fig. 1.22 $\mu = -1$ (*left*), $\mu = 1$ (*right*)
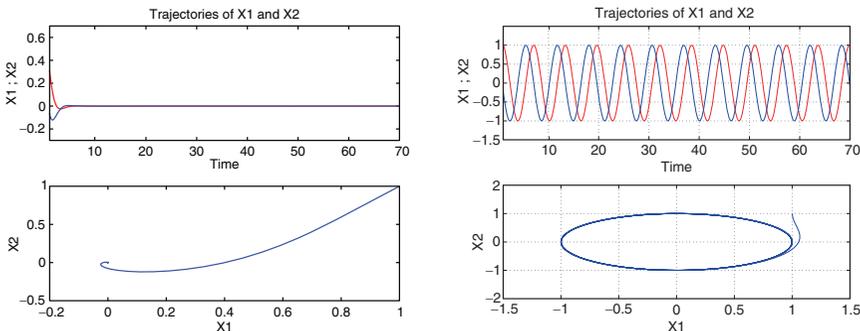
Fig. 1.23 *Left:* $\mu = -1$ and $(x_1(0), x_2(0)) = (1, 1)$. *Right:* $\mu = 1$ and $(x_1(0), x_2(0)) = (1, 1)$

Fig. 1.24 Supercritical bifur-
cation diagram
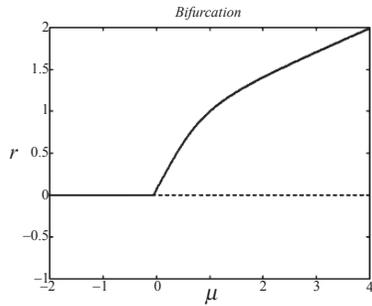
At this stage, we analyze this system in rectangular $(x_1 - x_2)$ coordinates. The only steady-state (fixed-point or equilibrium) solution to the initial system is

$$x_s = \begin{bmatrix} x_{1s} \\ x_{2s} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{1.107}$$

Linearizing the initial system, we have

$$\partial f_1/\partial x_1 = \mu - 3x_1^2 - x_2^2, \quad \partial f_1/\partial x_2 = 1 - 2x_1x_2, \\ \partial f_2/\partial x_1 = -1 - 2x_1x_2, \quad \partial f_2/\partial x_2 = \mu - x_1^2 - 3x_2^2, \tag{1.108}$$

the Jacobian is written

$$A = \begin{bmatrix} \mu - 3x_1^2 - x_2^2 & 1 - 2x_1x_2 \\ -1 - 2x_1x_2 & \mu - x_1^2 - 3x_2^2 \end{bmatrix} \tag{1.109}$$

which is, for an equilibrium solution of the origin:

$$A = \begin{bmatrix} \mu & 1 \\ -1 & \mu \end{bmatrix}. \tag{1.110}$$

The characteristic polynomial, from $det(\lambda I - A) = 0$, is

$$\lambda^2 - 2\mu\lambda + \mu^2 + 1 = 0 \tag{1.111}$$

which has as eigenvalues (roots): $\lambda = \mu \pm \dfrac{\sqrt{4\mu^2 - 4(\mu^2 + 1)}}{2} = \mu \pm j$. We see that when $\mu < 0$, the complex values are stable (negative real portion); when $\mu = 0$, the eigenvalues lie on the imaginary axis of the complex plane; and when $\mu > 0$ the complex values are unstable (positive real portion). The transition of eigenvalues from the left-half plane to the right-half plane is shown hereafter (see Fig. 1.25).

This example was for a supercritical Hopf bifurcation, where a stable limit cycle was formed whose diagram is given hereafter (Fig. 1.26).

Consider as an exercise for the reader to show the formation of a subcritical Hopf bifurcation, where an unstable limit cycle is formed.
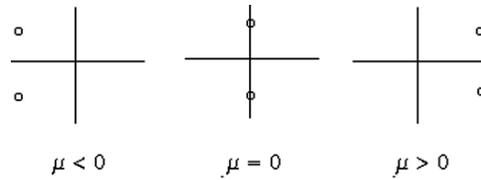
Fig. 1.25 Locations of eigenvalues in the complex plane. When the eigenvalues pass from the left side to the right side of the complex plane, a Hopf bifurcation occurs.
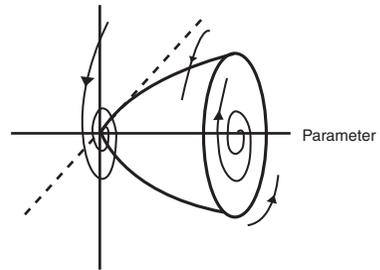
Fig. 1.26 Supercritical Hopf bifurcation

Previously, we have found that the Hopf bifurcation occurs when the real portion of complex eigenvalues became zero. In the previous example, the eigenvalues crossed the imaginary axis with zero slope, i.e., parallel to the real axis. In the general case, the eigenvalues will cross the imaginary axis with non-zero slope.

We have also to make it clearer how an analysis of the characteristic polynomial of the Jacobian (A) matrix can be used to determine when a Hopf bifurcation can occur. In case of a two-state system, the characteristic polynomial has the form:

$$a_2(\mu)\lambda^2 + a_1(\mu)\lambda + a_0(\mu) = 0, \tag{1.112}$$

where the polynomial parameters, a_i, are a function of the bifurcation parameter μ. (Note that it is common for $a_2 = 1$.) Suppose that the $a_i(\mu)$ parameters do not become 0 for the same value of μ, it is possible to show that a Hopf bifurcation occurs when $a_1(\mu) = 0$.

1.12.3.1 Higher Order Systems ($n > 2$)

Previouly, we dealt with Hopf bifurcations of two-state systems but Hopf bifurcations can occur in any order system ($n \geq 2$); they key is that two complex eigenvalues cross imaginary axis, while all other eigenvalues remain negative (stable). This is shown in Fig. 1.27 for the three state case.

In Fig. 1.27, we show the location of eigenvalues in complex plane as a function of μ. A Hopf bifurcation occurs as the eigenvalues pass from the left-hand side to the righthand side of the complex plane.

$$\mu < 0 \qquad\qquad \mu = 0 \qquad\qquad \mu > 0$$
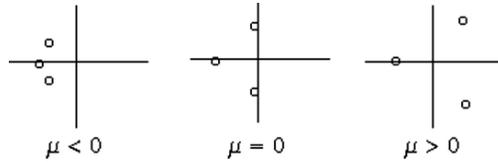
Fig. 1.27 Locations of eigenvalues. Similarly, a Hopf bifurcation occurs when the eigenvalues pass from the left to the right side.
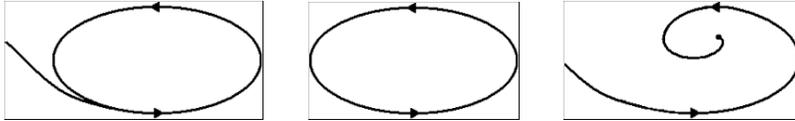
Fig. 1.28 Convergence to periodic solution (*left*). Periodic solution (*center*). Convergence to a fixed-point (*right*)

1.13 Poincaré–Bendixson Theorem

In a space with two dimensions, we can show that *the orbits of a nonlinear system move (in spiral) towards a closed curve or a limit-cycle.* And this, even when the solutions of the system are not known.

Theorem 1.4 (Poincaré–Bendixson). *Let us suppose that an orbit $x(x_0,t)$ of a system with two equations:*[15] $dx/dt = f(x)$, $x = (x_1,x_2)^T$, $f = (f_1,f_2)^T$ *remains in a compact "domain" $D \subset \mathbb{R}^2$ for all $t \geq 0$, then either:*

1. *$x(x_0,t)$ is a periodic solution of the system.*
2. *Or $x(x_0,t)$ tends towards a periodic solution.*
3. *Or $x(x_0,t)$ tends towards a fixed point.*

Illustration of the three types of solution enunciated in this theorem (Fig. 1.28).

1.13.1 Bendixson Criterion

Theorem 1.5 (Bendixson Criterion). *If in a "domain" which does not have hole,*[16] *$D \subset \mathbb{R}^2$ the expression*[17] *$Div f \equiv \sum_{i=1}^{2} \partial f_i/\partial x_i$ is not identically null and does not change sign, then the system $dx/dt = f(x)$ does not have a periodic orbit contained in D.*

---

[15] Definition from Dang-Vu and Delcarte (2000).

[16] Simply connected domain: i.e. domain without hole or separate parts in the domain.

[17] (*Div* : divergence).

1.14 Center Manifold Theorem

Suppose a fixed point $\bar{x}$ of the system $\dot{x} = f(x)$ with $x \in \mathbb{R}^n$ and if f is of class C^r on D which is included in $\mathbb{R}^n$, the linearization of the system above around the fixed point $\bar{x}$ can be written

$$\frac{d\xi}{dt} = \xi \qquad (1.113)$$

with $\xi = x - \bar{x}$. And J is the Jacobian matrix such that:

$$J \equiv Df(\bar{x}) = \left\| \frac{\partial f_i}{\partial x_j} \right\|_{x=\bar{x}}. \qquad (1.114)$$

If we write $\xi = u e^{\lambda t}$, $u \in \mathbb{R}^n$ and if we then replace it in $d\xi/dt = J\xi$ then we obtain the eigenvector and eigenvalue equations:

$$(J - \lambda I)u = 0, \qquad (1.115)$$
$$|J - \lambda I| = 0. \qquad (1.116)$$

Given:

(a) $(u_1, \ldots, u_s)$: the eigenvectors of J correspond to the eigenvalues λ of which the real part is *negative*.
(b) $(v_1, \ldots, v_u)$: the eigenvectors of J correspond to the eigenvalues λ of which the real part is *positive*.
(c) $(w_1, \ldots, w_c)$: the eigenvectors of J correspond to the eigenvalues λ of which the real part is *null,*

$$\text{with } s + u + c = n, \qquad (1.117)$$

with s for stable, u for unstable and c for center, constant or central, and consider also:

(1) E^s the vectorial subspace generated by $(u_1, \ldots, u_s)$.
(2) E^u the vectorial subspace generated by $(v_1, \ldots, v_u)$.
(3) E^c the vectorial subspace generated by $(w_1, \ldots, w_c)$

$$\text{with } \mathbb{R}^n = E^s \oplus E^u \oplus E^c. \qquad (1.118)$$

Then, it is possible to write the following *Center manifold theorem*:

Theorem 1.6 (Center manifold). *There exist manifolds of C^r class, stable W^s, unstable W^u, center W^c (central), which are tangent respectively to E^s, E^u and E^c at the fixed point $\bar{x}$. These manifolds are invariant with respect to the flow ϕ_t of the system $\dot{x} = f(x)$. W^s et W^u are single, however W^c is not necessarily single.*

Under the action of the flow, we have

$$\phi_t(W^s) \subset W^s, \qquad (1.119)$$
$$\phi_t(W^u) \subset W^u, \qquad (1.120)$$
$$\phi_t(W^c) \subset W^c, \qquad (1.121)$$

Fig. 1.29 Stable, unstable and
center manifold
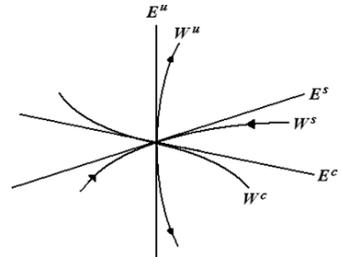
and we have:

$$lim_{t\to+\infty}\phi_t(x) = \bar{x} \text{ for all } x \in W^s, \tag{1.122}$$
$$lim_{t\to-\infty}\phi_t(x) = \bar{x} \text{ for all } x \in W^u. \tag{1.123}$$

The behavior in the neighborhood of $\bar{x}$ can be depicted as follows (Fig. 1.29).

1.15 Definitions of Chaos

The term chaos associated with an interval map was first formally introduced by
Li and Yorke in 1975, where they established a simple criterion for chaos in one-
dimensional difference equations (ref. to Li and Yorke 1975). Chaos in the sense of
Li–Yorke can be given by:

Definition 1.12 (Chaos in the sense of Li–Yorke). A continuous map $f : I \to I$,
where I is the unit interval ($I = [0,1]$), is a chaos in the sense of Li–Yorke if
there is an uncountable set $S \subset I$ such that trajectories of any two distinct points
x,y in S are *proximal* and *not asymptotic*, i.e. $\liminf_{n\to\infty} d(f^n(x), f^n(y)) = 0$ and
$\limsup d(f^n(x), f^n(y)) > 0$.

The requirement of uncountability of S in this definition (i.e. for continuous maps
of the interval, but not in a general compact metric space) is equivalent to the con-
dition that S contains two points, or that S is a perfect set (i.e. nonempty, compact
and without isolated points). Chaos in the sense of Devaney (1989):

Definition 1.13 (Devaney's definition of chaos – 1989). Let V be a set. A contin-
uous map $f : V \to V$ is said to be chaotic on V if

1. f is topologically transitive: for any pair of open non-empty sets $U, W \subset V$ there
   exists a $k > 0$ such that $f^k(u) \cap W \neq \varnothing$.
2. The periodic points of f are dense in V.
3. f has sensitive dependence on initial conditions: there exists a $\delta > 0$ such that,
   for any $x \in V$ and any neighborhood N of x, there exists a $y \in N$ and an $n \geq 0$ such
   that $|f^n(x) - f^n(y)| > \delta$.

There is still no unified, universally accepted, and rigorous mathematical definition of chaos. Among the diversity of proposed definitions, those of Li–Yorke and Devaney seem to be the most popular. Huang and Ye proved that any chaotic map in the sense of Devaney from a compact metric space to itself is chaotic in the sense of Li–Yorke. It has been also proved that if a continuous map from a compact metric space to itself has positive entropy, then it is chaotic in the sense of Li–Yorke. There exists a recurrent complaint about the lack of an universally accepted definition of chaos. Many papers refer to density of orbits (DO), sensitive dependence on initial conditions (SDIC) and topological transitivity (TT) as the main characteristics of chaos. It was also pointed out that density of orbits and topological transitivity are *topological* characteristics, whereas sensitive dependence on initial conditions is a *metric* one. It is well known that density of orbits and topological transitivity, when both satisfied, imply sensitive dependence on close initial conditions. We also known that SDIC alone do not imply neither density of orbits nor topological transitivity.

A redundancy was found in Devaney's definition of chaos. Banks et al. (1992) prove that the conditions (1) and (2) imply (3) in any metric space V. Furthermore, Assaf and Gadbois (1992) show that for general maps this is the only redundancy: (1) and (3) do not imply (2), and (2) and (3) do not imply (1). But, if we restrict to maps on an interval, then one obtains a stronger result:

Proposition 1.1. *Let U be a, not necessarily finite, interval and $f : U \rightarrow U$ a continuous and topologically transitive map. Then* (1) *the periodic points of f are dense in U and* (2) *f has sensitive dependence on initial conditions.*

The result (1) was given by Block and Coppel (1992), and an intuitive proof of the proposition is given by Vellekoop and Berglund (1994).

From Sharkovskii (1964) to Li–Yorke (1975): Chaos in the sense of Li–Yorke was a corollary of general Sharkovskii's theorem. We'll show that the first theorem below is the consequence of the second one.

Theorem 1.7 (Period three implies chaos). *If $f : I \rightarrow I$ is a continuous map having a periodic point of period* 3*, then f has periodic points of all periods.*

Definition 1.14. A point $x \in I$ is said periodic of period $n > 1$ if $f^n x = x$ and $f^i x \neq x$ for $i = 1, \ldots, n-1$, It is said that f is a fixed point if $fx = x$.

Above we denote $fA = f(A)$ for $A \subseteq I$. Thus when a point is periodic, its orbit is finite, but in addition, f acts bijectively on this orbit as a cycle. In order to give the Sharkovskii's theorem we need to define a specific ordering on the set of strictly positive natural numbers $\mathbb{N}^*$:

Definition 1.15 (Sharkovskii's ordering). Sharkovskii's ordering on $\mathbb{N}^*$ is the ordering $\succ$ defined by $3 \succ 5 \succ \cdots \succ 3 \cdot 2 \succ 5 \cdot 2 \succ \cdots \succ 3 \cdot 2^n \succ 5 \cdot 2^n \succ \cdots \succ 3 \cdot 2^{n+1} \succ 5 \cdot 2^{n+1} \succ \cdots \succ 2^n \succ 2^{n-1} \succ \cdots \succ 2^2 \succ 2 \succ 1$ and it is a total order.

Theorem 1.8 (Sharkovskii 1964). *Let $f : I \to I$ be continuous map having a periodic point of period n. If $n \succ m$ in the Sharkovskii ordering then f has a periodic point of period m.*

Example 1.1. $\bar{x} = f(x) = rx(1-x)$, $0 < r < 4$, it is easy for $r > 1$ to verify that the fixed point $x^* = (r-1)/r$ is locally asymptotically stable, i.e. $|f'(x^*)| < 1$ provided $1 < r < 3$. We want to show that for $1 < r < 3$, the fixed point x^* attracts each point in the interval $(0,1)$. First, we show that there are no n-periodic points of the map f for $n \geq 2$. By Sharkovskii's theorem, it suffices to show that there are no 2-periodic points. Suppose, on the contrary, there exists $x \in (0,1)$, $x \neq x^*$ such that $f^2(x) = x$. Then x satisfies $1 = r^2(1-x)(1-rx+rx^2)$ or $(x-x^*)(rx^2 - (r+1)x + ((r+1)/r)) = 0$. Then $0 < x = (r+1+(r^2 - 2r - 3)^{1/2})/2r < 1$. However the condition $1 < r < 3$ implies $r^2 - 2r - 3 = (r-3)(r+1) < 0$. This leads to a contradiction. Since f has no 2-periodic points in $(0,1)$, it follows that f has no nonwandering points excluding x^* and consequently x^* is a global attractor.

1.16 Invariant Sets and Attractors

From G.D. Birkhoff (1927) to J. Milnor, many authors have defined the notions of attracting set and attractor, among the most known definitions of an attractor there are those of R. Thom (1978), C. Conley (1978), Guckenheimer and Holmes (1983), D. Ruelle (1981), and J. Milnor (who also defined weak attractor). Hereafter, we'll define these notions which are common to flows and maps.

Definition 1.16 (Invariant set). For a flow ϕ_t [or a map g] defined on $U \subset \mathbb{R}^n$, a subset $S \subset U$ is said *invariant* if

$$\phi_t(S) \subset S, \ \forall t \in \mathbb{R} \ [\text{or } g^n(S) \subset S, \ \forall n \in \mathbb{Z}]. \qquad (1.124)$$

Definition 1.17 (Homoclinic orbit). Let S be an invariant set of the flow ϕ_t, p is a point in the phase space of the flow and let us suppose that $\phi_t(p) \to S$ when $t \to +\infty$, and $t \to -\infty$, then the orbit of p is said to be *homoclinic* to S.

Definition 1.18 (Heterocilinic orbit). Let S_1 and S_2 be two disjoint invariant sets of a flow ϕ_t and let us suppose that $\phi_t(p) \to S_1$ as $t \to +\infty$, and $\phi_t(p) \to S_2$ as $t \to -\infty$, then the orbit of p is said to be heteroclinic to S.

Definition 1.19 (Accumulation point). The ω-limit of x for a flow ϕ_t is the set of *accumulation points* of $\phi_t(x), t \to +\infty$.

Definition 1.20 (Set of accumulation points). In the same way, the α-limit of x for a flow ϕ_t is the set of *accumulation* points of $\phi_t(x), t \to -\infty$.

For example, if we consider a differential system with two fixed points, $\bar{x}_1$ (stable) and $\bar{x}_2$ (unstable) and no other limit. Then we say that for such a system, $\bar{x}_1$ is a ω-limit and $\bar{x}_2$ is a α-limit of the flow. The notion of "invariant" has an important role in dynamical systems, and this concept naturally leads to that of "attracting set".

Definition 1.21 (Attracting set). A compact set $A \subset U$ is an attracting set if:

- A is invariant under ϕ_t.
- A has a shrinking neighborhood, i.e. there is a neighborhood V of A, such as for all $x \in V$, $\phi_t(x) \in V$, for all $t \geq 0$ and $\phi_t(x) \to A$ for $t \to +\infty$.
  A basin of attraction, for an attracting set, is defined by

$$W = \bigcup_{t \leq 0} \phi_t(V). \qquad (1.125)$$

W is also called a *stable manifold* of A. We are able to define, in the same way, a "repelling" set and its unstable manifold by replacing t by $-t$ and n by $-n$. We'll give a stronger definition of an attractor in one of following sections.

Definition 1.22 (Transitive topology of a set). A set A is an attractor for the flow ϕ_t, if it satisfies the properties exposed in the preceding definition, i.e. if it is an attracting set, and if moreover it is topologically transitive.

Let A be a compact invariant set, A is said topologically transitive for a flow ϕ if for two open sets $U, V \subset A$, $\phi_t(U) \cap V \neq \varnothing$, for all $t \in \mathbb{R}$. The transitivity implies (and is implied by) the existence of a "dense *orbit*".

1.16.1 Definition of an Attractor

Definition 1.23 (Attractor – Ruelle 1981). Let A be a closed compact set of the phase-space (or of the flow). It is supposed that A is an invariant set (i.e. $\phi_t(A) = A$, for any t). It is said that A is stable if for any neighborhood U of A, there exists a neighborhood V of A such that any solution $x(x_0, t) \equiv \phi_t(x_0)$ will remain in U if $x_0 \in V$. If moreover:

$$\bigcap_{t \geq 0} \phi_t(V) = A \qquad (1.126)$$

and if there exists a *dense orbit* in A, then A is an *attractor*.

When A is an attractor, the set $W = \bigcup_{t < 0} \phi_t(V)$ is called "basin of attraction" of A. This set is the set of points *whose trajectories asymptotically converge towards A*. The simplest attractor is the *fixed point*. A second type of attractor for a vector field is the limit cycle: it is a closed trajectory which attracts all the close orbits (Fig. 1.30). The third type is often called *the strange attractor*.

1.16.2 Strange Attractor

This attractor was defined by Ruelle and Takens (1971). It is characterized by:

1. In the phase space, the attractor is of *null volume*.
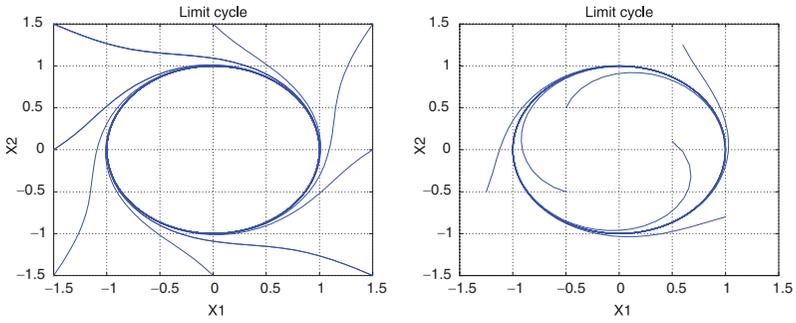
Fig. 1.30 Limit cycles
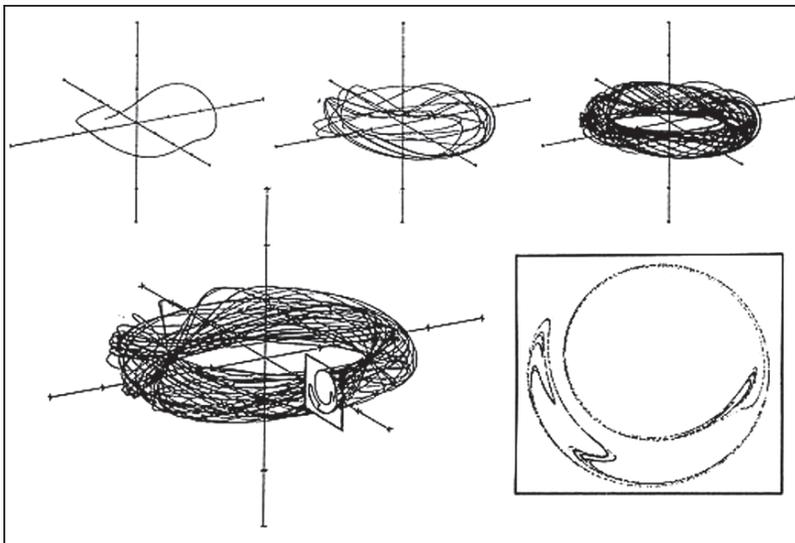
Fig. 1.31 Archetype of strange attractor and an associated Poincaré section

2. The *dimension D of the attractor is fractal (non-integer)* $2 < D < n$, where n is the dimension of the phase space.
3. There is "sensitive dependence on initial conditions" *(SDIC)*: two trajectories (of the attractor) initially very close will diverge.

An illustration of strange attractors can be given by means of the well-known James Gleick picture (see Gleick 1987) of a three-dimensional attractor and a Poincaré section (Fig. 1.31).

1.17 Some Nonlinear Dynamical Systems with Their Associated Attractors

Duffing attractor: The dynamical system can be written as follows:

$$\dot{x} = y, \; \dot{y} = x - x^3 + \varepsilon(a\cos\theta - by), \quad \dot{\theta} = 1. \tag{1.127}$$

It is periodic in θ and the Euler approximation is shown in Fig. 1.32 for $\varepsilon a = 0.4$ and $\varepsilon b = 0.25$ (see Guckenheimer and Holmes, 1983, pp. 82–91, 191).

Chua autonomous circuit: The dynamical system is written: $dx/dt = \alpha(-x + y - h(x), dy/dt = x - y + z, dz/dt = -\beta y$, where $h(x) = bx + 0.5(a-b)(|x+1| - |x-1|)$, α arbitrary, $\beta = 15$, $a = -1.3$, $b = -0.7$ (Fig. 1.33).

Hénon attractor: $x_{n+1} = 1 - ax_n^2 + y_n$, $y_{n+1} = bx_n$, with $a = 1.4, b = 0.3$ (Fig. 1.34).

Colpitts oscillator (simple chaotic generator): The dynamical system is written: $dx/dt = K K_1(x-z) - x, dy/dt = 2K_2 f(y,z), dz/dt = 2K_1(x-z) - 2K_2 f(y,z)$, where $f(y,z) = \{(z - y, \text{ if } z - y \leqslant 1) \text{ and } (1 \text{ otherwise})\}$, k arbitrary, $K_1 = 11$, $K_2 = 0.9$ (Fig. 1.35).

Ikeda attractor: $x_{n+1} = k + \beta(x_n \cos\omega_n - y_n \sin\omega_n)$, $y_{n+1} = \beta(x_n \cos\omega_n - y_n \sin\omega_n)$, $\omega_n = 0.4 - \alpha/(1 + x_n^2 + y_n^2)$, with $\alpha = 5.4$, $\beta = 0.9$, $k = 0.92$ (Fig. 1.36).

Rössler attractor: $\dot{x} = dx/dt = -(y+z)$, $\dot{y} = dy/dt = x + ay$, $\dot{z} = dz/dt = b + z(x - c)$, with $\alpha = 0.15$, $b = 0.20$, $c = 10$ (Fig. 1.37).
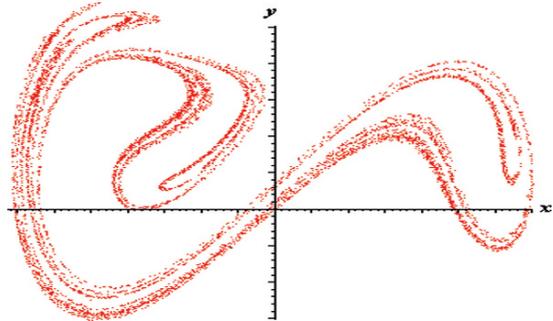
Fig. 1.32 Duffing attractor
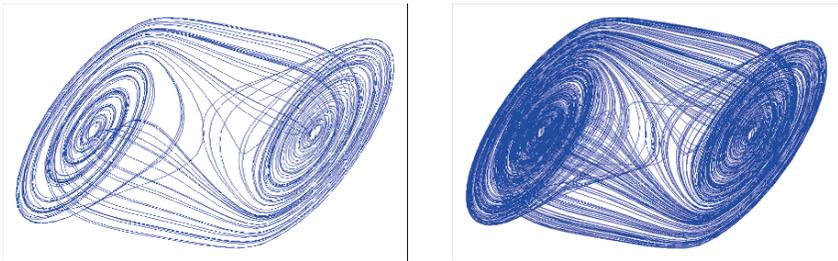
Fig. 1.33 Attractor in progress (*left*). Chua attractor (*right*)
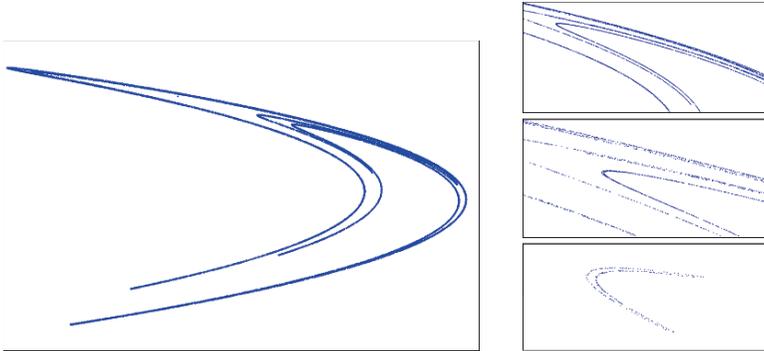
Fig. 1.34 Henon attractor. $x = [-2,\ldots,2], y = [-2,\ldots,2]$ (*left*). Enlargements (*right*)
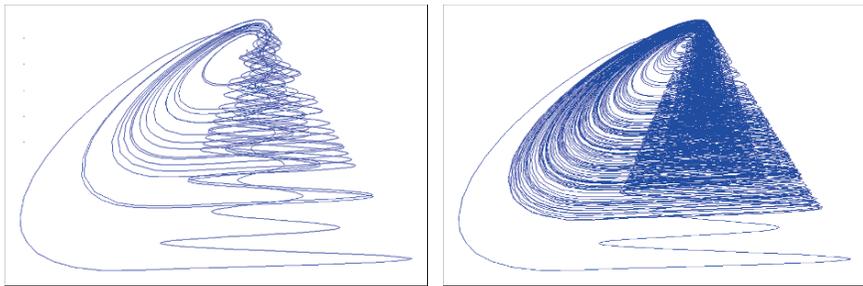
Fig. 1.35 Attractor in progress (*left*). Colpitts attractor (*right*)
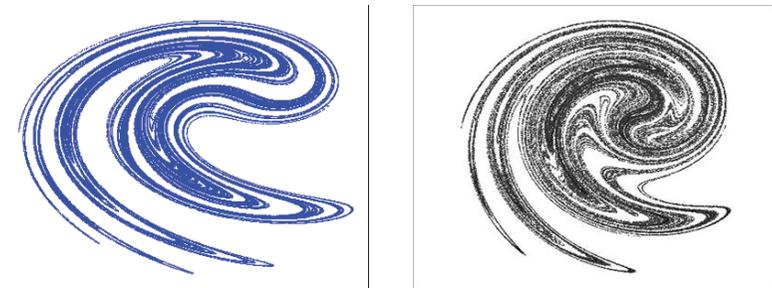
Fig. 1.36 Ikeda map $[(-0.4,\ldots,1.6),(-2,\ldots,1)]$ (*left*). Ikeda map variant (*right*)

Rayleigh oscillator: The dynamical system is written: $\varepsilon dx/dt = y + f(-x), dy/dt = -x - a - b\sin(\omega t)$, where $f(x) = -x(1-x^2)$, $\varepsilon = 0.1$, $a = 0.56946$, $b = 5e-4$, $\omega = 1.3946$ (Fig. 1.38).

Moon pendulum is in connection with Duffing equations: $x'' + mx' - \frac{1}{2}(1-x^2)x = a\cos(at)$, with $m = 0.15, a = 0.15$. *Mackay–Glass attractor*: $\dfrac{dz}{dt} = \dfrac{ax(t+T)}{1+[x(t+T)]^c} - bx(t)$, with $a = 0.2$, $b = 0.1$, $c = 10.0$, $T = 31.8$. (Figures not shown here.)
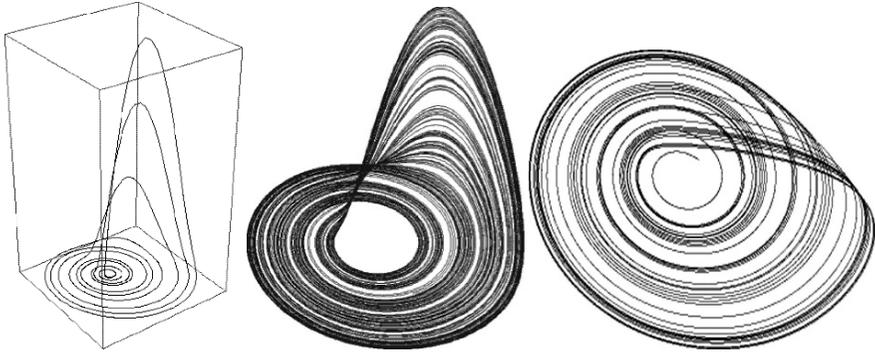
Fig. 1.37 In progress (*left*). Rössler attractor, 3D (*center*). In a plane, 2D (right)

Fig. 1.38 Rayleigh attractor
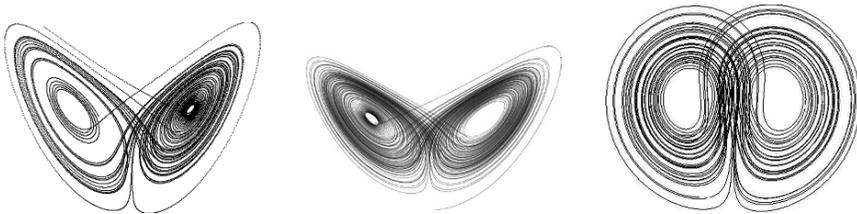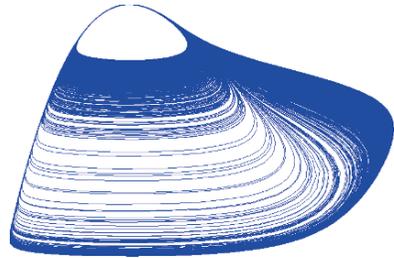
Fig. 1.39 In progress (*left*). Lorenz attractor, 3D (*center*). In a plane, 2D (*right*)

Lorenz attractor: $\dot{x} = dx/dt = \sigma(y-x)$, $\dot{y} = dy/dt = x(\rho - z) - y$, $\dot{z} = dz/dt = xy - \beta z$, with $\sigma = 16$, $\rho = 45.92$, $\beta = 4.0$ (Fig. 1.39).

Standard map (also known as the *Chirikov–Taylor map*): Standard Map $\{x_{n+1} = x_n + y_n$, $y_{n+1} = y_n - 2\pi\varepsilon^2 \sin(2\pi x_{n+1})\}$, for example: $x + y$, $y - 0.971635 \sin(2\pi x)/2\pi$, we obtain Fig. 1.40 where the closed loops correspond to stable regions with fixed points or fixed periodic points at their centers. The hazy regions are unstable and chaotic.

Arrowsmith–Place attractor and map: Given the map $\varphi : x_1 = x + y_1$, $y_1 = y + ax(x-1)$, with: $0 < a < 4$. Given the inverse map $\varphi^{-1} : x = x_1 - y_1$, $y = y_1 - ax(x-1)$, with: $0 < a < 4$ (Fig. 1.41).
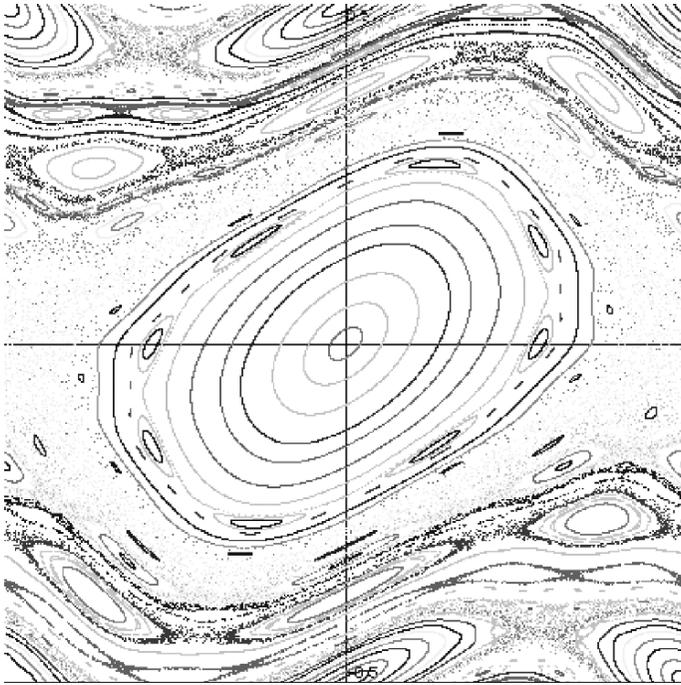
Fig. 1.40 Standard map (i.e. Chirikov–Taylor map)
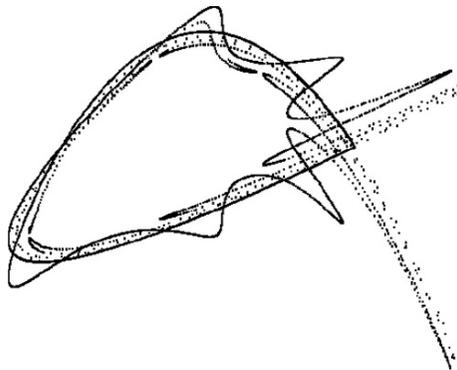
Fig. 1.41 Arrowsmith–Place attractor. Axis: $[(-1, \ldots, 2), (-1, \ldots, 1)]$
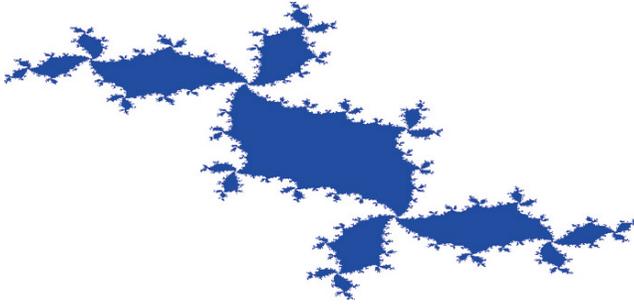
Fig. 1.42 Julia attractor for $c = -0.17 + i0.78$. Axis: $[(-1.5,\ldots,1.5),(-1.5,\ldots,1.5)]$

Julia attractor: $z_{n+1} = z_n^2 + c$, $z_n = a_n + ib_n$, $c = a_0 + ib_0$ with in the case above: $c = -0.17 + i0.78$ (Fig. 1.42). Some other known values for the parameter c:

$c = -0.12375 + 0.56508i$ (basin with an attractor)

$c = -0.12 + 0.74i$ (three attractors)

$c = -0.11 + 0.6557i$ (three attractors with thin spirale)

$c = -0.194 + 0.6557i$ (three attractors, Cantor Set)

$c = 0 + i$ (filament dendrite)

$c = -0.125 + 0i$ (parabolic case: between two and four attractors)

$c = 0.11031 - 0.67037i$ (Dust of Fatou),

$c = 0.27334 + 0.00742i$ (other parabolic case)

Mandelbrot attractor: We give a quick view of the Mandelbrot set, which belongs to the set of complex quadratic polynomials:

$$z_0 = 0, \quad z_{n+1} = z_n^2 + c. \tag{1.128}$$

Hereafter, we present the Mandelbrot set in the complex plane. This set is delimited by the following segments: The axis of the real part $= [-2,1]$ and the axis of the imaginary part $= [-1.5,1.5]$ (Fig. 1.43).

Attractors of quadratic maps: The dynamics results from the following system based on quadratic maps:

$$x_{n+1} = \theta_1 + \theta_2 x_n + \theta_3 x_n^2 + \theta_4 x_n y_n + \theta_5 y_n + \theta_6 y_n^2, \tag{1.129}$$
$$y_{n+1} = \theta_7 + \theta_8 x_n + \theta_9 x_n^2 + \theta_{10} x_n y_n + \theta_{11} y_n + \theta_{12} y_n^2. \tag{1.130}$$

It is possible by varying the value of the coefficients θ_i (with $i = 1, 2, 3, \ldots, 12$) of the quadratic map terms between -1.2 and 1.2 to depict in Fig. 1.44 a very short sample of strange attractors that such a quadratic system is able to provide among 25^{12} possible.
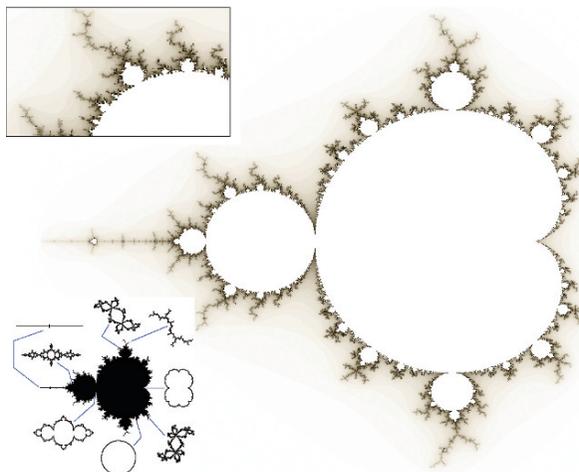
Fig. 1.43 Mandelbrot set: $f(z) = z^2 + c$
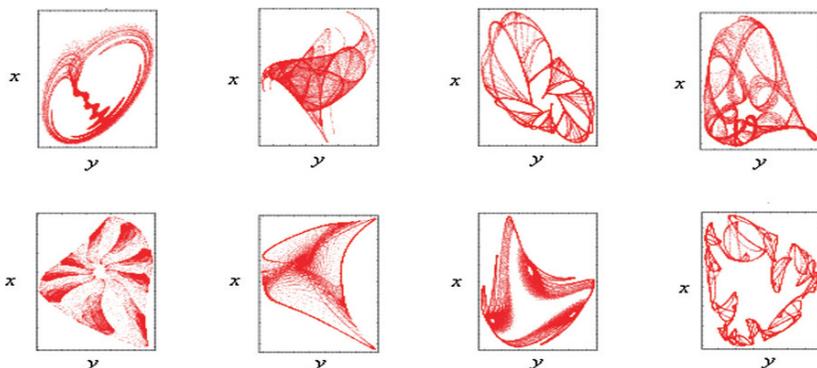
Fig. 1.44 Sample of strange attractors from the quadratic map

Hénon–Heiles attractor:

$$H = \frac{1}{2}(p_x^2 + p_y^2 + x^2 + y^2) + x^2 y - \frac{1}{3}y^3, \tag{1.131}$$

$$H = \frac{1}{2}(p_x^2 + p_y^2 + x^2 + y^2) + x^2 y - \frac{1}{3}y^3, \tag{1.132}$$

$$\dot{x} = \frac{\partial H}{\partial p_x} = p_x, \tag{1.133}$$

$$\dot{y} = \frac{\partial H}{\partial p_y} = p_y, \tag{1.134}$$

$$\dot{p}_x = -\frac{\partial H}{\partial x} = -x - 2xy, \tag{1.135}$$

$$\dot{p}_y = -\frac{\partial H}{\partial y} = -y - x^2 + y^2. \tag{1.136}$$

The trajectory of this four-dimensional system spreads in a "hypersurface":

$$H = E = cte. \tag{1.137}$$

In spite of the multiplicity of the possible combinations, a certain number of (limit) representative cases are highlighted. Each case presented hereafter shows a trajectory for different initial conditions:

$$\{p_x, p_y, y, x = (2E - p_x^2)^{1/2}\}, \tag{1.138}$$

and for

$$E = H : \{0.06, 1/12, 1/8, 1/6\} : \tag{1.139}$$

– Basic case: $E = 1/12$, $x = 0$, $y = 0.02$, $\dot{y} = p_y = -0.08$.
– "Hilborn" case: $E = 0.06$, $x = 0$, $y = 0$, $\dot{y} = p_y = 0$.
– Complex loop: $E = 1/12$, $x = 0$, $y = 0$, $\dot{y} = p_y = 0.123$.
– Isolated archipelago case: $E = 1/8$, $x = 0$, $y = 0$, $\dot{y} = p_y = 0.08$.
– Chaotic attack: $E = 1/8$, $x = 0$, $\dot{y} = -0.15$, $\dot{y} = p_y = 0$.
– Almost complete chaos: $E = 1/6$, $x = 0$, $y = -0.15$, $\dot{y} = p_y = 0$.

For example, in the Hilborn case, we show planes which make it possible to liberate structures of "toric appearance" which constitute the attractor. We then show intermediate structures which can be observed through the "archipelago" case, the "loops" case and the "chaotic attacks" case.

Hilborn $(x, dx/dt)$ Hilborn (x, y) Hilborn $(x, dy/dt)$

Archipelago $(x, dx/dt)$ Archipelago (x, y) Archipelago $(y, dy/dt)$

Loop $(x, dx/dt)$ Loop (x, y) Loop $(y, dy/dt)$

Chaotic attack (x, dx / dt)   Chaotic attack (x, y)   Chaotic attack (y, dy / dt)
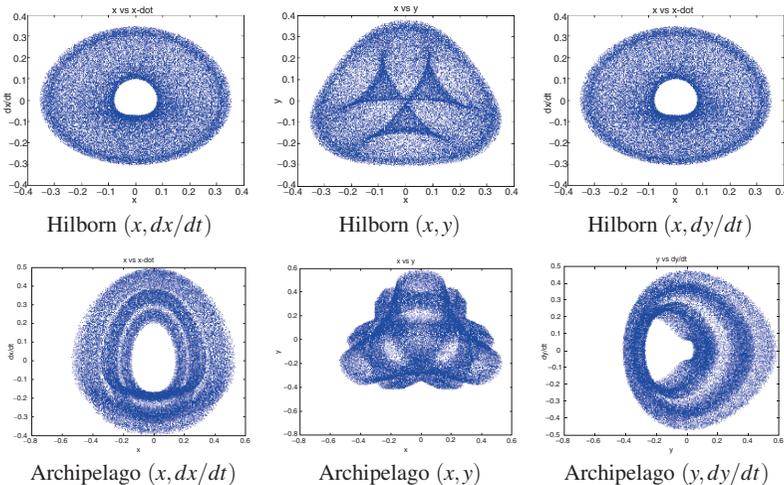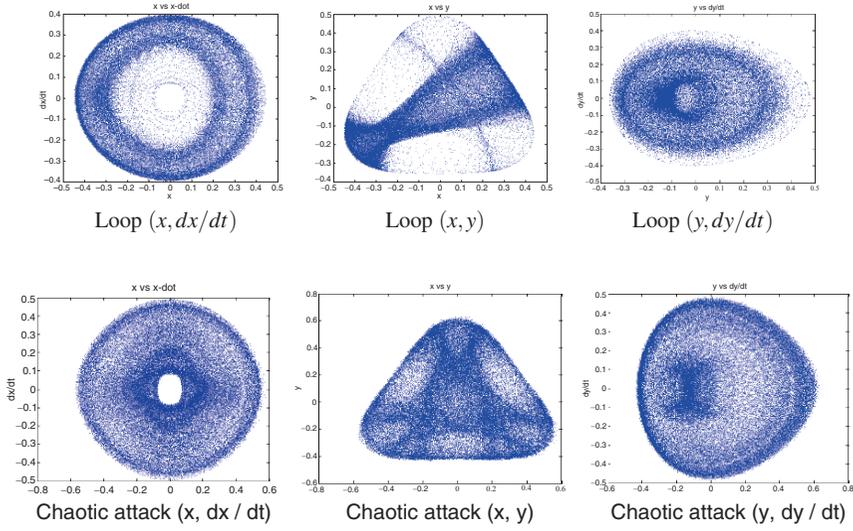
Now, we will show the graphs of all Poincaré sections (y, p_x) for which the conditions have been explicitly posed above:

Basic case
(torus–plane intersection)

Hilborn case

Loop of chaos

Isolated archipelago

Chaotic attack

Almost complete chaos
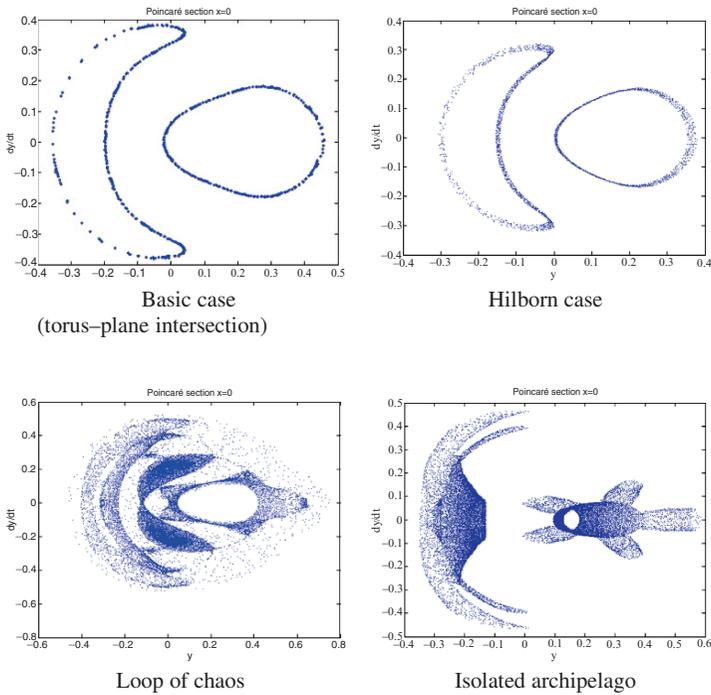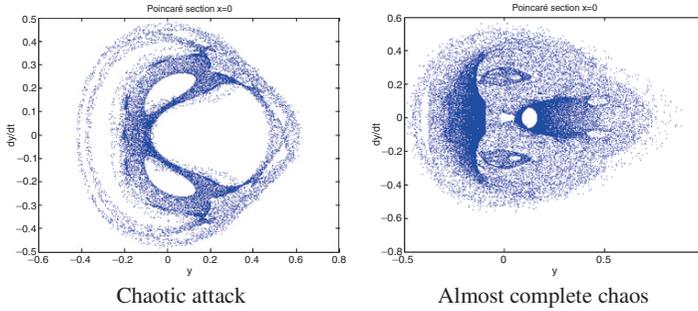
If we progressively change the initial conditions and increase E (which is assimilated to the energy of the model), the toric structures are partially replaced by chaos until a simultaneous coexistence.

1.18 Conservative and Dissipative Systems

A system is known as conservative if the "volume" of the phase space is preserved in the course of time. It is possible to study the contraction towards zero of the volume of the phase space without having to solve the differential system $\dot{x} = \frac{dx}{dt} = f(x)$, and this, by using the following theorem of the divergence (see Liouville's theorem and lemma about the divergence of a system).

Theorem 1.9 (Liouville Divergence). [18] *Given the flow ϕ_t of the system $\dot{x} = f(x)$, V a volume of phase space at time $t = 0$, $V(t) = \phi_t(V)$ the image of V by ϕ_t we have*

$$\frac{dV(t)}{dt} = \int_V Div \ f \ dx_1 \cdots dx_n, \quad with \quad Div \ f \equiv \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i} \tag{1.140}$$

Definition 1.24 (Conservative or dissipative system). A system is conservative if $dV/dt = 0$. A system is dissipative if $dV/dt < 0$.

In other words, a system is conservative if the flow associated with the differential system $\dot{x} = f(x)$ preserves the volume (or hyper-volume), i.e. by respecting:

$$\sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i} = 0. \tag{1.141}$$

In the case of a *Map* of the system, $x \longmapsto g(x)$, i.e. in the *state space*, we say that the system is conservative if:

$$|\det D_x g(x)| = 1, \tag{1.142}$$

---

[18] Theorem demonstration Dang-vu and Delcarte (2000).

where $|\det D_x \phi_t(x)| = 1$ with $D_x \phi_t = [\partial \phi_t^i / \partial x_i]$ is the matrix of partial derivatives. If under the action of the flow, or of the map, the volumes remain constant, we can say that these systems do not have *areas or regions of attraction in the phase space*. Thus, they cannot have asymptotically stable fixed-points, limit-cycles, or strange-attractors. However, the chaotic or very complex behaviors can exist in conservative models. This is the case in the Hamiltonian systems. The conservative systems are relatively rare in Economics. We will be able to quote, however, the application to cyclic economic growth problems, that Goodwin (1967) constructs from the A. Lotka and V. Volterra equations. Indeed, his works showed that such systems have a conservative behavior. One of the *applications* of the theorem above is related to the observation that all the *maps* are not always derived from "flows" generated by differential equations, it is thus sometimes necessary to preserve *the orientation of the flow*. Thus, it is necessary to control the behavior of this type of maps by

$$\det[D_x g(x)] > 0, \tag{1.143}$$

where $D_x g = [\partial g^i / \partial x_i]$ is the matrix of partial derivatives.

1.19 Hamiltonian and Optimal Growth Model

Definition 1.25 (Hamiltonian). A Hamiltonian system with n degrees of freedom is a system of equations of the motion of the form:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \tag{1.144}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad \text{with} \quad i = 1, 2, \ldots, n, \tag{1.145}$$

where $H = H(q, p, t)$ is the Hamiltonian. The phase-space of the system is $\mathbb{R}^{2n}$. Thus, it comes

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} + \sum_{i=1}^{n} \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} + \sum_{i=1}^{n} \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} \tag{1.146}$$

$$= -\sum_{i=1}^{n} \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} + \sum_{i=1}^{n} \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t}. \tag{1.147}$$

If the Hamiltonian H does not depend explicitly on time, it will remain identical in the course of time: $H(q, p) = \text{Constant}$.

1.19.1 The Optimal Growth Model with Infinite Horizon

The model can be formulated very briefly in the following way:

$$\max_{k} \int_0^\infty U(k,\dot{k})e^{-\sigma t}\,dt \tag{1.148}$$

with $(k,\dot{k}) \in S \subset \mathbb{R}^{2n}$, $k(0) = k_0$. U is the concave function of utility. k is the stock of capital, $\dot{k}$ is the net investment. The set S corresponds to the technological restrictions, and σ is the rate of "discount" ("actualization"). If we introduce an auxiliary vector $q \in \mathbb{R}^n$, then we can write

$$H(k,q) = \max_{k;(k,\dot{k})\in S} \{U(k,\dot{k}) + q\dot{k}\}, \tag{1.149}$$

where q can be assimilated to the "capital price". To solve our model, it must satisfy the following differential system:

$$\dot{k} = \partial H/\partial q, \tag{1.150}$$
$$\dot{q} = -(\partial H/\partial k) + \sigma q. \tag{1.151}$$

It can be assimilated to a Hamiltonian system with an additional element which is "σq" representing a perturbation. This system was frequently used by the economists to describe the mechanism of a Hamiltonian system, where $H(k,q)$ is the Hamiltonian function. The hypotheses concerning the *concavity* of U and of $H(k,q)$ condition the characteristics of obtained solutions. However, usually, a Hamiltonian function is regarded as concave in numerous fields. Consequently, the solution-points and their stability differ noticeably.

1.20 Torus and Combination of Basic Frequencies

The usual torus embedded in a three-dimensional space has a hole (see figures below). However, tori can also have multiple holes (the term n-torus is used for a torus with n holes). One of the more common uses of n-dimensional tori is in *dynamical systems. A fundamental result states that the phase space trajectories of a Hamiltonian system with n degrees of freedom and possessing n integrals of motion lie on an n-dimensional manifold which is topologically equivalent to an n-torus* (Tabor). (Moreover, we known that the quasi-periodic solution of a differential system can take the shape of a torus T^2.) The radius from the center of the hole to the center of the torus tube is R and the radius of the tube is l. The equation in Cartesian coordinates for a torus azimuthally symmetric about the z-axis is $(R - \sqrt{x^2 + y^2})^2 + z^2 = l^2$. The parametric equations are:

$$x = (R + l\cos\alpha)\cos\theta, \tag{1.152}$$
$$y = (R + l\cos\alpha)\sin\theta, \tag{1.153}$$
$$z = l\sin\alpha, \tag{1.154}$$

where R, l have constant lengths and $\alpha, \theta \in [0, 2\pi)$. $x = R\cos\theta$ and $y = R\sin\theta$ define a circle in the horizontal plane, while $u = l\cos\alpha$ and $v = l\sin\theta$ define another circle in a (u,v)-*plane*. This plane is mobile. Then, we write the cartesian equation (i.e. we have to remove α and θ). The two first equations above give $x^2 + y^2 = (R + l\cos\alpha)^2$, i.e. $l\cos\alpha = \sqrt{x^2 + y^2} - R$. By merging with the third one, we obtain: $z^2 + (\sqrt{x^2 + y^2} - R)^2 = l^2$.
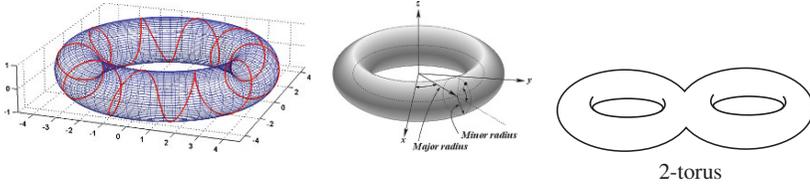
2-torus

If f_1 and f_2 represent both basic frequencies of the motion, then the torus is a combination of the frequencies f_1 and f_2. *The form of the Poincaré section of a solution torus T^2 depends on the ratio f_1/f_2.* When we observe the form of the intersection points of the torus and Poincaré section, we note that these points progressively describe a *closed curve* in the plane. Besides, this closed curve can cut itself according to the type of combination of basic frequencies f_1 and f_2.

(a) If *the ratio f_1/f_2 is irrational, the trajectory of the torus* (and not that of the curve on the Poincaré sectional surface) *is never closed* on itself *and covers the surface of the torus in a dense way*. When it intersects the Poincaré plane this torus shows a *continuous closed curve*.

(b) If *the ratio f_1/f_2 is rational, the trajectory is not dense on the surface of the torus*. The torus which intersects the "Poincaré sectional surface" exhibits *a curve which is not continuous*.

1.21 Quasiperiodic Route to Chaos (Ruelle Takens), and Landau T^n Tori

1.21.1 Description of Both Alternative Scenarios

The route towards chaos by quasi-periodicity is one of the scenarios showing the transition from the fixed point to chaos. It exhibits a series of bifurcations which lead to an increasingly complex dynamics. In order to simplify, we can consider that all the bifurcations are Hopf bifurcations. It is known that during a Hopf bifurcation, the *angle θ between two successive points* of the trajectory takes values between 0 and π : $\theta \subset [0, \pi]$. *If the angle is a rational fraction of π,* the trajectory passes by a finite number of points and *passes by the same point again* after a finite time.

This trajectory is strictly periodic. In the most general case where "the angle is not a rational fraction of π", *the cycle is densely covered*, and the trajectory never passes again by the same point. The trajectory is then known as "pseudo-periodic". We usually observe this type of trajectory after a first Hopf bifurcation; In each bifurcation, a new periodicity is superposed on the one already in place. According to the number of frequencies which are present in the spectrum, it is called a T^1 torus (i.e. a limit-cycle), a T^2 torus (i.e. two superposed cycles), a T^3 torus (i.e. three superposed cycles), etc. Thus, from a fixed point, two successive Hopf bifurcations lead to a dynamics which is called a T^2 torus. This torus corresponds to a *closed surface* in the phase-space, *densely covered* by the points of the dynamics. Thus we observe a scenario by successive bifurcations:

$$\text{Fixed point} \rightarrow \text{Limit cycle} \rightarrow \text{T}^2 \text{ Torus} \rightarrow \text{T}^3 \text{ Torus} \rightarrow, \text{etc.}$$

Nevertheless, as the dynamics becomes more and more complex, a tendency to the "coupling of frequencies" (or *resonance*) appears. When the dynamics is a T^2 torus generated by two limit cycles, the two involved cycles tend to enter in a rational ratio (ω_1/ω_2), i.e. to be *synchronized*. This tendency to coupling (or fixing) becomes increasingly strong as we change the control parameter in the direction which leads to more complexity. The transition from the torus to the coupling of frequencies does not constitute a bifurcation in a strict sense, because the dynamics is not qualitatively modified. Simply, the relationship between the two principal frequencies of the attractor is a rational ratio, and the torus is not densely any more traversed. The probability that the two frequencies are in a rational ratio increases as we approach the border of chaos (and is 1 at the border of chaos). The coupling of basic frequencies is well described with simple maps, such as the *Arnold map*. It precedes a new bifurcation which plunges the dynamics into a chaotic regime. The scenario of cascade of bifurcations is thus interrupted by the frequency coupling and the irruption of chaos. Thus, we have the following scenario:

$$\text{Fixed point} \rightarrow \text{Limit cycle} \rightarrow \text{T}^2 \text{ Torus} \rightarrow (\text{Coupling of frequencies}) \rightarrow \text{Chaos}$$

Figure 1.45 shows the Ruelle–Takens (1971) route to chaos, and the Landau (1944) route towards T^n torus. The *quasiperiodic route to chaos* was highlighted by Ruelle and Takens (1971). Furthermore, we know that there are different kinds of transition to reach chaos known as *canonical:*

1. "Period-doubling" *in unidimensional or multidimentional models*
2. "Intermittency" (i.e. explosive route) (Bergé et al. 1987)
3. "Saddle connection" or "Blue sky catastrophe", which is closely related to "transverse homoclinic orbits" (including obviously the *Smale horseshoe approach*) in theory and in practice.
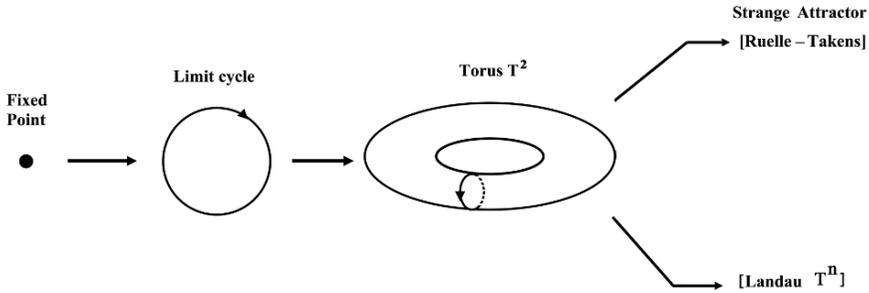
Fig. 1.45 Ruelle–Takens route to chaos (or Landau T^n torus)
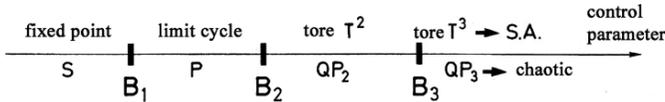
Fig. 1.46 S: Stationary regime. P: Periodic regime. QP_2: Quasiperiodic regime with two frequencies. QP_3: Quasiperiodic regime with three frequencies. SA: Strange attractor

1.21.2 Experimental Illustrations

1.21.2.1 Ruelle–Takens (R–T)

Ruelle–Takens theory (R–T) has been introduced in 1971, then detailed in 1978 by Ruelle–Takens–Newhouse (R–T–N). This theory called into question the previous Landau mechanism in which an infinity of Hopf bifurcations would be necessary to generate the turbulence. On the contrary, Ruelle–Takens have considered that a small number of bifurcations was sufficient to produce chaotic behaviors. The experiment used was that of Reynolds. Given a laminar flow, then by increasing the Reynolds number, the system loses its stability and becomes oscillating at the frequency f_1. The same process is repeated two times, thus we obtain successively three Hopf bifurcations at the frequencies f_1, f_2, f_3. Then according to R–T–N, the corresponding torus T^3 can become (under conditions) unstable and be replaced by a strange attractor. The behavior is no more quasiperiodic with three frequencies (tore T^3) but clearly chaotic. Here is the diagram of successive bifurcations B_1, B_2, B_3, which lead to the chaos of the Ruelle–Takens theory (see Fig. 1.46).

Experimental Illustration by the Rayleigh–Benard Instability (R–B)

In the Rayleigh–Bénard experiment, the fluid is the water whose temperature is so that the Prandtl number is equal to 5. The power spectrum of a quasi-periodic signal for three immeasurable frequencies observed in the R–B instability is shown. By increasing the ratio Ra/Ra_c the stationary convection loses its stability to become

periodic starting from Ra/Ra$_c$ = 30 (first Hopf bifurcation). For Ra/Ra$_c$ = 39.5 the previous periodic regime loses its stability to become quasiperiodic with two basic frequencies (second Hopf bifurcation). From Ra/Ra$_c$ = 41.5 a third bifurcation occurs with three basic frequencies. The frequency of each ray is indexed as $f = m_1 f_1 + m_2 f_2 + m_3 f_3$ (where m_i are integers) and cannot be indexed as $f = m'_1 f'_1 + m'_2 f'_2$. The incommensurability is verified because the ratios f_1/f_2, f_1/f_3, f_2/f_3 vary continuously with Ra/Ra$_c$. If there were synchronization (e.g. f_1/f_2 rational), then a threshold coupling would occur (i.e. f_1/f_2 would remain constant in a finite domain of Ra/Ra$_c$).
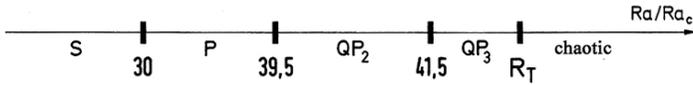
Beyond $R_T = $ Ra/Ra$_c \simeq 43$, the broad band noise becomes high, with a simultaneous widening of peaks, then the regime is chaotic. (Note that it is however difficult to distinguish the threshold R_T of chaotic regime and the one of the quasiperiodic regime which are sometimes viewed as confused.)

1.21.2.2 Transition to Chaos Starting from a Torus T^2 (C–Y)

Curry and Yorke Model

In the previous case, the torus T^3 is transformed into a strange attractor. The route to chaos can appear from the destabilization or destruction of a torus T^2 (with two frequencies). Taking into account the previous case, we could suppose that the dimension of T^3 is the minimal dimension to make possible the occurrence of strange attractors. But there is no incompatibility between both cases. Indeed, if the chaos occurs from a torus T^2, this is because there is another degree of freedom manifests itself, no longer in the form of a third frequency, but by a progressive abandonment of T^2 by the trajectories, which is equivalent to a destruction of this torus. The understanding of this route to chaos results from numerical experiments on a two-dimensional iterated map representing the Poincaré section of a tridimensional flow. This map must satisfy the following criteria:

- The map must be nonlinear.
- The map must be contracting (ref. to dissipative systems).
- The map must have a Hopf bifurcation.

    This Hopf bifurcation (for the tridimensional flow) by the Poincaré section corresponds to a limit cycle bifurcation. Curry and Yorke have created a model satisfying these criteria. This a composition of two homeomorphisms of $\mathbb{R}^2$: $\varphi = \varphi_1 \circ \varphi_2$. By using polar coordinates (θ, σ), the map φ_1 defines the $(k+1)$th iterate function of kth by: $\varphi_1 = \{\sigma_{k+1} = \varepsilon \log(1 + \sigma_k)$ and $\theta_{k+1} = \theta_k + \theta_0\}$ where ε is the control parameter. The map φ_2 (in cartesian coordinates) gives the map (x, y) as

relation between two successive iterates $\varphi_2 = \{x_{k+1} = x_k$ and $y_{k+1} = y_k + y_k^2\}$. $\varphi_1, \varphi_2, \varphi_1^{-1}, \varphi_2^{-1}$ are continuous and so φ an homeomophism. The map φ_1 corresponds to a *dissipative system*. The existence of *nonlinearities* result from the *logarithmic term* in φ_1 and from the *quadratic term* in φ_2. The following diagram summarizes the different stages of the route to chaos according to the Curry and Yorke model (see diagram).
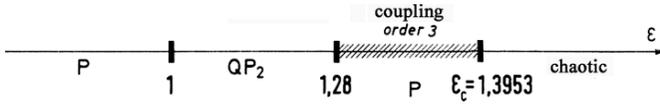
After the critical threshold ε_c a strange attractor occurs and the regime is chaotic (although there are windows of periodic behaviors).

1.21.3 Circle Map, Mode-Locking and Arnold Tongue

Let us define some essential concepts aiming to explain the Arnold tongue notion which is closely connected to the "quasiperiodic route to chaos". These concepts are the following: The *winding number, mode-locking* and *circle map*.

A *map winding number*[19] can be described as follows: The *winding number* $W(\theta)$ of a map $f(\theta)$ with an initial value θ is defined as follows:

$$W(\theta) = \lim_{n \to \infty} \frac{f^{(n)}(\theta) - \theta}{n}. \tag{1.155}$$

Such a number corresponds to the *average increase* in the *angle* θ per unit time (*average frequency*). A dynamical system with a *rational winding number* $W = p/q$ is *mode-locked*. By contrast, a dynamical system with an *irrational winding number* is *quasiperiodic*. Moreover, we know that the rational numbers are a set of zero measure on any finite interval, thus almost all winding numbers will be irrational, consequently *almost all maps will be quasiperiodic*.

A *mode locking* is a *phenomenon* in which a system being forced at an irrational period undergoes rational, periodic behavior which persists for a finite range of forcing values. This can happen for strong *couplings* between *natural* and *forcing oscillation* frequencies. Such a phenomenon can be explained by means of the *circle map* when, after iterations of the map, the *new angle differs from the initial value by a rational number*

$$\theta_{n+q} = \theta_n + p/q.$$

---

[19] *Winding number:* The number of times a closed curve in the plane passes around a given point in the counterclockwise direction.

This constitutes the *unperturbed circle map* with the *map winding number*

$$\Omega = p/q.$$

When Ω is not a *rational number*, the behavior is *quasiperiodic*.

The *circle map* is a *chaotic map* showing a number of interesting chaotic behaviors. The concept dates back to Andrey Kolmogorov and concerns a simplified system for driven mechanical rotors or also a simplified model of the phase-locked loop in electronics. The circle map exhibits *certain regions of its parameters where it is locked to the driving frequency* (phase-locking or mode-locking in electronics), these are called the *Arnold tongues*. The *circle map* was used among other things to study the dynamical behavior of electrocardiograms (EKG, ECG). The *circle map* is a *one-dimensional map* which maps a circle onto itself

$$\theta_{n+1} = f(\theta) = \theta_n + \Omega - \frac{K}{2\pi}\sin(2\pi\theta_n), \tag{1.156}$$

where θ_{n+1} is calculated *mod 1* and K is a constant. Ω and K are *two parameters*, where Ω can be regarded as an externally applied frequency, and K as a strength of *nonlinearity* (K : coupling strength, Ω : driving phase and may be interpreted as a driving frequency). The *circle map shows unexpected behaviors* as a function of parameters. It is connected to the *standard map* notion

$$I_{n+1} = I_n + \frac{K}{2\pi}\sin(2\pi\theta_n), \tag{1.157}$$

$$\theta_{n+1} = \theta_n + I_n + \frac{K}{2\pi}\sin(2\pi\theta_n), \tag{1.158}$$

for I and θ calculated *mod 1*. The writing of θ_{n+1} as follows

$$\theta_{n+1} = \theta_n + I_n + \frac{K}{2\pi}\sin(2\pi\theta_n) \tag{1.159}$$

gives the circle map with $I_n = \Omega$ and $K = -K$. The Jacobian of the circle map is

$$\frac{\partial\theta_{n+1}}{\partial\theta_n} = 1 - K\cos(2\pi\theta_n), \tag{1.160}$$

thus the *circle map is not area-preserving*. The *unperturbed circle map* is

$$\theta_{n+1} = \theta_n + \Omega. \tag{1.161}$$

In the case where Ω is *rational*, then it is known as the *map winding number* and Ω is written

$$\Omega = W \equiv p/q, \tag{1.162}$$

and *gives a periodic trajectory* because θ_n will return to the same point – at most – every q map orbits. In the case where Ω is irrational, then the behavior is quasiperiodic. If K is non-zero, then the behavior can be periodic in some finite region surrounding each rational Ω. This periodic behavior due to an *irrational forcing*

is known as *mode locking*. It is possible to represent the plane of the parameters (K, Ω) with the regions of periodic mode-locked parameter space plotted around rational Ω values (see map winding numbers), then the regions are seen to widen upward from 0 at $K = 0$ to some finite width at $K = 1$. *The region surrounding each rational number is known as an Arnold tongue*. For $K = 0$, the *Arnold tongues* are an *isolated set* of measure zero. For $K = 1$, they constitute a *Cantor set* of dimension $d \approx 0.08700$. For $K > 1$, the *tongues overlap* and the *circle map* becomes *non-invertible*. Let Ω_n be the parameter value of the circle map for a cycle with map winding number $W_n = F_n/F_{n+1}$ passing with an angle $\theta = 0$, where F_n is a Fibonacci number. Then the parameter values Ω_n accumulate at the rate (see Feigenbaum)
$$\delta \equiv \lim_{n \to \infty} \frac{\Omega_n - \Omega_{n-1}}{\Omega_{n+1} - \Omega_n} = -2.833.$$

Arnold tongue: In short, to specify the *Arnold tongue* concept let us consider the *circle map* again. In the case where K is non-zero, then the behavior is *periodic* in some finite region surrounding each rational Ω. Such periodic behavior due to *an irrational forcing* is known as *mode locking*. It is possible to represent, in the (K, Ω)-plane, the regions of periodic *mode-locked* parameter space plotted around rational Ω values (map winding number), then the regions are seen to widen upward from 0 at $K = 0$ to some finite width at $K = 1$.

The region surrounding each rational number is known as an Arnold tongue: For $K = 0$, the *Arnold tongues* are an *isolated set* of measure zero. For $K = 1$, they constitute a *Cantor set*. Usually, an *Arnold tongue* is known as a *resonance zone* emanating out from rational numbers in a two-dimensional parameter space.

Arnold tongues and Quasiperiodic route to chaos: The circle map is an essential concept to study the transition to chaos, in particular relating to the *quasiperiodic route to chaos*, and relating to Hopf or Neimark bifurcations of discrete-time dynamical systems. Let us consider in the complex plane, for a two-dimensional parameter system, the eigenvalue $\lambda(\gamma)$ of the Jacobian matrix J_γ. If one modifies only one parameter, this corresponds to moving along an arc in the complex (Im, Re)-plane (see Fig. 1.47b). In this complex plane, θ represents the position along the circum-
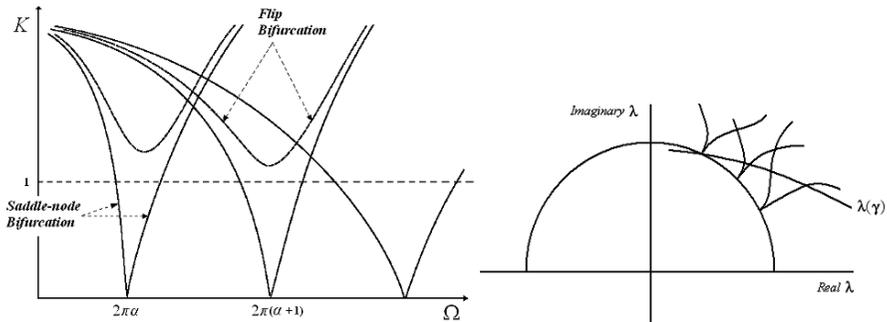
Fig. 1.47 (**a**) Arnold tongues in (k, Ω)-plane. (**b**) Arnold tongues in (Im,Re)-plane

ference of the circle and K corresponds to the normalized amplitude. Let us consider again a circle map

$$\theta_{n+1} = f(\theta) = \theta_n + \Omega - \frac{K}{2\pi} \sin(2\pi\theta_n), \qquad (1.163)$$

with mod 1. (In general we know that: $f(\theta + 1) = 1 + f(\theta)$; for $|K| < 1, f(\theta)$ is a diffeomorphism; for $|K| = 1$, $f^{-1}(\theta)$ is non-differentiable; for $|K| > 1$ there is no unique inverse of $f(\theta)$.) For such a circle map, let us describe the possible behaviors:

- If $K > 1$ the map $f(\theta)$ is *non-invertible* and *chaos can occur*, while knowing that chaotic and non-chaotic regions are intertwined in the (K, Ω)-plane.
- $K = 1$ is the *frontier between invertible and non-invertible circle maps*. The non-mode-locked intervals between Arnold tongues constitute a Cantor set with zero measure.
- If K is *small* (close to the bifurcation), the *circle map tends* to a *single attracting fixed-point* if Ω is a *rational number p/q* (same conclusion for more or less broad intervals of Ω close to p/q). The region surrounding each rational number is known as an Arnold tongue. The periodic windows are limited by two arcs of *saddle-node bifurcations* (see Fig. 1.47a). When each region widens upward starting from the bifurcation point, the regions or the windows tend to overlap themselves, and the more the coupling is strong, the more the *mode-locking phenomenon* tends to appear. Such a phenomenon is also known as the *synchronization* of frequencies.

The space of the parameters (K, Ω) is a very interesting framework to study the *transition to chaos*. In such a framework, either we vary only one parameter, or we vary two parameters simultaneously. *Firstly*, when we vary a parameter moving thus along an arc in the (K, Ω)-plane, the *mode-locking happens if K is increased beyond 1*. In such a case, the transition to chaos arises only by means of one of the canonical (codimension-1) routes to chaos,[20] i.e. the blue-sky catastrophe (saddle-connection), intermittency and period-doubling. *Secondly*, when we vary both parameters simultaneously K and Ω, a point on the line $K = 1$ can be attained, appertaining to the *non-mode-locking Cantor set*, outside the Arnold tongues. In this context, it is difficult to analyze the consequence of an increase in K beyond 1, as regards to the transition from quasiperiodic behavior towards chaos; although experiments seem to conclude that the transition can occur.

1.22 An Approach of KAM Theory: Invariant Torus and Chaos

The subject of the KAM theory[21] (Kolmogorov–Arnold–Moser) is often illustrated by the Hénon–Heiles model[22] which has this possibility to provide "invariant tori" and "chaos" simultaneously. The behaviors of the Hénon–Heiles model were

---

[20] See section entitled: "Transitions and Routes to Chaos".

[21] Dang-Vu and Delcarte (2000). Also cf. Verhulst (1996).

[22] Or by numerical calculations of Ollengren, Contopoulos.
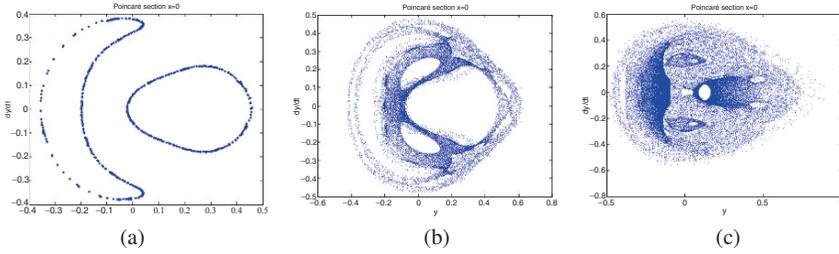
Fig. 1.48 (**a**) Torus–plane intersection. (**b**) Chaotic attack. (**c**) Chaos and tori persistence

described previously in the section about the attractors according to parameter setting on $E = H$ and initial conditions. In Fig. 1.48a ($E = 1/12$), the two closed curves correspond to the intersection of the (two-dimensional) torus surface with the Poincaré section. Figure 1.48b ($E = 1/8$) shows the chaotic attack of the toric structure. Figure 1.48c shows a structure with a chaotic dominant and the persistence of the toric structure.

The topics of KAM theory consider the Hamiltonian systems and their behaviors. A Hamiltonian system with n degrees of freedom ($ddl = n$) is a system of equations of the motion of the form:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \tag{1.164}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad \text{with} \quad i = 1, 2, \ldots, n, \tag{1.165}$$

where $H = H(q, p, t)$ is the hamiltonian, and it comes

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} + \sum_{i=1}^{n} \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} + \sum_{i=1}^{n} \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} \tag{1.166}$$

$$= -\sum_{i=1}^{n} \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} + \sum_{i=1}^{n} \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} + \frac{\partial H}{\partial t} \tag{1.167}$$

$$= \frac{\partial H}{\partial t}. \tag{1.168}$$

If H does not depend on time, it is preserved in the course of time, i.e. $H(q, p) =$ Constant. If the Hamiltonian is integrable, we can write the following general canonical transformation:

$$q_i = q_i(\theta_j, J_j), \tag{1.169}$$
$$p_i = p_i(\theta_j, J_j), \quad \text{with} \quad i, j = 1, 2, \ldots, n. \tag{1.170}$$

Thus, the new Hamiltonian system does not depend on θ anymore, and is written

$$\dot{\theta}_i = \frac{\partial H(J)}{\partial J_i} = \omega_i(J), \tag{1.171}$$

$$\dot{J}_i = -\frac{\partial H(J)}{\partial \theta_i} = 0. \tag{1.172}$$

This new system has the following solutions:

$$\theta_i = \omega_i t + \theta_i(0), \tag{1.173}$$

$$J_i = J_i(0). \tag{1.174}$$

J_i are constants and θ_i are the variables known as *angular*. The solutions of these models move on tori of the type T^n. The solution on the torus is known as *quasi-periodic* with n frequencies $\omega_1, \omega_2, \ldots, \omega_n$, if there is not n-tuple $m = (m_1, m_2, \ldots, m_n)$ with the $m_i \in \mathbb{Z}$ and $m_1 \omega_1 + \cdots + m_n \omega_n = 0$. The trajectory will not be closed and will cover the torus, and we will say that the trajectory is *dense*. On the other hand, if there is a n-tuple, the trajectory will be closed on itself and we will say that the trajectory is *periodic* (we will be able to find a m and a ω_0 such that $m\omega_0 = \omega$). *To simplify* the presentation, we will select a Hamiltonian with two degrees of freedom $(ddl = 2)$ which is written

$$H(J_1, J_2) = E = \text{ Constant.} \tag{1.175}$$

Such a Hamiltonian system thus has the following solutions-trajectories:

$$\theta_i = \omega_i(J_1, J_2)t + \theta_i(0), \quad \text{where } J_1 \text{ and } J_2 \text{ are constant.} \tag{1.176}$$

The trajectories will be on a torus of dimension 2, noted T^2. These trajectories will be

- *Periodic,* of *period* q, if $\alpha(J_1, J_2) = \dfrac{\omega_1(J_i)}{\omega_2(J_i)} = \dfrac{p}{q}$.
- *Dense* on the torus, if $\alpha(J_1, J_2)$ is *irrational*. $\left[\alpha(J_1, J_2)\text{ is called } the rotation number.\right]$

Let us imagine a Poincaré section[23] coming to cut the flow of trajectories, the Poincaré section can cut the trajectories at different time periods denoted $\Delta t = 2\pi/\omega_2$, and for this same period the angular variable θ_1 increases $\omega_1 \Delta t = 2\pi\alpha(J_1) = \Delta\theta_1$ (Fig. 1.49). We can pose the Poincaré map in the following way (with $\alpha'(J) \neq 0$):

$$P_0 : \begin{aligned} J_{n+1} &= J_n, \\ \theta_{n+1} &= \theta_n + 2\pi\alpha(J_{n+1}). \end{aligned} \tag{1.177}$$

This *map preserves the areas* because:

$$\frac{\partial(J_{n+1}, \theta_{n+1})}{\partial(J_n, \theta_n)} = \begin{vmatrix} 1 & 0 \\ 2\pi\alpha'(J_n) & 1 \end{vmatrix} = 1. \tag{1.178}$$

Now, in this integrable Hamiltonian $H(J)$, let us use a small perturbation δ which depends on (θ, J). The resulting Hamiltonian will be written

$$H(\theta, J) = H(J) + \delta H_1(\theta, J). \tag{1.179}$$

---

[23] Poincaré section: $\theta_2 = 0$.
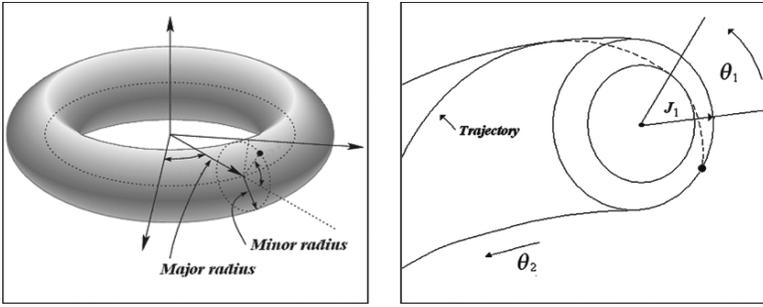
Fig. 1.49 A torus T^2 (*left*). Poincaré section of the torus (*right*)

Consequently, the Poincaré map posed above is written

$$P_\delta : \begin{aligned} J_{n+1} &= J_n + \delta f(J_{n+1}, \theta_n), \\ \theta_{n+1} &= \theta_n + 2\pi\alpha(J_{n+1}) + \delta g(J_{n+1}, \theta_n), \end{aligned} \qquad (1.180)$$

with the functions $f(J, \theta)$ and $g(J, \theta)$ which are taken as periodic of period 2π. The conservation property of areas (surfaces) described above must be respected, we must have

$$\left| \frac{\partial(J_{n+1}, \theta_{n+1})}{\partial(J_n, \theta_n)} \right| = 1. \qquad (1.181)$$

i.e. the functions f and g must satisfy

$$\frac{\partial f}{\partial J} + \frac{\partial g}{\partial \theta} = 0. \qquad (1.182)$$

1.22.1 KAM Torus: Irrational Rotation Number

Let us consider the case where $f(J, \theta) = f(\theta)$ and $g(J, \theta) = 0$, consequently the condition $\partial f/\partial J + \partial g/\partial \theta = 0$ is verified, and thus the Poincaré map of the perturbated Hamiltonian is written

$$P_\delta : \begin{aligned} J_{n+1} &= J_n + \delta f(\theta_n), \\ \theta_{n+1} &= \theta_n + 2\pi\alpha(J_{n+1}). \end{aligned} \qquad (1.183)$$

The question now is to find the conditions making that this map has an "invariant orbit":

$$\text{Invariant orbit: } J = J(\theta). \qquad (1.184)$$

If such is the case, the trajectory obtained will be spread on a torus called the *KAM torus*. Thus, the map P_δ becomes

$$J(\theta + 2\pi\alpha) = J(\theta) + \delta f(\theta). \qquad (1.185)$$

This map can be solved by using the *Fourier series* expression applied to f and J:

$$f(\theta) = \sum_{-\infty}^{+\infty} a_k e^{\theta k i}, \tag{1.186}$$

$$J(\theta) = \sum_{-\infty}^{+\infty} b_k e^{\theta k i}, \tag{1.187}$$

where a_k and b_k are the *Fourier coefficients*:

$$J(\theta + 2\pi\alpha) = J(\theta) + \delta f(\theta), \tag{1.188}$$

$$\sum_{-\infty}^{+\infty} b_k e^{(\theta + 2\pi\alpha)ki} = \sum_{-\infty}^{+\infty} b_k e^{\theta k i} + \delta \sum_{-\infty}^{+\infty} a_k e^{\theta k i}. \tag{1.189}$$

We can extract the expression of the Fourier coefficient b_k function of a_k with $k \neq 0$:

$$b_k = \frac{\delta a_k}{e^{2\pi k i \alpha} - 1}. \tag{1.190}$$

The *invariant orbit* $J = J(\theta)$ becomes

$$J(\theta) = b_0 + \sum_{k \neq 0} \frac{\delta a_k e^{\theta k i}}{e^{2\pi k i \alpha} - 1}. \tag{1.191}$$

It is thus advisable to check the convergence of the expression above now. If we consider that f is of class C^N, knowing that the absolute value of the nth derivative of f with respect to the angular variable θ is lower or equal to a bound noted M:

$$\left| d^N f / d\theta^N \right| \leq M. \tag{1.192}$$

We can write:[24]

$$|a_k| \leq \frac{M}{|k|^N}. \tag{1.193}$$

The rotation number $\alpha(J)$ must be irrational to observe a *KAM* torus, i.e. there must exist constants $c > 0$ and $m \geq 2$ which satisfy the following conditions:[25]

$$\text{Conditions:} \quad \left| \alpha - \frac{l}{k} \right| \geq \frac{c}{k^m}, \tag{1.194}$$
$$\text{i.e.} \quad |k\alpha - l| \geq \frac{c}{k^{m-1}}.$$

It is known that $(\sin x \geq 2x/\pi)$ when $0 \leq x \leq \pi/2$, consequently we can write the following succession of equalities and inequalities:

$$\left| e^{2\pi k i \alpha} - 1 \right| = 2 \left| \sin \alpha \pi k \right| = 2 \left| \sin \pi (\alpha k - l) \right| \geq 4 |k\alpha - l| \geq \frac{4c}{k^{m-1}}. \tag{1.195}$$

---

[24] Ref: Weierstrass approximation theorem: Any continuous function is uniformly approached by polynomial functions.

[25] With $k \in \mathbb{Z}^+$ and $l \in \mathbb{Z}$.

Thus starting from the expression which precedes and from the expressions described above $b_k = \dfrac{\delta a_k}{e^{2\pi k i \alpha} - 1}$ and $|a_k| \leq \dfrac{M}{|k|^N}$, we deduce for $k \neq 0$:

$$|b_k| \leq \frac{M_\delta}{4c\,|k|^{N-m+1}}. \tag{1.196}$$

Consequently

$$J(\theta) = b_0 + \sum_{k \neq 0} \frac{\delta a_k e^{\theta k i}}{e^{2\pi k i \alpha} - 1}$$

converges if $N - m > 0$. Since $m \geq 2$, it is thus for a value of $N : N > 2$, that $J(\theta)$ converges and that we can observe an orbit of *KAM*. Thus, when the *conditions* described above $\left|\alpha - \dfrac{l}{k}\right| \geq \dfrac{c}{k^m}$ or $|k\alpha - l| \geq \dfrac{c}{k^{m-1}}$ are *satisfied*, we observe the *presence of a KAM torus*. If the *conditions are not satisfied*, the *torus will demolish itself*. The ultimate torus which disappears is the torus having the rotation number according to

$$\alpha(J) = \frac{\omega_1(J)}{\omega_2(J)} = \frac{(\sqrt{5} - 1)}{2}. \tag{1.197}$$

The numerical value of $\alpha(J)$ thus plays an important part in the nature of solutions of the system. The approach of these numerical values opened developments concerning the "algebraic number" concept, which will not be evoked here. In the framework of this paragraph, we will not describe either the case where the rotation number $\alpha(J) = p/q$ is rational, for which there is *resonance* between p and q, which leads to the description of groups of *Poincaré–Birkhoff points* with *hyperbolic* or *elliptic* forms, which will be described later in the Hyperbolicity framework.

1.23 Approach of Dynamical Systems by Means of Pendulums and Oscillators

A *simple pendulum (also called a bob pendulum)* is a pendulum consisting of a single spherical (or point) mass attached to a wire of negligible weight. This simple gravity pendulum[26] will swing back and forth under the influence of gravity over its central (lowest) point (Fig. 1.50).

A *physical pendulum* is a *generalization* of the *simple* pendulum. An example would be a bar rotating around a fixed axle. A simple pendulum can be taken as a specific case of a physical pendulum with moment of inertia. Usually, the system can be mathematically described as follows:

$$I = ml^2, \tag{1.198}$$

---

[26] Pendulum: Plural can be written *pendulums* or *pendula*.
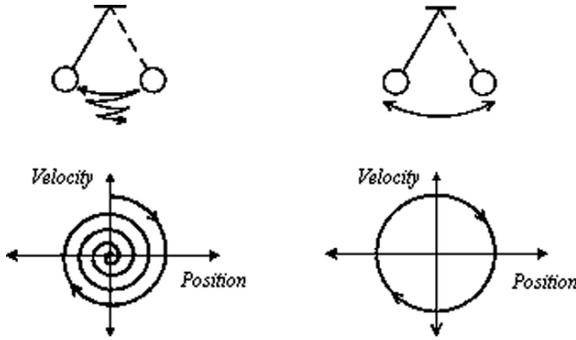
Fig. 1.50 Damped and undamped pendulums

where m is the mass and l is the wire length. The equation of motion of a physical pendulum can be found from the torque Υ on it,

$$\Upsilon = -mgl\sin(\omega) = I\beta = I\frac{d^2\omega}{dt^2},\tag{1.199}$$

where g is the gravitational acceleration, I is the moment of inertia, β is the angular acceleration, and ω is the angle of the wire measured from the downward vertical. Then

$$\frac{d^2\omega}{dt^2} + \frac{mgl}{I}\sin(\omega) = 0.\tag{1.200}$$

It is possible to define the *resonant frequency* as follows:

$$\theta_0 = \sqrt{\frac{mgl}{I}}\tag{1.201}$$

then the equation $\frac{d^2\omega}{dt^2} + \frac{mgl}{I}\sin(\omega) = 0$ can be written in the following simple form:

$$\ddot{\omega} + \theta_0^2\sin(\omega) = 0.\tag{1.202}$$

Two coupled first-order ordinary differential equations which follow make it possible to write the *equation of motion*:

$$\dot{x} = y,\tag{1.203}$$
$$\dot{y} = -\theta^2\sin(x).\tag{1.204}$$

This system of two equations provides the *phase portrait* as it is possible to observe in Fig. 1.51.
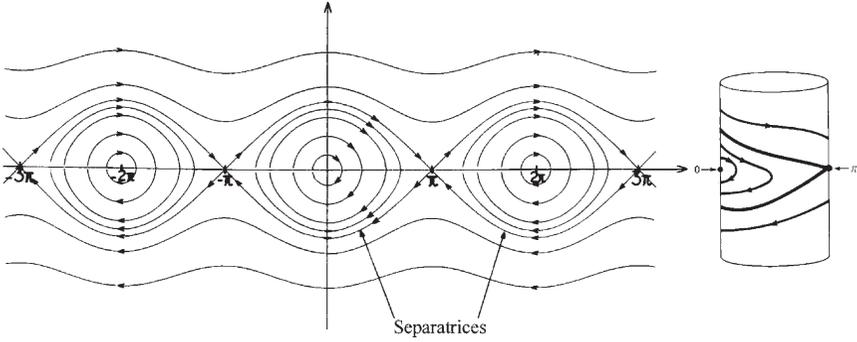
Fig. 1.51 Phase portrait of the pendulum

The *equation of motion* can also be written through the energy point of view. The gravitational potential energy and the kinetic can be written

$$V = -mgl \ \cos(\omega), \qquad (1.205)$$
$$T = (1/2)I\dot{\omega}^2. \qquad (1.206)$$

$V\left(\frac{1}{2}\pi\right) = 0$ is the zero-point of potential energy, and the sign has been chosen so that the potential energy is higher when the "bob" is higher. Then the *Lagrangian* can be written as follows:

$$\pounds = T - V = (1/2)I\dot{\omega}^2 + mgl \ \cos(\omega). \qquad (1.207)$$

If we calculate the derivatives, it comes

$$\frac{\partial \pounds}{\partial \dot{\omega}} = I\dot{\omega}, \qquad (1.208)$$

$$\frac{d}{dt}\left(\frac{\partial \pounds}{\partial \dot{\omega}}\right) = I\ddot{\omega}, \qquad (1.209)$$

$$\frac{\partial \pounds}{\partial \omega} = -mgl \sin(\omega). \qquad (1.210)$$

Then, the (Euler) differential equation can be written almost as at the beginning of the paragraph:

$$\ddot{\omega} + \theta_0^2 \sin(\omega) = \ddot{\omega} + \frac{mgl}{I} \sin(\omega) = 0. \qquad (1.211)$$

Moreover, using the *Hamiltonian* writing, the momentum is written

$$\psi_\omega = \frac{\partial \pounds}{\partial \dot{\omega}} = I\dot{\omega}, \qquad (1.212)$$

then the Hamiltonian is written

$$H = \dot{\omega}\psi_\omega - \pounds = (1/2)I\dot{\omega}^2 - mgl\,\cos(\omega) \tag{1.213}$$

$$= \frac{\psi_\omega^2}{2I} - mgl\,\cos(\omega), \tag{1.214}$$

then the motion equation are written

$$\dot{\omega} = \frac{\partial H}{\partial \psi_\omega} = \frac{\psi_\omega}{I} \tag{1.215}$$

$$\dot{\psi}_\omega = -\frac{\partial H}{\partial \omega} = -mgl\sin(\omega). \tag{1.216}$$

Remark 1.5 (Simple harmonic oscillator). The equations representing the simple harmonic oscillator as simply as possible are written as follows

$$\dot{x} = y, \tag{1.217}$$
$$\dot{y} = -\theta^2 x. \tag{1.218}$$

Such a system can provide multiple curves of motions in a phase portrait, which correspond to the multiple initial conditions of the system. Then the multiple "phase curves" corresponding to different initial conditions are represented in the same "phase plane" which gives the "Phase portrait" for a simple harmonic oscillator with various initial conditions (Fig. 1.52).

Remark 1.6 (Phase portrait). A phase portrait can be understood as a plot of the different phase curves corresponding to the different initial conditions in the same phase plane.[27]
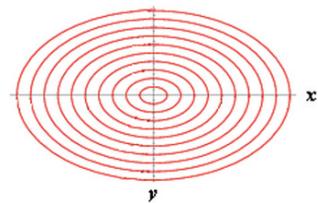
Fig. 1.52 A simple harmonic oscillator

---

[27] Phase plane: As seen previously, a phase plane is a phase space with two dimensions, i.e. corresponding to a map or a function with two degrees of freedom.

1.24 Navier–Stokes Equations of Flows, Attractors and Invariant Measures

1.24.1 Navier–Stokes Equations: Basic Model

The Navier–Stokes equations initially have sought to model the fluid flows (see Reynolds experiment) and their behaviors. According to the physical parameters of the experiment, these flows can be *laminar* or *convective and turbulent*. We give an illustration of a convective and turbulent flow hereafter (Fig. 1.53).

A reproduction of the Reynolds experiment makes it possible to highlight the Reynolds number r_e which corresponds to a critical value: $r_e = (VD)/v_{isc}$, where V is the speed, v_{isc} is viscosity, and D is the diameter of the tube receiving the flow. For $r_e \leq 2{,}300$, the device exhibits a laminar flow, we speak of a *laminar regime* and for $r_e > 2{,}300$ the device exhibits a turbulent flow, we speak of a *turbulent regime*. Hereafter, a picture of a real turbulent regime (Fig. 1.54).

Generally the movement of fluids in $\mathbb{R}^2$ or $\mathbb{R}^3$ is defined by a function $t \rightarrow v(t)$, where $v(t)$ are the speeds which belong to the *fields of flow speeds* (the speed of fluid is related with its viscosity). For a measure of a flow of a tridimensional fluid, Navier–Stokes equation can be written in the following way:

$$\frac{\partial v}{\partial t} + (v \cdot \nabla)v = -\nabla p + \frac{1}{r_e}\nabla^2 v, \tag{1.219}$$

with r_e the *number of Reynolds* and p the *hydrostatic pressure* (∇ partial derivatives, gradient). The solution of this equation with *nonlinear partial derivatives* provides

Fig. 1.53 Illustration of a convective and turbulent flow
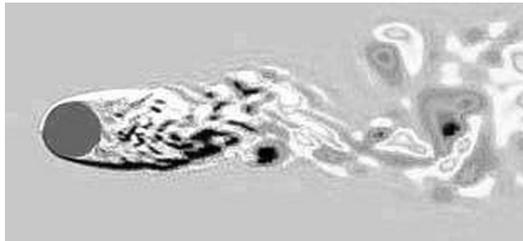
Fig. 1.54 Experiment of a real turbulent flow

a "field of speed" $v(x,t)$. We can observe the movements of microscopic elements (of particles) of the fluid. The movements of fluid particles are influenced by the convection phenomenon due to the field of speeds and are influenced by the molecular phenomenon of diffusion. The movements of (fluid) particles are influenced by the *convection* phenomenon, defined by the equation:

$$\dot{x} = v(x,t), \quad x \in \mathbb{R}^3. \tag{1.220}$$

The field of speeds can be given starting from a *function of the flow* denoted $\psi(x_1,x_2,t)$ with

$$v(x_1,x_2,t) = \left(\frac{\partial \psi}{\partial x_2}, -\frac{\partial \psi}{\partial x_1} \right). \tag{1.221}$$

Therefore, the equations of the motion of fluid particles become

$$\dot{x}_1 = \frac{\partial \psi}{\partial x_2}(x_1,x_2,t), \tag{1.222}$$

$$\dot{x}_2 = -\frac{\partial \psi}{\partial x_1}(x_1,x_2,t). \tag{1.223}$$

These equations constitute a *Hamiltonian system* where the function ψ is the Hamiltonian (H). This type of problems has opened the path to many developments and is central as regards fluid flows. An example of a flow function can be given by

$$\psi(x_1,x_2,t) = \psi_0(x_1,x_2) + \theta \psi_0(x_1,x_2,t), \tag{1.224}$$

with

$$\psi_0(x_1,x_2) = -x_2 + r_e \cos(x_1)\sin(x_2),$$
$$\psi_1(x_1,x_2,t) =$$
$$\frac{\gamma}{2}\left[\left(1 - \frac{2}{\lambda}\right)\cos(x_1 + \lambda t + \omega) + \left(1 + \frac{2}{\lambda}\right)\cos(x_1 - \lambda t - \omega)\right]\sin(x_2).$$

$\lambda > 0$, ω is a phase, and (r_e, γ, θ) are parameters (amplitude depending on the temperature) with $0 < \theta < 1$. The fluid particles motion equations are written

$$\dot{x}_1 = \frac{\partial \psi_0}{\partial x_2}(x_1,x_2) + \theta\frac{\partial \psi_1}{\partial x_2}(x_1,x_2,t), \tag{1.225}$$

$$\dot{x}_2 = -\frac{\partial \psi_0}{\partial x_2}(x_1,x_2) - \theta\frac{\partial \psi_1}{\partial x_2}(x_1,x_2,t). \tag{1.226}$$

1.24.2 Navier–Stokes Dynamics: Invariant Ergodic Measures, Characteristic Exponents and Hilbert Spaces

The differentiable Dynamics allowed to better foresee chaotic phenomena. In physics (aerodynamics or hydrodynamics), they also allowed in particular in Fluid Dynamics to approach turbulence phenomena which are very complex to define and to describe. It is possible to approach the turbulence phenomenon in dynamical systems by means of the *Mode* concept. The modes can be assimilated to *a collection of independent oscillators* . Each mode is periodic and each state is represented by an *angular variable*. In such a representation, a global dynamical system is *quasiperiodic*, i.e. it corresponds to *a superposition of periodic motions. A dissipative system becomes increasingly turbulent when the number of modes exciting the system increases*, i.e. when the number of oscillators necessary to describe the system increases. This analysis method which is in accordance with the *intuition* and *Fourier analysis* (i.e. the frequential decomposition) was *changed with the nonlinear dynamical systems*. In particular, the concept of the number of *excited modes* is replaced by new concepts, as the number of nonnegative characteristic exponents or the "dimension of information".

For example in the Fourier analysis, *the power spectrum* of a *Strange Attractor* is known as *continuous*, and we analyze this observation as corresponding to *an infinite number of modes*, i.e. with *an infinite collection of independent oscillators*. But we underlined previously that this is a way of reasoning which is only valid in *Linear theory, which must necessarily then take place in a phase space of an infinite dimension*, since we have an *infinite number of oscillators* and each one of these oscillators is spread in its space. Consequently, if we have to face a *continuous power spectrum*, there are two possibilities:

- Either we are studying a *system which visits an infinite number of dimensions in the phase space (such a reasoning* is valid only in *Linear theory)*.
- Or we have to face a *Nonlinear evolution of an attractor of finite dimension*.

Two realities are possible and the second case is frequent in practice.
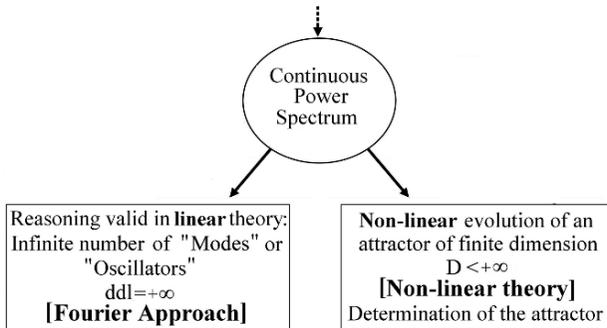
There exist methods that allow to discriminate the *effective dimension* of the attractor that we have to face. And it was shown that for hydrodynamic systems (see Grassbeger and Procaccia 1983b), although the system is of an infinite dimension (an infinite number of oscillators and a infinite number of degree of freedom), its effective dimension (of the attractor) is finite.

1.24.2.1 Invariant Ergodic Measure

One of the techniques with which we measure the dimension but also other dynamic quantities of a system is the *Ergodic theory*. This theory says that the *temporal average* is equal to the *spatial average*. The measure of such an average is invariant, we can write

$$\rho[f^{-t}(E)] = \rho(E), \quad t > 0, \tag{1.227}$$

where E is a subset of points of $\mathbb{R}^m$ and $f^{-t}(E)$ is a unit obtained by the evolution in the course of time t of each point in E. There can be many invariant measures in the dynamical systems but all are not relevant.

An invariant probability measure ρ is decomposable in different parts and each part is itself invariant. *Otherwise* ρ is *indecomposable* or *ergodic*. It is important to consider that in the dynamical systems which interest us the measure is not only invariant but also Ergodic, i.e. if ρ is ergodic, then the *ergodic theorem* affirms that for any function φ:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi[f^{-t}x(0)]dt = \int \rho(dx)\varphi(x), \tag{1.228}$$

for almost all the initial conditions $x(0)$ respects the measure ρ.

1.24.2.2 Invariant Measure of Probability, Dirac Delta Function and Attractor

An attractor A, strange or not, gives a global image of the long term behavior of a dynamical system. A way of specifying this image is to observe the measure of probability ρ of A, which describes how different parts of A are visited (in particular the frequency of visit) by an orbit $t \to x(t)$ which defines the system.

ρ can be defined like the temporal average of Dirac delta functions $\delta_{x(t)}$ at the points $x(t)$,

$$\rho = \lim_{t \to \infty} \frac{1}{T} \int_0^T dt \, \delta_{x(t)}, \tag{1.229}$$

For a continuous function φ we can write ρ:

$$\rho(\varphi) = \int \rho(dx)\varphi(x) \tag{1.230}$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \, \varphi[x(t)]. \tag{1.231}$$

The measure is invariant under the action of the dynamical system, i.e. invariant in the course of time. This invariance can be expressed in the following way:

$$\rho(\varphi \circ f^t) = \rho(\varphi). \tag{1.232}$$

1.24.2.3 Characteristic Exponent and Floquet Theory

If we evoke the Floquet theory and the stability of solutions of a system, we remember that after the action of the *map T* (i.e. after one period T), the initial point P_0 plus a small variation called δ, i.e. $P_0 + \delta$, is near P_0. The spacing between the two points is written: $T(P_0 + \delta) - P_0 \simeq M\delta$ with δ which tends obviously towards zero. The behavior of eigenvalues of the matrix M defines the stability of solutions. If we successively apply the map T a number of times equal to p (i.e. after p times one period T), then we can write: $T^p(P_0 + \delta) - P_0 \simeq M^p\delta$, the matrix M *eigenvalues* are also called *characteristic exponents*. M can be also expressed in the $Df(\bar{x})$ form, where $\bar{x}$ is the solution of the system. (For a linearized system) it is necessary to study the stability of solutions by analyzing the eigenvalues of the matrix e^{TR}, which is called the "characteristic Multiplier of Floquet", because (we remember that) the solutions of the new linearized system are written: $y = e^{TR}y_0$, where R is the matrix which determines the system of which it is necessary to study the eigenvalues to analyze stability).

1.24.2.4 Characteristic Exponent, Ergodic Measure and Measure of Entropy

Let f be a diffeomorphism and μ an *ergodic measure* (for f) with a compact support. We note $T_x f$ the matrix of partial derivatives, if f is a diffeomorphism of $\mathbb{R}^m$ (respectively the tangent map of f on $\mathbb{R}$). Then, by taking into account the multiplicative ergodic theorem of Oseledec, we can decompose the space $\mathbb{R}^m$ (resp. tangent space) into a direct sum $W_n^{(1)} \oplus \cdots \oplus W_n^{(s)}$ such that:

$$\lim_{k \to \pm\infty} \frac{1}{k} \log \left\| T_x f^k u \right\| = \lambda^{(r)}, \quad \text{if } u \in W_x^{(r)}. \tag{1.233}$$

The number $\lambda^{(r)}$ is called *characteristic exponent*. And the largest characteristic exponent denoted $\lambda^{(s)}$ satisfies

$$\lim_{k \to +\infty} \frac{1}{n} \log \| T_x f^n \| = \lambda^{(s)}, \tag{1.234}$$

since one supposes μ *ergodic,* the characteristic exponents are constant. If it is considered that the asymptotic behavior of the orbit $(f^n x)$ is described by the measure μ, and if the largest characteristic exponent $\lambda^{(s)}$ is strictly positive, then we are faced with a sensitivity to the initial conditions. If $\lambda^{(s)}$ is strictly negative, then the support of μ is an attracting periodic orbit.

We can reformulate the characteristic exponent in the following way. Given an ergodic invariant measure ρ, the following equation is considered

$$x_{n+1} = f(x_n), \tag{1.235}$$

where $x_i \in \mathbb{R}$, then one considers a small perturbation $\delta_x(0)$ and two initial points noted $x(0)$, $x(0)'$. The distance between these two points after a time N is written

$$x(N) - x(N)' = f^N(x(0)) - f^N(x(0)') \tag{1.236}$$

$$\approx \left(\frac{d}{dx}(f^N)(x(0)) \right) (x(0) - x(0)'), \tag{1.237}$$

with $f^N(x) = f(f(\cdots f(x) \cdots))$ and with $\frac{d}{dx}(f^N)(x(0)) = \frac{d}{dx}f(x(N-1)) \times \frac{d}{dx}f(x(N-2)) \cdots \frac{d}{dx}f(x(0))$.

In the case of m variables, i.e. for $x \in \mathbb{R}^m$, we replace, to describe the system, the derivative $(d/dx)f$ by the Jacobian matrix evaluated in x : $D_x f = (\partial f_i / \partial x_j)$. If it is supposed that all the factors in the expression above have similar sizes, we can imagine that df^N/dx increases or decreases *exponentially* with N. We can define the average rate of growth in the following way:

$$\lambda = \lim_{N \to +\infty} \frac{1}{N} \log \left| D_{x(0)} f^N \delta_x(0) \right|, \tag{1.238}$$

With the Oseledec theorem, this limit exists for almost all $x(0)$ respecting the invariant measure ρ. If ρ is *ergodic,* the *largest* λ is independent of $x(0)$. This number λ_1 is called the *largest Lyapunov exponent* of the map f which respects the measure ρ. The Lyapunov exponents, i.e. the characteristic exponents, give useful limits about the dimension of attractors.

Consider the quantity $h(\rho)$ which is the *Entropy measure* of ρ. This quantity can be limited in terms of characteristic exponents in the following way: $h(\rho) \leq \sum$ positive characteristic exponents, the more often in practice we have (Pesin formula):

$$h(\rho) = \sum \text{positive characteristic exponents.}$$

1.24.2.5 Haussdorff Dimension

The Haussdorff dimension is a general term to evoke the various mathematical definitions of the dimension for fractals sets.[28] Usually, it is said that the dimension of a set commonly corresponds to the quantity of information necessary to specify its points with enough precision.

---

[28] See definition of the Haussdorff dimension in the appendix.

Let Q be a compact set and we suppose that $N(\varepsilon)$ balls of radius ε are necessary to cover Q. Consequently a dimension $dim_D Q$, also called capacity of Q, is written

$$\dim_D Q = \lim_{\varepsilon \to 0} \sup \log \frac{N(\varepsilon)}{|\log(\varepsilon)|}. \tag{1.239}$$

This quantity is a bit weaker than $N(\varepsilon)\varepsilon^d \to finite$, which means that the volume of the set Q is finite of size d. By retaking the Takens theorem,[29] we can write

$$m \geq 2 \dim_D Q + 1, \tag{1.240}$$

m being real co-ordinates (or the dimension of the reconstructed phase space).

To express the *Haussdorff dimension*: $\dim_H Q$, we do not suppose that Q is compact. Moreover, we determine the "information dimension" of a probability measure $\rho : \dim_H \rho$, corresponding to the minimum of the Haussdorff dimension of sets Q for which $\rho(Q) = 1$. Consider a ball $B_x(r)$ of radius r in x and the measure ρ. Then let us suppose that the following limit exists

$$\lim_{r \to 0} \frac{\log \rho[B_x(r)]}{\log(r)} = \xi. \tag{1.241}$$

The existence of this limit implies that it is constant by the ergodicity of ρ. And ξ is equal to the dimension of information $\dim_H \rho$. In practice, we choose N regularly spaced points x_i and we estimate $\rho[B_{x_i}(r)]$, with N large, by[30]

$$\frac{1}{N} \sum_{j=1}^{N} \Theta \left[r - |x_j - x_i| \right]. \tag{1.242}$$

The expression above is also written

$$C(r) = \frac{1}{N^2} \sum_{ij}^{N} \Theta \left[r - |x_j - x_i| \right] \tag{1.243}$$

and the *information dimension* becomes

$$\lim_{r \to 0} \frac{\log C(r)}{|\log(r)|}. \tag{1.244}$$

1.24.2.6 Navier–Stokes Equations and Haussdorff Dimensions, Exponents and Entropy

We are interested in the measures, the limits and the relationships between Haussdorff dimensions, Lyapunov exponents and Entropy. In this framework, consider the

---

[29] See Takens theorem, and the one of Manes, or Broomhead and King.

[30] Where $\Theta(\cdot)$ is a simple function, cf. Grassberger and Procaccia (1983a, 1983b). $\Theta(\cdot) = [1 + sgn \cdot]/2$.

Navier–Stokes equations, and their conditions (d is the density which is constant, v the kinematic viscosity of the device, (v_i) the field of speeds belonging to Ω, p the hydrostatic pressure and g an external field of force):

$$\frac{\partial v_i}{\partial t} = -\sum_j v_j \partial_j v_i + v \triangle v_i - \frac{1}{d} \partial_i p + g_i, \qquad (1.245)$$

$$\sum_j \partial_j v_j = 0. \qquad (1.246)$$

$\sum_j \partial_j v_j = 0$ are the conditions of incompressibilities of the fluid.

We consider these same Navier–Stokes equations in a limited domain $\Omega \in \mathbb{R}^d$ with $d = 1$ or $d = 2$. For any *invariant measure* ρ, we have the following relationships between the *dissipation of energy* ε and the *ergodic quantities*:[31]

$$h(\rho) \le \sum_{\lambda_i \ge 0} \lambda_i \le \frac{B_d}{v^{1+d}} \left\langle \int_\Omega \varepsilon^{(2+d)/4} \right\rangle \qquad (1.247)$$

$$\dim_{H\rho} \le B'_d \frac{|\Omega|^{2/(d+2)}}{v^{d/2}} \left\langle \int_\Omega \varepsilon^{(2+d)/4} \right\rangle^{d/(d+2)} \qquad (1.248)$$

B_d and B'_d are constant and $h(\rho)$ the measure of entropy.

1.24.2.7 Nonlinear Operators, Navier–Stokes Equations and Hilbert Space

If we select the usual map which describes in discrete time the evolution of a dynamics:

$$x_{n+1} = f(x_n), \qquad (1.249)$$

that we can write in a continuous mode:

$$\frac{dx}{dt} = F(x), \qquad (1.250)$$

and we introduce a nonlinear operator of the temporal evolution of the dynamics f^t with $t \in \mathbb{R}$ or $\mathbb{N}$, such that $t \ge 0$; with the properties $f^0 = identity$, $f^s f^t = f^s + f^t$. The variable evolves in the phase space noted M which is $\mathbb{R}^m$, but which can be of infinite dimension, or which can be a "manifold", such as a toric or spherical structure for example. In this connection, the Banach or Hilbert spaces are of a major use in Fluid Mechanics (aerodynamics or hydrodynamics). If the space M is linear, we define the linear operator $D_x f^t$ which corresponds to the matrix of partial derivatives of f at the point x. And if we write $f^1 = f$, we obtain

$$D_x f^n = D_{f^{n-1}x} f \cdots D_{fx} f D_x f. \qquad (1.251)$$

---

[31] Ruelle (1982b, 1984) and cf. Lieb (1984).

If we consider again the case of Navier–Stokes equations and their associated conditions (d is the density which is constant, v the viscosity of the system, (v_i) the field of speeds belonging to à Ω, p the pressure and g an external field of force):

$$\frac{\partial v_i}{\partial t} = -\sum_j v_j \partial_j v_i + v \triangle v_i - \frac{1}{d}\partial_i p + g_i, \qquad (1.252)$$

$$\sum_j \partial_j v_j = 0. \qquad (1.253)$$

The condition $\sum_j \partial_j v_j = 0$ is called the condition of "incompressibility". In addition, it is specified that v_i is of "free divergence" and we constrained v_i with $v_i = 0$ on $\partial \Omega$. It will be noted that fields of vectors (v_i) (of free divergence) are orthogonal with the gradients. Thus we can eliminate the pressure p from the initial equation by the *orthogonal projection* of the equation on the fields of freely divergent vectors (v_i). Consequently we obtain a similar equation with

$$\frac{dx}{dt} = F(x). \qquad (1.254)$$

with M which is a Hilbert space of fields of square integrable vectors, which are orthogonal with the gradients.

1.24.2.8 Dirichlet Norm and Navier–Stokes Equations

Consider the following Navier–Stokes equations (the density is not specified in the system and p is not considered as a constant any more):

$$\frac{\partial v_i}{\partial t} + \sum_{j=1}^{d} v_j \frac{\partial v_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} + v \triangle v_i + g_i, \qquad (1.255)$$

with the condition:

$$\sum_{i=1}^{d} \frac{\partial v_i}{\partial x_i} = 0. \qquad (1.256)$$

with $i = 1,\ldots,d$. And with $d =$ dimension of the phase space ($d = 2$ or 3). With $x \in \mathbb{R}^d$ and the time t.

- The left part of the principal equation above is the expression of the acceleration of a fluid particle.
- The right part of the principal equation above contains the gradient of pressures p, contains a "dissipative term" with the Laplace operator which is multiplied by a constant v (the viscosity or kinematic viscosity) and g always corresponds to an external force.

The fluid is always included in an region (area) $\Omega \in \mathbb{R}^d$, and its speed or velocity is imposed on $\partial \Omega$; this is the condition $v_i = a_i$, and the initial conditions are written:

$v_i = v_{i0}$ for $t = 0$. Now consider $v = a + u$ in the initial Navier–Stokes equation, we can define the following norms:

$$|u| = \left[\int_\Omega dx \sum_i u_i^2 \right]^{1/2}, \tag{1.257}$$

$$\|u\| = \left[\int_\Omega dx \sum_{ij} \left(\frac{\partial u_i}{\partial x_j} \right)^2 \right]^{1/2}, \tag{1.258}$$

which is the *Dirichlet norm* (Ruelle 1980).

1.25 The Three-Body Problem (H. Poincaré)

As a preliminary, we have to underline that, if chaos can appear in a model with only *one equation* in a *discrete-time* model, it requires *three-equations* with a *continuous time* model: it is one of the fundamental subjects treated in the famous "three-body problem". When he studied this question, Henri Poincaré highlighted the concepts of "Poincaré map", "first-return map" and "Poincaré section" which are presented elsewhere in different sections of this book. The problem is to compute the mutual gravitational interaction of three masses. This problem is surprisingly difficult to solve, even in the case of the "restricted three-body problem", corresponding to the (simple) case of three masses moving in a common plane. One of the questions in this context was to demonstrate that the *solar system* as modeled by Newton's equations *is dynamically stable*. This question was a generalization of the *famous three body problem*, which was considered one of the most difficult problems in mathematical physics. The three body problem consists of nine simultaneous differential equations. *The difficulty was to demonstrate that a solution in terms of invariants converges*. First, the two-body problem was studied by *Kepler* and implied at the time famous names of the physics and mathematics such as *Newton* and *Poincaré*, and today the subject continues to occupy scientists. During the study of this problem, one of the question was to analyze the existence of stable equilateral triangle configurations corresponding to *Lagrange points*. The approximate calculations were applied in particular to the Earth–Moon–Sun system, using series expansions involving thousands of algebraic terms. Starting from a basic setup, the three-body system conserves standard mechanical quantities like energy and angular momentum. Poincaré understood that the three-body problem could not be solved in terms of algebraic formulas and integrals. He studied possible trajectories for three-body systems and identified the famous *sensitive dependence on initial conditions*, then he also created a topology to describe the motions of the system and its orbits. Figure 1.55 depict *"idealized" orbits* of motions of two stellar objects in a simple plane.

Hereafter, the figures depict different trajectories given by the system resulting from different initial conditions. At the bottom the two bodies are initially at rest,
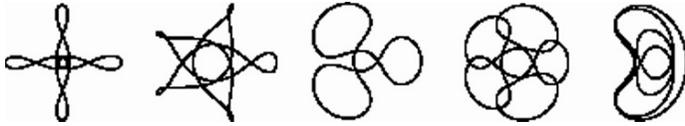
Fig. 1.55 Idealized orbits of motions of two stellar objects

the body at the top is given progressively larger rightward velocities.[32] *In fact, one of the bodies escapes from the other two, like t or sometimes* $t^{2/3}$. Generally this happens fast, but sometimes all three bodies exhibit complex and apparently random behavior. "The delay before escaping is reminiscent of resonant scattering" (Wolfram 2002).
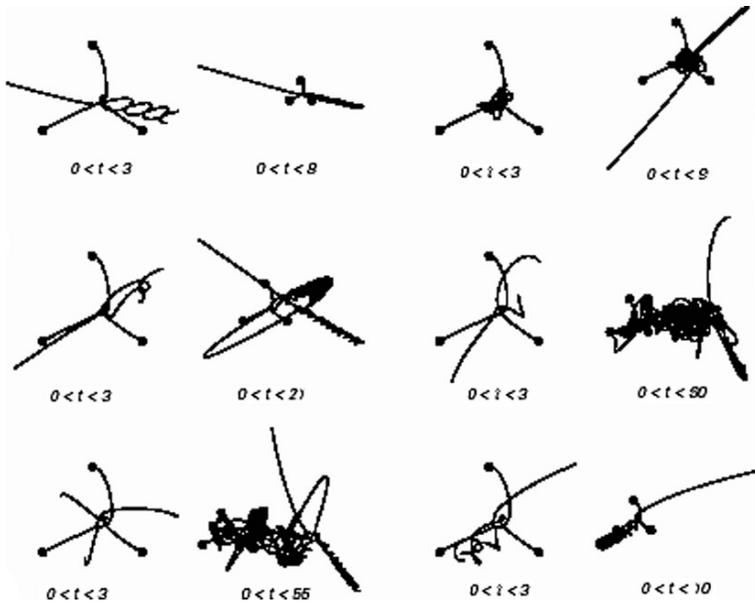
A dynamical system can be written through a *continuous time* system of n differential equations of the first-order which can basically be noted

$$\dot{x}(t) = f(x(t),t), \tag{1.259}$$

where $\dot{x}(t) \in \mathbb{R}^n$ and $x(t) \in \mathbb{R}^n$ and f indicates a field of vectors. We know that the system above is called the *flow* and when f does not depend explicitly on time t, the system of differential equation is written

$$\dot{x}(t) = f(x(t)). \tag{1.260}$$

---

[32] Ref: Quotation of Wolfram (2002, p. 972, Section: Chaos theory and randomness from initial conditions).

This system is known as an autonomous flow. We can be interested in the flows of lower size or equal to 3. Since in the general case it is difficult, even impossible, to find an analytical solution for the non-autonomous and even the autonomous systems above, we consider the trajectory φ in the phase space formed by the points $x(t)$ of $\mathbb{R}^n$ for each value of t. For $n \geqslant 3$ the study of the trajectory φ in the phase space $\mathbb{R}^n$ is irksome and for $n = 3$, Poincaré proposed a new method for the study of curve φ, which substitutes the *intersection points* of this curve with the plane P suitably selected. It is the starting point of the *Poincaré section* method. Remark: The position of the "idealized" planets corresponds to the following differential equation $\dot{x}(t) = -x(t)/(x(t)^2 + (1/2(t + \alpha \sin(2\pi t)))^2)^{3/2}$ where α is the eccentricity of the elliptical orbit (here with $\alpha = 0.1$).

1.26 The Poincaré Section

The Poincaré cut method allows to considerably simplify the flow analysis of dynamical systems by bringing several advantages:

(1) *Reduction of the study space of the flow*. The Poincaré section makes it possible to pass from a flow in $\mathbb{R}^3$ (for example) *to a map of the plane on itself.* Or more generally to pass from a *space $\mathbb{R}^n$* to *a space $\mathbb{R}^{n-1}$*.

(2) *Substitution* of differential equations of the system with algebraic equations corresponding to the map of the plane on itself, of the type $P = T(P)$, as we have just seen it.

(3) Considerable *reduction of the quantity of data*, reduced to those of the Poincaré section. In some cases, we can pass from a complete trajectory to a very small number of points. Let us quote for example the case of a system which is spread in time on 10,000 iterations, where we have to face a periodic regime of period-four, then we thus pass from an analysis of 10,000 points to a group of four points in the plane of the section. Thus, we can ignore almost all the points of the trajectory of the phase-space except for a group of "solutions" points of the system.

(4) *Fast and precise diagnosis of the periodic* or aperiodic behavior of the system, with a minimum number of numerical calculation, whereas in the phase-space, it can be extremely difficult to identify the nature of the behavior of a model, in particular in the case of several dependent variables. Indeed sometimes the *curve tangle of trajectories in the plane of variables is illegible*, and we can see nothing of the dynamics that we visualize except its complexity.

(5) Lastly, let us specify that the way in which the points of the map occupy the sectional surface of Poincaré makes it possible to identify an aperiodic behavior very quickly. *If the points are distributed on a surface, we have to face a flow behavior which is aperiodic and even chaotic. If on the other hand the points are distributed on a traditional curve, this is because we faced a quasiperiodic dynamics or an aperiodic dynamics but strongly dissipative.* Because of a "too

fast" contraction of surfaces in the phase space, which prevents from identifying the "lateral extension" of the Poincaré curve of the attractor, we must look further into the analysis.

The *Poincaré section* is written as a *map* from the plane to itself $x_i(k+1) = T(x_i(k))$ in the form of an iteration, which allows to easily handle the flows which are initially very voluminous and to highlight (by the iterated map method) relevant elements of the model and its properties.

1.26.1 Periodic Solution

For a periodic solution, the writing of the map is $P_0 = T(P_0) = T^2(P_0) = \cdots$. If P_0 is the periodic solution of the model, i.e. if the solution is reduced to a unique point (see Fig. 1.56) or a small number of points, then the map denoted T iterates the same point (which is a fixed point of the model). This very interesting writing allows to treat the problem of the periodic solution stability of a model. In Fig. 1.56 the intersection of the trajectory with the Poincaré plane is reduced to a unique point P_0.

Thus we consider again the approach relating to the stability of a system by studying the derivative of the map, written in the form of what we call the Floquet matrix:

$$\text{Floquet matrix: } M = \left[\frac{\partial T}{\partial x_i}\right]_{x_i^0} \quad \text{with} \quad i = 1, 2. \qquad (1.261)$$

This is the derivative of the map T with respect to the variable: x, whereas the usual approach studies the derivative ("temporal slope") of the variable with respect to the time in its differential system. We can explain that the point P_0 after the action of the map T (i.e. after one period, or after a turn) is a new point that we can write $P_0 + \varepsilon$ which is at a certain distance of P_0. After one iteration, this distance is equal to $M\varepsilon \simeq T(P_0 + \varepsilon) - P_0$ where $\|\varepsilon\| \to 0$. As observed previously several times, the eigenvalues of M express the stability of the trajectory of a system. After k periods or iterations, we write $T^k(P_0 + \varepsilon) - P_0 \simeq M^k \varepsilon$. Thus the initial distance ε is itself multiplied by M^k. And if the *eigenvalues of M have values lower than* 1, such that they are inside the unit circle of the complex plane, then the periodic trajectory of the model is linearly stable, because the distance with P_0 tends to decrease. On the other hand, if one of the eigenvalues is >1, the distance increases exponentially, the solution is unstable.
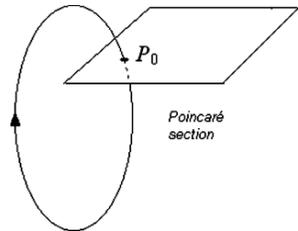
Fig. 1.56 Poincaré section of a limit cycle. In the simple case selected here, the section is reduced to a unique point

1.26.2 Quasiperiodic Solution

We can imagine a *biperiodic dynamics* with two basic frequencies f_1 and f_2, then we know that the associated attractor is a T^2 torus. This object conciliates two different cyclic trajectories which are rolled up and superposed on a volume which is obviously spread in a three-dimensional coordinate system. A first cycle longer which would be for example f_1 is used as a base for the form of the attractor, and around this cycle a second cycle f_2 of a "smaller diameter" is rolled up which moves to the limits of the first revolution. The set constitutes a "tubular form" closed on itself. The *Poincaré section* of such an object or such a dynamics can be:

1. An *ellipse* and even a circle which magnificently *summarizes* the double orbit or the double periodicity of system solutions. In fact, this ellipse is made up of points which are as many intersections of the trajectory of dynamics with the Poincaré section. These points of impact in the plane summarize the nature of the attractor.
2. *Cycloids*, for example, "distorted ellipses" that have the form of an eight (or others) and are combinations (i.e. harmonics) of the two basic frequencies.

The geometrical forms that we identify on the surface of the plane which is the Poincaré section, are the result of the ratio between the two basic frequencies f_1/f_2. This ratio can be rational or irrational. If the ratio is irrational, the trajectory does not close on itself and covers, in a *dense* way, the surface of the torus. In such a case, both frequencies are said incommensurable, then the closed curves are *continuous*. Each point is the transform of another point of the same curve. The iterative process never takes the same point twice, and thus ends up cover in continuous manner, at the same time the whole of the orbit in the Poincaré section, but also ends up cover the surface of the torus in the plane of higher size than $\mathbb{R}^3$. If the closed orbit observed in the Poincaré section is denoted "O", we can write: $T(O) = O$, because it is important to note that each point of this orbit is the transform by the "first-return map" of another point of the same curve. On the contrary, if the ratio f_1/f_2 is rational, the Poincaré section shows a finite set of points. But here, the points do not form a continuous curve, and *do not constitute a dense trajectory on the Poincaré sectional surface.* It is said that there is *coupling of two frequencies f_1 and f_2.* We know that $P_{k+1} = T(P_k) = T(T(P_{k-1})) = T^2(P_{k-1}) = \cdots$ with, $i = 1, 2$. Thus, we can write $P_i = T^j(P_i)$.

Figure 1.57 depicts a torus T^2 and its Poincaré section which shows the closed curve O. Here, the ratio of f_1 and f_2 is supposed to be irrational.

The appearance of the Poincaré section depends on the ratio f_1/f_2. If the ratio is irrational, the trajectory does not close on itself and covers, in a dense way, the surface of the torus. In such a case, both frequencies are said incommensurable, then the closed curves are *continuous*. Each point is the transform of another point of the same curve, this one is invariant by the map T.

Fig. 1.57 Torus T^2 and its
Poincaré section
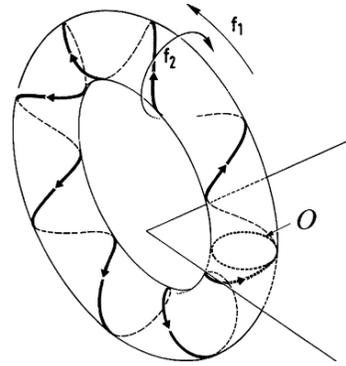
1.26.3 Aperiodic Solution

We can comment these solutions by saying that they are obviously generated by complex regimes, which produce complex objects, as for example strange attractors which are more difficult to describe. It will be simply noted that the solutions of this type of dynamics are dissipative and that they undergo a *contraction of volumes,* until they exist in the Poincaré section only by points distributed on *a segment*, or *arcs of curves*. This means that these solutions can be summarized significantly, inside a unidimensional space. Thus, the unidimensional map, which is *the first return map,* appears very useful for the study of this type of dynamics. This is approached in one of the following sections.

1.26.4 Some Examples

Here are some typical Poincaré sections in the Fig. 1.58. Figure 1.58a shows a quasiperiodic regime for an irrational ratio f_1/f_2. Figure 1.58b shows also a quasiperiodic regime. Figure 1.58d shows a quasiperiodic regime for a rational ratio f_1/f_2 equal to 3/5, which corresponds to a 5-period cycle. Figure 1.58e shows a chaotic regime.

1.27 From Topological Equivalence of Flows Towards the Poincaré Map

1.27.1 Rotation Number, Orientation-Preserving Diffeomorphism and Topological Equivalence of Flows

As a preliminary, before outlining the Poincaré map and associated concepts, which are related to the topologically conjugacy concept, we introduce some useful and
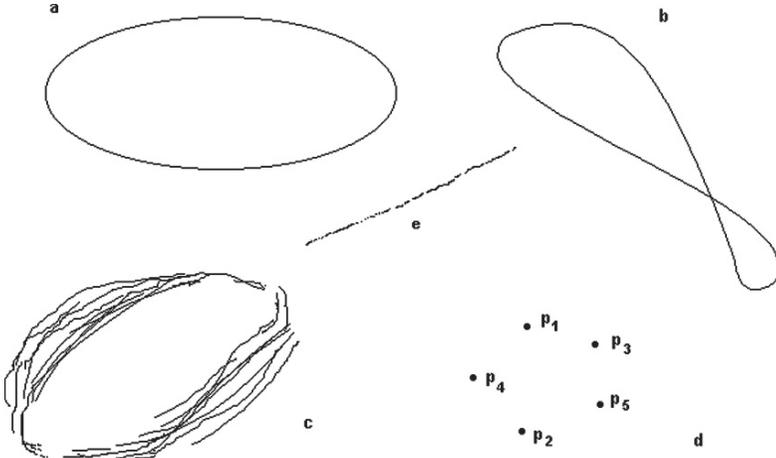
Fig. 1.58 Sketches of some typical orbits intersecting the Poincaré section: (**a**, **b**) Quasi-periodic orbits. (**c**, **e**) Chaotic orbits. (**d**) 5-period cycle

important notions such that: the *diffeomorphism of the circle and the rotation number*, the *orientation-preserving diffeomorphism* and the *topological equivalence of flows*.

First, let S^1 be a *circle which is an elementary differential manifold* and let us consider the example of *pure rotations which are* diffeomorphisms on S^1. If we note θ the angular displacement (associated with the radius), it is possible to simulate the *rotation "$\mathfrak{R}$"* by σ which is written as follows (θ corresponds to units of 2π):

$$\mathfrak{R}_\sigma(\theta) = (\theta + \sigma)\bmod 1. \tag{1.262}$$

If $\sigma = p/q$, $p, q \in \mathbb{Z}$ and relatively prime, then it comes

$$\mathfrak{R}_\sigma^q(\theta) = (\theta + p)\bmod 1 = \theta, \tag{1.263}$$

then it is possible to write that *every point of the circle is a periodic point of period-q*, so the orbit of any point is a q-cycle. By contrast, if σ is *irrational* then

$$\mathfrak{R}_\sigma^m(\theta) = (\theta + m\sigma)\bmod 1 \neq \theta, \tag{1.264}$$

for any θ, and the orbit of any point covers the *circle densely*. A basic example is given hereafter concerning *pure rotation $\mathfrak{R}_\sigma$* with $\sigma = p/q = 2/5$ (Fig. 1.59). The orbit of θ revolves two times (i.e. $p = 2$) before coming back to θ on the fifth iteration.

The "lift" notion of a map $h : S^1 \rightarrow S^1$ must be introduce here. The context enables us to consider that h is a *homeomorphism*, furthermore let us consider a continuous function $\hat{h} : \mathbb{R} \rightarrow \mathbb{R}$ such that:

$$\pi(\hat{h}(x)) = h(\pi(x)) \tag{1.265}$$
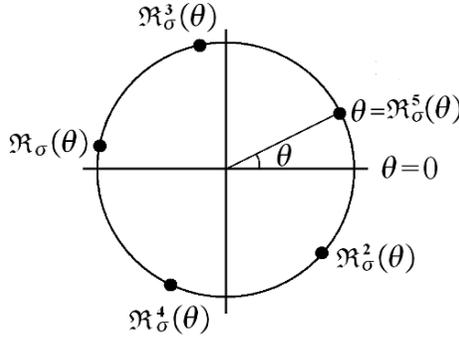
Fig. 1.59 Orbit of a pure rotation $\mathfrak{R}_\sigma$ with $\sigma = p/q = 2/5$

with

$$\pi(x) = x \bmod 1 = \theta. \tag{1.266}$$

Thus $\hat{h}$ is called a *lift* of $h : S^1 \to S^1$ onto $\mathbb{R}$. See the commutative diagram, figure hereafter:
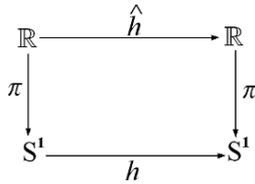
Proposition 1.2 (Lift of the orientation-preserving homeomorphism). *Let $\hat{h}$ be a lift of the orientation-preserving homeomorphism $h : S^1 \to S^1$. Then $\hat{h}(x+1) = \hat{h}(x) + 1$ for every $x \in \mathbb{R}$.*

At this stage we must highlight that the *Conjugacy is* a fundamental aspect of the topology in connection with the concepts of "equivalence relation". They make it possible to identify when flows have the same behavior (for developments see "Hyperbolic nonlinear fixed point" section).

Definition 1.26 (Topologically conjugate diffeomorphisms). Two diffeomorphisms $g, v : M \to M$, are topologically conjugate (or C^0-conjugate) if there is a homeomorphism $h : M \to M$, such that

$$h \cdot g = v \cdot h. \tag{1.267}$$

The topologically Conjugacy for two *flows* ϕ_t $\varphi_t : S \to S$ is defined similarly with the preceding definition and the preceding equation is replaced by $h \cdot \phi_t = \varphi_t \cdot h$ with $t \in \mathbb{R}$. The *definition* above shows that the homeomorphism h *takes each orbit* of g (or ϕ_t) *onto an orbit* of v (or φ_t) *preserving the parameter* $p(t)$: that means $g^p(x) \overset{h}{\to} v^p(h(x))$ for each $p \in \mathbb{Z}$, or with the flows $\phi_t(x) \overset{h}{\to} \varphi_t(h(x))$ for each $t \in \mathbb{R}$.

Let us recall that *rational and irrational rotations* are discriminated by means of the *rotation number*. Thus, for any homeomorphism $h : S^1 \to S^1$ it is possible to write the following definition.

Definition 1.27 (Rotation number). The rotation number $\alpha(h)$ of a homeomorphism $h : S^1 \to S^1$ is given by

$$\alpha(h) = \left(\lim_{n \to \infty} \frac{\hat{h}^n(x) - x}{n} \right) \tag{1.268}$$

$\mathrm{mod}\, 1$, where $\hat{h}$ is a lift of h.

Remark 1.7 (Rotation number of pure rotation). Let us consider $\hat{\mathfrak{R}}_\delta(x) = x + \delta$ which is the lift of the pure rotation $\mathfrak{R}_\delta(\theta) = (\theta + \delta)\,\mathrm{mod}\, 1$. Then $\hat{\mathfrak{R}}_\delta^n(x) = x + n\delta$ and $\alpha(\mathfrak{R}_\delta) = \delta$ (i.e. the rotation number of the pure rotation is equal to δ). The relation between the pure rotation $\mathfrak{R}_\delta$ and its lift $\hat{\mathfrak{R}}_\delta$ is depicted on the following commutative diagram:
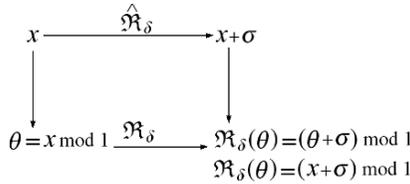
Proposition 1.3 (A "rational" rotation number defines "periodic" points). *A diffeomorphism $g : S^1 \to S^1$ has "periodic points" if and only if its rotation number $\alpha(g)$ is "rational".*

The circle diffeomorphism with the *rational number* $p/q \in \mathbb{Q}$ possesses an *even number* of *period-q cycles*. Moreover, for $\hat{g}^q(x)$ we know[33] that $\hat{g}^q(1) = \hat{g}^q(0) + 1$ and due to this, if a fixed point x_0^* occurs then at least one further fixed point x_1^* must occur. Consider, for example $(q = 3)$, $\hat{g}^3(x)$ a circle diffeomorphism with a stable 3-cycle and an unstable 3-cycle. Furthermore when x_0^* is stable, then x_1^* is unstable with a stable 3-cycle. The following figures show for $\hat{g}^3(x)$ the *alternation of the stable and unstable points along the bisector as well as along the circle*. It is easy to observe (figure at right) the periodic points of the map g on the circle for the lift $\hat{g}^3$ (figure at right):

---

[33] It is an extension of the proposition about a lift of the orientation-preserving homeomorphism (see previously).
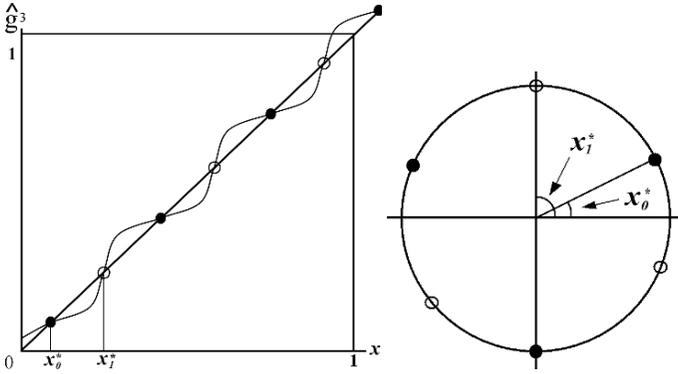
Theorem 1.10 (Orientation-preserving diffeomorphism topologically conjugate to the pure rotation). [34] *If an orientation-preserving diffeomorphism $g : S^1 \to S^1$ is of class C^2 and $\alpha(g) = \mu \in \mathbb{R}\backslash\mathbb{Q}$, then it is topologically conjugate to the "pure rotation" $\mathfrak{R}_\mu$.*

At this stage, we want to highlight the *equivalence of flows*. Then, let us consider a homeomorphism h which is used to select successive points in the orbit of one map g onto those of another map v. We are interested in the description of the equivalence of two flows of maps g and v which are supposed to have similar behaviors. However, by analogy since we are interested in the equivalence of flows, let us consider the following flows ϕ_t and φ_t in order to state the definition that follows:

Definition 1.28 (Topologically equivalent flows). Two flows are called topologically equivalent (or C^0) if there is a homeomorphism h taking orbits of ϕ_t onto those of φ_t preserving their orientation.

We know that the equivalence requires only the preservation of the orientation, so it is possible to write $h(\phi_t(x)) = \varphi_{\tau_y(t)}(y)$, with $y = h(x)$ and $\tau_y(t)$ is an increasing function of t for every y.

1.27.2 Poincaré Map (First Return Map) and Suspension

Let us consider the flow map $\phi_t = M \to M$ which is a diffeomorphism for each fixed t. A diffeomorphism can be obtained from a flow if we take its *time-τ map*, $\phi_\tau = M \to M$, $\tau > 0$. The orbits of ϕ_τ are constrained to follow the trajectories of the flow because $\{\phi_\tau^m(x)|m \in \mathbb{Z}\} = \{\phi_{m\tau}(x)|m \in \mathbb{Z}\} \subseteq \{\phi_t(x)|t \in \mathbb{R}\}$; i.e. the dynamics of ϕ_τ are highly influenced by the flow ϕ and are not typical of those of diffeomorphisms on M.

Moreover, it is interesting to highlight that, while the orbits of x under the action of ϕ_{τ_1} and ϕ_{τ_1} (with $\tau_1 \neq \tau_2$) behave similarly for $x \in M$, the *two maps are not*

---

[34] See Denjoy's theorem and Arnold (1983), Nicketi (1971).

necessarily of the same topological type. Another way of obtaining a diffeomorphism from a flow is to construct its *Poincaré map*. Then, let us consider the flow ϕ on M with a "vector field" $\mathbf{X}$ and suppose that Σ is a codimension-1 submanifold of M satisfying:

1. Every orbit of ϕ meets for arbitrarily large positive and negative times.
2. If $x \in \Sigma$ then $\mathbf{X}(x)$ is not tangent to Σ. Then, Σ is said to be a *global cross section* of the flow. Let $y \in \Sigma$ and $\tau(y)$ be the least positive time for which $\phi_{\tau(y)}(y) \in \Sigma$.

Definition 1.29 (Poincaré map). The poincaré map, or first return map, for Σ is defined to be

$$\mathbf{P}(y) = \phi_{\tau(y)}, y \in \Sigma. \tag{1.269}$$

Let us note that by construction $\mathbf{P} : \Sigma \to \Sigma$ is a diffeomorphism and $\dim \Sigma = \dim M - 1$. In opposition to time-τ maps (see before), we expect these diffeomorphisms to reflect the properties of flows in one higher dimension.

Coordinates (φ, θ)
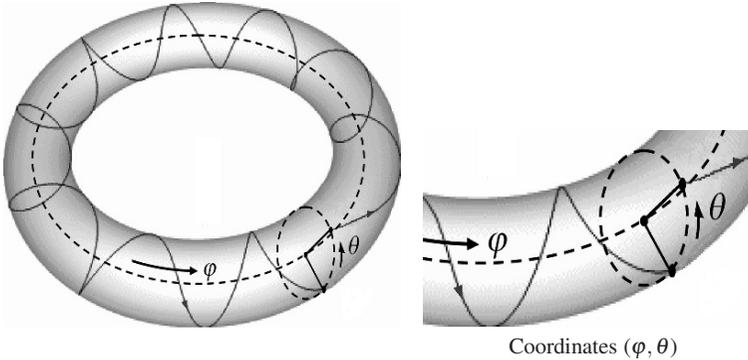
Example 1.2 (Flow on the torus). A flow on the torus can be written as follows:

$$\dot{\theta} = a, \quad \dot{\varphi} = b, \ a, b > 0, \tag{1.270}$$

with θ and φ (which are depicted in the previous figure). The solution of the equation above are written

$$\theta = at + \theta_0, \ \text{and} \ \varphi = bt + \varphi_0 \tag{1.271}$$

reduced mod 2π, then $\{\theta, \varphi\}$ first return to $\{\theta_0, \varphi_0\}$ when $\{t = t_\theta, t = t_\varphi\}$, where $\{at_\theta = 2\pi, bt_\varphi = 2\pi\}$. Then, if $a/b = p/q$, $p, q \in \mathbb{Z}^+$ and $qt_\varphi = pt_\theta$ and the orbit through (θ_0, φ_0) returns to this point after q revolutions around the torus in the φ-sense and p in the θ-sense. Thus, if a and b are "rationally" related, then every point of T^2 is a periodic point, it means that *every point lies on a closed orbit*. On the contrary, if a and b "are not rationally" related, then the orbit through (θ_0, φ_0)

never returns to that point even if it can be very close. When we select a φ constant such that $\varphi = \varphi_0$, we obtain a "global section of the torus" if Σ is a circle S^1 with coordinate θ. Due to the fact that the orbit of the flow first return to $\varphi = \varphi_0$ after $t_\varphi = 2\pi/a$ and $\theta = at + \theta_0$, then it is possible to write that the Poincaré map $\mathbf{P} : S^1 \to S^1$ is a rotation by $2\pi a/b$. Obviously there are flows for which there is no global section. Then, it is not possible to state that every flow corresponds to a diffeomorphism if we use a Poincaré map. But the opposite is true, indeed every diffeomorphism g is the Poincaré map of a flow, and is called a "suspension"[35] of g. Thus, in particular since Smale (1967) and Arnold (1968), we know that the outcome of a diffeomorphism $(\dim M - 1)$ have an "analog" for the flow in one higher dimension $(\dim M)$.

Definition 1.30 (Suspension). The flow $\phi_t(x, \theta) = (g^{(t+\theta)}(x), t + \theta - [T + \theta])$ with $x \in M, \theta \in [0, 1]$ and $[\, \cdot \,]$ which indicates the integer part of its contents "$\cdot$," defined on a compact manifold $\bar{M}$ by identification of $(x, 1)$ and $(g(x), 0)$ in the topological product $M \times [0, 1]$, is called the suspension of the diffeomorphism $g : M \to M$.

Poincaré map: Briefly, the Poincaré map $\mathbf{P} : \Sigma \longrightarrow \Sigma$ is a procedure used to eliminate a dimension of the system and thus a continuous system is transformed into a discrete one. (Usually, it is considered as a surface that transversely intersects a given orbit. For systems subjected to periodic forcing, Poincaré section may be represented by a surface that corresponds to a specific phase of the driving force. On this base, we have a *stroboscopically* sample of the system outputs.) Thus, a Poincaré section is used to construct a $(n - 1)$-dimensional discrete dynamical system, i.e. a Poincaré map, of a continuous flow given in n dimensions. This reduced system of $n - 1$ dimensions preserves many properties, e.g. periodicity or quasi-periodicity, of the original system (Fig. 1.60).

If we reason in a system with three degrees of freedom, these systems can exhibit a periodic cycle or a chaotic attractor. A Poincaré section Σ is assumed to be a part of a plane, which is placed within the 3D phase-space of the continuous dynamical system such that either the periodic orbit or the chaotic attractor intersects the Poincaré section. Thus, the Poincaré map is defined as a discrete function $\mathbf{P} : \Sigma \longrightarrow \Sigma$ which associates consecutive intersections of a trajectory of the 3D flow with Σ.

1.28 Lyapunov Exponent

We mentioned previously that the sensitive dependence on initial conditions was the characteristic of a chaotic system. This observation can be expressed by the fact that two points, or two trajectories, initially very close deviate exponentially and this in a finite number of points or iterations, sometimes even very short. In such a system, the forecast is thus impossible, except maybe over very short durations. The most effective tool to identify such processes, that they come from dynamical systems or from experimental series, is the Lyapunov characteristic exponent (LCE).

---

[35] *Suspension:* Roughly, the space join of a topological space E and a pair of points S^0, $\Sigma(E) = E * S^0$.
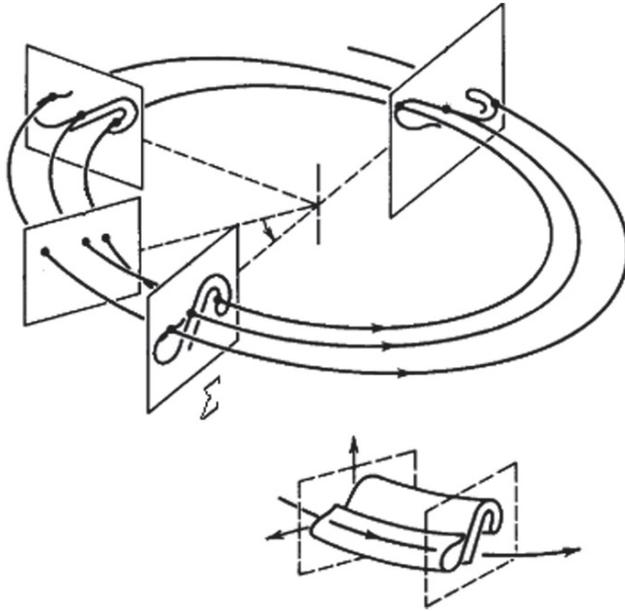
Fig. 1.60 Poincaré sections according to a stroboscopic sampling

1.28.1 Description of the Principle

Consider two nearby points $x_1(0)$, $x_2(0)$ at the moment $t = 0$ of a model, starting points of two trajectories in the phase space, the distance between these two points is denoted $d(0)$. At a moment t, i.e. after the movement of points along their respective trajectory, we again measure the distance between both points denoted $d(t)$. To use another terminology, we can say that we *applies a flow* ϕ_t to both initial points and after a lapse of time t, we measure $d(t)$. Then, we calculate the evolution of these two distances (ratio $d(0)/d(t)$) by the term $e^{\chi t}$. For a t which tends towards the infinite, χ converges towards a limit. This limit is the *Lyapunov characteristic exponent*. If $\chi > 0$, we say that *under the action of the flow, the close orbits diverge exponentially*. It is also said that the *Lyapunov characteristic exponent* measures the speed of divergence of a system.
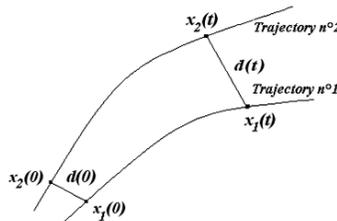
1.28.2 Lyapunov Exponent Calculation

The Lyapunov exponent $\lambda(x_0)$ measures the gap of trajectories. Let $x_0, x_0 + \varepsilon$ be two close points, then we write the Lyapunov exponent by: $\varepsilon e^{n\lambda(x_0)} = |f^n(x_0 + \varepsilon) - f(x_0)|$. For the limits of ε and n, we have

$$\lambda(x_0) = \lim_{n \to \infty} \lim_{\varepsilon \to 0} \frac{1}{n} \log \left| \frac{f^n(x_0 + \varepsilon) - f(x_0)}{\varepsilon} \right| = \lim_{n \to \infty} \frac{1}{n} \log \left| \frac{d f^n(x_0)}{dx} \right|.$$

It is pointed out that $x_i = f^i(x_0)$ and $f^n(x_0) = f(f^{n-1}(x_0))$, consequently $d f^n(x_0)/dx = f'(x_{n-1})f'(x_{n-2})\cdots f'(x_1)f'(x_{n-1})$. $\lambda(x_0)$ is written

$$\lambda(x_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \left| f'(x_i) \right|.$$

1.28.3 Other Writing and Comment

If we consider two points X_0 and $X_0 + \Delta x_0$, *each one of them generates an orbit in the phase space* of an arbitrary model. Incidentally, we can say that these orbits can be regarded as the parametric functions of a variable which is the time. If we use one of the two orbits as reference orbit, then the separation between the two orbits will be also function of time. Because "significant dependences" can appear in some parts of the system, as in the logistic equation, with attracting points or attracting periodic points. The separation of orbits which has the form $\Delta x(X_0, t)$ is also function of the localization of the initial value. In a system with attracting fixed points or attracting periodic points, $\Delta x(X_0, t)$ *decreases asymptotically over time*. If a system is unstable, then the orbits diverge exponentially. For chaotic points, the function $\Delta x(X_0, t)$ will behave in an erratic way. Thus usually we study the average exponential rate of the divergence of two orbits initially close by using the formula:

$$\lambda = \lim_{\substack{t \to +\infty \\ |\Delta X_0| \to 0}} \frac{1}{t} \log \frac{|\Delta x(X_0, t)|}{|\Delta X_0|}.$$

The number called Lyapunov exponent is used to distinguish the majority of the varieties of orbit types. The Lyapunov exponent measures the sensitive dependence on initial conditions estimating the exponential divergence of orbits. These exponents are used as a dynamical diagnostic tool for chaotic system analysis and can also be used for the calculation of other invariant quantities as the attractor dimension. A qualitative picture of dynamics is given by means of the signs of these exponents. The positive Lyapunov exponents define directions of local instabilities in the dynamics and any system containing at least one positive exponent has a chaotic behavior. A behavior with more than one positive exponent is called an hyperchaos (see Savi and Pacheco 2002; Machado 2003). *For a dynamical system*

which can be linearized, the calculation of Lyapunov exponents is carried out from the algorithm by Wolf et al. (1985). For time series, the calculation is more difficult. In practice, there are two different types of algorithms: (1) Trajectories, real space or direct methods (see Wolf 1985; Rosenstein 1993; Kantz 1994); and (2) perturbations, tangent space or the Jacobian matrix method (see Sano and Sawada 1985; Eckmann 1986; Brown 1991; Briggs 1990). See figure below:
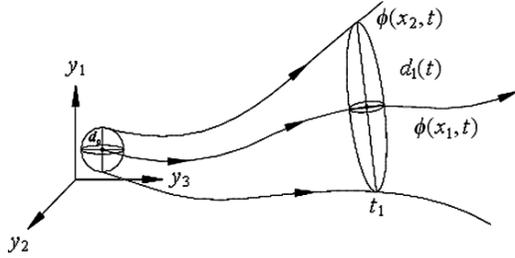
To give an example of Lyapunov exponent calculation, let us consider a D-sphere of states which is transformed by a system in a D-ellipsoid. The Lyapunov exponent concept is related to the expanding and contracting character of different directions in phase space. The divergence of two nearby orbits is analyzed by the variation of the relation between the initial D-*sphere* and the D-*ellipsoid*. This variation can be written: $d(t) = d_0 \beta^*$, where d is the diameter and β is a reference basis. When the *Lyapunov exponent* λ is *negative* or vanishes, trajectories do not diverge. When the *Lyapunov exponent is positive,* the trajectories diverge, *characterizing chaos*. In a chaotic regime, there is a local exponential divergence of nearby orbits and thus, suitable algorithms are necessary to calculate Lyapunov exponents (Wolf 1985; Parker and Chua 1989). They calculate the average of this divergence considered on different points of the trajectory. Then, when the distance $d(t)$ becomes large, a new $d_0(t)$ is defined to calculate the divergence as follows: $\lambda = \dfrac{1}{t_n - t_0} \sum_{k=1}^{n} \log_\beta \left(d(t_k)/d_0(t_{n-1}) \right)$. The attractor dimension may be calculated from the Lyapunov spectrum considering the Kaplan–Yorke conjecture (Kaplan and Yorke, 1983).

1.28.4 Interpretation of λ

(1) $\lambda < 0$: Here the system generates a stable fixed point or a stable periodic orbit. Such a negative value characterizes a dissipative or non-conservative system. The more the exponent is negative, the more the stability is large. A superstable fixed point will have an exponent which tends towards $-\infty$. (2) $\lambda = 0$: A system with such an exponent is conservative; the system shows a "Lyapunov stability". (3) $\lambda > 0$: Here the *orbit is unstable and chaotic*. The initial very close points diverge to arbitrary values over time. Here a graphic would be similar to a cloud of points without a distinct trajectory.

1.29 Measure of Disorder: Entropy and Lyapunov Characteristic Exponent

(1) The word entropy has important physical implications as the amount of "disorder" of a system. The main contributions came from Shannon, Kolmogorov, Sinai, Renyi. The *Shannon* entropy definition for a variable X is given by

$$H(X) = -\sum_x P(X)\log_2[P(x)] \qquad (1.272)$$

bits, where $P(x)$ is the probability that X is in the state x. $P\log_2 P$ is defined as 0 if $P = 0$. The joint entropy of variables $X_1,\ldots,X_n$ is then defined by $H(X_1,\ldots,X_n) = -\sum_{x_1}\cdots\sum_{x_n}P(x_1,\ldots,x_n)\log_2[P(x_1,\ldots,x_n)]$. Furthermore, it is possible to describe the *Kolmogorov entropy* (or Kolmogorov–Sinai). Divide phase space into D-dimensional hypercubes of content ε^D. Given $P_{i_0,\ldots,i_n}$ the probability that a trajectory is in hypercube i_0 at $t = 0$, i_1 at $t = T$, i_2 at $t = 2T$, etc. Then define $K_n = h_K = -\sum_{i_0,\ldots,i_n} P_{i_0,\ldots,i_n} \ln P_{i_0,\ldots,i_n}$, where $K_{N+1} - K_N$ is the information needed to predict which hypercube the trajectory will be in at $(n+1)T$ given trajectories up to nT. *Kolmogorov entropy* is then defined by

$$K \equiv \lim_{T\to 0}\lim_{\varepsilon\to 0+}\lim_{N\to\infty}\frac{1}{NT}\sum_{n=0}^{N-1}(K_{n+1}-K_n). \qquad (1.273)$$

The Kolmogorov entropy is related to Lyapunov characteristic exponents by

$$h_K = \int_P \sum_{\sigma_i>0}\sigma_i d\mu. \qquad (1.274)$$

(2) The Lyapunov characteristic exponent (LCE) provides the rate of exponential divergence from perturbed initial conditions. To study the behavior of an orbit around a point $\mathbf{X}^*(t)$, let us perturb the system and write $\mathbf{X}(t) = \mathbf{X}^*(t) + V(t)$ with $V(t)$ the average deviation from the unperturbed trajectory at time t. In a *chaotic region*, the LCE σ is independent of $\mathbf{X}^*(0)$. It is given by the *Oseledec theorem*, which states that: $\sigma_i = \lim_{t\to\infty}\frac{1}{t}\ln|V(t)|$. For an n-dimensional map, the *Lyapunov characteristic exponents* are given by

$$\sigma_i = \lim_{N\to\infty}\ln|\lambda_i(N)| \qquad (1.275)$$

(with $i = 1,\ldots,n$) where λ_i is the *Lyapunov characteristic number*. One Lyapunov characteristic exponent is always 0, since there is never any divergence for a perturbed trajectory in the direction of the unperturbed trajectory. The larger the LCE, the greater the rate of exponential divergence and the wider the corresponding separatrix of the chaotic region. For the standard map, an analytic estimate of the width of the chaotic zone by Chirikov gives

$$\delta I = Be^{-AK^{-1/2}}.$$

Since the Lyapunov characteristic exponent increases with increasing K, some relationship likely exists connecting both. Given a trajectory (as a map) that has initial conditions (x_0, y_0) and a nearby trajectory with initial conditions $(x', y') = (x_0 + dx, y_0 + dy)$. The distance between trajectories at iteration k is then $dk = |(x' - x_0, y' - y_0)|$, and the mean exponential rate of divergence of the trajectories is defined by: $\sigma_1 = \lim_{k \to \infty} \frac{1}{k} \ln(d_k/d_0)$. For an n-dimensional phase space (map), there are n Lyapunov characteristic exponents $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. However, because the largest exponent σ_1 will dominate, this limit is useful only for finding the largest exponent. Since d_k increases exponentially with k, after a few steps the perturbed trajectory is no longer nearby. Then it is necessary to renormalize frequently every t steps. Defining $r_{k\tau} \equiv d_{k\tau}/d_0$, we can then calculate $\sigma_1 = \lim_{n \to \infty} \frac{1}{n\tau} \sum_{k=1}^{n} \ln r_{k\tau}$. The calculation of the second smaller Lyapunov exponent can be done by observing the evolution of a two-dimensional surface. It will behave as $e^{(\sigma_1 + \sigma_2)t}$, so σ_2 can be obtained if σ_1 is known. To find smaller exponents, the process can be repeated.

For *Hamiltonian systems*, the LCEs exist in additive inverse pairs, so if σ is an LCE, then so is $-\sigma$. One LCE is always 0. For a one-dimensional oscillator (with a two-dimensional phase space), the two LCEs therefore must be $\sigma_1 = \sigma_2 = 0$, so the motion is *quasiperiodic* and cannot be *chaotic*. For *higher order Hamiltonian systems*, there are always at least two 0 LCEs, but other LCEs can enter in plus-and-minus pairs ℓ and $-\ell$. If they (too) are both zero, the motion is *integrable* and *not chaotic*. If they are nonzero, the positive LCE ℓ results in an exponential separation of trajectories, which corresponds to a *chaotic region*. (It is not possible to have all LCEs negative, thus it is why convergence of orbits is never observed in Hamiltonian systems).

Given a *dissipative system*. For an arbitrary n-dimensional phase space, there must always be one LCE equal to 0, since a perturbation along the path results in no divergence. The LCEs satisfy $\sum_i \sigma_i < 0$. Therefore, for a two-dimensional phase space of a dissipative system, $\sigma_1 = 0$, $\sigma_2 < 0$. For a three-dimensional phase space, there are three possibilities:

(1) $\sigma_1 = 0$, $\sigma_2 = 0$, $\sigma_2 < 0$: (integrable).
(2) $\sigma_1 = 0$, $\sigma_2, \sigma_2 < 0$: (integrable).
(3) $\sigma_1 = 0$, $\sigma_2 > 0$, $\sigma_3 < -\sigma_2 < 0$: (chaotic).

1.30 Basic Concepts of Nonlinear Theory Illustrated by Unidimensional Logistic Equation: The Paradigm of a Nonlinear Model

An idea commonly admitted associates the simplicity of a system with its predictability. In the same way, the complexity is associated with the ideas of *unpredictability* and *chaotic behavior*. However, this association appears inaccurate. The logistic equation dynamics is a proof of this.

1.30.1 A Simple Dynamic Equation Which Contains a Subjacent "Deterministic Chaos"

Beyond the demographic model (P.F. Verhulst 1844, R.M. May 1976), this function presents numerous advantages. Indeed, we are faced with the *simplest* dynamical model, depending on only one variable, and yet it can have a *highly unpredictable behavior*. It can exhibit a *deterministic chaos* that is interesting for the comprehension of complex dynamics in all the fields of sciences. Deterministic chaos is a term used to describe a behavior that is seemingly (and numerically) chaotic, whereas in reality, it is clearly determined by the simple equation of the model and by its initial conditions. Thus, we know today, since the discovery of the mechanism of this function, *that simple deterministic systems can have apparently random behaviors*. It is thus not essential to look for chaotic behavior in multidimensional dynamical systems to develop the concepts of nonlinear dynamics. The logistic equation is written in the discrete case: $x_{n+1} = \alpha x_n(1 - x_n)$. It expresses an autonomous, unidimensional and nonlinear model of population growth with α as *the fertility parameter*. This parameter can vary between $0.00\ldots1$ and 4. This model, easy to handle, is a rather rich representation of problems that we meet in nonlinear dynamics. Thus, it will be used to study many concepts in connection with nonlinearities. Moreover, this demographic function relates to a set of competencies concerning Economics. The logistic equation is also called *quadratic map* or *Verhulst equation*.

1.30.2 Fixed Points

The fixed point is a value of the variable which remains unchanged, whatever the iterations we carry out from this same fixed point. Of course, there can be several fixed points in a model. As regards the logistic equation, we can calculate the fixed points with a fertility parameter fixed at $\alpha = 0.5$:

(1) For $\alpha = 0.5$, the equation becomes: $x_{n+1} = (0.5)x_n(1 - x_n)$, and there is only one fixed point for $x = 0$. Indeed for $\alpha = 0.5$, and $x_0 = 0.2$, after less than ten iterations, the trajectory of the "orbit" of the model converges towards zero, the fixed point of the parameterized model.
(2) For $\alpha = 1$, i.e. for $x_{n+1} = x_n(1 - x_n)$, whatever the initial value of the model, they converge all toward zeros $x_1 = x_2 = x_3 = x_4 = x_5 = \cdots = 0$. Considering what precedes, this fixed point is called *attracting fixed point* for this equation. i.e. there is *an attractor constituted by only one point*. However convergence is slower, indeed, even after 100 iterations, convergence does not succeed completely.
(3) For $\alpha = 2$, i.e. for $x_{n+1} = 2\,x_n(1 - x_n)$, which is always a quadratic equation. Here we find two fixed points, i.e. $x = 0$ and $x = 0.5$. But, here the fixed points are not two attractors. Only $x = 0.5$ is *attracting and stable*, whereas the second $x = 0$ is an *unstable* point or *repelling*.

(4) From $\alpha = 3$, i.e. for $x_{n+1} = 3\, x_n(1 - x_n)$, the answers of the model become complicated. There are still two fixed points, but they are unstable.

(5) For values of $\alpha \in [3, 3.4985...]$, all the trajectories tend towards two fixed values, which are related to α. It is a *first bifurcation* from which we have to face an attractor composed of two unstable points. Whatever the initial value, the variable will tend towards a cycle made up of two values, taking alternatively one or the other value. We are in a *period-doubling*.

(6) For values of $\alpha \in [3.4985, 3.54409, ...]$, a *new bifurcation with a new period-doubling occurs.* Here the attractor is composed of four values, and like in the preceding interval, the variable successively takes alternatively one of these four values.

(7) Moreover, the phenomenon is reproduced twice until obtaining 8 periods, corresponding to the limit value $\alpha = 3.569946...$ Consequently the system becomes *aperiodic*. Around this limit, for values of α with a high number of decimals, we observe a *Cantor set*, which is constituted by a great number of points, our new attractor.

(8) Starting from $\alpha = 3.569946..$, the model has a chaotic mode (regime), i.e. the outputs are "randomly" distributed in the segment [0,1].

Behaviors in the space of motion. Here are behaviors for $\alpha \in [2.8, 4]$:

(a) View of the convergence towards the fixed point, before $\alpha = 3$:

$\alpha = 2.8$ $\alpha = 2.96$ $\alpha = 2.99$
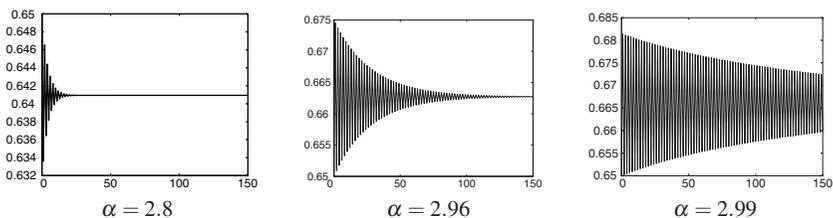
(b) *Period-2:* View of the behavior, after the first period-doubling.

$\alpha = 3.01$ $\alpha = 3.13$ $\alpha = 3.55$
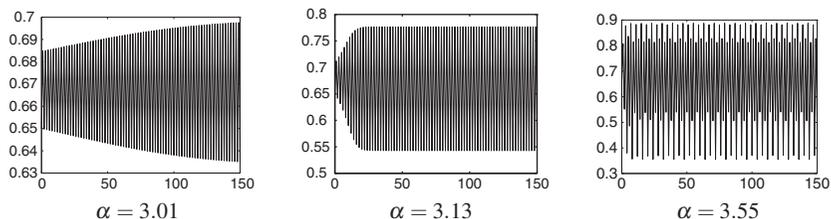
(c) View of the behavior in the "chaotic zone". Some singularities (whose period-icities) are observable, e.g. $\alpha = 3.58, \alpha = 3.59, \alpha = 3.745, \alpha = 3.835$:

$\alpha = 3.58$

$\alpha = 3.59$

$\alpha = 3.62$

$\alpha = 3.64$

$\alpha = 3.65$

$\alpha = 3.67$

$\alpha = 3.685$

$\alpha = 3.73$

$\alpha = 3.745$

$\alpha = 3.805$

$\alpha = 3.82$

$\alpha = 3.835$

$\alpha = 3.865$

$\alpha = 3.88$

$\alpha = 3.895$

$\alpha = 3.91$

$\alpha = 3.955$

$\alpha = 3.98$
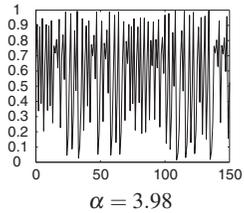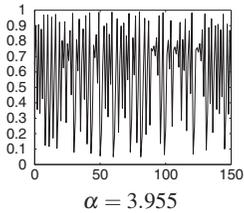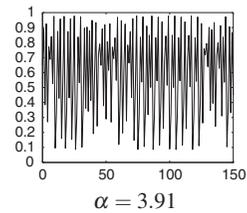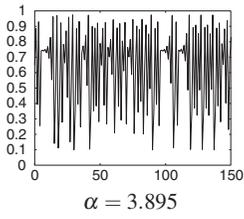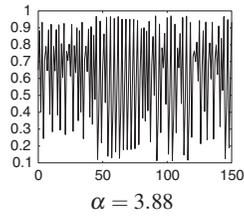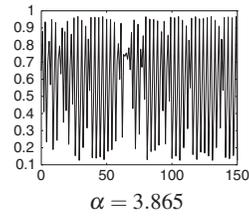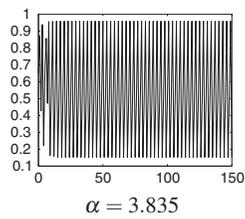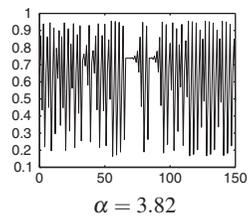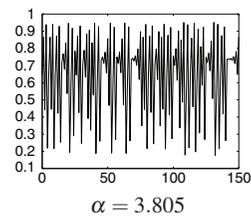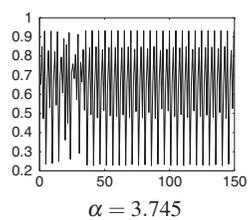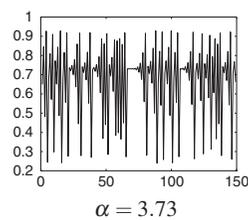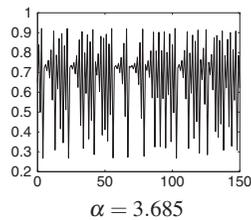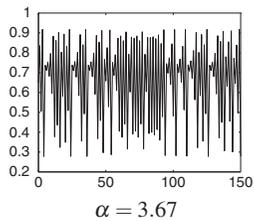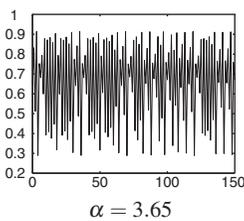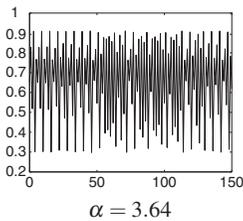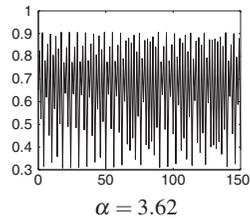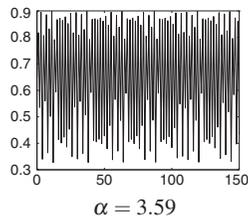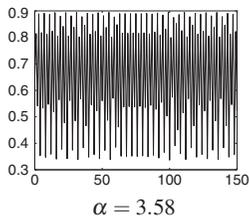
Illustrations for various initial values and for $\alpha = [0.5, 4]$.
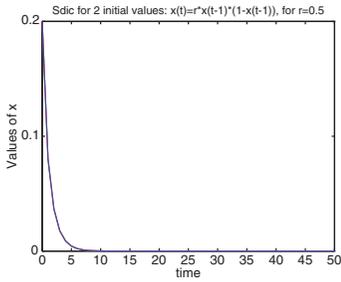
$\alpha = 0.5$
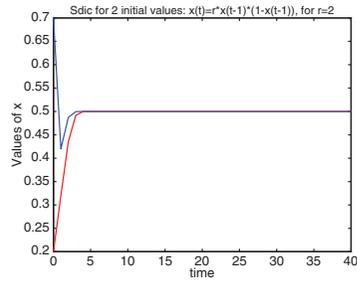
$\alpha = 2$ for $x_0 = (0.2$ and $0.7)$
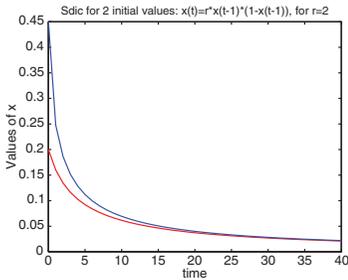
$\alpha = 2$ for $x_0 = (0.2$ and $0.45)$
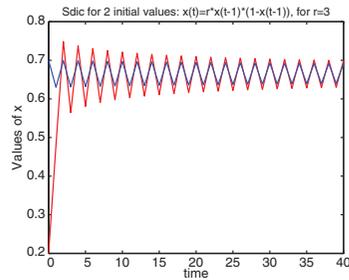
$\alpha = 3$ for $x_0 = (0.2$ and $0.7)$
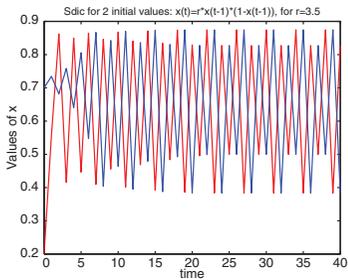
$\alpha = 3.5$ for $x_0 = (0.2$ and $0.7)$
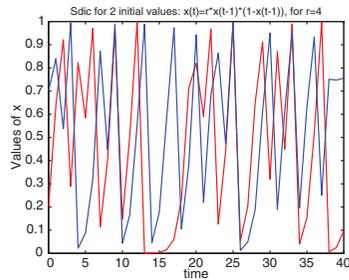
$\alpha = 4$ for $x_0 = (0.2$ and $0.7)$
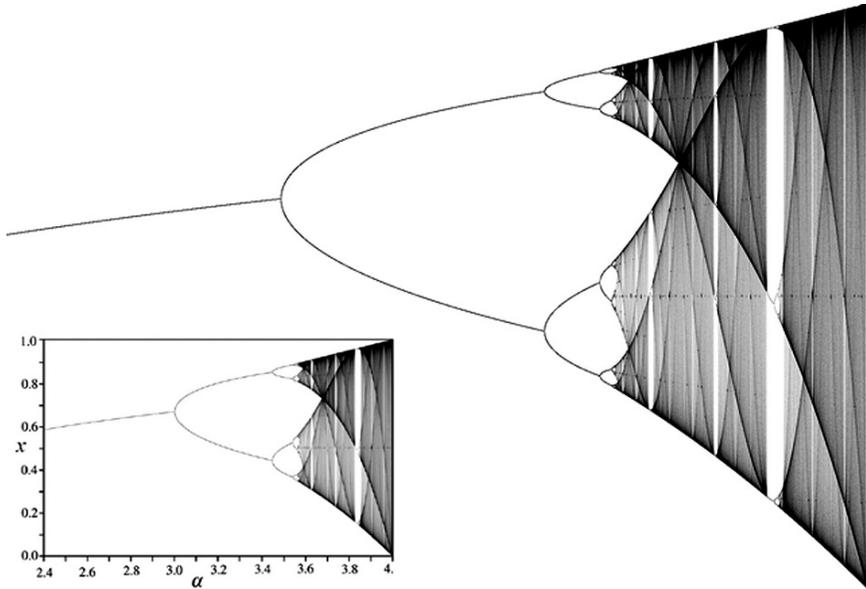
Fig. 1.61 Logistic orbit for $\alpha = 2.6,\ldots,4$
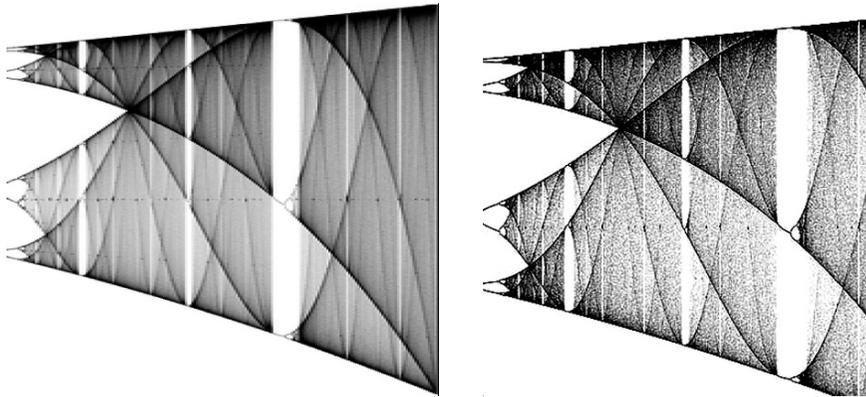
Fig. 1.62 Enlargement of $\alpha = 3.4,\ldots,4$ (*left*). $\alpha = 3.4,\ldots,3.9$ (*right*)

1.30.3 Logistic Orbit

1.30.3.1 Trajectory of the Model According to Fertility Parameter α

This trajectory is the response of the system *when we vary the fertility parameter* for values ranging between two and four. Thus the (α, x)-plane expresses the diversity of outputs of the model according to value change of the parameter (Figs. 1.61–1.63).
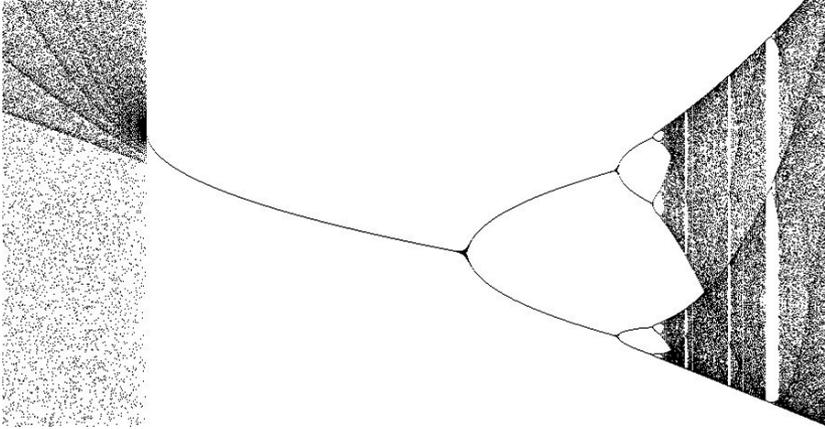
Fig. 1.63 Enlargement of $\alpha = 3.822,\ldots,3.856$

Inside this orbit composed of a great number of points, we observe the behavior of the trajectory which reproduces at a smaller scale the bifurcation phenomenon visible on a normal scale (see Fig. 1.63).

1.30.4 Sensitive Dependence on Initial Conditions

The sensitive dependence on initial conditions (SDIC) *characterizes the chaotic regime*. In the chaotic regime, if we start from two very close points, after a lapse of time (sometimes very short), we can observe that the points are in very distant positions. This is one of the essential characteristics of deterministic chaos. The difference between the two values can be close to the value of the variable itself. Thus, unless we identify with a quasi-infinite precision the value of the initial condition, the long-term evolution of the system is impossible to predict. A graphic illustration below (Fig. 1.64).

The convergence of both behaviors is observed if the calculated series is lengthened. We also observe the predictable behavior of variables, in this *stable regime*. I.e. for a value of $\alpha = 3$.

Above on the left (Fig. 1.65), for $\alpha = 3.56$, the convergence of both simulations is very fast, contrary to the preceding simulation. Lastly, the behavior of two simulations in a chaotic regime, for two very close initial values (0.60 and 0.65) over a short or a long period, shows very clearly that the trajectories are different and do not converge. To underline the divergence of two extremely close initial values, we simulate the behavior of the system for 50 iterations with the following initial values $x_0 = 0.4001$ and $x_0 = 0.4$ (Fig. 1.66). We observe that the divergence starts a little before the tenth iteration.
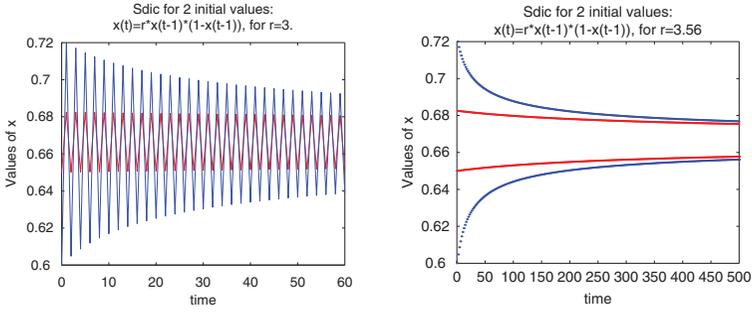
Fig. 1.64 *Left:* $\alpha = 3$; $x_0 = (0.60$ and $0.65)$. *Right:* Points only
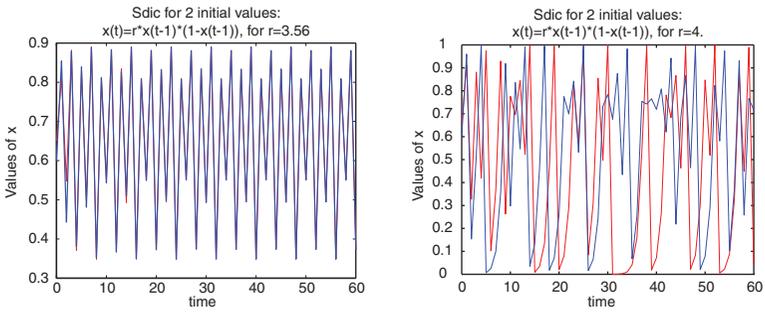
Fig. 1.65 *Left:* $\alpha = 3.56$; $x_0 = (0.60$ and $0.65)$. *Right:* $\alpha = 4$; $x_0 = (0.60$ and $0.65)$
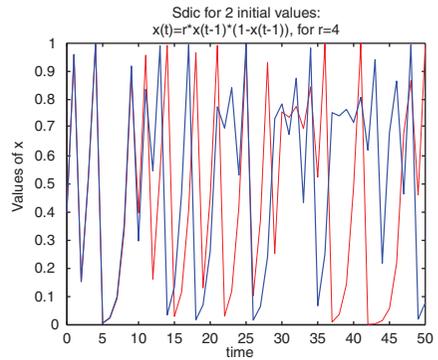
Fig. 1.66 $x_0 = 0.4001$ and $x_0 = 0.4$
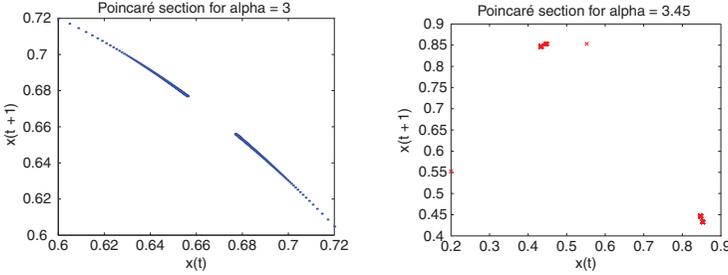
Fig. 1.67 Section for $\alpha = 3$ (*left*). $\alpha = 3.4$. n-period cycle (*right*)
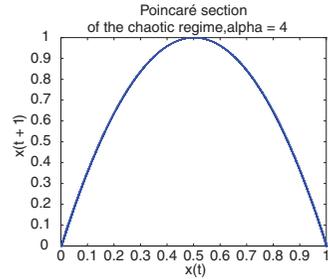
Fig. 1.68 $\alpha = 4$. Chaotic
regime
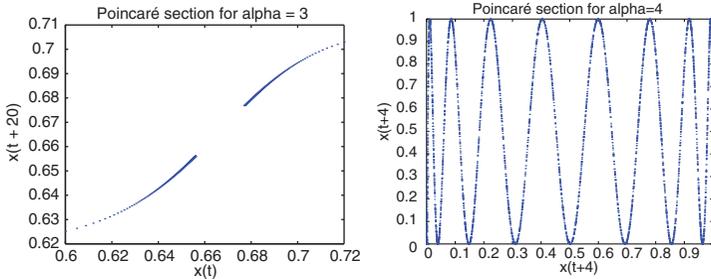
Fig. 1.69 *Left:* $\alpha = 3$ and $[x(t), x(t+20)]$. *Right:* $\alpha = 4$ and $[x(t), x(t+4)]$

1.30.5 Poincaré Sections of the Logistic Equation

Figures 1.67–1.69 are constructed with 5,000 points (iterations). It is advisable to observe the impact of the trajectory in the plane $[x(t), x(t+1)]$. In the first graph, we observe the convergence step by step along the trajectory to the fixed point. In the second, we will notice that the number of different impacts is very small, in spite of the great number of calculated points. Normally for alpha equal to 3.4, we are located in a stable cycle with two fixed points once the convergence was done. At this stage we did not clean the trajectory of points before the "cruising regime" is reached. We will also notice the parabolic form of the Poincaré section when we are in a chaotic regime. Lastly, we produce sections with higher lags.

Note that for a lag of four "steps" the shape of the orbit is sinusoidal, whereas it was parabolic. Obviously, if we observed the points in the plane of the initial trajectory, we would have a *group of points ("a cloud of points")* without any distinct form.

1.30.6 First-Return Map

The first return map written in a generic way $x_{t+1} = g(x_t)$ is often used for the analysis of dynamics and (in spite of its simple form) can reliably reflect aperiodic or chaotic behaviors, as we evoked it in the paragraph relating to the Poincaré section. The logistic equation, which is an unidimensional nonlinear model, can be studied via this first return map, and efficiently explain the aperiodic or chaotic behaviors. This *first return map* is sometimes also called *Poincaré map*. It allows relevant analyses in the study of intermittency, convergence and divergence phenomena. It allows to highlight the solutions of a model: i.e. fixed points, stable, or unstable, etc. In this analysis type, for which we show a figure below, *there exists a "bi-univocal" link between the stability of a fixed point and the tangent slope of the curve at this fixed point, if the modulus* $|.|$ *of the slope is lower than 1.* This type of graph allows to study the iterations of the model starting from an initial condition x_0. The first iteration $x_1 = g(x_0)$ is obtained by seeking the intersection of this curve with the vertical of the abscissa x_1. The procedure continues in this way until the fixed point or the divergence. We will use this type of map again in the section relating to the Cobweb diagram. First-return maps of the logistic model are depicted in Fig. 1.70 for an aperiodic regime $\alpha = 4$ and for $\alpha = 2.95$.
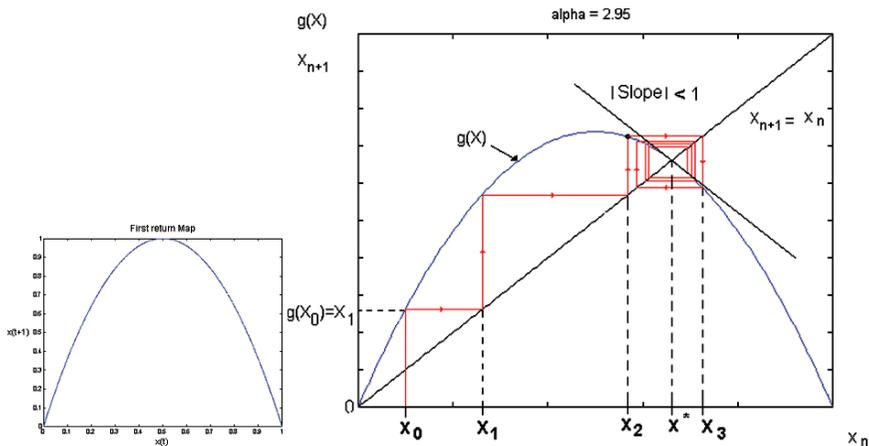
Fig. 1.70 $\alpha = 4$ (*left*). $\alpha = 2.95$ (*right*)

1.30.7 Solutions and Stability of the Model

The quadratic map or logistic equation (Verhulst 1844, May 1976) is written $x_{k+1} = \alpha x_k (1 - x_k)$ where x is the "population" variable nomalized on $[0,1]$. For this discrete time equation, by finding the solution as k tends towards the infinite, we can determine the steady-state behavior. Consider $x_{k+1} = \alpha x_k - \alpha x_k^2$, as we approach the steady-state solution $x_{k+1} = x_k$, (if x^* denotes the solutions). So we can write: $x^* = \alpha x^* - \alpha x^{*2}$, or

$$\alpha x^{*2} - (\alpha - 1)x^* = 0. \tag{1.276}$$

We find (via quadratic formula) the steady-state (fixed-point) solutions

$$x^* = 0, \quad x^* = \frac{\alpha - 1}{\alpha}. \tag{1.277}$$

Thus, if we suppose that the initial population is zero, it will remain at zero under the action of the system. For a non-zero initial condition, we will be able to expect a convergence (steady-state) to $\alpha - 1/\alpha$. For four *symptomatic* values of α, let us show the non-zero solutions that can be expected:

$\alpha :$	2.95	3.20	3.50	3.75
$x^* :$	0.6610	0.6875	0.7143	0.7333

1.30.7.1 Stability Theorem of Fixed Point solutions

By definition we know that x^* is a *fixed point* if it is solution of $x^* = g(x^*)$ or $g(x^*) - x^* = 0$. Consequently, the *stability theorem* is written: Theorem (Stability) x^* is a stable solution of $x^* = g(x^*)$, if $\left| \frac{\partial g}{\partial x} \right| < 1$, when evaluated at the point x^*.

1.30.8 Stability Theorem Applied to Logistic Equation

Remember that the map is written $g(x) = \alpha x(1 - x) = \alpha x - \alpha x^2$, thus:

$$\frac{\partial g}{\partial x} = \alpha - 2\alpha x = \alpha(1 - 2x), \tag{1.278}$$

or by changing the notation: $g'(x^*) = \alpha - 2\alpha x^* = \alpha(1 - 2x^*)$. If we apply the theorem of stability, the following conditions must be satisfied:

$$|\alpha(1 - 2x^*)| < 1, \tag{1.279}$$

consequently x^* is a stable solution. In the preceding section, the general solutions were calculated

$$x_1^* = 0, \quad x_2^* = \frac{\alpha - 1}{\alpha}. \tag{1.280}$$

It remains, therefore, to numerically study the *stability* for the values of α wished. If we take the values of α stated in the preceding table again, which are interesting for the model, we find:

α	x_1^*	$|g'(x_1^*)|$	Stability for x_1^*	x_2^*	$|g'(x_2^*)|$	Stability for x_2^*
2.95	0	2.95	Unstable	0.6610	0.9499	Stable
3.20	0	3.20	Unstable	0.6875	1.2000	Unstable
3.50	0	3.50	Unstable	0.7143	1.5000	Unstable
3.75	0	3.75	Unstable	0.7333	1.7500	Unstable

1.30.9 Generalization of the Stability of (Point) Solutions of the Quadratic Map: Generic Stability

- *Generic stability of x_1^*:*
  x_1^* being equal to zero and $g'(x_1^*) = \alpha - 2\alpha(0) = \alpha$, then $|g'(x_1^*)| = |\alpha|$. The stability criterion being $|g'| < 1$, as long as $-1 < \alpha < 1$, then x_1^* is a stable solution.
- *Generic stability of x_2^*:*
  x_2^* being equal to $(\alpha - 1)/\alpha$, and $g'(x_2^*) = \alpha - 2\alpha(\alpha - 1)/\alpha = \alpha - 2(\alpha - 1) = -\alpha + 2$, thus $|g'(x_2^*)| = |-\alpha + 2|$. Consequently, there is stability when $1 < \alpha < 3$ and instability outside.

1.30.10 Bifurcation Diagram

The set of solutions for $1 < \alpha < 4$ can be illustrated by the Fig. 1.71.

A change of stability appears for x_1^* at $= 1$. For x_2^*, changes of stability appear at $\alpha = 1$ and $\alpha = 3$. The values of α, for which the characteristics of stabilities change, correspond to bifurcation points. At $\alpha = 1$, the bifurcation is called *transcritical*, because there is an *exchange of stability between x_1^* and x_2^** (subject of the graph Fig. 1.71). This diagram based on a linear stability analysis should not be confused with the orbit diagram that highlights the periodic behavior resulting from the resolution of the nonlinear algebraic equation but cannot exhibit unstable solutions unlike the bifurcation diagram.

1.30.11 Monotonic or Oscillatory Solution, Stability Theorem

From the stability criterion previously defined, we can refine the analysis by characterizing the behavior and trajectory of the model through the analysis of $\partial g / \partial x$.
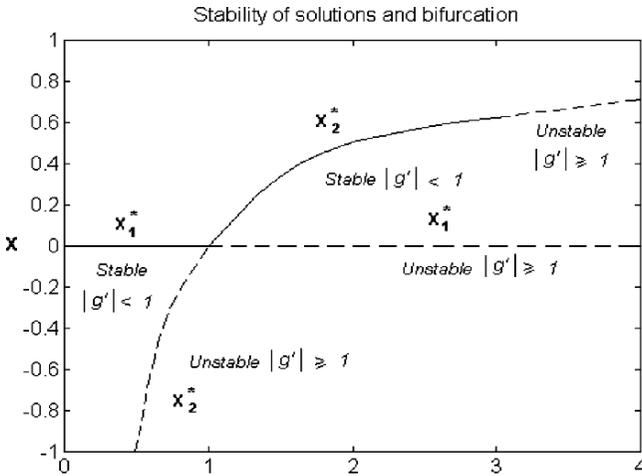
Fig. 1.71 Bifurcation diagram stable solutions (*solid lines*), unstable solutions (*dashed lines*)

$\partial g/\partial x$	Stability	Characteristics of motion
$\partial g/\partial x < -1$	Unstable	Oscillatory
$-1 < \partial g/\partial x < 0$	Stable	Oscillatory
$0 < \partial g/\partial x < 1$	Stable	Monotonic
$1 < \partial g/\partial x$	Unstable	Monotonic

This approach constructed from the analysis of the stability cannot be used directly to study the periodic behaviors, which are developed in the section about the period-doublings and the harmonic cascade.

1.30.12 Lyapunov Exponent Applied to the Logistic Map

1.30.12.1 Some Lyapunov Exponents of the Logistic Map

Below, we use the logistic equation and we produce an output (500 points) for different values of α. For each outcome produced, we calculate the Lyapunov exponent by eliminating the first 50 values of the signal (i.e. convergence phase) during the iterations to avoid perturbations in the calculation (Figs. 1.72–1.74).

Fast divergence of trajectories is observed in the last cobweb diagram.

1.31 Coupled Logistic Maps and Lce's

(A) In a one-dimensional map $x_{n+1} = f(x_n)$, the function f can be any linear or nonlinear map (e.g. *logistic map*). A two-dimensional discrete-time system can be obtained by coupling two such one-dimensional maps (von Bremen et al. 1997):
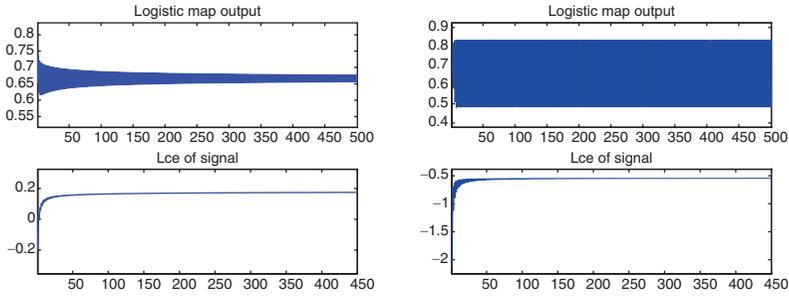
Fig. 1.72 For $\alpha = 3$ (*left*). For $\alpha = 3.3$ (*right*)
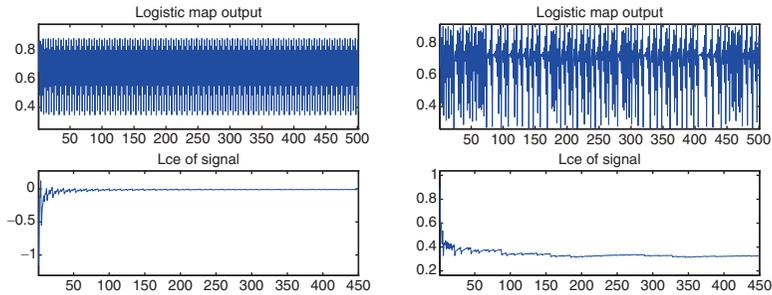
Fig. 1.73 For $\alpha = 3.565$ (*left*). For $\alpha = 3.7$ (*right*)
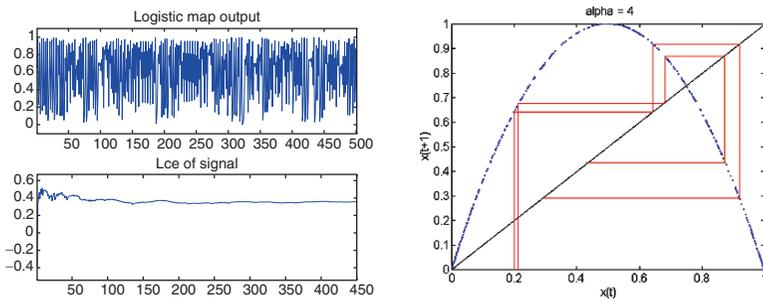
Fig. 1.74 For $\alpha = 4$ (*left*). For $\alpha = 4$ (*right*)

$\{x_{n+1} = df(x_n) + (1 - d)f(y_n), \; y_{n+1} = (1 - d)f(x_n) + df(y_n)\}$ where d is the coupling parameter. This form of coupling often arises in physical systems. One can think of $f(x_n)$ and $f(y_n)$ as simulating the population dynamics of a particular species (e.g. the species can be any biological or chemical species, or even scalar fields such as temperature) at two adjacent locations. If, after every time increment, only a fraction d of these species remains in the same location and the rest migrate to the other location, their dynamics is described by the system above. The coupling parameter d can vary between 0 and 1. If $d = 1$, there is no coupling and if $d < 1$, there is coupling. In short, the mapping described by the system above is
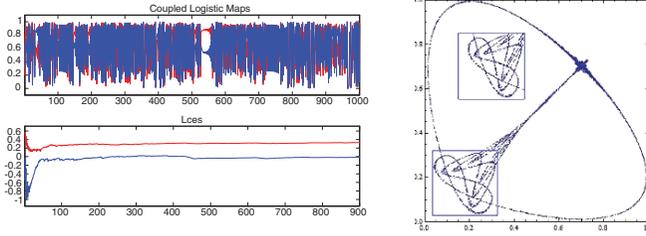
Fig. 1.75 (**a**) $\alpha = 3.6$, $d = 0.2$, $x_0 = 0.2$, $y_0 = 0.3$, Lce's. (**b**) $\delta_x = 1.0776$, $\delta_y = 1.0812$

denoted $(x_{n+1}, y_{n+1}) = M(d) \circ (x_n, y_n)$, to explicitly indicate the dependence on the parameter d.

The Lyapunov exponents are given by $\lambda_i = \lim_{n \to \infty} \log_e (\|\mu_i\|/n)$ where μ_i are the eigenvalues of the product of the Jacobian matrices at every iteration. The two-dimensional system above has two Lyapunov exponents. The Jacobian matrix for this system is

$$J_n[M(d)]_{(x_n, y_n)} = \begin{bmatrix} df'(x_n) & (1-d)f'(y_n) \\ (1-d)f'(x_n) & df'(y_n) \end{bmatrix}$$

(see Fig. 1.75a, 1,000 iter.).

(B) The coupled logistic map (see Fig. 1.75b) can also be given, in a different way, by the equations $\{x_{n+1} = \delta_x(3y_n + 1)x_n(1 - x_n), y_{n+1} = \delta_y(3x_n + 1)y_n(1 - y_n)\}$. This system exhibits chaotic behaviors when δ_x and δ_y lie in the neighborhood of the region $[1.032, 1.0843]$. For certain ranges of the values of the parameters and the initial conditions, the dynamics can also converge to periodic orbits.

1.31.0.2 Lyapunov Exponent of the Trajectory of the Logistic Orbit

Here are (1) the orbit of the logistic model and (2) a continuous graph equivalent to the table of critical points for $\alpha \in [3, \ldots, 4]$:
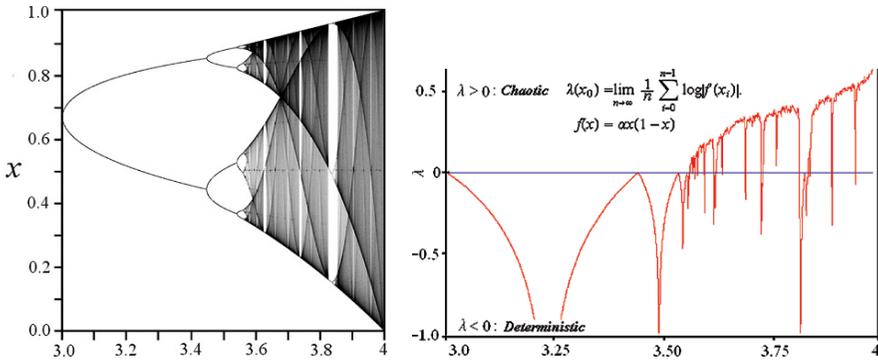
1.31.0.3 Critical Points of the Model

Calculation of some Lyapunov exponents for some values of the Logistic equation when the number of iterations is equal to 4,000 and $x_0 = 0.5$.

α	λ	Comment
1	$-0.005112\ldots$	Start stable fixed point
1.99	$-6.643\ldots$	
1.999	$-9.965\ldots$	
2	Calculator error*	
2.001	$-9.965\ldots$	Superstable fixed point
2.01	$-6.643\ldots$	
3	$-0.003518\ldots$	Start 2-cycles
3.236067977\ldots	$-19.43\ldots$*	Superstable 2-cycles $(1+\sqrt{5})$
3.449489743\ldots	$-0.003150\ldots$	Start stable 4-cycles $(1+\sqrt{6})$
3.5699456720	$-0.002093\ldots$	Start of chaos (Hofstadter)
3.56994571869	$+0.001934\ldots$	Start of chaos (Dewdney)
3.828427125\ldots	$-0.003860\ldots$	Start stable 3-cycles $(1+\sqrt{8})$
3.9	$+0.7095\ldots$	Back into chaos
4	$+2$	End of chaos

1.31.1 Period-Doubling, Bifurcations and Subharmonic Cascade

The "route to chaos" of the quadratic function is done by successive period-doublings, i.e. bifurcations. The method previously presented within the framework of the stability study and the fixed points is limited. It could predict that a *particular fixed point was unstable*, but could not identify the type of periodic behavior that could appear. In this section, we will show how to find these bifurcation points and period-doublings and their respective branches.

Period-2: When a period-doubling appears, the value of the output at the moment k is equal to the value of the output at the moment $k-2$. In fact, during its movement the point takes positions alternatively on the two branches of the bifurcation, running "step by step" from one branch to another. This phenomenon can be expressed in the following way:

$$x_k = x_{k-2}, \quad \text{or} \quad x_{k+2} = x_k \tag{1.281}$$

Knowing that: $x_{k+1} = g(x_k) = \alpha x_k(1-x_k)$, then we can write

$$x_{k+2} = g(x_{k+1}), \quad x_{k+2} = g(g(x_k)), \quad x_{k+2} = g^2(x_k) \tag{1.282}$$

For the logistic (or quadratic) equation, g^2 will be written in the following way:

$$x_{k+2} = \alpha x_{k+1}(1-x_{k+1}) \tag{1.283}$$

and if we replace: $x_{k+1} = \alpha x_k(1 - x_k)$, in the equation above, we find $x_{k+2} = \alpha[\alpha x_k(1 - x_k)][1 - (\alpha x_k(1 - x_k))]$ and since $x_k = x_{k-2}$, we write $x_k = \alpha[\alpha x_k(1 - x_k)][1 - (\alpha x_k(1 - x_k))]$ this equation becomes $x_k = \alpha^2[-\alpha x_k^4 + 2\alpha x_k^3 - (1+\alpha)x_k^2 + x_k]$ i.e. $g^2(x_k) = \alpha^2[-\alpha x_k^4 + 2\alpha x_k^3 - (1+\alpha)x_k^2 + x_k]$. We have to face a polynomial of degree 4, and we can find the solutions of this equation by a graphical representation of g^2 and x. This is done in the figure shown below on the base of a parameter value $\alpha = 3.2$. The precise observation of the figure makes it possible to find the following four solutions (x_s) for this behavior of period-2.

$$x_s = 0, \quad x_s = 0.5130, \quad x_s = 0.6875, \quad x_s = 0.7995. \tag{1.284}$$

In the figure, we can also see that the solutions $x_s = 0$ and $x_s = 0.6875$ are *unstable, since the slopes of $g^2(x_s)$ are larger than 1:*

1. Indeed a solution of period-2 is *stable* if $\left|\partial(g^2(x_s))/\partial(x)\right| < 1$.
2. Besides, a solution for $x = g(x)$ will always appear as one of the solutions of $x = g^2(x)$.
3. If a solution is unstable for $x = g(x)$, then it will be also unstable for $x = g^2(x)$.
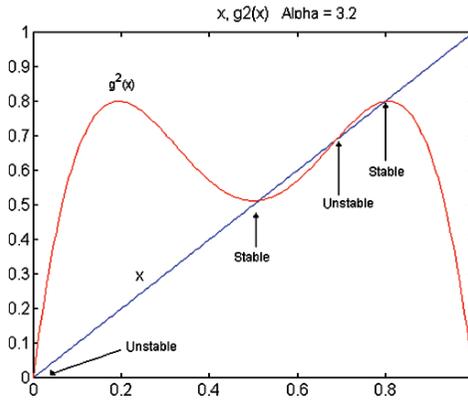
To verify the remarks which have been just stated, we will observe in the graph hereafter that $x = 0.6875$ is at the same time solution of $x = g(x)$ and $x = g^2(x)$.
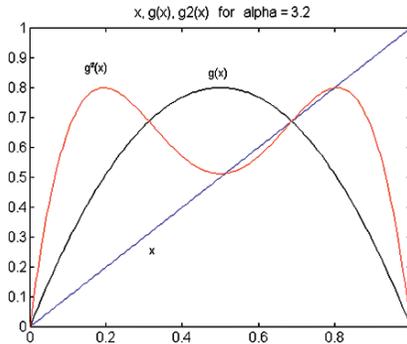
Period-1: We can show the solution of $x = g^2(x)$ when $\alpha = 2.95$ of which we know that it is single and asymptotically stable. This solution is $x = 0.6610$, when the number of iterations is large, i.e. tends towards the infinite. In other words, asymptotically $x_{k+2} = x_{k+1} = x_k = 0.6610$.

x, g(x), g2(x) for alpha = 2.95
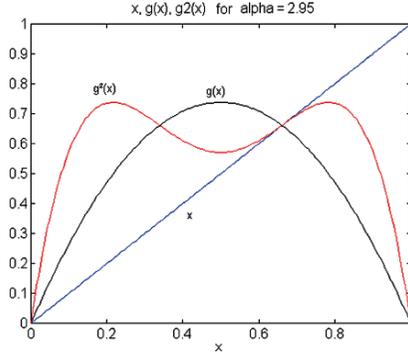
The singularity: $\alpha = 3$. To this value of α corresponds a bifurcation point, since the absolute value of the slope of $g(x)$ and of $g^2(x)$ is equal to 1, for $x = 0.66667$. The figure below is a transition between the two figures which precede. Even if the difference is not easy to see, we will observe the tangency of $g^2(x)$ and $x = x$.

x, g(x), g2(x) for alpha=3.0
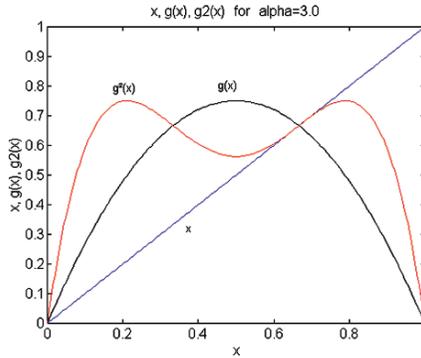
Period-4: When a period-4 appears, the value of the output at the moment k is equal to the value of the output at the moment $k - 4$. In fact, during its displacement the point takes positions alternatively on the four branches of the bifurcation. This phenomenon can be expressed in the following way:

$$x_k = x_{k-4}, \quad \text{or} \quad x_{k+4} = x_k. \tag{1.285}$$

Knowing that: $x_{k+1} = g(x_k) = \alpha x_k(1 - x_k)$, then we can write

$$
\begin{aligned}
x_{k+4} &= g(x_{k+3}) \\
x_{k+4} &= g(g(x_{k+2})) \\
x_{k+4} &= g(g(g(x_{k+1}))) \\
x_{k+4} &= g(g(g(g(x_{k+2})))) \\
x_{k+4} &= g^4(x_k).
\end{aligned}
\tag{1.286}
$$

For the logistic equation, g^4 is a polynomial of order eight, with eight solutions, which are shown in the graph, calculated below from $\alpha = 3.5$:

x, g4(x) for alpha = 3.5
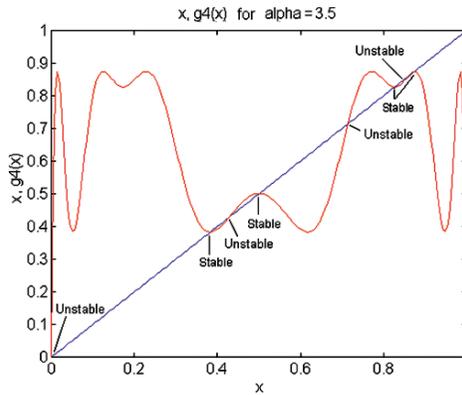
It will be noticed below, that even unstable solutions of $x = g(x)$ are also solutions of $x = g^2(x)$ and of $x = g^4(x)$.

x, g(x), g2(x), g4(x) for alpha=3.5
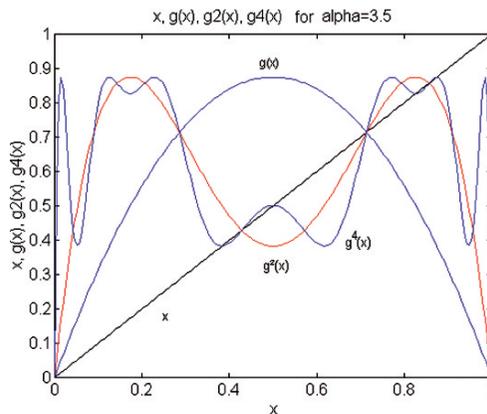
Period-n: By extension of what was done for period-2 and period-4, we can write: $x_k = x_{k-n}$, or $x_{k+n} = x_k$ and $x_{k+n} = g^n(x_k)$. g^n *is a polynomial of degree 2n, which admits 2n solutions, whose n of them are stable.*

1.31.1.1 Iterative Functions and Fixed Points (Case $\alpha = 3.56$)

We already evoked in another section the occurrence of fixed points in the logistic model. In short it is known that for:

- $0 < \alpha \leq 1$: The single *fixed point* is $x = 0$ and it is *asymptotically stable*.
- $1 < \alpha < 3$: The point $x = 0$ is unstable and the fixed point is equal to $x = 1 - 1/\alpha$. It is asymptotically stable and its *domain of attraction* on x is $0 < x < 1$.

On the other hand, for the values which follow, the analysis is much less elementary:

- $\alpha > 3$: By increasing the values of α from 3 to 4, we observe periodic orbits which appears with period-doublings 2, 4, 8,... At each value of α for which a period-doubling appears, there is a bifurcation. And there is an infinite number of bifurcations and period-doubling until the value $\alpha_\infty = 3.569946\ldots$

Fixed Points and Periodic Points of Period-k, for $\alpha = 3.56$

For a fixed α equal to 3.56 for example, if we call k the *number of "periods"* of a trajectory, for $k = 1,\ldots,9$. Then we can write for $g(x) = (3,56)x(1-x)$, with $x \in [0,1]$:

$$g^k(f(x)) = g(g(g(\cdots g(x)))) = g^k\left((3,56)x(1-x)\right). \qquad (1.287)$$

The periodic points of period k are identified when $x = y$, i.e. during the intersection of g^k with the diagonal x. The fixed points of g are indicated by the symbol "•" at the intersection of curves.
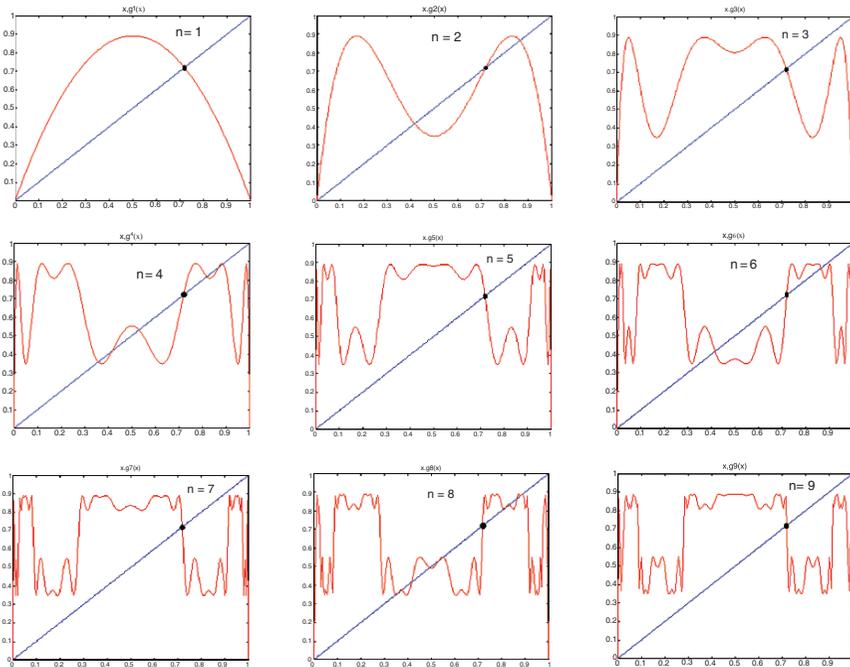
Fixed points are noted in the graphs by • at the intersection of x and g^k.

Fixed Points and Periodic Points of Period-k for $\alpha = 3.99027$

$g^k(x); k = 1, \ldots, 10; g(x) = \alpha x(1 - x)$; with $\alpha = 3.99027$.
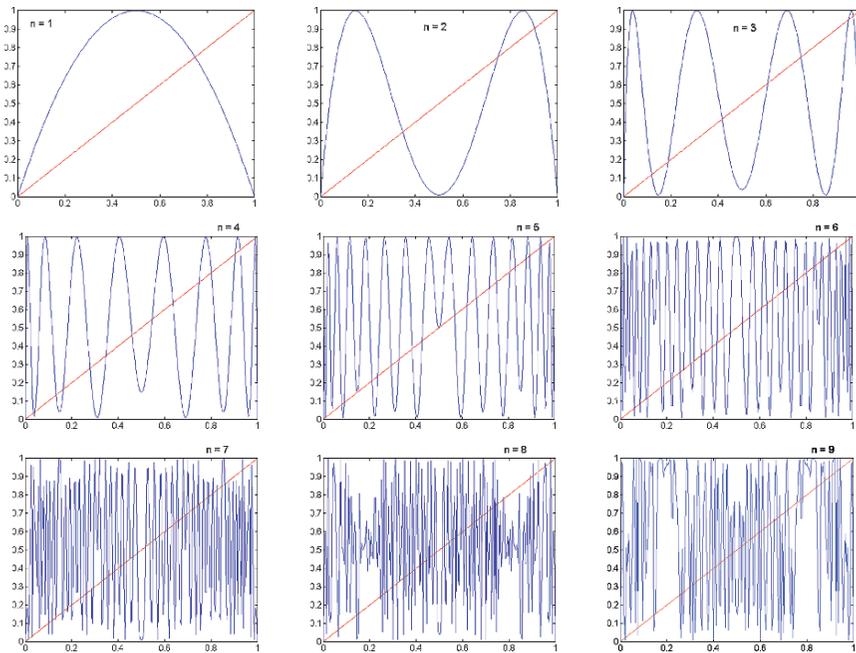
1.31.2 Subharmonic Cascade, Accumulation Point

1.31.2.1 Subharmonic Cascade

When the period-doubling phenomenon consecutive to a bifurcation is repeated an "infinite" number of times, we are faced with a cascade of bifurcations. This type of cascade occurs when they are associated with an subharmonic instability. In a very general way, it is possible to speak of *the notion of subharmonic instability of a periodic solution*. Here, we must understand the way in which a periodic equilibrium can have its stability called into question. Indeed, a periodic solution can see its equilibrium called into question, if the control parameter modification intervenes with time intervals separated by an integer number of half-periods $T/2$, that is to say $N \times (T/2)$, by knowing that the destabilization is optimal if it is located exactly at the half-period. Thus the suitable variation of the control parameter of a system must "theoretically" facilitate the generation of subharmonic instabilities giving place to successive period-doublings 2T, 4T, 8T, etc., i.e. subharmonic cascades. The subharmonic term refers to the fact that the stimulation of the system

intervenes inside the basic period which constitutes the harmonic. Thus, we say that it is subharmonic.

1.31.2.2 Accumulation Point of a Subharmonic Cascade

Let's take the generic case of a function of the type: $x_{t+1} = 4\alpha x_t (1 - x_t)$ and pose the following iterated functions: $x_{k+1} = g(x_k) = 4\alpha x_k (1 - x_k)$ and $x_{k+2} = g^2(x_k) = g(g(x_k))$. We describe the operation of iterated functions in another section of the chapter, but we can say right now that we have *a bifurcation-point* between a period-1 and a period-2 *if the absolute value of the slope of $g(x)$ and $g^2(x)$ is equal to 1, at the point of intersection* of $g(x)$ of $g^2(x)$ and $x = x$. See the graph below.
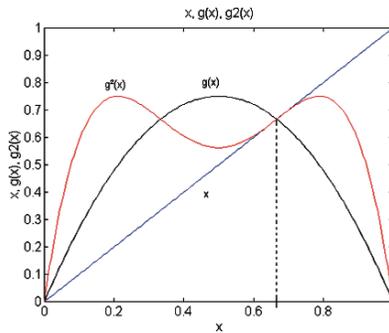
Within the framework of the bifurcation points study for the quadratic equation above, we list the values of the parameter α for which there is change of period. These points are gathered in the table below.

Period	α at bifurcation point	For 4α
$1.2^0 = 1$		
	$\alpha_1 = 0.75$	$4\alpha = 3$
$1.2^1 = 2$		
	$\alpha_2 = 0.86237\dots$	$4\alpha = 3.4495\dots$
$1.2^2 = 4$		
	$\alpha_3 = 0.88602\dots$	$4\alpha = 3.5441\dots$
$1.2^3 = 8$		
	$\alpha_4 = 0.89218\dots$	$4\alpha = 3.5687\dots$
$1.2^4 = 16$		
	$\alpha_5 = 0.8924728\dots$	$4\alpha = 3.5699\dots$
$1.2^5 = 32$		
	$\alpha_6 = 0.8924835\dots$	$4\alpha = 3.5699\dots$
$\vdots$		
$1.2^\infty = \infty$	$\alpha_\infty = 0.892486418\dots$	$4\alpha = 3.5699\dots$

Observe this remarkable phenomenon which shows for an increasing series 1.2^p (with p an integer) of period, that the parameter converges towards an accumulation point $\alpha_\infty = 0.892486418\ldots$.

1.31.3 Stable Cycles and Super-Stable Cycles

1.31.3.1 Stable or Unstable Cycles

It is said that a set of points $x_1, x_2, \ldots, x_{p-1}$ constitutes a *stable p order if* $g(x_i) = x_{i+1}$ for $i = 0, 1, 2, \ldots, p-2$ and $g(x_{p-1}) = x_0$. This means that each point of a cycle of p order is a fixed point for $g^p : g^p(x_i) = x_i$ for $i = 0, 1, 2, \ldots, p-1$ and is not a fixed point for g^k if $k < p$. If we consider a periodic point x of a cycle of p order, if the eigenvalues of the $Dg^p(x)$ matrix are *inside the unit circle*, then *the cycle is stable*. If one of the eigenvalues has its absolute value higher than 1, then the *cycle is unstable*.

1.31.3.2 Attracting, Repelling and Super-Attracting Cycles

If $g^p(x_i) = x_i$ we can write the derivative of g^p at the point x_0:

$$(g^p)'(x_0) = g'(x_{p-1}) \cdots g'(x_1) g'(x_0). \tag{1.288}$$

However, it is known that $x_0 = x_p$ thus we can say that all the $(g^p)'(x_i)$ are equal to the same value noted m_p, for $i = 0, 1, \ldots, p-1$. where m_p is called the *multiplier of cycle. We will say* that a set of points $x_1, x_2, \ldots, x_{p-1}$ is:

Behaviors	Conditions
Attracting (or stable)	if $\lvert m_p \rvert < 1$,
Repelling (or unstable)	if $\lvert m_p \rvert > 1$,
Indifferent	if $\lvert m_p \rvert = 1$,
Super-attracting (or super-stable)	if $m_p = 0$.

1.31.4 Cobweb Diagram

1.31.4.1 Outputs Field of the Model

Before studying the dynamics of the model by a cobweb diagram, the most basic observation that we can do, is the increase of the domain in which the variable of the quadratic map is spread. The observation of diagrams below makes it possible to note the increase in the field of outputs that the model gives during the variation of the fertility parameter α, which passed from 1 to 4 (Fig. 1.76).
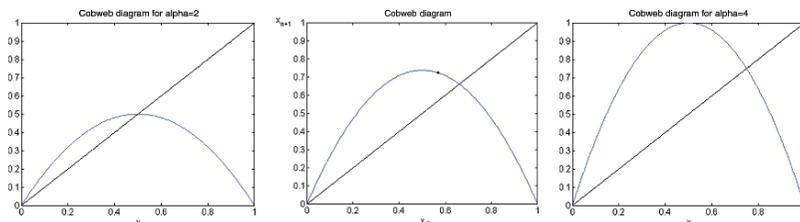
Fig. 1.76 $\alpha = 2$ *(left)*. $\alpha = 2.95$ *(center)*. $\alpha = 4$ *(right)*

Indeed for $\alpha = 4$, the responses occupy Cobweb diagram the totality of the segment $[0,1]$, whereas for $\alpha = 1$, the field of responses occupies the segment $[0,0.25]$. For $\alpha = 2$, the answer belongs to $[0,0.5]$. And for the segment of responses $[0,0.75]$, $\alpha = 2.95$. Thus, the studied dynamics will be spread in more or less vast fields.

1.31.4.2 Dynamics of the Logistic Map

It is very interesting to study the behavior of a system with a unique discrete variable, by using the Cobweb diagram. This is a frequently used tool in economics to observe the dynamics of behaviors. It makes it possible, in particular, to identify solutions of the model, fixed points, divergent trajectories, convergent trajectories, transitory movements, oscillatory motions, and "chaotic" behaviors. The "Cobweb" clearly *does not present the "temporality"* of the model, but their "dynamics" by *visualizing* the *periodicities, stabilities, asymptotic stabilities, explosive motions, transitory oscillations, intermittencies or behaviors*. For the logistic equation, this diagram is obtained by tracing two curves in a graph, i.e. x and $g(x) = \alpha x(1-x)$. The solutions of the model are indicated by the intersection of both curves specifying thus the *fixed points* of the system.

We position an initial condition x_0 in the graph and more exactly on the layout of the curve x, then we observe the response of the logistic model $g(x)$ for this "input". Then, by a simple projection, the obtained output projected on the curve x re-becomes an "input", which will be the subject of a new iteration under the action of $g(x)$. Thus, with the second iteration, we obtain a new output, which in its turn is reintroduced in the loop as an input by the projection on the curve x. The latter is ready again to be the subject of a third iteration, and so on until the end of the selected number of iterations. In the case $\alpha = 2.95$, the curve $g(x) = x_{n+1} = 2.95x_n(1 - x_n)$ seems a reversed parabola. From the initial condition $x_0 = 0.1$, a value x_1 is obtained by tracing a vertical line to the curve $g(x)$, where $g(x_0) = 0.265$ is found, then we trace a horizontal line to the curve $x = x$, since $x_1 = g(x_0)$, consequently $x_1 = 0.265$. After this, a vertical line is traced to the curve $g(x)$ to obtain $g(x_1) = 0.575$, then, a vertical line is traced to the curve $x = x$, consequently, $x_2 = 0.575$. This progression, step by step is depicted in the figure below:
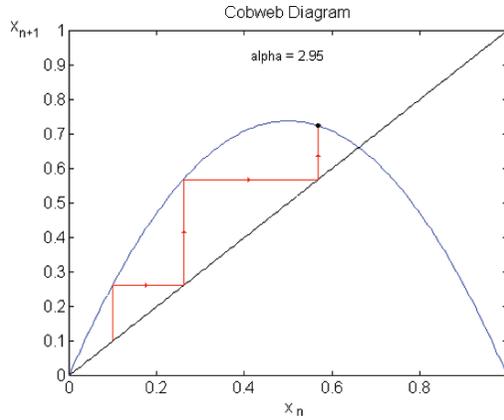
Period-n with n = 1,2,3,4 and Chaotic Regime

Period-1: The figure below shows a process which *converges* towards a *fixed point* $x_\infty = 0.661$, when $\alpha = 2.95$. From the initial condition $x_0 = 0.1$, and before the fixed point, we observe a transitory motion which converges.
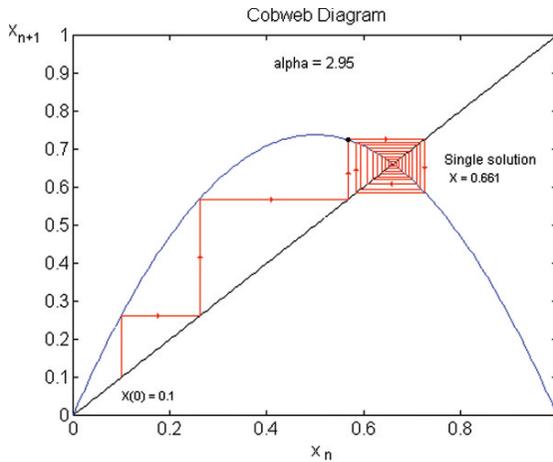
The oscillation between 0.5130 and 0.7995 corresponds to a *transitory motion*.

Period-2: The two figures below show that for a value of $\alpha = 3.2$, the iterative process can adopt different behaviors, according to the selected initial value:

(a) For $x_0 = 0.5130$, and $\alpha = 3.2$, there is an oscillation between $x = 0.5130$ and 0.7995, *without transitory motion*.

Cobweb Diagram
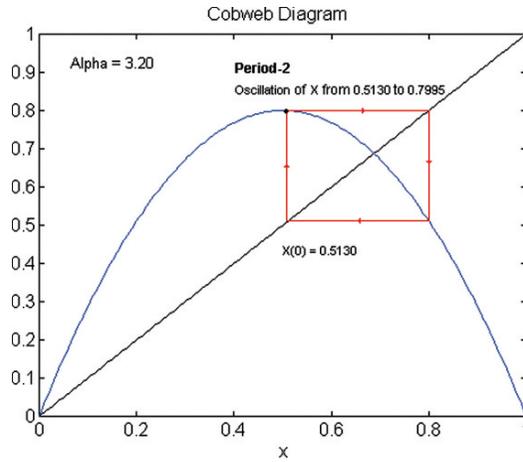
(b) For $x_0 = 0.1$, and $\alpha = 3.2$, there is an oscillation between $x = 0.513, 0.7995$, *with a transitory motion.*

Cobweb Diagram
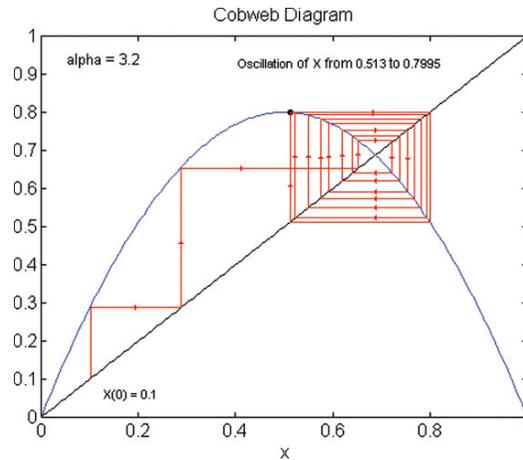
Period-4:

(a) For $x_0 = 0.1$, and $\alpha = 3.5$, there is an oscillation between $x = 0.3828, 0.8269$, 0.5009 and 0.8750, *with a transitory motion.*

Cobweb Diagram
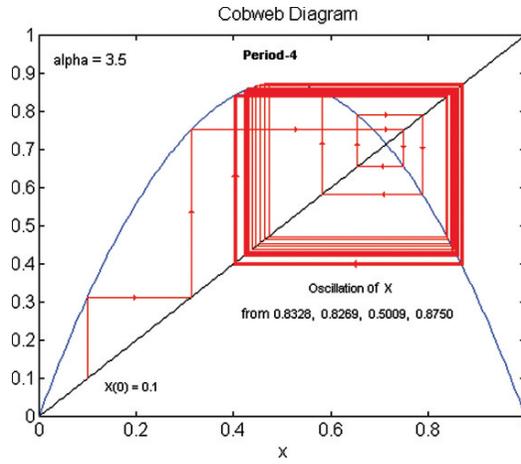
(b) For $x_0 = 0.3828$ and $\alpha = 3.5$, there is an oscillation between 0.3828, 0.8269, 0.5009, an 0.8750, *without transitory motions*.

Cobweb Diagram
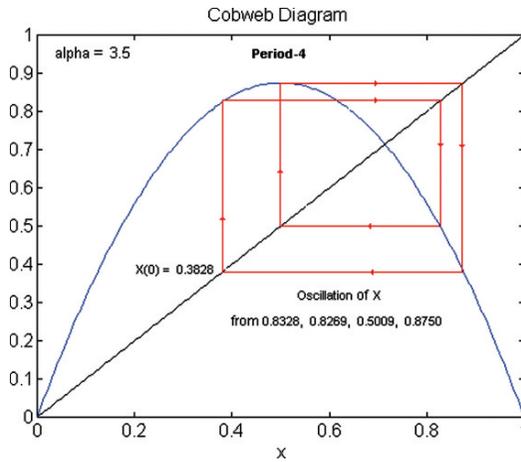
Chaotic regime: For $\alpha = 3.75$, the behavior is chaotic and the responses are transitory.

Cobweb Diagram
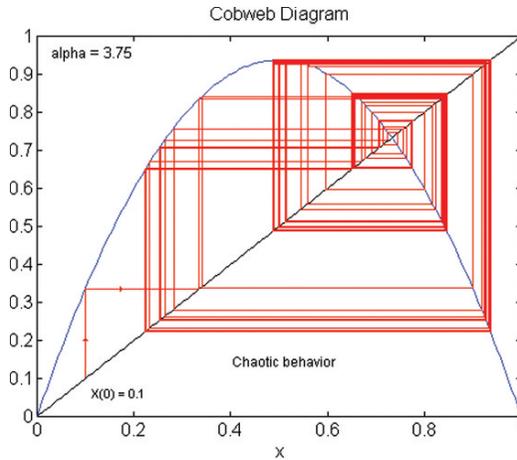
Period-3: For $\alpha = 3.83$, the response of the model is alternatively $0.15615, 0.50466$, and 0.957417, i.e. a period-3 which implies *a chaotic behavior*.

Cobweb Diagram
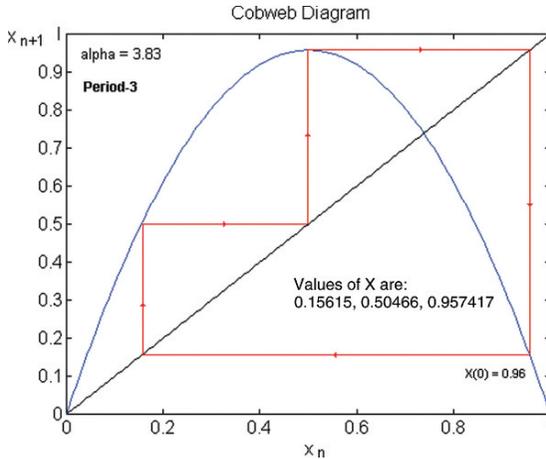
1.31.5 *Bifurcation Measure or Feigenbaum Constant*

For the logistic equation, the numerical calculations show that for a rather large number of iterations n, the computed values satisfy the following ratio:

$$\delta = \lim_{n \to \infty} \frac{\alpha_n - \alpha_{n-1}}{\alpha_{n+1} - \alpha_n}. \tag{1.289}$$

Calculated in a precise way, this constant δ also called the *universal constant*, has a value equal to $4.66920161\ldots$. During this route to chaos, the model of the logistic equation, shows period-doublings which give place to bifurcations

(period-2, period-4, period-8, period-16, etc.). Feigenbaum stated that any *one hump function* (graph resulting from the calculation of the *iterated function* $g(x)$ for a quadratic map) *has a cascade of bifurcations* for which the Feigenbaum number is refound. The number is calculated by comparing the values of α at each successive bifurcation. *Values of α at the bifurcation points:*

n	Period	α
1	2	3.0
2	4	3.44949
3	8	3.544090
4	16	3.564407
5	32	3.568759
6	64	3.569692
∞	∞	3.56995

From the table above and from the values of α, we can try to compute the value of the Feigenbaum constant for $\alpha_{n-1} = 3.564407, \alpha_n = 3.568759, \alpha_{n+1} = 3.569692$:

$$\delta = \lim_{n \to \infty} \left(\frac{3.568759 - 3.564407}{3.569692 - 3.568759} \right) = 4.6645. \quad (1.290)$$

The computed value is close to the exact value, although slightly different.
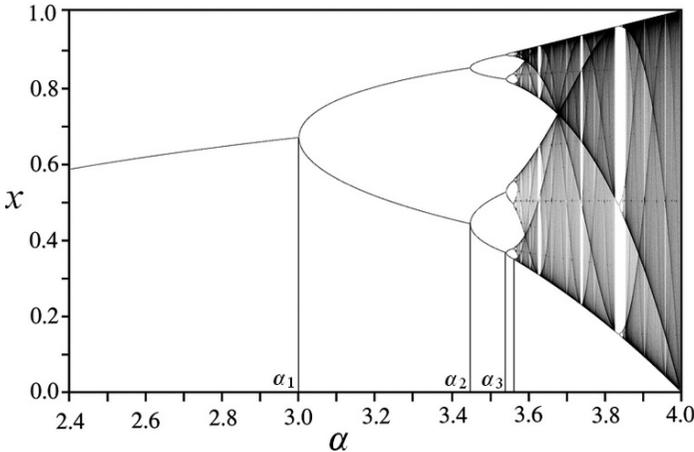
1.31.6 Iterative Functions of the Logistic Equation

1.31.6.1 Iterations from 1 to 10

Let us take the logistic equation $x_{t+1} = \alpha x_t (1 - x_t)$ again and let us imagine the iterated function (n times) of this equation, as we could see it several times in the
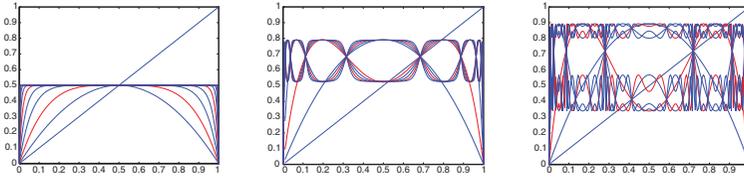
Fig. 1.77 $\alpha = 2$ (*left*). $\alpha = 3.1$ (*center*). $\alpha = 3.57$ (*right*)

chapter: $f^n(x(t)) = f(f(f(\ldots(x(t))\ldots)))$. It is possible to represent in a figure the result of each 10 first iterations for a panel of some values of the parameter α of the logistic model (Fig. 1.77).

One easily observes the *complexification* of the structure of the different curves f^1 to f^{10}, which emphasizes the transition of the system to a chaotic regime.

Remark 1.8 (One-hump function). The term "one-hump" function is rarely used whereas these functions are frequently used. They are iterated functions of which the graph results from the calculation of the iterated function $g(x)$ for a quadratic map, which has a parabolic form and explains the "one hump" name: $x_{n+1} = g(x_n)$. We will also use this term in one of the up coming chapters concerning the logistic equation, that we will couple with a function delay, within the more general framework of Propagation theory of information in Economics.

1.32 The Bifurcation Paradox: The Final State is Predictable if the Transition is Fast Enough

1.32.1 Probability of a Final State and Speed of Transition

The final state of a system is predictable if the phase of transition is rather fast. Indeed, the probability of occurrence of a final state depends on two elements: (1) the speed of transition and (2) the noise level in the system. The *logistic equation is* used here as a support. The different changes of state of the system pass by bifurcation-points which precede the period-doublings. *We are interested more particularly in the first bifurcation of the logistic equation.* The works about this subject shows that the probability of transition towards a given final state varies between 1 and 1/2. For the probability equal to 1 the transition is done quickly and without any noise added to the system, the probability of 1/2 corresponds to a slow transition. Thus we are interested in the behavior of the model around the bifurcations. Generally, when a system undergoes a bifurcation, it can be transformed into different new states of equilibrium at the exit of the bifurcation point. If noise is added to the model, as seen previously, the probability to anticipate one of both states is equal to 1/2, i.e. comparable to randomness. If there are (as for the first bifurcation of the logistic equation) two new possible equilibrium states, which "differ only in phase",

i.e. ABABAB(serie1) or BABABA(serie2), then their respective probabilities are equal to 1/2:

$$P(\text{series 1}) = P(\text{series 2}) = 1/2. \tag{1.291}$$

The purpose here is to notice that in real dynamical systems, the final state is also determined by the *transition speed through the bifurcation point*. Indeed, when the parameter of the system is modified quickly, the *noise* does not have time to influence the final state, and the transition occurs as if there was no noise at all. A non-perturbated transition, with a "finite" speed allows a normal forecast of the final state. In this case, the probability of each state depends on the initial conditions of the system. Thus the probability of the appearance of a state will be given and be equal either to 1 or to 0. We have $P(state1) = 1$ and $P(state2) = 0$, or $P(state1) = 0$ and $P(state2) = 1$. This work aims to describe the relation between: (1) the rate of change of the control parameter, (2) the noise level, (3) and the predictability of the final state.

The principle is to determine for a given level of noise the transition speed necessary so that the state which follows the transition has a perfectly determined probability, null or equal to 1.

1.32.2 Variation of the Control Parameter of the Perturbated Logistic Equation

Given the logistic équation: $x_{n+1} = \alpha x_n (1 - x_n)$, this equation is perturbated with a control parameter which varies with time α_n: $x_{n+1} = \alpha_n x_n (1 - x_n) + \varepsilon(n, \gamma)$ where $\varepsilon(n, \gamma)$ is a pseudo random process centered at zero, uniformly distributed between $-\gamma$ and $+\gamma$, and with a variance $\sigma^2 = \langle \varepsilon^2 \rangle = \gamma^2/3$. The first period-doubling of the model occurs around $\alpha = 3$, and it is this bifurcation which is studied. We define the behavior of α in the following way:

$$\begin{array}{ll} \alpha_0 & \text{for}: n < N_1 \\ \alpha_n = \alpha_0 + \Delta\alpha_{n-N_1} V & \text{for}: N_1 \leq n < N_2 \\ \alpha_f & \text{for}: n > N_2 \end{array}$$
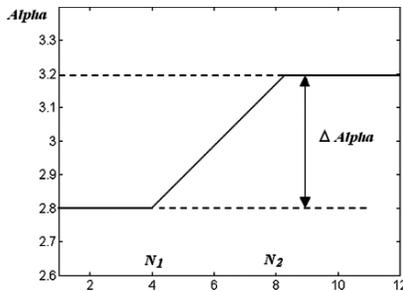
That means α_n is constant in α_0 until the period of time equal to N_1, then it evolves step by step to N_2, and is constant equal to α_f afterwards. We also use: $\alpha_0 = 2.8$, $\alpha_f = 3.2$, thus $\Delta\alpha_{n-N_1} = 0.4$. $1/V = N_2 - N_1$ which is the number of steps necessary to complete the transition. V is then regarded *as the transition speed* of the system. When the noise is equal to zero, and after the parameter α reached $\alpha_f = 3.2$, the system will be in a periodic regime of period-2, in which x_n alternates between a high value and a low value, corresponding to the two branches of the bifurcation. In this mode, there are two possible situations or states that we can note f_1 and f_2 for each state. For f_1, x_n takes the low values for an odd index and the high values for an even index. It is the reverse for $x_n(f_1) = x_{n+1}(f_2)$. Suppose now that the system is perturbated, with (1): a speed V evolving on $0.0002,\dots,1$. (transition time from 1 to 5,000 steps), (2): $9 \times 10^{-12} < \gamma < 9 \times 10^{-2}$ and a variance equal to $\gamma^2/3$, (3): and the variance of the logistic equation for $\alpha = 3.2$ is equal to 0.020. Here we give the SNR (Signal to Noise Ratio) of the system:

$$SNR = 10 \, \log_{10}\left[\tfrac{0.0205}{\gamma^2/3}\right] = -20 \, \log_{10}\gamma - 12.1. \tag{1.292}$$

Let us take an initial condition $x_0 = (\alpha_0 - 1)/\alpha_0 = 0.6429$ which is in fact the fixed point of the logistic equation of the origin. When the speed $V = 1$, if the parameter is changed abruptly from α_0 to α_f, then the final state obtained is always f_1. From this observation, it is possible to calculate the occurrence probability of f_1, when we vary the speed of the system. Here we will not have all the intermediate results obtained in the works whose authors are quoted in margin.[36] However we can give the most important results. A particular analysis of the probability was done for two different speeds and with a scale of noise from $-150\,\mathrm{dB}$ to $-70\,\mathrm{dB}$. Both speeds are $V = 0.004$ and 0.008, (i.e. respectively 250 steps and 125 steps) (Fig. 1.78).

For $V = 0.004$ the *critical noise is equal to 128 dB. For $V = 0.008$ the critical noise is equal 108 dB.* For a very low level of noise, i.e. for a high SNR, the probability P_1 of f_1 is 1. When the level of noise increases, P_1 moves towards 0.5 which is the random probability of f_1. From there, if we choose a probability of 0.75 and if we refer on the scale of the noise level, we can find the critical point: $\sigma_{critical}$. Under this critical value of noise it is said that the system is mainly deterministic,
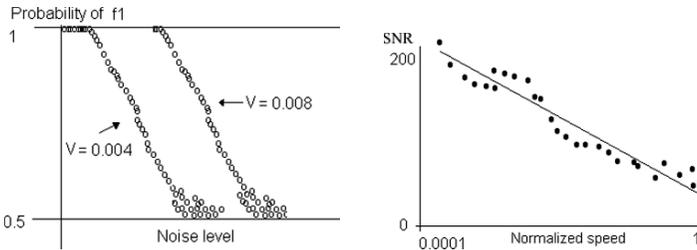
Fig. 1.78 Probability of f_1 (*left*). Log–log scale (*right*)

[36] Cf: O.Y. Butkovskii, Y.A. Kravtsov, and J.S. Brush. NAG000.NAG300.

i.e. the probability P_1 is closer to 1 than of 0.5. For noise values which are higher than the critical value, the system is known as mainly random or stochastic, i.e. the probability P_1 is close to 0.5. The critical noise level $\sigma_{critical}$:

$$SNR_{critical}(dB) = 10.27 - 52.02 \log_{10} V \qquad (1.293)$$

or $\log_{10}(V_{critical}) = 0.1974 - 0.01922\, SNR(dB)$. Using the noise variance $\sigma^2_{critical} = \langle \varepsilon^2 \rangle = \gamma^2/3$ instead of SNR (see SNR equation), we may rewrite the equation $SNR_{critical}(dB) = 10.27 - 52.02 \log_{10} V$, as

$$\sigma^2_{critical} = cV^a \qquad (1.294)$$

with $a = 5$.[37] If the speed is strictly higher than its critical value for a given noise level, the probability P_1 will be close to 1. If the speed is lower than its critical value, the probability of occurrence of the state f_1 will be close to 0.5. It is thus particularly recommended to very quickly change the control parameter of a system, rather than slowly and carefully. The change speed of the parameter has very important consequences on the response of the system. Indeed, a fast change of the value can bring about an evolution from a stable state towards another stable state, whereas a too slow change of the parameter value can destabilize the system towards a chaotic regime. The forecast probability of a state after the bifurcation point depends on a relation between the speed V of the control parameter change and the noise level σ^2 of the system. Under of a critical value $\sigma^2_{critical}$, the probability of forecast is close to 1. Above this critical value, there is *equipropability* of both solutions. This remark can have technical applications in Economics or Social Sciences.

1.33 Hyperbolicity and Kolmogorov Capacity Dimension

This dimension is an instrument quite as important as the Lyapunov exponent, it allows:

– *To distinguish between deterministic chaos and random motion*
– *To characterize the nature of the attractor* which we have to face, which is chaotic or "strange"

   We can approach the concept in the following way:

• Let us suppose a set of points P, in a space of dimension d.
• If we imagine, for example, a Poincaré section with only one delay (i.e. a map of the first return), which intersects the trajectory of an arbitrary dynamics, the dynamics is lied on this plane.

---

[37] A connection seems interesting with the developments concerning the Hurst exponent and the (R/S) Rescaled Range Statistics.

- We imagine then, a cube (or a "hypercube" of unspecified dimension) noted ε, and we measure the $N(\varepsilon)$ *number of ε which are necessary to cover the set of points P*, whence comes its other name: "Box counting".

Then, D, the capacity dimension of the set of points P is equal to

$$D = \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(\frac{1}{\varepsilon})}. \tag{1.295}$$

We can illustrate this point by choosing a single point for the set P. Consequently the number of ε necessary to cover P is $N(\varepsilon) = 1$ and the dimension D is equal to 0. In the case where P is a segment of which by convention the length is equal to 1, then $N(\varepsilon) = \frac{1}{\varepsilon}$ and $D = 1$. And in the case where P is a simple plane, by convention of a side equal to 1, then $N(\varepsilon) = \frac{1}{\varepsilon^2}$ and $D = 2$. And so on.

The subtlety lies in the fact that if we meet *traditional geometrical objects*, the *capacity dimension D* does not differ from the usual *Euclidean dimension*, i.e. amongst other things that D is an *integer*. This is not always the case, for example in *the Cantor set* that we evoked within the framework of the evolution of the logistic orbit and its fixed points. Indeed, a Cantor set can be defined in the following way.

1.33.1 The Cantor Set

We select a segment length equal to 1 and we divide it into three sub-segments of equal length. We can remove the central sub-segment, then again subdivide the two remaining sub-segments in three equal sub-segments, inside which we remove the central sub-segment. We repeat the operation n times and obtain a triadic Cantor set (Figs. 1.79 and 1.80).
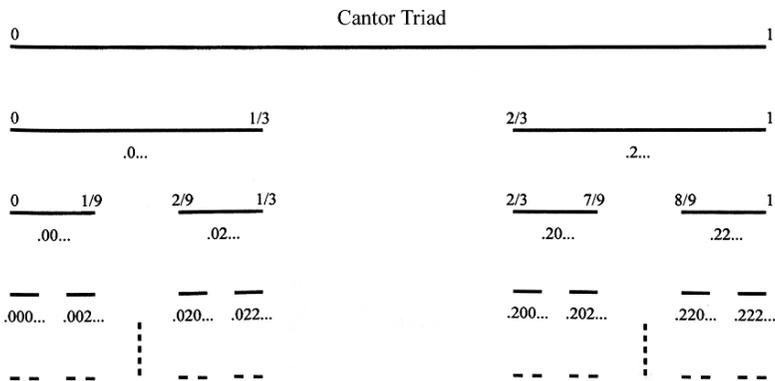
Fig. 1.79 "Cantor dust" and "triadic Cantor set"
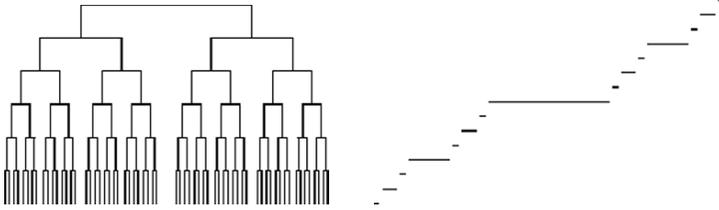
Fig. 1.80 Diagram of a "Cantor space" (*left*). Cantor function (*right*)

(The Cantor set[38] is sometimes also called the triadic Cantor set, the Cantor comb or no middle third set. The Cantor function[39] is a particular case of a devil's staircase.) If n is finite, the Cantor set is composed of some number of sub-segments and its dimension D is equal to 1, as in the case mentioned above of a segment. On the other hand, if n is very large, tending towards the infinite, the sub-segments become points, and we obtain a Cantor set with only points. Consequently, the capacity dimension must be calculated. If we choose a "hypercube" with one dimension, whose side is equal to ε, there will be $N(\varepsilon) = 1$ for $\varepsilon = 1$ and $N(\varepsilon) = 2$ for $\varepsilon = 1/3$. If this statement is generalized, there will be $N(\varepsilon) = 2^n$ for $\varepsilon = (1/3)^n$. For $n \to \infty$ or $\varepsilon \to 0$ we calculate the dimension D:

$$D = \lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \frac{\log 2^n}{\log 3^n} \simeq 0.63 \qquad (1.296)$$

It is observed that the dimension of the set is lower than 1, and is thus not an integer. Indeed the geometrical set that we made is more complicated than the set or the usual objects that we meet in Euclidean spaces. We can say that this set P is an object larger than a point but smaller than a segment. The capacity dimension and the Haussdorff dimension are methods which we gather under the term "*fractal dimension*", because they are adapted to study fractal objects, i.e. *objects whose dimension does not correspond to an integer.* The more the dimension is high, the more the space is "filled" or "blackened" *by the points of the* set P, produced by the flow of the system. A non-integer fractal dimension indicates that the orbit of the system tends to fill or "blacken" less than an integer value. It is said for example that a Brownian motion has a capacity dimension equal to 2, meaning that for a very long motion, the plane in which the points move will tend to be filled and blackened by the trajectory (see paragraph on Brownian motion). The relevance of such matter appears when we analyze chaotic dynamics on non-invertible maps of one dimension, for example, under some conditions of the control parameter.

[38] *Remark*: It is possible to apply the Continuous wavelet transform of the devil's staircase corresponding to the "uniform triadic Cantor set"; and it is interesting to notice that the aspect of the "cantor space" (see above) is similar to the skeleton of a wavelet transform, i.e. the set of all the maxima lines (Arneado et al. 1999, p. 352).

[39] *Cantor function*: The Cantor function $f(x)$ can be defined as a function on $[0,1]$ such that for values of x on the Cantor set, i.e. $x = \sum_i 2 \cdot 3^{-n_i}$, then $f(x) = \sum_i 2^{-n_i}$ which is thus extended to other values by knowing that f is a monotonic function.

1.33.2 Finite System and Non-Intersection of Phase Trajectories

1.33.2.1 Bi-Periodical Regime and Coupling of Frequencies

An important principle (Bergé et al. 1988, p. 121) to point out is the following: When a dynamical system is described by a finite number of ordinary equations, the trajectories which represent the system cannot cut themselves (except in a singular point where they converge). Indeed, in a system with an infinite number of differential equations the same initial condition positioned on a point of intersection could have different trajectories. Besides this shows us the concept of sensitive dependence on initial conditions in a new way, by weakening it. Consider the example of a torus T^2 on which there is a trajectory made up of two basic frequencies ω_1 and ω_2, we known that the ratio ω_1/ω_2 can be rational or irrational:

(1) In the case of an irrational ratio, we face a bi-periodic (or quasi-periodic) regime (and the trajectory segments cover in a dense way the rectangle of the unfolded torus, or equivalently the torus surface).
(2) In the case of a rational ratio, we face a periodic regime (and the trajectory describes a finite number of parallel segments).

In this context, the only structural instability coming from a bi-periodical regime will result from the *coupling of frequencies ω_1 and ω_2* (or synchronization) during which the ratio stops to be rational and becomes rational, as seen above. This means that the attractor of the type T^2 cannot be the expression of a pure chaotic regime, i.e. of an aperiodic regime. *(The system that has produced this attractor is of a dimension higher or equal to 3. This remark is important for what follows).*

1.33.3 Hyperbolicity: Contradiction Between Dissipative System and Chaos Solved by the Capacity Dimension

Here we will try to describe some *properties* of *aperiodic attractors*. The concept of *Kolmogorov non-integer dimension* can be connected to the concept of chaos for a "*dissipative*" dynamical system, which is a priori an antinomy. We can understand it and we will try to explain it, by considering that their complex behavior is the result of an *"infinite" repetition of "stretching" and "folding* of the beam of orbits" under the action of a flow. We illustrate the principle by the graphs below, in order do better to approach the handling in a purely intuitive way. It is known that any attractor, whatever the dimension (even low), which has a sensitive dependence on initial conditions, will have a chaotic regime. This type of attractor contains two paradoxes. It is sometimes said that this is the essential reason for which chaos in systems with low dimension was discovered so late.

1.33.3.1 The First Paradox

The first paradox for the attractor is constituted by the *contradiction the attraction that implies a "contraction of trajectories (folding)"* and the SDIC which on the other hand "requires" a divergence (stretching) of trajectories. A minimum dimension of the flow seems necessary in order to allow this type of attractor. We know that if we want that SDIC occurs, it is necessary that the flow spreads in a three-dimensional space (important, see preceding section Bergé et al. 1988, p. 121). We thus take the minimalist example of a three-dimensional flow graphically representable to evoke the problems. This three-dimensional flow contracts in the direction of (YY') and diverges in the direction of (XX') (see Fig. 1.81a). Thus, we can say that an initial flow will generate a kind of two-dimensional plane inside which the divergence of the SDIC (sensitive dependence on initial conditions) is spread, this is the ABCE plane of Fig. 1.81b. This first step corresponds to the stretching of the flow. At this stage, we must ensure that the flow can continuously act in a bounded (three-dimensional) space. This is the aim of the folding.[40] We observe that the CE segment has doubled its length compared to AB, thus a folding can be practiced on CE, by dividing it of half into D in its center and folding the two extremities C and E on themselves. Thus the CD segment of the same length that AB can be brought closer and "pasted" on AB. The result of this folding is represented on Fig. 1.81d. Thus, we have built a three-dimensional flow in which the SDIC exists in a finite space. We *observe that from a starting space of dimension*
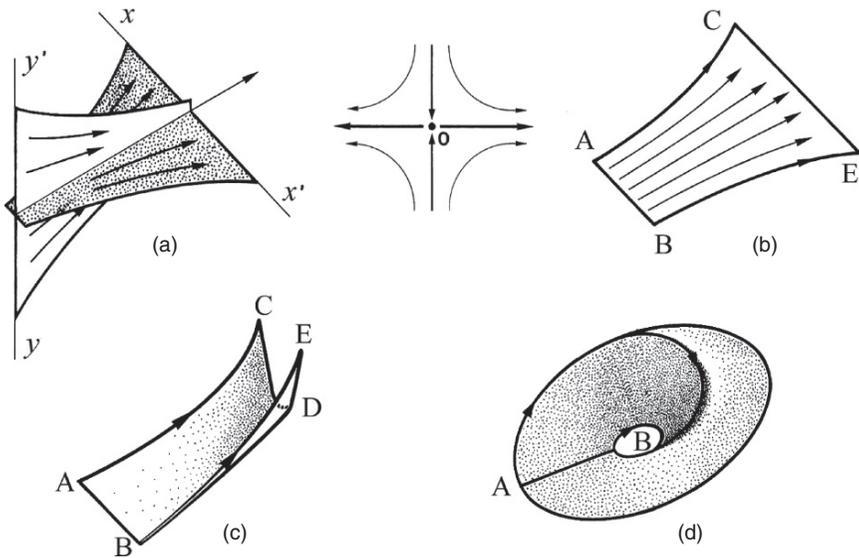
Fig. 1.81 Different construction stages of a strange attractor

---

[40] Successive foldings of flow give the sheeted structure characteristic of strange attractors.

*three, the contraction of the attractor and the expansion of the SDIC are recon-
ciled by means of this so particular work on phase-space of the system.* If a cut of
Fig. 1.81a is practiced, we can observe that the flow which undergoes the attraction
exists in a vertical plane (YY') and the flow which undergoes the expansion exists
in a vertical plane (XX'). And we could even say that the cut looks like a kind of
saddle point, repelling horizontally in relation to the center and attracting vertically,
the other elements of the cut tending asymptotically towards the axes in the form
of hyperboles, sometimes towards the axes (XX') and towards (YY'). This is why
it is said that the reconciliation of "opposites" is carried out by means of the *hyper-
bolicity* concept. *The folding must be done so that the flow acts continuously in the
space*.

1.33.3.2 The Second Paradox

The second paradox between attractor and SDIC is as follows: *It is said that to allow
the SDIC to exist, (Bergé et al. 1988, p. 121) it is necessary that the dimension of the
attractor is higher than 2*. However we said that for a *dissipative system* for which
there is an attractor, there is a contraction of spaces, planes and volumes, i.e. *the
attractor must have a null dimension after the action of the flow over time*. i.e. if we
start from a starting system of dimension 3, at the arrival, we must reduce to a space
lower than 3. An attractor containing a chaotic regime must thus have a dimension
such as $2 < D < 3$, where D is the dimension of the attractor and not necessarily that
of the system. *Indeed, because of the SDIC, the dimension must be higher than two
and because of the attractor, the dimension must be lower than three*. Consequently
how can we reconcile this contradiction, because attractors of this type exist in the
reality? This attractor type has a *non-integer fractal dimension*, we evoked it in the
section about capacity dimension. They are called strange attractors.

   Thus, to take the title given to the present section, we *reconcile* the *coexistence*
of a *dissipative system* and *chaos* if the *phase-space of a system* has *a dimen-
sion higher or equal to 3*. Thus, the attractor of the dynamical system will be
spread in $2 < D < 3$. Some systems "lodging" the attractor have Kolmogorov non-
integer dimensions contained between 2 and 3. It is known that *chaotic attractors,
belonging to (continuous-time) dissipative systems where the phase space is three-
dimensional at the beginning*, must have a capacity dimension of the type $2 < D < 3$.
If we imagine a set P which at the beginning is of three dimensions, i.e. a volume
and which *under the action of the system flow shrinks asymptotically*. This means
that the dimension D of *any attractor* of the system must be smaller than three.
Moreover, to get *an attractor chaotic*, in the sense of the sensitive dependence on
initial conditions (SDIC), it is necessary that its dimension is higher than 2. Indeed,
in a space with two dimensions, there is no attractor more complex than a *limit cycle*
in $\mathbb{R}^2$. For a toric attractor of dimension 2, we will not be able to find more com-
plex than a quasi-periodic orbit. We can depict more quickly in a different way the
work of unfolding and folding shown above as follows, and in particular through
two figures below.
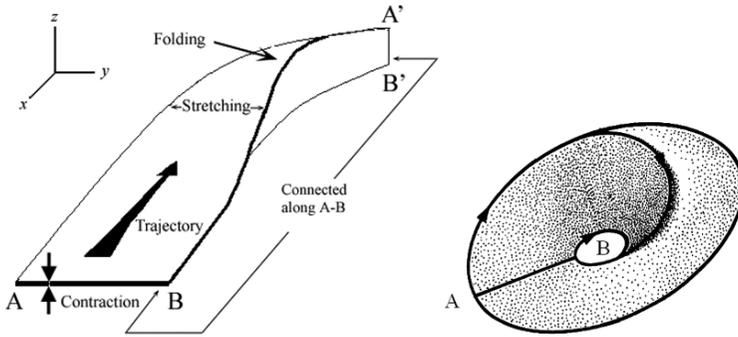
We can say that at the beginning the flow moves overall in only one direction (formalized by a large arrow), *but it tends to disperse or it is disseminated by the stretching and folding* (intermediate arrows) and pushed towards a third direction. This is the contraction (small arrows). The result of this work is that we transform a volume into a sheet with "almost" two dimensions. But, *although the close trajectories diverge exponentially, it is said that the motion of the system by construction is limited*. This is why the sheet, on which the flow takes place, must be entirely folded on itself. Thus, two extremities (ends) AB and $A'B'$ of the sheet must be folded, then must be joined and be brought together by the ends. However, since $A'B'$ has two distinct sheets and $A'B'$ is brought together with AB, which has only one sheet, in order to make the junction adjust well (which is a condition necessary to represent the system in the form of a continuous and invertible flow), the sheet on which there is the flow must have a thickness contained between 0 and 1 (i.e. a "book" of thickness), composed of an infinite number of sheets, each one of zero thickness. In fact, if we choose a two-dimensional Poincaré section which intersects the trajectories or the orbits, the set of intersection points resulting from it will have a fractal structure similar to the *Smale horseshoe*. When the compression (small arrows) is very strong, the resulting Poincaré map can appear unidimensional.

1.33.4 Chaotic Attractor in a System of Dimension 1

We can choose the example of *the logistic equation*, which is spread in a space of dimension 1. It is said that if a *chaotic behavior* is observable in such a model, in fact *its capacity dimension must be equal to 1*.[41] Strictly speaking, such dynamics cannot be named chaotic attractors, in the sense of the definition stated in the paragraph about attractors. Indeed, they do not have any contracting "neighborhood", such as we evoked it in the preceding section. We can conclude that a (discrete-time: resulting from iterated maps) dynamical system spreading itself in a space of dimension 1, which has a positive exponent ($\lambda > 0$), cannot be "dissipative". Thus,

---

[41] See concepts of "Ergodic invariant" and "Lebesgue measure".

we are reconciled with the contradiction which shocked us, stating in the preceding sections and paragraphs that:

– On the one hand, for $\lambda < 0$, we are faced with a dissipative system, with a convergence towards a stable fixed point, or a stable periodic orbit (several "fixed points"), characterizing a *non-conservative system*.
– On the other hand, chaos can exist in a dissipative dynamics.

The latter remark is false on $\mathbb{R}$ and possible in spaces higher than 2. The logistic equation is spread in a space of dimension 1, where we observe the behavior of a single variable. For this model, we used a vocabulary commonly admitted and used in many specialized works evoking the unidimensional dynamical systems. In this context, the term "chaotic attractor" appears to be inappropriate, so we will speak of chaotic behavior or attractor.

1.33.5 *Measure of the Complexity Level of Attractors*

We can intuitively consider the fact that the capacity dimension of a dynamics measures, in a certain manner, the complexity level of a time series. We can also consider that this dimension reveals *the number of subjacent operative variables* of a dynamic system, of which we do not necessarily have an algebraic construction at disposal. In other words, even if there are not the equations of the motion of the studied system, we can nevertheless have an "appreciation" of the number of variables implied in the evolution of the system. The *capacity dimension* is used to understand and measure the dimension of the *complex geometrical structure* of attractors and strange attractors. Beyond this crucial function, the dimension D of a complex dynamics can be used to characterize the limits of an attractor, i.e. the bounds which separate the different zones of the phase-space which are under the influence of the attractor(s). In practice, the calculation of D is not always easy to do, in particular when D is higher than 2. A slightly different approach was developed by Grassberger and Procaccia.

1.34 Nonlinearity and Hyperbolicity

1.34.1 *Homoclinic Tangle and Smale Horseshoes Map*

Firstly, let us recall the definitions of "invariant sets" and "homoclinic or heteroclinic orbits". These concepts are common to flows and maps.

Definition 1.31 (Invariant set). For a flow ϕ_t [or a map g] defined on $U \subset \mathbb{R}^n$, a subset $S \subset U$ is said *invariant* if

$$\phi_t(S) \subset S, \; \forall t \in \mathbb{R} \quad [\text{or } g^n(S) \subset S, \; \forall n \in \mathbb{Z}]. \tag{1.297}$$

Definition 1.32 (Homoclinic orbit). Let S be an invariant set of the flow ϕ_t, p is a point in the phase space of the flow, and let us suppose that $\phi_t(p) \to S$ when $t \to +\infty$, and $t \to -\infty$, then the orbit of p is said to be *homoclinic* to S.

Definition 1.33 (Heteroclinic orbit). Let S_1 and S_2 be two disjoint invariant sets of a flow ϕ_t, and let us suppose that $\phi_t(p) \to S_1$ as $t \to +\infty$, and $\phi_t(p) \to S_2$ as $t \to -\infty$, then the orbit of p is said to be *heteroclinic* to S.

It is important to highlight that the "homoclinic tangle" is a *topological structure similar to the "Smale horseshoe map"*. We briefly present a homoclinic tangle. Let p be the intersection point, with p' ahead of p on one manifold and p'' ahead of p of the other. The mapping of each of these points $g(p')$ and $g(p'')$ must be ahead of the mapping of p, $g(p)$.

The only way this can happen is if the manifold loops back and crosses itself at a new homoclinic point. Then, another loop must be formed, with $g^2(p)$ another homoclinic point. Since $g^2(p)$ is closer to the hyperbolic point than $g(p)$, the distance between $g^2(p)$ and $g(p)$ is less than that between p and $g(p)$. The area (or volume) preservation requires the area (or volume) to remain the same, so each new curve, which is closer than the previous one, must extend further. The loops become longer and thinner. The set of curves leading to a *dense area* of *homoclinic points* is known as a *homoclinic tangle*.
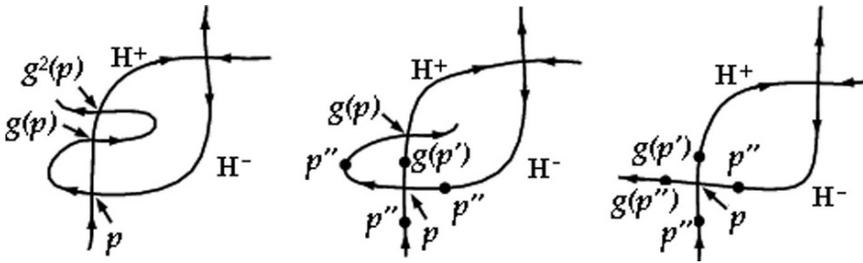
Homoclinic points appear where chaotic regions touch a *hyperbolic fixed-point*.

1.34.2 Smale Horseshoe: Structural Stability

Stephen Smale (1965) built a *diffeomorphism* $f : \mathbb{R}^2 \to \mathbb{R}^2$, with very complex dynamics, which admits an infinity of periodic orbits of arbitrarily large periods. To illustrate it as simply as possible, we are only interested in a *diffeomorphism* of the rectangle $\Delta = [0,1]^2$ on its image. The construction is carried out (see figure which follows) by a *composition* $(f = \varphi \circ E)$ of a *hyperbolic linear map* $E : (x,y) \to (3x, y/3)$ with a *nonlinear transformation* φ. These elements are defined such that:

$$f(\Delta^0) = \Delta_0, \quad f_{|\Delta^0} : (x,y) \to (3x, y/3) \tag{1.298}$$

$$f(\Delta^1) = \Delta_1, \quad f_{|\Delta^1} : (x,y) \to (-3x+3, -(y/3)+1) \tag{1.299}$$
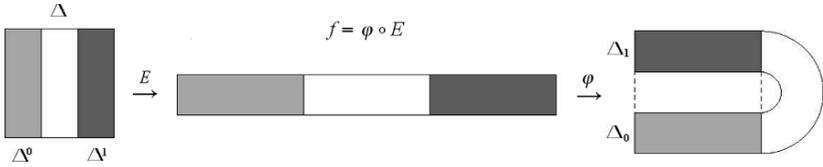
In order to understand the iteration of f it is interesting to seek a subset Φ of Δ *invariant by the diffeomorphism*. Let us notice that:
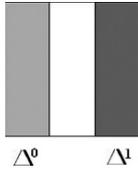
is equal to $f^{-1}(\Delta) \cap \Delta = \{x \in \Delta | f(x) \in \Delta\} = \Delta^1 \cup \Delta^0$, and
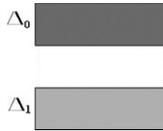
is understood as being equal to $f(\Delta) \cap \Delta = \{x \in \Delta | f^{-1}(x) \in \Delta\} = \Delta_1 \cup \Delta_0 = f(\Delta^1) \cup f(\Delta^0)$. By means of iteration we obtain
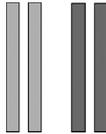
which can be understood as being equal to

$$= f^{-2}(\Delta) \cap f^{-1}(\Delta) \cap \Delta = \{x \in \Delta | f(x) \in \Delta,\ f^2(x) \in \Delta\} \qquad (1.300)$$

$$= f^{-1}(\Delta^1 \cup \Delta^0) \cap (\Delta^1 \cup \Delta^0) = \cup \Delta^{\theta_0, \theta_{-1}},\ \ \theta_0, \theta_{-1} \in \{0,1\} \qquad (1.301)$$

with $\Delta^{\theta_0, \theta_{-1}} = \Delta^{\theta_0} \cap f^{-1}(\Delta^{\theta_{-1}})$. Similarly, it is possible to write

$$f^{-n}(\Delta) \cap \cdots \cap \Delta = \cup \Delta^{\theta_0, \dots, \theta_{-n}},\ \ (\theta_0, \dots, \theta_{-n}) \in \{0,1\}^{n+1} \qquad (1.302)$$

and

$$\Delta^{\theta_0, \dots, \theta_{-n}} = \bigcap_{i=0}^{n} f^{-i}\left(\Delta^{\theta_{-i}}\right). \qquad (1.303)$$

It is known that the width of vertical rectangles decreases exponentially and $\bigcap_{i=0}^{n} f^{-i}(\Delta)$ corresponds to the union of 2^n vertical rectangles, and if $n \to \infty$, we obtain $\bigcap_{i=0}^{\infty} f^{-i}(\Delta) = K \times [0,1]$ where K is the *triadic Cantor set*. Similarly, through the image of horizontal rectangle unions it is obtained
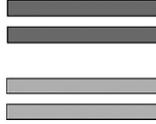
which can be understood as being equal to

$$= \Delta \cap f(\Delta) \cap f^2(\Delta) = \cup \Delta_{\theta_1,\theta_2}, \quad \text{with } (\theta_1,\theta_2) \in \{0,1\}^2. \qquad (1.304)$$

By extension, it is possible to write:

$$\Delta \cap f(\Delta) \cdots \cap f^n(\Delta) = \cup \Delta_{\theta_1,\dots,\theta_n}, \quad \text{with } (\theta_1,\dots,\theta_n) \in \{0,1\}^n, \qquad (1.305)$$

and

$$\Delta_{\theta_1,\dots,\theta_n} = \bigcap_{i=1}^{n} f^i\left(\Delta^{\theta_i}\right). \qquad (1.306)$$

As previously the *infinite intersection* gives

$$\bigcap_{i=0}^{\infty} f^i(\Delta) = [0,1] \times K,$$

where K is always the *triadic Cantor set. Then, the set*

$$\Phi = \bigcap_{-\infty}^{+\infty} f^i(\Delta) \qquad (1.307)$$

is an *invariant set* by the diffeomorphism f. Moreover, $\Phi = K \times K$. And it is possible to introduce the following very interesting equality and sets:

$$\Delta(\theta_{-n},\dots,\theta_0,\theta_1,\dots,\theta_n) = \Delta^{\theta_0,\dots,\theta_{-n}} \cap \Delta_{\theta_1,\dots,\theta_n}. \qquad (1.308)$$

providing a *bijective map* $g: \Sigma_2 = \{0,1\}^{\mathbb{Z}} \to K \times K$. If Σ_2 is provided with the *topological product*, it was proven that g is a *homeomorphism*. Respecting the reasoning which precedes, briefly, a simplified Smale's Horseshoe diagram can be also given as follows:

Δ $\Delta \cap f(\Delta)$ $\Delta \cap f(\Delta) \cap f^2(\Delta)$

$\Delta \cap f^{-1}(\Delta)$ $\Delta \cap f^{-1}(\Delta) \cap f^2(\Delta)$ $f^2(\Delta) \cap f^{-1}(\Delta) \cap \Delta \cap f(\Delta) \cap f^2(\Delta)$
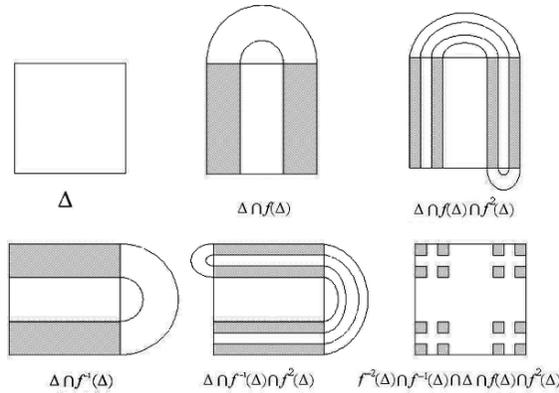
As seen above, the *invariant set* is written

$$\Phi = \bigcap_{-\infty}^{+\infty} f^i(\Delta) \tag{1.309}$$

such a map describes a *Cantor set* with the following properties:

– Φ contains a *countable set of periodic orbits of arbitrarily long periods*.
– Φ contains an *uncountable set of (bounded) non-periodic orbits*.
– Φ contains a *dense orbit*.

Henceforth, whenever the Poincaré map of a hyperbolic periodic orbit has a transverse homoclinic point, this provides a horseshoe map, which alternately insures the existence of a strange attractor and chaotic dynamics. Poincaré came across this chaotic behavior in studying the three-body problem a long time before the beginnings of modern chaos theory.

1.34.3 Hyperbolic Set (Anosov Diffeomorphisms)

Within the framework of the horseshoe, let us suppose that H is an open set of $\mathbb{R}^k$ and $f : H \to \mathbb{R}^k$ a C^1-diffeomorphism on its image and let us suppose that there exists a (f-invariant compact) *subset* $\Phi \subset H$. It is possible to write the following definition:

Definition 1.34 (Hyperbolic set). The set Φ is said hyperbolic for the map f if there exist $C > 0$, $\lambda \in]0,1[$ and for any $x \in \Phi$ a decomposition $\mathbb{R}^k = E_x^s \oplus E_x^u$ into stable and unstable subspaces such that:

(1) For any $v \in E_x^s$, $\|D_x f^n(v)\| \leq C\lambda^n \|v\|$, $(n \geqslant 0)$.
(2) For any $v \in E_x^u$, $\|D_x f^{-n}(v)\| \leq C\lambda^n \|v\|$, $(n \geqslant 0)$.
(3) $D_x f(E_x^s) = E_{f(x)}^s$, $D_x f(E_x^u) = E_{f(x)}^u$.

(It could be shown that the "*horseshoe set* Φ" is a hyperbolic set, by choosing for example $\lambda = 1/3$.)

Definition 1.35 (Anosov diffeomorphism). A diffeomorphism of the "torus" $f: \mathbf{T}^k \to \mathbf{T}^k$ is called "Anosov diffeomorphism" if $\mathbf{T}^k$ is a hyperbolic set for f.

1.34.4 Symbolic Dynamics

Dynamical systems are often ciphered (or coded) by means of symbolic dynamics. The Smale horseshoe is a particular case of symbolic dynamics use. A general case can be written briefly as follows

$$\mathbf{\Omega}_N = \{\theta = (\ldots, \theta_{-1}, \theta_0, \theta_1, \ldots) \, | \, \theta_i \in \{0, \ldots, N-1\}\} \qquad (1.310)$$

and can be provided with a given topology by means of the following distance:

$$d_\lambda(\theta, \theta') = \sum_{n=-\infty}^{+\infty} \frac{|\theta_n - \theta'_n|}{\lambda^{|n|}}, \quad \lambda > 1. \qquad (1.311)$$

The dynamics can be given by the *shift*[42]

$$\begin{aligned} \sigma : \mathbf{\Omega}_N &\to \mathbf{\Omega}_N, \\ \sigma(\theta) &= \theta', \quad \theta'_n = \theta_{n+1}. \end{aligned} \qquad (1.312)$$

Main cases are not *defined* on all the words of the *alphabet* but only on a *subset characterized by a matrix* which can be *noted B* and the technique is the following: given $B = (b_{ij})$, $0 \le i, j \le N-1$, $b_{ij} = 0$ or 1, then let us consider

$$\mathbf{\Omega}_B = \{\theta \in \mathbf{\Omega}_N \, | \, b_{\theta_n \theta_{n+1}} = 1, \text{ for } n \in \mathbb{Z}\}. \qquad (1.313)$$

Definition 1.36 (Topological Markov Chain). The restriction $\sigma_{|\mathbf{\Omega}_B} = \sigma_B$ is called a "topological Markov Chain", it is also called a "subshift of finite type".

1.34.5 Properties of the Smale Horseshoe Map

Changing slightly the notations and using the more usual ones in order to show different aspects of the Smale horseshoe map, it is possible to write the following *statements* and *detailed properties*:

- The horseshoe map f is an invertible map on $\mathbb{R}^2$.
- $S = [0,1] \times [0,1]$: Unit square in $\mathbb{R}^2$.

---

[42] See "Bernouilli shift".

- $I_0 = S$.
- $I_n = f(I_{n-1}) \cap S, \quad n = 1, 2, 3, \ldots$.
- $I_n = f^{-1}(I_{n+1}) \cap S, \quad n = -1, -2, -3, \ldots$.
- $I_+ = \bigcap\limits_{n \geq 0} I_n$: vertical lines.
- $I_- = \bigcap\limits_{n \leq 0} I_n$: horizontal lines.
- $I = I_+ \cap I_-$: Invariant set.
- V_0 and V_1 : left and right vertical strips.
- H_0 and H_1 : upper and lower horizontal strips.
- $f(H_i) = V_i$ with $i = 0, 1$.
- $V_{\theta_1 \theta_2 \theta_3 \cdots \theta_n} = V_{\theta_1} \cap f(V_{\theta_2 \theta_3 \cdots \theta_n})$, after n iterations of f (with $\theta_i \in \{0,1\}$).
- $H_{\theta_0 \theta_{-1} \theta_{-2} \cdots \theta_{-n}} = H_{\theta_0} \cap f^{-1}(H_{\theta_{-1} \theta_{-2} \cdots \theta_{-n}})$, after n iterations of f.
- $V_{\theta_1 \theta_2 \cdots}$: Vertical line belonging to I_+.
- $H_{\theta_0 \theta_{-1} \theta_{-2} \cdots}$: Horizontal line belonging to I_-.
- $\ldots, \theta_{-3}, \theta_{-2}, \theta_{-1}, \theta_0, \theta_1, \theta_2, \theta_3, \ldots$: is a bi-infinite sequence.
- $\mathbf{E}$: space of bi-infinite sequences.
- $\mathscr{B}_L$: *Bernoulli left-shift* on $\mathbf{E}$, which can be written as follows (for $\beta_i = \theta_{i+1}$)

$$\mathscr{B}_L(\ldots, \theta_{-2}, \theta_{-1}, \theta_0, \theta_1, \theta_2, \ldots) = (\ldots, \beta_{-2}, \beta_{-1}, \beta_0, \beta_1, \beta_2, \ldots), \qquad (1.314)$$

The *Smale horseshoe map properties* can be written:

1. The set I is an *invariant set*, with respect to S, of the horseshoe map.
2. Vertical and horizontal lines I_+ and I_- are each Cantor-like sets.
3. $\mathscr{B}_L^{-1}$ on $\mathbf{E}$ *is "equivalent" to f on I*, i.e. under the action of f a point y belonging to I ($y \in I$) is the same as a *Bernoulli right-shift* on the corresponding sequence in $\mathbf{E}$.
4. Let us consider 2 initially very close points in I, with respect to a metric and provided with the distance defined on $\mathbf{E}$:

$$d(\theta, \beta) = \sum_{i=-\infty}^{\infty} \frac{|\theta_i - \beta_i|}{2^{|i|}}, \qquad (1.315)$$

   move away quickly as f is iterated. This property is regarded as the sensitive dependence on initial conditions.
5. The horseshoe map contains a *countable set of periodic orbits of arbitrarily long periods (possibly infinite)*.
6. The horseshoe map contains an *uncountable set of (bounded) non-periodic orbits*.
7. The horseshoe map contains a *dense orbit*.

The last properties *(4) (5) (6) (7) are used* to characterize chaotic dynamics in this framework.

1.34.6 Folding and Unfolding Mechanism: Horseshoe and Symbolic Dynamics (Symbolic Coding)

It is fundamental to understand that it is the *folding and unfolding mechanism* which allows the description of a *chaotic system* in terms of *symbolic dynamics*, as it is possible to observe on the horseshoe map f. Let us suppose that we are in a plane, then the Symbolic dynamics acts from the plane to the plane and can be *geometrically explained*. As it was seen previously by means of mathematics, the starting unit square is initially stretched in a direction, then flattened in another. The long ribbon, which is thus obtained, is then folded in the horseshoe shape so as to intersect the initial square in two horizontal bands H_0 and H_1. *In the case which is depicted in the figure below:* The initial unit square is horizontally stretched, then vertically flattened. One of the halves of the long ribbon, which is thus obtained, is then *folded* on the other. The invariant set of the map f can be ciphered (or coded) symbolically by means of the partition of this set in two horizontal bands $\{H_0, H_1\}$, or similarly in two vertical bands $\{V_0, V_1\}$.
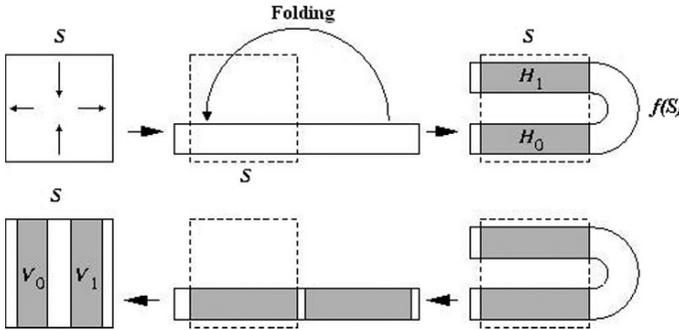
By means of the map f, there exists Φ an *invariant set* of points, and the points belonging to it *never leave the unit square*. For each iteration, the orbit of an initial point of the invariant set Φ is necessarily inside one or the other of the two bands, respectively associated with symbols 0 and 1. This method is called *Symbolic Coding* of trajectories. Each point p of the invariant set Φ is associated with a bi-infinite symbolic sequence which is noted $\Sigma(p)$ and built with symbols indicating what are the bands successively visited by the orbit of the point under the action of the horseshoe map f:

$$\Sigma(p) = \cdots \theta_j \cdots \theta_{-2}\theta_{-1} \cdot \theta_0\theta_1\theta_2 \cdots \theta_i \cdots, \quad \text{with } \theta_i = i \text{ if } f^i(p) \in H_i \quad (1.316)$$

The θ_0 symbol (after the point $\cdot$) indicates in which band the point p is located. The symbolic sequence (1) is broken up into two sequences, the first sequence is the *forward sequence*: $\Sigma_+(p) = \theta_0\theta_1\theta_2 \cdots \theta_i \cdots$ corresponding to the future of p, and second is the *backward sequence* written: $\Sigma_-(p) = \theta_{-1}\theta_{-2} \cdots \theta_j \cdots$ corresponding to its past. The process can be extended to another partition based on two vertical ribbons V_0 and V_1 observed on the preceding figure. The most important subject

to understand is that the equivalence or *correspondence* between the *points of the invariant set* and *infinite binary sequences is a bijection*. Indeed, a simple measurement of the state of the system makes it possible to determine the system. The purpose is to know in which band it is. By means of the partition $\{V_0, V_1\}$, the points whose orbit remains in the unit square can be associated with a bi-infinite symbolic sequence $\Sigma(p) = \cdots \theta_{-2}\theta_{-1} \cdot \theta_0\theta_1 \cdots$.
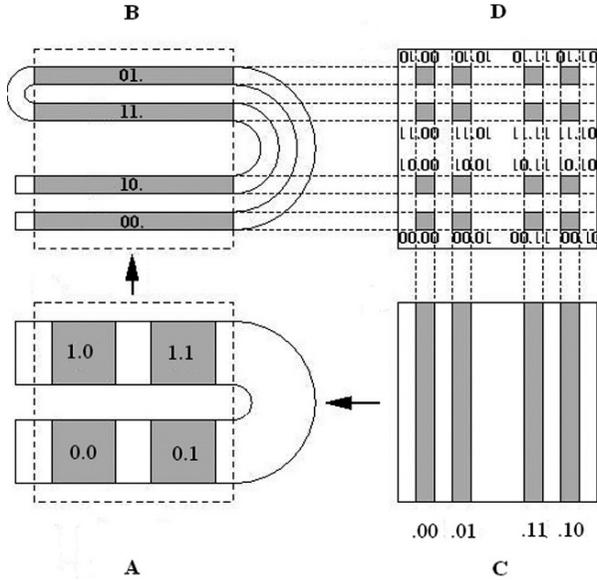
The figure above depicts the mechanism to create the set of points such that the n_+ first symbols of $\Sigma_+(p)$ and the n_- first symbols of $\Sigma_-(p)$ coincide.

$$A : n_- = 1, \quad n_+ = 1,$$
$$B : n_- = 2, \quad n_+ = 0,$$
$$C : n_- = 0, \quad n_+ = 2,$$
$$D : n_- = 2, \quad n_+ = 2.$$

The symbols of the backward sequence make it possible to localize a point vertically, and the symbols of the forward sequence make it possible to localize a point horizontally. Obviously, the *point is localized with a better precision* when the *number of known symbols increases*.

1.34.7 Smale–Birkhoff Homoclinic Theorem

Closely related to the Smale (horseshoe map) topology[43] and to the hyperbolicity concept, it is interesting to present the following fundamental theorem.

---

[43] See also homoclinic tangle.

Theorem 1.11 (Smale–Birkhoff homoclinic theorem). *Let f be a "diffeomorphism" (C^1) and suppose p is a "hyperbolic fixed point". A homoclinic point is a point $q \neq p$ which is in the stable and unstable manifold. If the stable and unstable manifold intersect transversally[44] at q, then q is called transverse. This implies that there is a homoclinic orbit $\gamma(q) = \{q_n\}$ such that $\lim_{n\to\infty} q_n = \lim_{n\to-\infty} q_n = p$. Since the stable and unstable manifolds are invariant, we have $q_n \in W^s(p) \cap W^u(p)$ for all $n \in \mathbb{Z}$. Furthermore, if q is transversal, so are all q_n since f is a diffeomorphism.*

The following figure is a *sketch* of behaviors in this framework (with p the *hyperbolic fixed point* and q the *homoclinic point* which is called *transverse*):
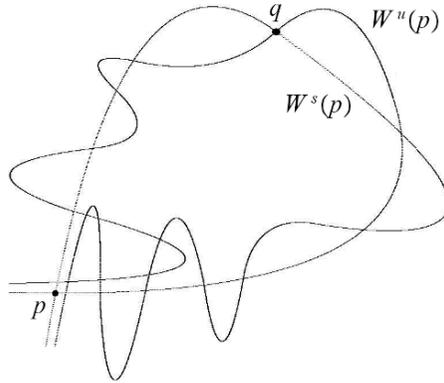
Example: Let us consider the following map φ:

$$x_1 = x + y_1, \quad y_1 = y + ax(x-1), \tag{1.317}$$

for $0 < a < 4$ *(Arrowsmith–Place dynamics)*. There are fixed points at $(x,y) = (0,0)$ and $(1,0)$ and the point $(0,0)$ is non-hyperbolic. The fixed point $(1,0)$ is an *hyperbolic saddle point* with its invariant stable and unstable manifolds. At $(0,0)$ the linear approximation is conjugate to an anticlockwise rotation through angle θ, with $2\sin\theta = (a(4-a))^{1/2}$ and $2\cos\theta = (2-a)$. Similarly, at $(1,0)$ the linearization gives: $s = u\{-k \pm (a(4+a)^{1/2})/2\}$ with (u,s) local coordinates at $(1,0)$. Then, in order to complete the *homoclinic tangle*, consider the inverse map φ^{-1}:

$$x = x_1 - y_1, \quad y = y_1 - ax(x-1), \tag{1.318}$$

The result of this device is shown in Fig. 1.82; we can observe the *homoclinic tangle* with the *non-hyperbolic point* at $(0,0)$ and the *hyperbolic saddle point* at $(1,0)$.

---

[44] see Arnold about *Transversality*.
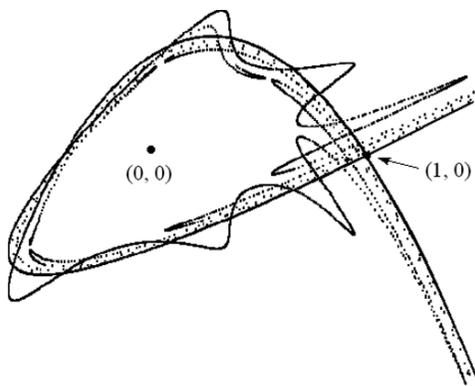
Fig. 1.82 Non-hyperbolic point at $(0,0)$ and hyperbolic saddle point at $(1,0)$

1.34.8 Hyperbolicity and Hartman–Grobman Theorem: Hyperbolic Nonlinear Fixed Points

Let us consider as a preliminary the following definitions and theorems concerning the *properties* of *flows* and *diffeomorphisms*. We will use the following notations: discrete dynamical systems are written as the iteration of a function or *map* **g** as follows:

$$\mathbf{x}_{t+1} = \mathbf{g}(\mathbf{x}_t), \quad t \in \mathbb{N} \text{ or } \mathbb{Z}. \tag{1.319}$$

In the case where t is continuous, the dynamics are written by a differential equation as follows:

$$\frac{d\mathbf{x}}{dt} = \dot{\mathbf{x}} = \mathbf{X}(\mathbf{x}) \tag{1.320}$$

and the *solution* of this equation can be described by a *flow* ϕ_t with *velocity* given by the *vector field* **X**. Moreover, **x** gives the state of the system and the values taken by **x** depict the state of the system or the phase space. The phase space can be Euclidean, or *non-Euclidean* (see *differential manifolds*).

Definition 1.37 (Homeomorphism). $h : U \rightarrow V$ is a homeomorphism if it is injective and continuous and if its inverse is continuous.

Definition 1.38 (Diffeomorphism). A map $\mathbf{g} : U \rightarrow V$ is a *diffeomorphism* if it is a *homeomorphism and if* **g** *and the inverse map* $\mathbf{g}^{-1}$ *are differentiable*. Any diffeomorphism is a homeomorphism, but the converse is false.

Let us recall some basic elements concerning diffeomorphisms, vector fields and differentiable maps. Let us suppose that U is an open subset $\mathbb{R}^n$ and V is an open subset $\mathbb{R}^m$. A *function* $g : U \rightarrow \mathbb{R}$ is said to be of *class* C^r if it is r-fold *continuously differentiable*. For $\mathbf{g} : U \rightarrow V$ the coordinates in U and V can be written $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$. Then **g** can be written contingent on component functions $g_i : U \rightarrow \mathbb{R}$, with $y_i = g_i(x_1, \ldots, x_n)$ and $i = 1, \ldots, m$. The map **g** is a C^r-*map* if g_i is C^r for

each $i = 1, \ldots, m$. The map $\mathbf{g}$ is *differentiable* if it is a C^r-*map* (for $1 \leqslant r \leqslant \infty$). The map $\mathbf{g}$ is said to be *smooth* if it is C^∞. (If a map is continuous and not differentiable it is called C^0-map). Then, we can write the definition of a diffeomorphism as follows:

Definition 1.39 (Diffeomorphism). [45] $\mathbf{g} : U \rightarrow V$ is a diffeomorphism if it is a bijection and both $\mathbf{g}$ and $\mathbf{g}^{-1}$ are differentiable mappings. $\mathbf{g}$ is called a C^k-diffeomorphism if both $\mathbf{g}$ and $\mathbf{g}^{-1}$ are C^k-maps.

Let us notice that a *bijection* $\mathbf{g} : U \rightarrow V$ is a *diffeomorphism* if and only if $m = n$ and the matrix of *partial derivatives*

$$\mathbf{Dg}(x_1, \ldots, x_n) = \left[\frac{\partial g_i}{\partial x_j} \right]_{i,j=1}^n$$

is non-singular at every $\mathbf{x} \in U$. Then, for $U = \mathbb{R}^2$ and $V = \{(x,y)|x,y > 0\}$, $\mathbf{g}(x,y) = (\exp(y), \exp(x))^T$ is a *diffeomorphism* due to

$$\text{Det } \mathbf{Dg}(x,y) = -\exp(x+y) \neq 0 \tag{1.321}$$

for each $(x,y) \in \mathbb{R}^2$.

From these recalls and developments, we consider the following notions and properties concerning diffeomorphisms, flows and the hyperbolicity.

Definition 1.40 (Hyperbolic linear diffeomorphism). A *linear diffeomorphism* $\mathbf{A} :$ $\mathbb{R}^n \rightarrow \mathbb{R}^n$ is said to be "hyperbolic" if it has no "eigenvalues" with "modulus" equal to unity.

Theorem 1.12 (Contraction, expansion). *If* $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ *is a "hyperbolic linear diffeomorphism", then there are subspaces E^s and $E^u \subseteq \mathbb{R}^n$ invariant under $\mathbf{A}$, such that $\mathbf{A}|E^s$ is a contraction, $\mathbf{A}|E^u$ is an expansion and*

$$E^s \oplus E^u = \mathbb{R}^n. \tag{1.322}$$

Let us consider a simple *linear diffeomorphism* $\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^p$, this linear diffeomorphism $\mathbf{g}$ is a *contraction* if all its *eigenvalues have modulus less than unity*, or it is an *expansion* if all its *eigenvalues have modulus greater than unity*. e.g. if the eigenvalues of $\mathbf{g}$ are distinct for any $\mathbf{x} \in \mathbb{R}^p$ with $k \in \mathbb{Z}^+$

$$\mathbf{g}^k \mathbf{x} = \mathbf{M}^{-1} \mathbf{D}^k \mathbf{M} \mathbf{x} \tag{1.323}$$

with $\mathbf{D} = [\lambda_i \delta_{ij}]_{i,j=1}^p$ knowing that $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. The ith column of $\mathbf{M}$ is an eigenvector of $\mathbf{g}$ and its eigenvalues λ_i. For $i = 1, \ldots, p$, if $|\lambda_i| < 1$ then $\mathbf{D}^k \rightarrow 0$ as $k \rightarrow \infty$. Moreover, the orbit (under the action of $\mathbf{g}$) of any point in $\mathbb{R}^p$ tends to the *origin* and thus shows a *contraction*. Conversely, for $i = 1, \ldots, p$, if $|\lambda_i| > 1$ then $\mathbf{D}^k \rightarrow 0$ as $k \rightarrow \infty$. Moreover, the orbit (under the action of $\mathbf{g}$) of any point in $\mathbb{R}^p$ expands away from the *origin* and thus shows an *expansion*.

---

[45] See also *diffeomorphism definition* in the Appendix.

Flow: From the point of view of the *flow*, let us consider the last theorem which precedes, in this case, the subspaces E^s and E^u are respectively identified as the eigenspaces of eigenvalues with modulus less than 1 and greater than 1. It is *fundamental* to consider that they are called *stable and unstable eigenspaces of* $\mathbf{A}$. Directly connected with what precedes, it is also fundamental to consider that *the direct sum of* E^s *and* E^u *gives the whole of* $\mathbb{R}^n$, because $\mathbf{A}$ is *hyperbolic*.[46] The expansion $E^u = \mathbb{R}^n$ and the contraction $E^s = \mathbb{R}^n$ are obviously hyperbolic. Moreover, if E^s, E^u are different from $\mathbb{R}^n$ the *diffeomorphism* is considered to be of *saddle-type*.

Let us consider a linear transformation $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^n$, it is possible to describe the corresponding linear *flow* on the same spaces as follows:

$$\phi_t(\mathbf{x}) = \exp(\mathbf{A}t)\mathbf{x}, \tag{1.324}$$

with $\exp(\mathbf{A}t)$ the exponential matrix which is defined as follows:

$$\exp(\mathbf{A}t) = \sum_{k=0}^{\infty} \frac{(\mathbf{A}t)^k}{k!}.$$

$\exp(\mathbf{A}t)\mathbf{x}$ has a vector field $\mathbf{A}\mathbf{x}$ and $\exp(\mathbf{A}t)\mathbf{x}_0$ is solution of the linear differential equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ passing through $\mathbf{x}_0$ at $t = 0$.

Definition 1.41 (Hyperbolic linear flow). The linear flow $\exp(\mathbf{A}t)\mathbf{x}$ is hyperbolic if $\mathbf{A}$ has no eigenvalues with zero real part.

Theorem 1.13. *Let* $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ *define a hyperbolic linear flow on* $\mathbb{R}^n$ *with dim* $E^s = n_s$. $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ *is said topologically equivalent to the system*

$$\dot{\mathbf{x}}_s = -\mathbf{x}_s, \quad \mathbf{x}_s \in \mathbb{R}^{n_s}, \tag{1.325}$$
$$\dot{\mathbf{x}}_u = -\mathbf{x}_u, \quad \mathbf{x}_u \in \mathbb{R}^{n_u}, \tag{1.326}$$

with $n = n_s + n_u$.

Let $\mathbf{A}$ *be a hyperbolic linear diffeomorphism with stable and unstable eigenspaces respectively* $E_{\mathbf{A}}^s$ *and* $E_{\mathbf{A}}^u$. *Define* $\mathbf{A}_i = \mathbf{A}|E_{\mathbf{A}}^i$, $i = s, u$. *Then* $\mathbf{A}_i$ *is said to be orientation-preserving (reversing) if* Det $\mathbf{A}_i > 0$ (Det $\mathbf{A}_i > 0$).

Conjugacy: A fundamental aspect of the topology relates to the concepts of "equivalence relation" and "*Conjugacy*". They make it possible to identify when flows have the same behavior.

Definition 1.42 (Topologically conjugate diffeomorphisms). Two diffeomorphisms $\mathbf{g}, \mathbf{q} : \mathbf{S} \to \mathbf{S}$, are topologically conjugate (or C^0-conjugate) if there is a homeomorphism $\mathbf{h} : \mathbf{S} \to \mathbf{S}$, such that

$$\mathbf{h} \cdot \mathbf{g} = \mathbf{q} \cdot \mathbf{h}. \tag{1.327}$$

---

[46] See Arnold.

The topologically Conjugacy for two *flows* ϕ_t $\varphi_t : \mathbf{S} \to \mathbf{S}$ is defined similarly with the preceding definition and the preceding equation is replaced by $\mathbf{h} \cdot \phi_t = \varphi_t \cdot \mathbf{h}$ with $t \in \mathbb{R}$. The *definition* above shows that the homeomorphism $\mathbf{h}$ *takes each orbit of* $\mathbf{g}$ *(or* ϕ_t*) onto an orbit of* $\mathbf{q}$ *(or* φ_t*) preserving the parameter* $p(t)$: that means $\mathbf{g}^p(\mathbf{x}) \xrightarrow{\mathbf{h}} \mathbf{q}^p(\mathbf{h}(\mathbf{x}))$ for each $p \in \mathbb{Z}$, or with the flows $\phi_t(\mathbf{x}) \xrightarrow{\mathbf{h}} \varphi_t(\mathbf{h}(\mathbf{x}))$ for each $t \in \mathbb{R}$.

Theorem 1.14 (Topologically conjugate). *Let* $\mathbf{A}, \mathbf{B} : \mathbb{R}^n \to \mathbb{R}^n$ *be hyperbolic linear diffeomorphisms. Then* $\mathbf{A}$ *and* $\mathbf{B}$ *are topologically conjugate if and only if:*

(1) $\dim E_{\mathbf{A}}^s = dim E_{\mathbf{B}}^s$ *(or equivalently* $dim E_{\mathbf{A}}^u = dim E_{\mathbf{B}}^u$*)*
(2) For $i = s, u$, $\mathbf{A}_i$ *and* $\mathbf{B}_i$ *are either both orientation preserving or both orientation-reversing.*

A *"nonlinear dynamical system"* is usually defined on a differentiable manifold, nevertheless the topological type of a fixed point is characterized by the restriction of the system to sufficiently small neighborhoods of points. Such neighborhoods can be selected to lie in a single chart so that only one representative of the system is involved. Then, in practice it is interesting to use and define diffeomorphisms and flows on open sets in $\mathbb{R}^n$. Let U be an open subset of $\mathbb{R}^n$ and $\mathbf{g} : U \to \mathbb{R}^n$ be a *nonlinear diffeomorphism* with *isolated fixed point* at $\mathbf{x}^* \in U$. The *linearization* of $\mathbf{g}$ at $\mathbf{x}^*$ is given by the well-known formula:

$$\mathbf{Dg}(\mathbf{x}^*) = \left[\frac{\partial g_i}{\partial x_j} \right]_{i,j=1 \, |_{\mathbf{x}=\mathbf{x}^*}}^n . \tag{1.328}$$

Definition 1.43 (Hyperbolic fixed point). A fixed point $\mathbf{x}^*$ of a diffeomorphism $\mathbf{g}$ is a hyperbolic fixed point if $\mathbf{Dg}(\mathbf{x}^*)$ is a hyperbolic linear diffeomorphism.

Theorem 1.15 (Hartman–Grobman). *Let* $\mathbf{x}^*$ *be a hyperbolic fixed point of the diffeomorphism* $\mathbf{g} : U \to \mathbb{R}^n$. *Then there exist a neighborhood* $N \subseteq U$ *of the point* $\mathbf{x}^*$ *and a neighborhood* $N' \subseteq U$ *containing the origin such that* $\mathbf{g}|N$ *is "topologically conjugate" to* $\mathbf{Dg}(\mathbf{x}^*)|N'$.

From the two preceding theorems it follows that here are $4n$ topological types of hyperbolic fixed points for diffeomorphisms: $\mathbf{g} : U \to \mathbb{R}^n$. They are depicted in the following invariant manifold theorem (which was already introduced previously in another way).

Theorem 1.16 (Invariant manifold). *Let* $\mathbf{g} : U \to \mathbb{R}^n$ *be a diffeomorphism with a hyperbolic fixed point at* $\mathbf{x}^* \in U$. *Then, on a sufficiently small neighborhood* $N \subseteq U$ *of* $\mathbf{x}^*$, *there exist local stable and unstable manifolds:*

$$W_{loc}^s(\mathbf{x}^*) = \{ \mathbf{x} \in U | \mathbf{g}^n(\mathbf{x}) \to \mathbf{x}^* \text{ as } n \to \infty \}, \tag{1.329}$$

$$W_{loc}^u(\mathbf{x}^*) = \{ \mathbf{x} \in U | \mathbf{g}^n(\mathbf{x}) \to \mathbf{x}^* \text{ as } n \to -\infty \}, \tag{1.330}$$

with the same dimension as E^s *and* E^u *for* $\mathbf{Dg}(\mathbf{x}^*)$ *and tangent to them at* $\mathbf{x}^*$.

It is possible to extend and generalize the preceding theorem which makes it possible to define *global stable and unstable manifolds* at $\mathbf{x}^*$ as follows:

$$W^s(\mathbf{x}^*) = \bigcup_{m \in \mathbb{Z}^+} \mathbf{g}^{-m}(W^s_{loc}(\mathbf{x}^*)), \qquad (1.331)$$

$$W^u(\mathbf{x}^*) = \bigcup_{m \in \mathbb{Z}^+} \mathbf{g}^m(W^u_{loc}(\mathbf{x}^*)), \qquad (1.332)$$

when $W^s(\mathbf{x}^*)$ and $W^u(\mathbf{x}^*)$ meet *transversely* at one point, they must do so infinitely many times and the result is *homoclinic tangle*. *The stated concepts above have a natural extension and generalization to the periodic points of* $\mathbf{g}$. Let us consider $\mathbf{x}^*$ belonging to a α-cycle of $\mathbf{g}$ then $\mathbf{x}^*$ is said to be a *hyperbolic point* of $\mathbf{g}$ *if it is a hyperbolic fixed point of* $\mathbf{g}^\alpha$. Then, the *orbit of the fixed point* $\mathbf{x}^*$ under the action of $\mathbf{g}$ *is referred to as a "hyperbolic periodic orbit"*. Then, the topological type of this orbit is determined by that of the corresponding fixed point of $\mathbf{g}^\alpha$. The last theorem above can be extended to $\mathbf{g}^\alpha$ (with α-cycle).

Flow: Obviously, it is possible to *apply* the previous *concepts and theorems* to *nonlinear flows*. Let us give the following definition and theorem, but before let us provide the convenient following basic elements:

- $\mathbf{X}$ is a *vector field*.
- U an *open subset* of $\mathbb{R}^n$.
- $\dfrac{d\mathbf{x}}{dt} = \dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$ is the *flow*, with $\mathbf{x} \in U$.
- Let $\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$ be such that $\mathbf{X}(\mathbf{x}^*) = 0$, $\mathbf{x}^* \in U$.
- Linearization of $\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$ at $\mathbf{x}^*$ is the linear differential equation written:

$$\dot{\mathbf{y}} = \mathbf{DX}(\mathbf{x}^*)\mathbf{y}, \qquad (1.333)$$

with $\mathbf{y} = (y_1, \ldots, y_n)^T$ are local coordinates at $\mathbf{x}^*$, and where

$$\mathbf{DX}(\mathbf{x}^*) = \left[\frac{\partial X_i}{\partial x_j} \right]^n_{i,j=1} \Big|_{\mathbf{x}=\mathbf{x}^*}. \qquad (1.334)$$

Definition 1.44 (Hyperbolic singular point of the flow). A singular point $\mathbf{x}^*$ of a vector field $\mathbf{X}$ is said to be hyperbolic if no eigenvalue of $\mathbf{DX}(\mathbf{x}^*)$ has zero real part.

Considering that $\mathbf{x}^*$ is a singular point of the vector field $\mathbf{X}$ then it is a fixed point of the flow $\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$. The definition above shows that $\mathbf{x}^*$ is a hyperbolic singular point of the vector field $\mathbf{X}$ if the flow $\exp(\mathbf{DX}(\mathbf{x}^*)t)\mathbf{x}$ of the linearization of $\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$ is hyperbolic in the sense of "hyperbolic linear flow" definition.

Theorem 1.17 (Hartman–Grobman (flow)). *Let* $\mathbf{x}^*$ *be a hyperbolic fixed point of* $\dot{\mathbf{x}} = \mathbf{X}(\mathbf{x})$ *with the flow* $\phi_t : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$. *Then there exists a neighborhood* N *of* $\mathbf{x}^*$ *on which* ϕ *is "topologically conjugate" to the linear flow* $\exp(\mathbf{DX}(\mathbf{x}^*)t)\mathbf{x}$. *Homoclinic tangle (Fig. 1.83), where stable and unstable manifolds of the saddle-point* $\mathbf{x}^*$ *intersect an arbitrary number of times (or infinitely).* $\mathbf{x}^*$ *is also the hyperbolic fixed point and* q *is the transverse homoclinic point.*

Fig. 1.83 Homoclinic tangle
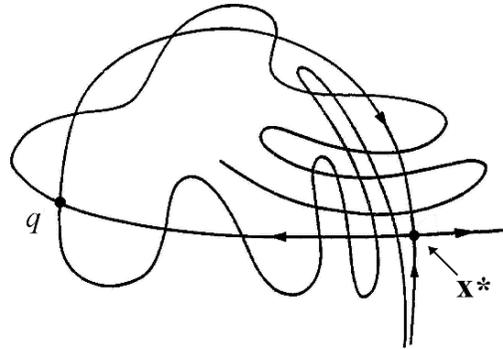
1.34.9 Hyperbolic Structure

An *invariant set* Φ of the horseshoe diffeomorphism $\mathbf{g} : U^2 \to U^2$ is said to be a *hyperbolic set* or to have a *hyperbolic structure* for $\mathbf{g}$. From the definitions exposed previously concerning the "hyperbolic fixed points" and "hyperbolic nonlinear fixed points" for a map $\mathbf{g}^\alpha$, there can be "derived" some important associated notions of "hyperbolic *periodic* orbit", "hyperbolic *invariant circle*", i.e. the *non-trivial* hyperbolic set (about hyperbolicity see statement and discussion by Arrowsmith and Place 1990). When we are involved in local behaviors of a map at a fixed point (concerning Euclidean or Banach spaces), the eigenvalues of the derivative map $\mathbf{Dg}^\alpha$ express the characteristics of the hyperbolic fixed point. However, if we analyze an invariant set Φ of a horseshoe diffeomorphism, this method cannot be applied to characterize the hyperbolicity. The called upon reasons in this framework are that we are "located" in *manifolds*[47] *and differentiable manifolds*[48] and also the *complexity of the invariant set* Φ does not make it possible to explain it by means of a fixed-point of $\mathbf{g}^\alpha$ (e.g. invariant set Φ can have aperiodic orbits which cannot deal with a fixed-point). Indeed, *in the invariant set* Φ, $\mathbf{x}$ *and* $\mathbf{g}(\mathbf{x})$ *are different points and this difference (or distance) must be taken into account in the hyperbolicity.* However, there is a way of making it by replacing the *derivative map* by the *tangent map*. Consequently, the notions of *hyperbolic nonlinear fixed points*, *invariant manifolds* and the *Hartman–Grobman theorem (*exposed previously) can be used and applied. Firstly, they can be applied to diffeomorphisms $\mathbf{g} : M \to M$ when M is a *n-dimension differentiable manifold* (which is not a subset of $\mathbb{R}^n$). Thus, as we said above, the *derivative map*

---

[47] *Manifold:* A *topological space* which is *locally Euclidean;* there are *four types*: *topological, piecewise linear, differentiable, or complex analytic functions of those in Euclidean space; intuitively, a surface.* (Furthermore, a manifold can have global topological properties, such as non-contractible loops, that distinguish it from the topologically trivial $\mathbb{R}^n$.) (See Appendix.)

[48] *Differentiable manifold:* Topological space with a maximal differentiable atlas; roughly speaking, a smooth surface (see Appendix for complete definitions). *Differentiable atlas:* A family of *embeddings* $h_i : E^n \to M$ of Euclidian space into a topological space M with the property that $h_i^{-1} h_j : E^n \to E^n$ is a differentiable map for each pair of indices i, j. *Atlas:* An atlas for a *manifold* is a *collection of coordinate patches* that *covers* the *manifold.*
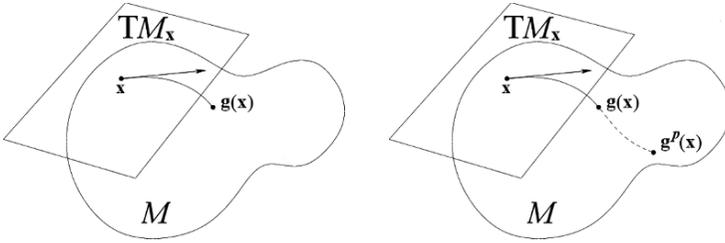
Fig. 1.84 $\mathbf{T}M_{\mathbf{x}}$: tangent space to M at $\mathbf{x}$ (*left*). And $\mathbf{x}$ orbit through the map $\mathbf{g}$ in M (*right*)

$\mathbf{Dg}$ is replaced by the *tangent map* $\mathbf{Tg}$. Thus, let us consider the *derivative map*

$$\mathbf{Dg}(\mathbf{x}^*) : \mathbb{R}^n \to \mathbb{R}^n \qquad (1.335)$$

which is usually used to analyze the fixed-point $\mathbf{x}^*$ of $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ (see preceding section about hyperbolic nonlinear fixed point) and let us replace it by

$$\mathbf{Tg}_{\mathbf{x}^*} : \mathbf{T}M_{\mathbf{x}^*} \to \mathbf{T}M_{\mathbf{x}^*} \qquad (1.336)$$

knowing that $\mathbf{T}M_{\mathbf{x}^*}$ is the *tangent space* to M at $\mathbf{x}^*$ (Fig. 1.84).

Since M is an n-dimension differentiable manifold, then $\mathbf{T}M_{\mathbf{x}}$ is regarded as an *equivalent class of curves* on M having the same tangent vector at $\mathbf{x}$.[49] Thus, let us consider $\rho(0) = \mathbf{x}^*$ the parameterized curve on M passing through $\mathbf{x}^*$ and more generally let us consider $\rho(t)$ to study $\mathbf{g}$ near $\mathbf{x}$. (with t belonging to $\mathbb{R}$ or to an interval of $\mathbb{R}$). The tangent vector at $\mathbf{x}^*$ is obtained by *differentiation* of $\rho(t)$ with respect to t. Such an approach can be carried out only if we use a *local chart*.

Let us recall briefly how to define a "local chart" and some very useful associated concepts: if we consider a manifold M of dimension n, then for any $\mathbf{x} \in M$ there is a neighborhood $U \subseteq M$ containing $\mathbf{x}$, and we consider a *homeomorphism* $\mathbf{h} : U \to \mathbb{R}^n$ with $V = \mathbf{h}(U) \subseteq \mathbb{R}^n$ from which the coordinates curves can be mapped back onto U. Then, $\mathbf{h}$ can be taken as the local coordinates on the *local surface* (U of M) which is also called a *patch*.[50] The pair $(V, \mathbf{h})$ is said a *chart* which can be taken to determine *differentiability* on the neighborhood U. Moreover, *several open sets* by means of the *patches* can be necessary to *cover*[51] the whole *manifold*. The following graphical example of the sphere is given to show a differentiable manifold and a patch (Fig. 1.85). Similar decompositions could be shown, e.g. the torus or the cylinder.

---

[49] For a report or discussion see Chillingworth (1976) and Arrowsmith and Place (1990).

[50] *Patch (or local surface):* A "patch", also called a "local surface", is a differentiable mapping $\mathbf{f}: U \to \mathbb{R}^n$, where U is an open subset of $\mathbb{R}^2$. More generally, if A is any subset of $\mathbb{R}^2$, then a map $: A \to \mathbb{R}^n$ is a patch provided that $\mathbf{f}$ can be extended to a differentiable map from U into $\mathbb{R}^n$, where U is an open set containing A. Here, $\mathbf{f}(U)$ (or more generally, $\mathbf{f}(A)$) is called the map trace of $\mathbf{f}$ (see Appendix).

[51] *Cover:* An element a of a partially ordered set covers another element b if a is greater than b, and the only elements that are both greater than or equal to b and less than or equal to a are a and b themselves. *Covering:* For a set E, a collection of sets whose union contains E.
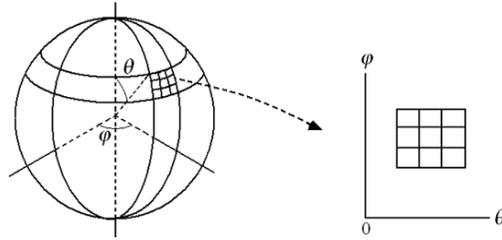
Fig. 1.85 A differential manifold (*sphere*) and patch of local (*polar*) coordinates
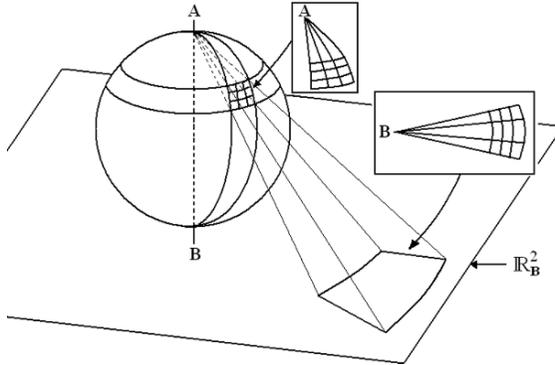
Fig. 1.86 Stereographic projection of the sphere

Besides, more generally to understand the framework, we have the opportunity to show the stereographic projection of the sphere with patches (Fig. 1.86).

Furthermore, consider a map $\mathbf{g} : U \to U$, from this map it is possible to build a new map $\check{\mathbf{g}} : V \to V$, such that $\check{\mathbf{g}} = \mathbf{h} \cdot \mathbf{g} \cdot \mathbf{h}^{-1}$ (see commutative diagram below):

$$
\begin{array}{ccc}
U & \xrightarrow{\ \mathbf{g}\ } & U \\
{\scriptstyle \mathbf{h}^{-1}}\big\uparrow & & \big\downarrow{\scriptstyle \mathbf{h}} \\
V & \xrightarrow{\ \check{\mathbf{g}}\ } & V
\end{array}
$$

and if it is also assumed that $\mathbf{g}$ is a C^r-*map* on V, we are thus provided with the preliminary elements which make possible to define a *local diffeomorphism* on a manifold M. This induces that when we want a global description of a manifold M by means of charts we need to cover it with a collection of the open sets U_i, each one being associated with a *local chart* $(V_i, \mathbf{h}_i)$ (see Arrowsmith and place 1990, Arnold 1973, Chillingworth 1976).

After these reminders, remember also that for M an *n-dimension differentiable manifold*, the *tangent vector* at $\mathbf{x}^*$ is obtained by *differentiation* of $\rho(t)$ with respect to t (remember that $\rho(0) = \mathbf{x}^*$ is the parameterized curve on M passing through $\mathbf{x}^*$ and in a generic way we write $\rho(t)$ to study $\mathbf{g}$ near $\mathbf{x}$). Such an approach can be carried out only if we use a *local chart* $(V_i, \mathbf{h}_i)$. It is possible to gather $\check{\mathbf{g}}, \check{\rho}$,

$(V_i, \mathbf{h}_i)$ and to provide the local elements to study the *differentiability* as follows:
$\check{\mathbf{g}}_i = \mathbf{h}_i \cdot \mathbf{g} \cdot \mathbf{h}_i^{-1}$, $\check{\rho}_i = \mathbf{h}_i \cdot \rho$,

$$(\mathbf{g} \cdot \rho)_i = \mathbf{h}_i \cdot (\mathbf{g} \cdot \rho). \tag{1.337}$$

These local elements verify: $(\mathbf{g} \cdot \rho)_i(t) = \check{\mathbf{g}}_i \cdot (\check{\rho}_i(t))$. With a manifold M (at least) of class C^1, the preceding equation can be differentiated to provide at $t = 0$:

$$(\mathbf{g} \cdot \rho)_i(0) = \mathbf{D}\check{\mathbf{g}}_i \cdot (\check{\rho}_i(0))\dot{\check{\rho}}_i(0). \tag{1.338}$$

$(\mathbf{g} \cdot \rho)_i(0)$ and $\dot{\check{\rho}}_i(0)$ are *i-representatives* of components of $\mathbf{T}M_{\mathbf{x}^*}$ in $(V_i, \mathbf{h}_i)$. Since we postulate being in presence of a *hyperbolic fixed-point*, (locally) this means that $\mathbf{D}\check{\mathbf{g}}_i(\check{\mathbf{x}}_i^*)$ has no eigenvalue with unit modulus (see hyperbolic fixed point definition and the associated definitions). The derivative map $\mathbf{D}\check{\mathbf{g}}_i(\check{\mathbf{x}}_i^*)$ is the local representative of the tangent $\mathbf{T}\mathbf{g}_{\mathbf{x}^*}$.
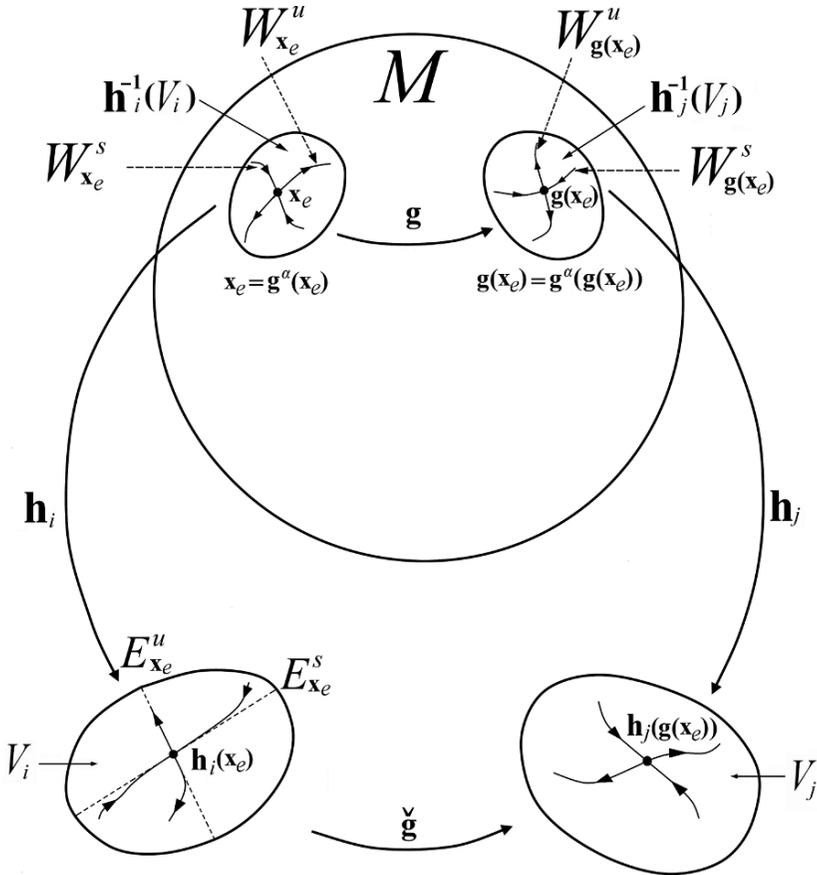
The preceding figure is used to illustrate the general framework. The invariant manifold theorem (see a previous section) makes it possible to provide the stable and unstable manifolds in each chart for $\mathbf{g}^\alpha$ concerning a hyperbolic periodic orbit. The subspaces or eigenspaces $E_{\mathbf{x}_e}^u$ and $E_{\mathbf{x}_e}^u$ are tangent to the images of $W_{\mathbf{x}_e}^u$ and $W_{\mathbf{x}_e}^u$. Let us observe in the figure the two charts $(V_i, \mathbf{h}_i)$ and $(V_j, \mathbf{h}_j)$. The homeomorphism $\mathbf{h}$ makes it possible to pass from a neighborhood of M to V (locally for i, j) and make it possible to map back onto this neighborhood of M (recalls: in a generic way, for any $\mathbf{x} \in M$ there are a neighborhood $U \subseteq M$ containing $\mathbf{x}$ and $\mathbf{h} : U \to \mathbb{R}^n$ with $V = \mathbf{h}(U) \subseteq \mathbb{R}^n$ from which the coordinates curves can be mapped back onto U, $\mathbf{h}^{-1}(V) = U$).

At the same time, it is possible to provide a *metric* for $\mathbf{T}M_{\mathbf{x}^*}$ *(and* $\mathbf{T}V_{\check{\mathbf{x}}_i^*}$*) which can be applied* to all $\mathbf{w}$ of stable and unstable eigenspaces. (We have chosen the lowercase notation $\mathbf{w}$ which differs from W or $\mathbf{W}$ and expresses a notion which is different but closely related of course.) We thus introduce the concept of *Riemannian structure* for differentiable manifolds. However, it is important to understand that (relevant hypothesis for a hyperbolic fixed-point) when $\mathbf{x}^*$ *lies in the overlap of two charts* $(V_i, \mathbf{h}_i)$ and $(V_j, \mathbf{h}_j)$, we must *choose a metric for each chart* and *not for all* $\mathbf{w} \in \mathbf{T}M_{\mathbf{x}^*}$. Each metric is selected on $\mathbf{T}V_{\check{\mathbf{x}}_i^*}$ and $\mathbf{T}V_{\check{\mathbf{x}}_j^*}$, and they respect $|\check{\mathbf{w}}_i|_i = |\check{\mathbf{w}}_j|_j$ where $\check{\mathbf{w}}_i$ is the i-representative of $\mathbf{w}$ and $\check{\mathbf{w}}_j$ is the j-representative of $\mathbf{w}$. Moreover, *it is of primary importance to understand that the common value defines* $\|\mathbf{w}\|_{\mathbf{x}^*}$ *for each* $\mathbf{w} \in \mathbf{T}M_{\mathbf{x}^*}$. For all points $\mathbf{x}$ of all overlaps of an *atlas*,[52] these types of metrics $\|\mathbf{w}\|_{\mathbf{x}}$ (positively definite) constitute a *Riemannian structure* for the manifold M. (Recall: an *Atlas* for a *manifold* is a *collection of coordinate patches that covers* the *manifold*.) Thus, provided with this Riemann structure, the manifold M can depict the hyperbolic characteristic of the hyperbolic fixed-point $\mathbf{x}^*$ by verifying that, *first (1):*

$$\mathbf{T}M_{\mathbf{x}^*} = E_{\mathbf{x}^*}^s \oplus E_{\mathbf{x}^*}^u, \tag{1.339}$$

with $E_{\mathbf{x}^*}^s$ is obviously the *stable eigenspace* and $E_{\mathbf{x}^*}^u$ is the unstable eigenspace of $\mathbf{Tg}_{\mathbf{x}^*}$; and *second (2),* there exist λ, γ, δ with $\lambda, \gamma > 0$, and $0 < \delta < 1$ such that:

$$\|(\mathbf{Tg}_{\mathbf{x}^*})^n(\mathbf{w})\|_{\mathbf{x}^*} > \lambda \delta^n \|\mathbf{w}\|_{\mathbf{x}^*} \qquad \text{for all } \mathbf{w} \in E_{\mathbf{x}^*}^s, \tag{1.340}$$

$$\|(\mathbf{Tg}_{\mathbf{x}^*})^n(\mathbf{w})\|_{\mathbf{x}^*} < \gamma \delta^{-n} \|\mathbf{w}\|_{\mathbf{x}^*} \qquad \text{for all } \mathbf{w} \in E_{\mathbf{x}^*}^u, \tag{1.341}$$

with $n \in \mathbb{Z}^+$.

Hereafter, the figure is a schematic definition of a *norm* $\|\cdot\|_{\mathbf{x}^*}$ on $\mathbf{T}M_{\mathbf{x}^*}$ in terms of norms on $\mathbf{T}V_{\check{\mathbf{x}}_i^*}$ and $\mathbf{T}V_{\check{\mathbf{x}}_j^*}$. The compatible norms on $\mathbf{T}V_{\check{\mathbf{x}}_i^*}$ and $\mathbf{T}V_{\check{\mathbf{x}}_j^*}$ are respectively $|\cdot|_i$ and $|\cdot|_j$.

---

[52] *Atlas:* for a *manifold* is a *collection of coordinate patches* that *covers* the *manifold*.

$$\|\mathbf{w}\|_{\mathbf{x}^*} = |\check{\mathbf{w}}_i|_i \qquad\qquad\qquad \|\mathbf{w}\|_{\mathbf{x}^*} = |\check{\mathbf{w}}_j|_j$$
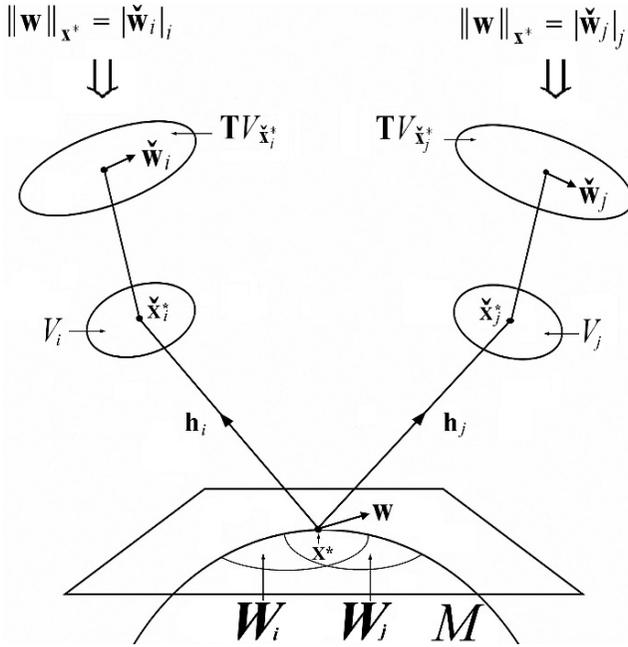
At this stage, we can also (explicitly) re-introduce the orbit and recurrence notions in the writings and notations. So, considering that the map $\mathbf{g}$ has a α-periodic orbit, the invariant set $\Phi^{(\alpha)} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\alpha-1}\}$, it comes $\mathbf{x}_e = \mathbf{g}^e(\mathbf{x}_0)$ is the fixed point of $\mathbf{g}^\alpha$. Then, it is possible to re-write the preceding conditions as follows, by verifying that, *first (1)*:[53]

$$TM_{\mathbf{x}_e} = E^s_{\mathbf{x}_e} \oplus E^u_{\mathbf{x}_e} \text{ and } \mathbf{Tg}_{\mathbf{x}_e}\left(E^s_{\mathbf{x}_e}\right) = E^s_{\mathbf{g}(\mathbf{x}_e)}, \quad \mathbf{Tg}_{\mathbf{x}_e}\left(E^u_{\mathbf{x}_e}\right) = E^u_{\mathbf{g}(\mathbf{x}_e)} \qquad (1.342)$$

with $E^s_{\mathbf{x}^*}$ is obviously the stable eigenspace and $E^u_{\mathbf{x}^*}$ is the stable eigenspace of $\mathbf{Tg}_{\mathbf{x}^*}$; and *second (2)*, there exist λ, γ, δ with $\lambda, \gamma > 0$, and $0 < \delta < 1$ such that:

$$\left\|\left(\mathbf{Tg}_{\mathbf{x}_e}\right)^n \mathbf{w}\right\|_{\mathbf{g}^n(\mathbf{x}_e)} > \lambda \delta^n \|\mathbf{w}\|_{\mathbf{x}_e} \qquad \text{for all } \mathbf{w} \in E^s_{\mathbf{x}_e}, \qquad (1.343)$$

$$\left\|\left(\mathbf{Tg}_{\mathbf{x}_e}\right)^n \mathbf{w}\right\|_{\mathbf{g}^n(\mathbf{x}_e)} < \gamma \delta^{-n} \|\mathbf{w}\|_{\mathbf{x}_e} \qquad \text{for all } \mathbf{w} \in E^u_{\mathbf{x}_e}. \qquad (1.344)$$

Definition 1.45 (Hyperbolic invariant set). An invariant set Φ is said to be hyperbolic for $\mathbf{g}$ (or to have a hyperbolic structure) if for each $\mathbf{x} \in \Phi$, the tangent space $TM_{\mathbf{x}^*}$ splits into two linear subspaces $E^s_{\mathbf{x}}, E^u_{\mathbf{x}}$ such that (1):

$$\mathbf{Tg}_{\mathbf{x}}\left(E^{s,u}_{\mathbf{x}}\right) = E^{s,u}_{\mathbf{g}(\mathbf{x})}, \qquad (1.345)$$

---

[53] For discussion and developments, ref. to Arrowsmith and Place (1990).

and (2) the inequations:

$$\left\| \left(\mathbf{Tg}_{\mathbf{x}_e} \right)^n \mathbf{w} \right\|_{\mathbf{g}^n(\mathbf{x}_e)} > \lambda \, \delta^n \left\| \mathbf{w} \right\|_{\mathbf{x}_e} \qquad \text{for all } \mathbf{w} \in E^s_{\mathbf{x}_e}, \tag{1.346}$$

$$\left\| \left(\mathbf{Tg}_{\mathbf{x}_e} \right)^n \mathbf{w} \right\|_{\mathbf{g}^n(\mathbf{x}_e)} < \gamma \delta^{-n} \left\| \mathbf{w} \right\|_{\mathbf{x}_e} \qquad \text{for all } \mathbf{w} \in E^u_{\mathbf{x}_e}, \tag{1.347}$$

with $x_e \mapsto x$, are satisfied for all positive integers n, and (3) the subspaces $E^s_{\mathbf{x}}$, $E^u_{\mathbf{x}}$ depend continuously on $\mathbf{x} \in \Phi$.

Theorem 1.18 (Insets and invariant manifold). *Let $\mathbf{g} : M \to M$ be a diffeomorphism on a compact manifold without boundary with a hyperbolic non-wandering set Ω. If the periodic points of $\mathbf{g}$ are dense in Ω, then Ω can be written as a disjoint union of finitely many basic sets Ω_i:*

$$\Omega = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_k. \tag{1.348}$$

Each Ω_i is "closed", "invariant" and contains a "dense orbit" of $\mathbf{g}$. Furthermore, the splitting of Ω into basics sets is unique and M can be decomposed as a disjoint union

$$M = \bigcup_{i=1}^{k} in\left(\Omega_i \right), \tag{1.349}$$

where $in(\Omega_i) = \{\mathbf{x} \in M | \mathbf{g}^m(\mathbf{x}) \to \Omega_i, m \to \infty\}$ is the inset[54] of Ω_i.

1.34.10 Homoclinic Orbit and Perturbation: Melnikov

Transverse homoclinic points can occur in the *Poincaré map* of some types of three-dimensional flows. In the literature, there are many studies about periodic solutions of systems with dimension larger than 2.[55] The *theory of periodic solutions* uses *functional analytic and topological methods.* (It is known that due to the translation property of solutions of autonomous equations, periodic solutions of these equations correspond with closed orbits in phase space.) More generally, let us consider non-autonomous equations. It is interesting to study the "so-called" T-mapping α_T. If we consider an equation $\dot{x} = f(t,x)$ in $\mathbb{R}^n$ and its solution $x(t;x_0)$ which begins at $t = t_0$ in x_0, then such a point x_0 gives $(t_0 + T; x_0)$ knowing T constant, and selected so that the solutions exists (with $0 \leq t - t_0 \leq T$. Moreover, the set $S \subset \mathbb{R}^n$ of points x_0 to which we can assign such a point $x(t_0 + T; x_0)$ is mapped by the T-mapping α_T into $\mathbb{R}^n$, $\alpha_T : S \to \mathbb{R}^n$. In the mapping α_T, T is taken as a parameter. x_0 is a fixed point of α_T if $\alpha_T (t_0; x_0) = x_0$ or $\alpha_T (t_0 + T; x_0) = x_0$. In such a way, let us consider in $\mathbb{R}^n$ the generic equation $\dot{x} = f(t,x)$ which is T-periodic $f(t + T, x) = f(t,x)$, with $t \in \mathbb{R}$ and $x \in \mathbb{R}^n$. It is known that the equation has a T-periodic solution if and only if the T-mapping α_T has a fixed point. The most important notion here is to understand that

---

[54] *Inset:* roughly speaking an inset is *a map within a map*, either at a smaller *scale* to show relative location, or a larger scale to show detail.

[55] Amann (1983), Hale (1969), Sansone and Conti (1964), Cesari (1971).

we show the fixed point without explicitly knowing the "solutions".[56] After these recalls, let us consider the following basic model:

$$\ddot{x} + x - x^2 = \varepsilon f(t,x), \tag{1.350}$$

where f is T-periodic in t. Then, we can consider the T-map of the $x, \dot{x}$-phase-plane. $\varepsilon \geqslant 0$ is a perturbation.[57] Below, for $\varepsilon = 0$, the left-figure shows the unperturbated phase-plane (which exhibits in particular a saddle-loop), and for $\varepsilon > 0$, the right figure shows the *invariant manifold* of point P_{ε} in the T-map ($\varepsilon > 0$) and the distance (function).

Distance at time t
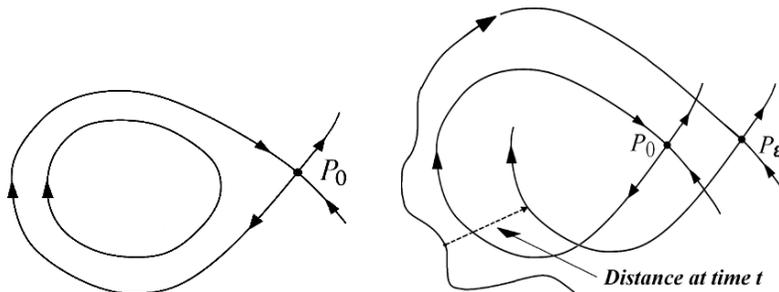
When a Poincaré map exhibits a *saddle-point* with a *closed loop*, an orbit starting on such a loop is said *homoclinic*. When $t \to +\infty$, the orbit approaches the saddle-point, i.e. a periodic solution, and when $t \to -\infty$, the *homoclinic orbit* approaches *the same periodic solution*. When the dimension of a map is higher than two, the *fixed point* is *"hyperbolic", i.e. the real parts* of the *associated eigenvalues are nonzero*. Moreover, *when a loop connects two saddle points* (or two hyperbolic fixed points in more dimensions), *the orbits on such a loop are said heteroclinic*. When $t \to +\infty$, they approach a periodic solution, and when $t \to -\infty$, they approach a periodic solution of different type.

Starting from (left-figure above) the saddle-loop, i.e. a *saddle-point* with a *closed loop*, when a small perturbation $\varepsilon > 0$ is introduced, the *saddle-point* is slightly *shifted* and becomes a *hyperbolic fixed point*. Then, the *saddle-loop* is *transformed into* a *stable* and *unstable manifold*. When these two manifolds intersect, we have at such an intersection point a homoclinic orbit as the point belongs to both manifolds. Consequently, when the two manifolds intersect, these manifolds usually intersect an infinite number of times, providing thus an infinite number of homoclinic orbits. Moreover, the map contains a *horseshoe map*. The figure relative to the transformation shows a *homoclinic tangle*.

The *Melnikov method* consists in *calculating the distance* and thus the *possibility for the stable and unstable manifold to intersect*. The method uses the *properties of unperturbated* solutions.[58]

---

[56] See also *Brouwer's theorem* and lemma, and see Amann (1983).

[57] see also the *Kolmogorov–Arnold–Moser theorem* (1954) outlined by Kolmogorov, which was subsequently proved in the 1960s by Arnold and Moser.

[58] See Guckenheimer and Holmes (1983) and Wiggins (1988).

Let us consider an *alternative approach* in order to show the Melnikov method (and let us change the notation). Thus, consider a (planar) *differential equation*:

$$\dot{\mathbf{x}} = \mathbf{g}_0(\mathbf{x}), \quad \text{at} \quad \mathbf{x} = \mathbf{0}, \tag{1.351}$$

which contains a *hyperbolic saddle-point* at $\mathbf{x} = \mathbf{0}$ and a *homoclinic saddle connection* Υ, as it is possible to observe in the figure below.
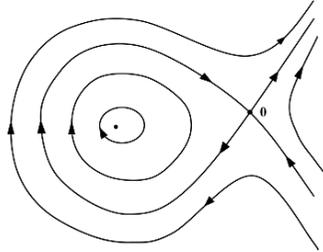
Let us assume the *flow* of such a dynamics in $\mathbb{R}^2 \times Q^1$ which is written

$$\dot{\mathbf{x}} = \mathbf{g}_0(\mathbf{x}), \quad \dot{\theta} = 1. \tag{1.352}$$

Then the *saddle-point is a saddle periodic orbit* ξ_0, with $\theta \in Q^1, \mathbf{x} \in \mathbb{R}^2$. It is supposed that the stable manifold $\mathbf{W}^s(\xi_0)$ intersects the unstable manifold $\mathbf{W}^u(\xi_0)$ (in a cylindrical surface $\Upsilon \times Q^1 \subseteq \mathbb{R}^2 \times Q^1$). However, this case is not automatic (indeed, through a Poincaré section Σ_θ, the stable and unstable manifolds of the fixed point usually do not intersect transversely). Provided with this device, then the Melnikov approach consists in applying small perturbations to the preceding system, it comes

$$\dot{\mathbf{x}} = \mathbf{g}_0(\mathbf{x}) + \varepsilon \mathbf{g}_1(\mathbf{x}, \theta), \quad \dot{\theta} = 1, \ (\text{with } \varepsilon \in \mathbb{R}^+), \tag{1.353}$$

with $\theta \in [0, 2\pi]$. In order to understand the influence of the perturbation, let us introduce the following proposition.

Proposition 1.4 (Same topological type). *Let $\mathbf{x}^*$ be a hyperbolic fixed point of the diffeomorphism $\mathbf{g} : U \to \mathbb{R}^n$. Then there is a neighborhood A of $\mathbf{x}^*$ in U and a neighborhood B of $\mathbf{g}$ in the set of C^1-diffeomorphisms $\mathbf{g} : U(\subseteq \mathbb{R}^n) \to \mathbb{R}^n$ with the C^1-norm such that every map $\mathbf{q} \in B$ has a unique hyperbolic fixed point $\mathbf{y}^* \in A$ of the same topological type as $\mathbf{x}^*$.*

It follows from the preceding proposition that when a sufficiently small value of ε is selected, the system also has a hyperbolic periodic orbit ξ_ε near ξ_0. But the invariant manifolds $\mathbf{W}^s(\xi_\varepsilon)$ and $\mathbf{W}^u(\xi_\varepsilon)$ need not intersect to design a cylinder (see Fig. 1.87). The *distance* between these two manifolds is the subject of the Melnikov function. Remember that $\mathbf{x}_0 \in \mathbb{R}^2$, Υ is the *saddle connection* of the unperturbated system and $\mathbf{x}_0$ belongs to this saddle connection Υ. The Poincaré section is dependent on $\theta : \Sigma_\theta$. (We select a *perpendicular section* called $\mathbf{L_p}$ to the *saddle connection* Υ in the *plane of the Poincaré section* Σ_θ). The *perturbated system*
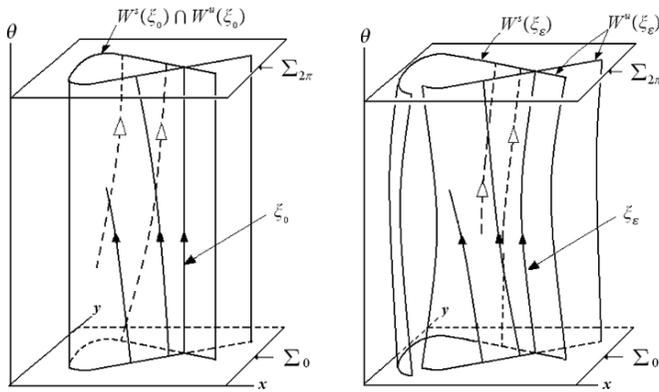
Fig. 1.87 Unperturbated system $\varepsilon = 0$ (*left*). Perturbated system $\varepsilon > 0$ (*right*)

is characterized by $\xi_\varepsilon, \mathbf{W}^s(\xi_\varepsilon), \mathbf{W}^u(\xi_\varepsilon)$, which is respectively the orbit, the stable manifold and the unstable manifold. We can observe the behavior of this system in the plane of the Poincaré section Σ_θ (similar to the Poincaré map[59]). The intersections are written: $\mathbf{W}^s(\xi_\varepsilon) \cap \Sigma_\theta$, $\mathbf{W}^u(\xi_\varepsilon) \cap \Sigma_\theta$, $\xi_\varepsilon \cap \Sigma_\theta = \mathbf{x}^*_{\varepsilon,\theta}$, where $\mathbf{x}^*_{\varepsilon,\theta}$ is the *hyperbolic saddle point* close to $\mathbf{x} = \mathbf{0}$.

---
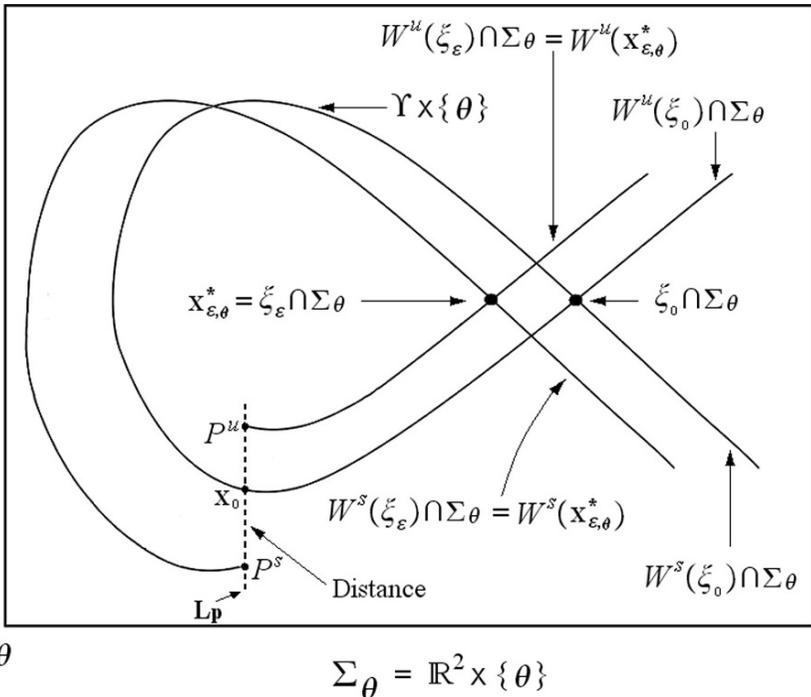
[59] The Poincaré map is written $\mathbf{P}_{\varepsilon,\theta} : \Sigma_\theta \longrightarrow \Sigma_\theta$.

The *distance* (relates to the Melnikov function) between $\mathbf{W}^s(\xi_\varepsilon)$ and $\mathbf{W}^u(\xi_\varepsilon)$ is calculated along $\mathbf{L_p}$. Usually, this distance varies according to θ if ε is strictly positive $(\varepsilon > 0)$. Along $\mathbf{L_p}$ it is possible to observe closest to $\mathbf{x}_0$ the intersection points P^u and P^s with the curves of each manifold $\mathbf{W}^s(\xi_\varepsilon)$ and $\mathbf{W}^u(\xi_\varepsilon)$. The *distance function is time-dependent,* thus the *Melnikov function* can be written as follows:

$$M(\theta) = \int_{-\infty}^{+\infty} \mathbf{g}_0(\mathbf{x}_0(t-\theta)) \wedge \mathbf{g}_1(\mathbf{x}_0(t-\theta), t) dt \qquad (1.354)$$

with

$$\Delta_\varepsilon(\theta, \theta) = \varepsilon M(\theta) + O(\varepsilon^2). \qquad (1.355)$$

$O(.)$ is a type of *asymptotic notation.*[60] Knowing that $\Delta_\varepsilon(t, \theta)$ is $|\mathbf{g}_0(\mathbf{x}_0(t-\theta))|$ times the component of the vector $[\mathbf{x}^u(t; \theta, \varepsilon) - \mathbf{x}^s(t; \theta, \varepsilon)]$ *perpendicular* to $\mathbf{g}_0(\mathbf{x}_0(t-\theta))$, and $\Delta_\varepsilon(\theta, \theta)$ can be written

$$\Delta_\varepsilon(\theta, \theta) = \varepsilon \int_{-\infty}^{+\infty} \mathbf{g}_0(\mathbf{x}_0(t-\theta)) \wedge \mathbf{g}_1(\mathbf{x}_0(t-\theta), t) dt + O(\varepsilon^2). \qquad (1.356)$$

Proposition 1.5 (Transverse intersection). *If $M(\theta)$ has simple zeros, then for sufficiently small $\varepsilon > 0$, $\mathbf{W}^u(\mathbf{x}_{\varepsilon, \theta}^*)$ and $\mathbf{W}^s(\mathbf{x}_{\varepsilon, \theta}^*)$ intersect transversely for some $\theta \in [0, 2\pi]$. Otherwise, if $M(\theta)$ is delimited away from zero, then $\mathbf{W}^u(\mathbf{x}_{\varepsilon, \theta}^*) \cap \mathbf{W}^s(\mathbf{x}_{\varepsilon, \theta}^*) = \varnothing$ for all θ.*

The Melnikov method is particularly interesting because it can be applied to the *Duffing attractor* whose dynamical system can be written as follows: $\dot{x} = y$, $\dot{y} = x - x^3 + \varepsilon(a\cos\theta - by)$, $\dot{\theta} = 1$, it is periodic in θ and the most typical behavior is shown in the figure hereafter for $\varepsilon a = 0.4$ and $\varepsilon b = 0.25$.[61]
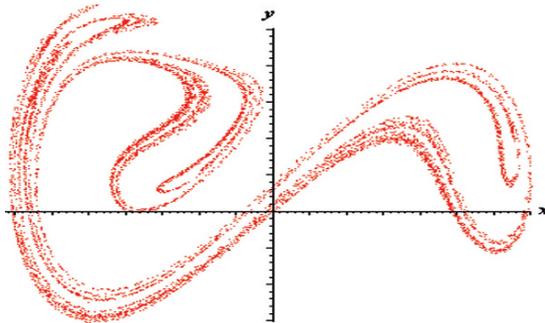
It is an interesting exercise to apply the Melnikov function and Poincaré map to the Duffing attractor; in order to observe the system behavior according to the parameter setting and in particular for different ε values (for discussions and developments,

[60] The symbols $O(x)$ (sometimes called the *O*-symbol, *O*-notation) are a type of *asymptotic notation* collectively known as *Landau symbols.*

[61] Using the Euler approximation.

see Arrowsmith and Place, Guckenheimer and Holmes). We know that the Duffing dynamical system can be written

$$\dot{x} = y, \quad \dot{y} = x - x^3 + \varepsilon(a\cos\theta - by), \quad \dot{\theta} = 1, \tag{1.357}$$

when $\varepsilon = 0$, the system becomes $\dot{x} = y$, $\dot{y} = x - x^3$, $\dot{\theta} = 1$. It can be written as a *Hamiltonian system* such that: $H(x,y) = [(y^2 - x^2 + x^4/2)]/2$ and with $H(x,y) = 0$, the system describes *two homoclinic orbits* that we can write $\Upsilon_0^{\pm}$, the saddle point is localized at $\mathbf{x} = \mathbf{0}$. For discussions about the case, when $\varepsilon > 0$ see Guckenheimer and Holmes and for the Melnikov function also see Arrowsmith–Place. The following figures briefly depict the typical behaviors of the Duffing oscillator by means of the Poincaré map: (1) Phase-portrait of the system for $\varepsilon = 0$ such that $\dot{x} = y$, $\dot{y} = x - x^3$, $\dot{\theta} = 1$, which exhibits *two homoclinic orbits* $\Upsilon_0^{\pm}$ and the stable and unstable manifolds with the saddle point at $\mathbf{x} = \mathbf{0}$; (2) Trajectory *sketch* of stable and unstable manifolds for $\varepsilon a = 0.10$, $\varepsilon b = 0.25$; (3) Trajectory *sketch* of stable and unstable manifolds for $\varepsilon a = 0.40$, $\varepsilon b = 0.25$; (4) Euler approximation of the (real) *attractor* for $\varepsilon = 0.40$, $\varepsilon b = 0.25$.
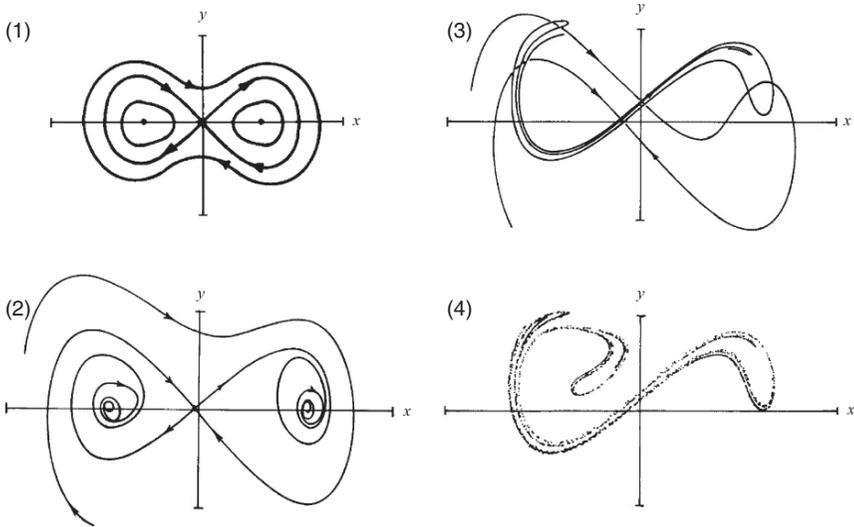
Within the general framework of this section about the Melnikov method which is related to the Smale-horseshoe map, now we want to highlight the *horseshoe map* and its associated dynamics. As shown in a previous section, the homoclinic tangle depicted in Fig. 1.88 (theoretical) expresses the stable and unstable manifolds of the saddle-point $\mathbf{x}^*$ which intersect an arbitrary number of times (or infinitely). And $\mathbf{x}^*$ is called the *hyperbolic fixed point* and q is the transverse homoclinic point.

Let us consider the domain Δ which is shown in Fig. 1.89.[62] By invariance Δ must maintain contact with $\mathbf{W}^u(\cdot)$ and with $\mathbf{W}^s(\cdot)$ under the action of iterations by

---

[62] *Domain*: The term domain is most commonly used to describe the set of values for which a map is defined.

Fig. 1.88 Homoclinic tangle
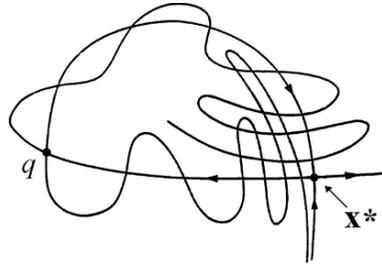
g. Figure 1.89 shows the domain Δ, then maps $\mathbf{g}(\Delta)$, $\mathbf{g}^2(\Delta)$, $\mathbf{g}^3(\Delta)$, $\mathbf{g}^4(\Delta)$, $\mathbf{g}^5(\Delta)$, $\mathbf{g}^6(\Delta)$, $\mathbf{g}^7(\Delta)$. The remaining iterates of Δ appear in Fig. 1.89 (right). Figure 1.89 (right) shows that $\mathbf{g}^7(\Delta)$ intersects Δ, and has the shape of a *horseshoe map*. Let us notice that $\mathbf{g}^7(\Delta)$ has *the shape of a horseshoe*. It can be demonstratedthat Δ includes an *invariant Cantor set* Φ on which $\mathbf{g}^7$ is *topologically conjugate* to a full shift on two "symbols" (see Wiggins 1988). Such elements imply that $\mathbf{g}^7$ has:

– A *countable infinity of periodic points* of all possible periods
– An *uncountable infinity of non-periodic points*
– A *dense orbit*, that is a point in Φ whose orbit under $\mathbf{g}^7$ approaches every point in Φ arbitrarily closely.

1.34.11 Shilnikov Phenomenon: Homoclinic Orbit in $\mathbb{R}^3$

Let us consider a three-dimensional phase flow of the following autonomous equation $\dot{x} = f(x)$, $x \in \mathbb{R}^3$.[63] Then suppose that $x = 0$ is an equilibrium point with a *complex pair of conjugate (unstable) eigenvalues* $\lambda, \bar{\lambda}$ and *real positive eigenvalue* $\gamma > 0$. It is supposed that $Re\lambda < 0, |Re\lambda| < \gamma$ and there exists one *homoclinic orbit*
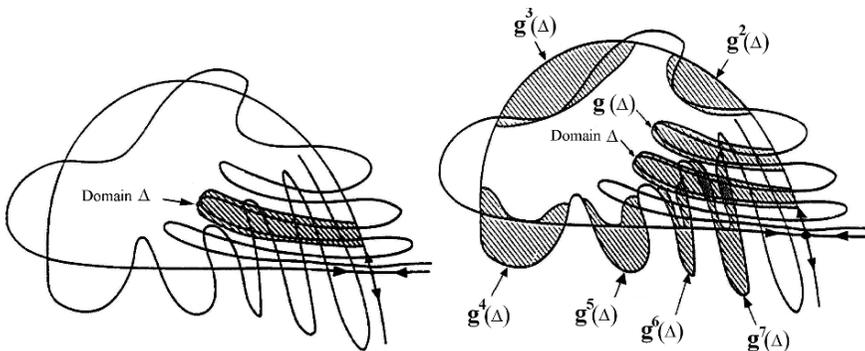
Fig. 1.89 Domain: Δ (*left*). Horseshoe map: $\mathbf{g}^\alpha(\Delta)$ (*right*)

[63] See for an introduction Wiggins (1988), and also Glendinning and Sparrow (1984), Arneado et al. (1985).

coming from $x = 0$ for $t \to -\infty$ which depicts a spiral towards the origin as $t \to +\infty$. Figure 1.90 describes a homoclinic orbit coming from the origin in $\mathbb{R}^3$ and spiralling back to it. Due to the existence of the homoclinic orbit, it is interesting to define a Poincaré (return) map in its neighborhood. Such a map implies *stretching*, *compressing* and *winding,* which is a consequence of the hyperbolicity and horseshoe map. Shilnikov demonstrated that there is a horseshoe structure and then chaos in the return map defined near the homoclinic orbit. The Rossler map (see Odes demonstrations) reproduces this type of structure. Such Shilnikov phenomena have also been studied in $\mathbb{R}^4$. This topic which consists in showing the existence of a horseshoe map in the flow near a homoclinic orbit have been applied by Devaney to Hamiltonian systems (1976). For such an application, it is necessary to have an orbit homoclinic to an equilibrium point in $\mathbb{R}^4$ with a *complex pair of unstable eigenvalues* or an *energy manifold* in $\mathbb{R}^6$ (see Verhulst and Hoveijn 1992).

The Shilnikov phenomenon was detected in a large number of applications (see, e.g. continuous-time model of inventory business cycles, Gondolfo 1983 and Lorentz 1989b). Shilnikov (1965, 1970) proved the existence of horseshoe-like dynamics in a three-dimensional (continuous-time) dynamical system in the presence of a homoclinic orbit. We have expressed the Shilnikov argument above, from the system described previously: $\dot{x} = f(x)$ with $x \in \mathbb{R}^3$ and with *complex pair of conjugate unstable eigenvalues*[64] $\lambda, \bar{\lambda}$ (with $Re\lambda < 0, |Re\lambda| < \gamma$) and with a *real positive eigenvalue* $\gamma > 0$. If we explicitly write the pair of complex conjugate eigenvalues as follows $a \pm ib$, (where a corresponds to $Re\lambda$ real part of the complex number), then we can write the theorem proved by Guckenheimer and Holmes (1983, p. 319) as follows:

Theorem 1.19 (Countable set of horseshoes). *If $|a| < \gamma$, then the flow ϕ_t, associated with the system $\dot{x} = f(x)$ ($x \in \mathbb{R}^3$) can be perturbated to ϕ_t', has a homoclinic orbit ξ' near ξ, and the return map of ξ' for ϕ_t' has a countable set of horseshoes.*

The theorem shows that the systems, for which the stated conditions hold, possess invariant chaotic sets. In addition, the *Smale–Birkhoff homoclinic theorem (see a previous section)* establishes the relation between *homoclinic* and *horseshoe-like invariant sets*. We know that this theorem states that the existence of transversal homoclinic orbits of a diffeomorphism implies the existence of *horseshoe-like* invariant set. The transversal homoclinic orbits mean the transversal intersections of stable and unstable manifolds of a hyperbolic fixed point.

Fig. 1.90 Shilnikov bifurcation and homoclinic orbit in $\mathbb{R}^3$

---

[64] With $Re\lambda < 0, |Re\lambda| < \gamma$.

1.35 Transitions and Routes to Chaos

We know that there exist different kinds of transition to reach chaos known *as canonical:*

- *Period-doubling* in unidimensional or multidimensional models by means of control parameter(s) change (SDP: sensitive dependence on parameters) (see Schuster 1989, p. 63). In the case of two control parameters, the parameter(s) change is related to the *codimension* notion.
- *Intermittency* (Bergé et al. 1987), i.e. explosion.
- *Saddle connection* (or *Blue sky catastrophe*) closely related, in theory and in practice, to *"transverse homoclinic orbits"* (*obviously including the* "Smale-horseshoe approach").

  Quasiperiodic route to chaos, as described in a previous section, is a fundamental intermediate step in the transition to chaos (see Ruelle and Takens 1971; Landau 1944; Newhouse, Ruelle and Takens 1978; Lorentz 1963). This approach dates back to Landau's work in Physics. Then, Ruelle–Takens have considered (unlike Landau) that a small number of bifurcations was sufficient to produce turbulent phenomena and chaos.

The first canonical transition type and the Ruelle–Takens route to chaos have already previously been outlined in other sections, then we are particularly interested in the intermittency and in the saddle-connection on which we will return.

1.35.1 Transition to Chaos Through Intermittency

It is possible to write that an *intermittent regime* is *a predictive criterion of disorder of a system*, i.e. a transition from a stable state towards chaos. A signal is called intermittent (for example periodic) when the signal has rare variations and generally of very high amplitude. These *rare fluctuations are localized in time and space*. As a preliminary, let us simply approach the concept of intermittency by means of low-dimension systems depending on a single parameter. In such cases, we know that there exists a critical value of α noted α_c before which the system has a stable limit cycle (i.e. stable fixed-points), and after which the behavior of the system is regular. When $\alpha > \alpha_c$, and in particular if α is still close to α_c the *oscillations* continue but they are *interrupted* from time to time by a different *irregular type of behavior* whose characteristics depend on α and whose *average frequency* depends on it. This *occurrence frequency* decreases when we approach α_c. And conversely. when it is far from α_c, then the *"memory" of oscillations disappears* little by little until being lost completely. Topologically, in terms of local bifurcations, it is possible to distinguish three different types of intermittencies, related to the loss of stability of "local" (i.e. localized or small) attractors, which are:

(1) The saddle-node bifurcation
(2) The subcritical Hopf or Neimark bifurcation
(3) The subcritical flip bifurcation

When there is destruction of a "local" attractor by a local bifurcation, this does not always imply transition to chaos. Indeed, the global (or larger) attractor can preserve its stability. At the beginning of this section, we chose to start the topic by means of low-dimension systems. So let us consider the typical case of the *unidimensional logistic equation* depending on a single parameter, which is written as follows:

$$x_{n+1} = f(x_n) = \alpha x_n (1 - x_n). \tag{1.358}$$

We know that the critical value of α is $\alpha_c \approx 1 + \sqrt{8}$. In this example, the involved type of intermittency is the first one, i.e. the saddle-node bifurcation. Roughly, before α_c, the system has a stable (quasi-)periodic behavior with one or multiple fixed-points, the behavior of the system is regular; then when $\alpha > \alpha_c$, and in particular if α is close to α_c, the oscillations continue but they are interrupted sporadically by a different irregular type of behavior. More exactly, when $\alpha = \alpha_c \approx 1 + \sqrt{8}$, then the logistic equation shows a period 3-cycle; and going further in the window, the equation shows flip bifurcations which lead to period 3^n cycles. For values of α greater than α_c, but close to α_c, the dynamics is regular, i.e. 3-periodic. When $\alpha < \alpha_c$ but still near α_c, the motion is regular and almost periodic but interrupted sporadically by explosion of chaos. Thus, the iterates of the map near the 3-cycle and the duration of regular behavior is an inverse function of the distance $|\alpha_c - \alpha|$. Such a phenomenon can be understood by considering the fixed point of the map $f^3(x_n)$ which corresponds to the 3-cycle (see Fig. 1.91b).

For $\alpha > \alpha_c$, the map $f^3(x_n)$ has one stable fixed point, and for $\alpha < \alpha_c$, the map $f^3(x_n)$ does not have fixed point. For $\alpha = \alpha_c$, the *fixed point is destroyed*, and through a fold bifurcation, *the stable and unstable fixed points fuses and disappear*. Moreover, for $\alpha < \alpha_c$ (see Fig. 1.91a) we can observe in the neighborhood *where the fixed point was, that there is a slowdown of the motion*. And after a certain number
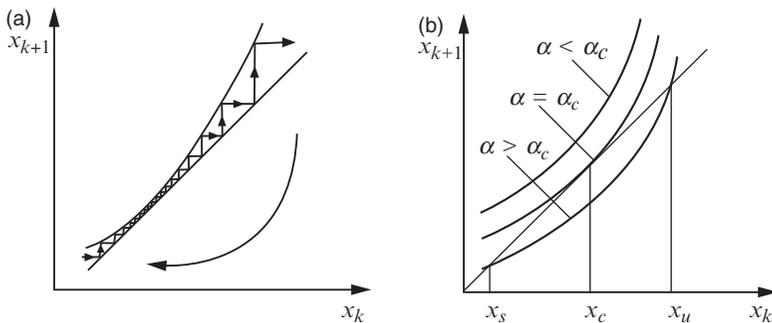
Fig. 1.91 (a) $\alpha < \alpha_c$ (b) Basic three cases

of iterations the system leaves the neighborhood and wanders away irregularly until it is reinjected in the channel between the map curve and the diagonal, etc.
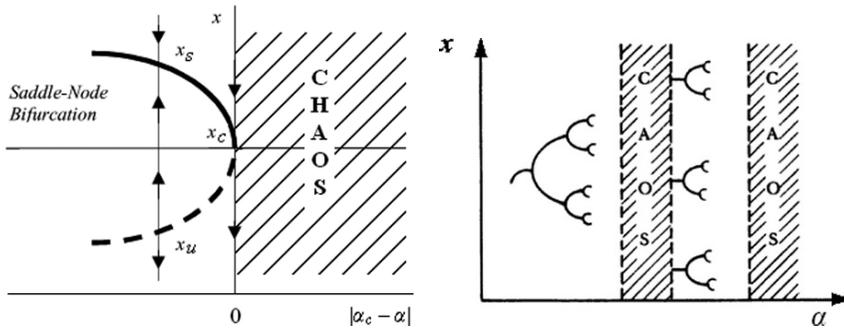
At this stage, we leave the preceding approach which used the logistic model (and the low dimension systems) to extend it. Besides, it is possible to say that an *intermittent regime* is *a predictive criterion of disorder of a system*, i.e. a transition from a stable state towards chaos. A signal is called intermittent (for example periodic) when the signal has rare variations and generally of very high amplitude. These *rare fluctuations are localized in time and space*. This pseudo-definition describes phenomena which *escape* and are *not explained by statistical and traditional econometric approaches* and in particular by the probability distributions. In Nature, intermittencies are observable, in particular in the fluid flow phenomena or meteorology for example. One of the approaches of this intermittency notion is based on the *Floquet theory* (see "Floquet theory" section). *This analysis approaches linear instability of the limit cycle* which helps to explain:

- The increase in the fluctuations of a dynamics from a periodic mode, and by contrast,
- The *return to a periodic mode* after *intermittent fluctuations* by means of the process *re-laminarization* or *re-injection*.

Commonly, three types of linear instability of periodic trajectories are evoked. They correspond to the three manners that the eigenvalues of a system have to cross the unit circle in the complex plane. Or more exactly, they correspond to the manners that the Floquet multipliers (e^{TR}) have to cross the unit circle. Since the multiplier is the result of the linearization of a system, $\dot{x} = f(x)$ around the periodic solution (see Floquet theory section). We linearize by means of $y = x - \bar{x}$, where $\bar{x}$ are the solutions of the system. Then we write $\dot{y} = \partial \bar{x}/\partial t (x - \bar{x})$. After that, we study the *stability* of solutions of the linearized system, *by analyzing the eigenvalues* of the matrix e^{TR} which is called the *Floquet characteristic Multiplier*. And it is reminded that the solutions of the new linearized system are written: $y = e^{TR} y_0$, where R is the matrix which determines the system whose eigenvalues are studied to analyze its stability. We are thus led back to the analysis of eigenvalues behavior in the unit circle. Some models describe better than others this phenomenon of intermittent modes; it is for example, the Lorenz model or the Rayleigh–Bénard model which
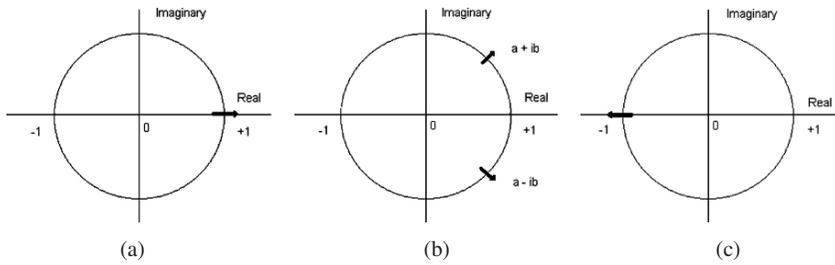
Fig. 1.92 (**a**) Type-1 intermittency. (**b**) Type-2 intermittency. (**c**) Type-3 intermittency

is a fluid dynamics model. The quality of the description depends on the position occupied by the intermittent mode in relation to the other modes of the described model, and in particular the control that we can have on the handling of different modes through the control parameter. It will simply be said in this paragraph that the *theory of the intermittency attempts to study the nonlinearities* and uses the *Taylor expansions* and exposes the *different processes of relaminarization*. We will provide very briefly the three categories of intermittency:

- Type 1 intermittency: The *eigenvalue* of the Floquet matrix *leaves the unit circle of the complex plane by* +1. It is usually said for this case that the intermittent transition occurs between two other regular modes, for example, from a periodic mode to a quasiperiodic mode (Fig. 1.92a).
- Type 2 intermittency: The *eigenvalue* of the Floquet matrix *leaves the unit circle through complex values*, or more exactly, *when two "complex conjugate Floquet multipliers" leave the circle unit of the complex plane*. Then, the unstable fluctuation increases in an exponential way (Fig. 1.92b).
- Type 3 intermittency: In this case, the crossing of the eigenvalue of the unit circle in the complex plane is done by −1, and the associated *bifurcation is subcritical* (Fig. 1.92c).

1.35.1.1 Type-1 Intermittency

Let us consider the linear approximation of the cubic recurrence equation:

$$x_{t+1} = f(x_t) = \beta + x_t + x_t^2. \tag{1.359}$$

We are interested implicitly in the evolution of the gap between two trajectories defined from the primitive (i.e. integral) of this function. The *type-1 intermittency occurs when the eigenvalue of largest modulus of the Floquet matrix crosses the unit circle*[65] *by the value* +1, which implies that $\beta > 0$. When $\beta = 0$, a saddle-node bifurcation appears (see figures below):

---

[65] *Unit circle:* A circle of radius 1, such as the one used to define the functions of trigonometry. The unit circle can also be taken to be the contour in the complex plane defined by $|z| = 1$, where $|z|$ is the *complex modulus*.
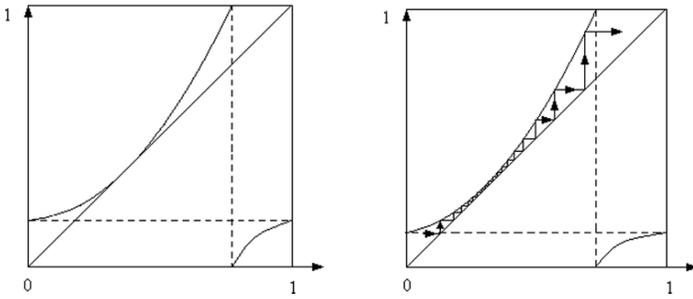
If we choose two different initial values, we observe an increase in amplitude of the gap between the two trajectories symbolized here by the dynamics inside the "channel". It expresses the phase of *laminarization*. However the nonlinear term blocks this divergence which butts against a limiting value, then a phenomenon of *explosion* occurs and *chaotic fluctuations* appear. After this phase, a new phase known as of *relaminarization* occurs characterized by a new cycle whose amplitude increases. This phenomenon is well understood when Fig. 1.93 (left) is analyzed:
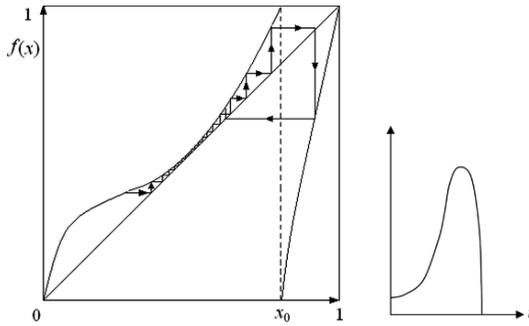
Fig. 1.93 Relaminarization (*left*). Distribution (*right*)

The curve representative of the function $x_{t+1} = f(x_t) = \beta + x_t + x_t^2$ separates into two branches whose presence announces the existence of a discontinuity implying jumps in the dynamics between two phases of periodic fluctuations. The phases of laminarization intervening with random intervals, in order to do predictions, it was necessary to seek to build the statistical distribution (see Fig. 1.93 right) of the average time-delay that elapses between the periodic modes and the turbulent modes. As an example, let us consider $x_{t+1} = f(x_t) = \beta + x_t + x_t^2$, if x_{t+1} and x_t are close, we can take the number of iterations whose index is written n, as a continuous variable which is expressed

$$dx(n)/dn = \beta + x^2 \tag{1.360}$$

whose generic solution is $x(n) = \beta^{1/2}\tan[\beta^{1/2}(n)]$. Then it is proved that the average duration of laminar phases diverges at the rate $\beta^{-1/2}$. The theoretic shape of the distribution is shown in Fig. 1.93 (right).

1.35.1.2 Type-3 Intermittency

In this case, the eigenvalue crossing of the unit circle is done in -1, and the associated *bifurcation is subcritical*. Let us consider

$$x_{t+1} = f(x_t) = -(1+\beta)x_t + \lambda x_t^2 + \delta x_t^3, \tag{1.361}$$

where λ, δ are parameters. We write the following iterations:

$$x_{t+2} = (1+2\beta)x_t + \lambda' x_t^2 + \delta' x_t^3, \quad \delta' = -2(\lambda^2 + \delta) \text{ and } \lambda' < \delta'. \tag{1.362}$$

To depict the type-3 intermittency, it is also possible to write the equation above with $\lambda' = 0$, which becomes $x_{t+2} = (1+2\beta)x_t + \delta' x_t^3$, $\delta' = -2(\lambda^2 + \delta)$ and $0 = \lambda' < \delta'$. In this case, like in the preceding case a saddle-node bifurcation appears at $\delta = 0$. Similar laminarization phenomena can be highlighted and the associated *bifurcation is subcritical*, the nonlinear terms reinforce the instability of the linear term.

1.35.1.3 Type-2 Intermittency

The *eigenvalue* of the Floquet matrix *leaves the unit circle by the two conjugate complex values*. Or more exactly, *when two "complex conjugate Floquet multipliers" leave the unit circle of the complex plane*. Then the unstable fluctuation increases in an exponential way. The complex eigenvalues $(a \pm ib)$ can also be written in polar coordinates as follows:

$$\xi_{1,2} = \rho e^{\pm i\theta}. \tag{1.363}$$

Let us suppose that for a given value of the parameter $\beta = \beta_c$, a Hopf bifurcation occurs. In this case the modulus is unitary. At each iteration, the Floquet matrix makes rotate each point of an angle of θ. Consider the complex number $z = a \pm ib$ and the equation

$$z_{t+1} = \rho e^{i\theta} z_t + \eta z_t^2 \tag{1.364}$$

Then, it comes: $|z_{t+1}|^2 = \rho^2|z_t|^2 + |\eta|^2|z_t|^4 + \rho|z_t^2|(e^{i\theta}\bar{z}_t\bar{\eta} + e^{-i\theta}z_t\eta)$ where $\bar{z}_t$ is the complex conjugate of z_t. Then, it is proved after some approximations that the dynamics depends on the element:

$$\theta_{t+1} = \rho\theta_t \left(\frac{Q}{2\rho^2}\theta_t^2 + 1\right) \tag{1.365}$$

with Q depending on η. This relation is of comparable nature than the relation considered in the type-3 intermittency.

1.35.2 Saddle Connections ("Blue Sky Catastrophes") and Reminder About the Stability Boundaries

Unlike the saddle-node or Hopf bifurcations, the saddle-connection bifurcations are *global bifurcations* (and they cannot be detected by analyzing the zeros of the vector field). Unlike in the case of intermittency, where we explained that the phenomenon of intermittency occurs through a local–global bifurcation, so that in such a situation there is a discontinuous change in the local stability properties of a fixed point, associated with global changes in the phase portrait of the dynamics.

Before describing heteroclinic and homoclinic saddle connections, the latter being of course the most important here, as a preliminary, let us point out the significance of a *separatrix*, which is a very useful notion within the framework of the present section. A *separatrix*[66] is phase curve, i.e. an *invariant manifold* which meets *a "hyperbolic fixed point"* (i.e. an *intersection of a stable and an unstable invariant manifold*) or *connects "the unstable and stable manifolds* of a *pair* of *hyperbolic or parabolic fixed points"*. The saddle-connections are more difficult to locate than other types of bifurcations. Usually, the stable separatrix of a saddle is not the unstable separatrix of the same or any other saddle. But this does generally occur for some parameter value in one-parameter family, and when it does, the phase portrait tends to undergo huge changes (some examples are given through the Figs. 1.94 and 1.95 by varying the control parameter). Figure 1.94 depicts a *Heteroclinic saddle connection* between two saddles.

If a stable and unstable separatrix of one saddle coincide, the saddle connection is called homoclinic (see Fig. 1.95 that depict Homoclinic saddle connection at bifurcation, it is possible to observe that the sink remains throughout, thus a limit cycle is created). Figure 1.95 depicts a *Homoclinic saddle connection* at bifurcation.

Usually, the *homoclinic loop* of a *homoclinic saddle connection* has a unique zero of the vector field. Such a zero can be a source or a sink, unless a Hopf bifurcation or a saddle-node bifurcation occurs for the same value of the parameter. If this zero is a source, then it will remain a source for nearby values of any parameters. However the global spiral behavior changes branches; if a *stable separatrix* goes "towards inside" before bifurcation, then an *unstable separatrix* will be present after bifurcation (see Fig. 1.95). Generally, the stable separatrix will stem from the source,
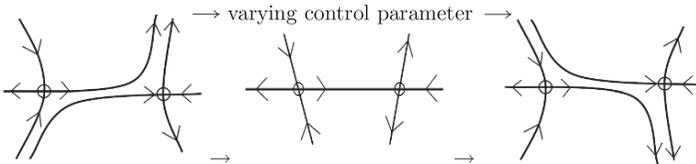
Fig. 1.94 Before bifurcation (*left*), at bifurcation (*center*), after bifurcation (*right*)

---

[66] Moreover, a separatrix indicates a boundary between phase curves with different properties (e.g. separatrix in the equation of motion for the pendulum occurs at the angular momentum where oscillation gives way to rotation).

$\longrightarrow$ varying control parameter $\longrightarrow$
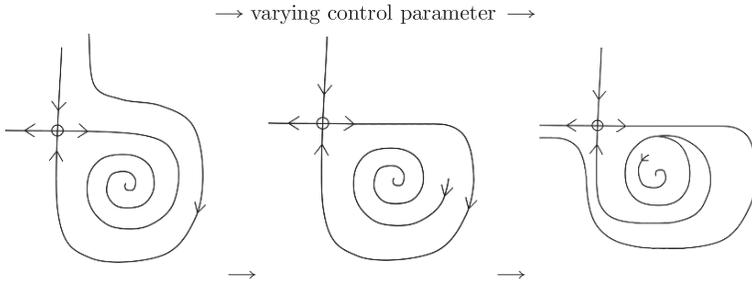
Fig. 1.95 Before bifurcation (*left*), at bifurcation (*center*), after bifurcation (*right*)

$\longrightarrow$ varying control parameter $\longrightarrow$
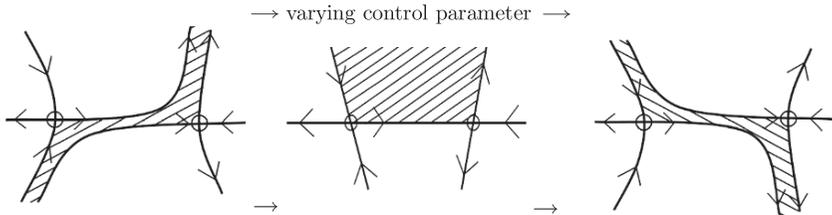
Fig. 1.96 Before bifurcation (*left*), at bifurcation (*center*), after bifurcation (*right*)

$\longrightarrow$ varying control parameter $\longrightarrow$
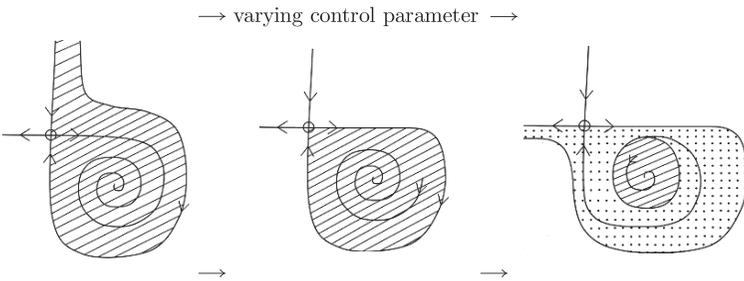
Fig. 1.97 Before bifurcation (*left*), at bifurcation (*center*), after bifurcation (*right*)

however the unstable separatrix cannot lead to a source and must behave differently. Frequently, a limit cycle will attract it, such a limit cycle was born in the homoclinic saddle connection. In such a case, that of a homoclinic saddle connection with a limit cycle, a Hopf bifurcation is frequently present in the neighborhood of the parameter space when the limit cycle disappears.

Obviously, the topic dealt here is the study of dynamical behaviors while changing the value of the parameter. This is equivalent to analyzing the (attracting or repelling) basins and the changes of basins. This analysis is summarized through Figs. 1.96 and 1.97. Figure 1.96 depicts the *change of basins* for a *Heteroclinic saddle connection*.

Figure 1.97 depicts the *change of basins* for a *Homoclinic saddle connection*.

Obviously, the topic of this section is the *Homoclinic saddle connection, in such a framework the occurrence of limit cycles is inescapable*. As we can observe in Fig. 1.98, it is possible to say that a homoclinic saddle connection is a loop
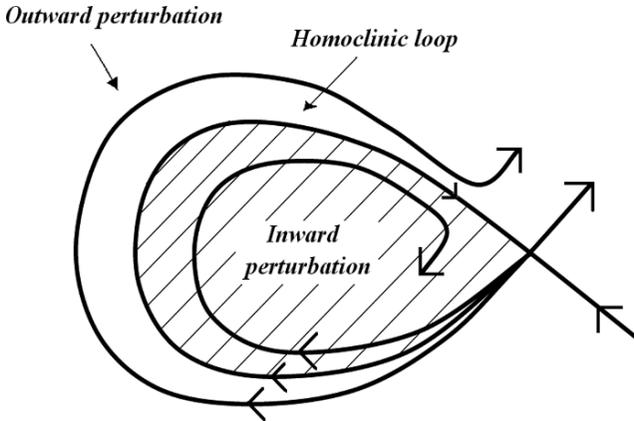
Fig. 1.98 Inward and outward perturbation of a homoclinic orbit

delimiting a zone of the plane which is depicted through *striations* (and called the *inward perturbation* in the figure). For the preparation and presentation of the following theorem, let us consider a *perturbation* f_δ of the equation $\dot{x} = f_\delta(x)$ at the bifurcation, for δ close to δ_0, *this perturbation will be called "inward" if the unstable separatrix, which at bifurcation led back to zero, now goes inside the loop zone, and will be called "outward" otherwise* (see Fig. 1.98). The saddle connection is called attracting when the trajectories go inside the spiral towards the homoclinic loop. Otherwise, the saddle connection is called repelling (see Figs. 1.98 and 1.99).

We can now enunciate the following theorem (Hubbard and West 1995b):

Theorem 1.20 (Homoclinic saddle connection). *Let* $\dot{x} = f_\delta(x)$ *be a family of differential equations depending on a parameter* δ, *and suppose that for some value* δ_0 *of the parameter, the vector field has a zero* x_0 *which is a saddle with a homoclinic saddle connection. Let* $\dot{y} = Ay$ *be the linearization of* $\dot{x} = f_{\delta_0}(x)$ *at* x_0:

1. *If* $trA > 0$, *the saddle connection is repelling.*
   If $trA < 0$, *the saddle connection is attracting.*
   If $trA = 0$, *no conclusion.*
2. *If the saddle connection is repelling, and* δ *is close to* δ_0 *with the perturbation outward, there will be for that value of* δ *a repelling limit cycle.*
3. *If the saddle connection is attracting, and* δ *is close to* δ_0 *with the perturbation inward, there will be for that value of* δ *an attracting limit cycle.*

Figure 1.99 depict attracting and repelling *homoclinic loops*.

Different numerical simulations can be gradually carried out by means of following examples. $\dot{x} = y$, $\dot{y} = x^3 - x + ax^2 - 0.1y$, for $a = 0.1$ or $a = 0.2$. Furthermore, by means of the system $\dot{x} = y$, $\dot{y} = x^3 - x + 0.2x^2 + (bx^2 - 0.2)y$. When $b \simeq 1.4$ *for which there is a homoclinic saddle connection*, it is possible to observe that there is a saddle point for $(x_0, 0)$ where x_0 is a positive solution of the equation $x^2 + 0.2x - 1 = 0$, that means $x_0 = -0.1 + \sqrt{1.01} \simeq 0.9$. So the following
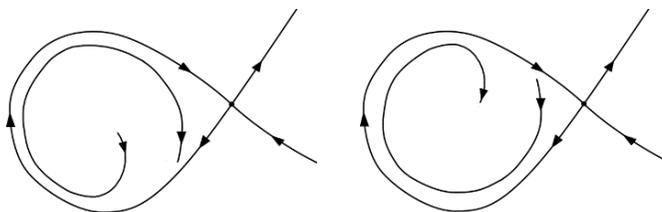
Fig. 1.99 Attracting (*left*), repelling (*right*)

linearization matrix:

$$A = \begin{bmatrix} 0 & 1 \\ 3x_0^2 - 0.4x_0 - 1 + 2bx_0y & bx_0^2 - 0.2 \end{bmatrix} \qquad (1.366)$$

has $trA = b_0^2 - 0.2 \simeq (1.4)(0.9)^2 - 0.2 > 0$. Then the saddle connection is repelling, and repelling limit cycles occur for the *outward perturbations* which appear for $b > b_0$.

Van der Pol oscillator case: The Forced pendulum is the quasi-general schematiza-tion of "the oscillator". These periodic systems are widespread in many scientific fields. They are used to represent an immense quantity of phenomena, e.g. the gravitation of an electron around an atom of crystal plunged in magnetic fields or a biological rhythm, like breathing or the contraction of the cardiac muscle, the alternation of waking and sleep, the breeding cycles of plants, etc. *The key for the analysis of these phenomena is the oscillator*. In particular, the van Pol oscil-lator is the typical modeling of oscillatory motions for these systems, and allows to highlight the behavior changes from periodic motions to chaotic motions. Then, an example of the "route to chaos" within the framework of saddle-connections, i.e. the blue sky catastrophe, can be given *in a two-dimensional flow*, by the analysis of the *van der Pol system* which can be written as follows:

$$\dot{x} = ay + \mu x(b - y^2), \qquad (1.367)$$
$$\dot{y} = -x + \gamma. \qquad (1.368)$$

In such a system, the parameter γ becomes a control criterion. Suppose that γ_c is a critical value, if γ is smaller than the critical value ($\gamma < \gamma_c$), there exist an unstable saddle point and a stable limit cycle. When the value of γ is increased, the limit cycle comes closer to the saddle point until they meet, at $\gamma = \gamma_c$, and they clash. At this stage, the stable and the unstable arm of the saddle coincide with the location of the cycle, so a homoclinic saddle connection occurs and the cycle disappears. When $\gamma > \gamma_c$, the orbits that were in the area of the saddle point wander away and if there exists another distant attractor, they move towards it.

In the saddle connections, the "blue sky catastrophe" can appear at the same time with a *hysteresis phenomenon* that does not exist in intermittency. Such a case can be easily described in the following way, if initially there exists a limit cycle, when

the value of the control parameter is decreased until a critical value, the limit cycle joins the saddle point, so a blue sky catastrophe occurs, and the limit cycle is thus demolished and the dynamics relocates towards the distant stable node. Conversely, when the value of the parameter is increased starting from a critical value, the node is not demolished while at the same time a limit cycle occurs and the node preserves its stability until a fold catastrophe appears, then the node disappears and the dynamics comes back to the limit cycle. The behavior of the system is thus different according to whether we increase or decrease the control parameter; this is a main characteristic of the hysteresis phenomenon.

As we said in the introduction of this section, *the saddle-connection bifurcations correspond to "global bifurcations"*, unlike in the intermittency which occurs through a "local–global bifurcation", so that in such a situation there is a discontinuous change in the local stability properties of a fixed point, associated with global changes in the phase portrait of the dynamics. In the *saddle-connections, in particular about differential systems of dimension* $\geqslant 3$*, the occurrence* of *blue sky catastrophes can imply more complex attractors* which lead to the *creation* of *chaotic attractor by means* of *"collision"* with *saddle-type structures*.[67] A description of these *catastrophic creation or destruction of horseshoe-like chaotic attractor* is given by the works of Abraham and Shaw (1988) to which we will be able to refer for a thorough analysis of this very interesting topic. These works propose stroboscopic sketches of blue sky catastrophe occurrences helping to visualize the phenomena.

1.35.2.1 Example of Blue Sky Catastrophe (Shilnikov–Turaev–Gavrilov) and Reminder About Stability Boundaries

As a preliminary, let us reintroduce in a new way, the concepts of *stability boundaries* of periodic orbits (Andronov and Leontovich 1937; Andronov et al. 1971). The *boundaries of stability regions of periodic orbits* in systems of *differential equations* can be distinguished by the characteristics of the bifurcation. Then two sub-groups (A) and (B) can be highlighted:

(A) The *first group* contains the following codimension-1 bifurcations:

1. *Saddle-node* (fold) *bifurcation*: *two periodic orbits* (one stable and one unstable) merge on the stability boundary and disappear beyond it.
2. *Period-doubling* (flip) *bifurcation*: a *multiplier* of the bifurcating periodic orbit is equal to (-1).
3. *Bifurcation* from a *periodic orbit* to a *two-dimensional invariant torus* (i.e. birth of an *invariant torus*): i.e. "a cycle loses its skin" (see Andronov).
   These cases lead to the analysis of the corresponding fixed-point of the Poincaré map on a cross-section transverse to the periodic orbit.

---

[67] Blue sky catastrophe: The attractor disappears "in the blue sky" after a collision with a saddle-type (unstable) periodic orbit. This is a global bifurcation. Abraham and Marsden (1978); Abraham and Stewart (1986).

(B) The *second group* contains the codimension-1 bifurcations which concern cases in which a *periodic orbit merges with an equilibrium state*:

4. First case is the *Andronov–Hopf bifurcation* in which a *periodic orbit shrinks into an equilibrium state* with *a pair of characteristic exponents* $\pm i\lambda$, (the estimation of the period T of such bifurcating orbit tends to $2\pi/\lambda$).

For the other cases, the *periodic orbit* "*adheres*" *to a homoclinic loop* to an *unstable equilibrium state* which can be:

5. A *saddle* with characteristic exponents in both open (left and right) half-planes,
6. A (*simple*) *saddle-node with one zero real exponent*. (Because of the disappearance of the vector field at the equilibrium state, the period of the bifurcating orbit tends to infinity as it approaches the homoclinic loop, while its perimeter remains of a finite length.)

In addition to this list of six cases, Shilnikov and Turaev (1995a, pp. 596–599) proposed a *new scenario* for the *saddle-node bifurcation* of periodic orbits, which can have the appearance of a stable periodic orbit of an infinitely large perimeter and infinitely long period. Because of such a "device", Abraham calls this case:

7. *A blue sky catastrophe*. Besides, this boundary can (under certain conditions) "separate" the Morse–Smale systems from systems with hyperbolic attractors in the parameter space (see Shilnikov L.P. and Turaev D.V. 1995b).

In the multi-dimensional framework, for the generic one-parameter families there exist seven such stability boundaries. There are two types that are conditioned by the existence or the non-existence of the periodic orbit at the critical moment. In the first case, the intersection of the periodic orbit with a local cross-section exhibits the fixed point of the Poincaré map, then it is possible to simplify the study by analyzing *the exit* of *multipliers* (of the fixed point) *from the unit circle*.

Bifurcation	Multiplier of the bifurcating periodic orbit	Comment
1. Saddle-Node (fold): (similar to the two-dimensional case)	$(+1)$	Periodic orbit *disappears*
2. Period-doubling (flip):	(-1)	Periodic orbit *does not disappear*, but *loses its stability*
3. Birth of a torus: (a stable two-dimensional invariant torus is born)	$(a \pm ib)$: pair of complex-conjugate multipliers	Periodic orbit *remains but loses its skin*
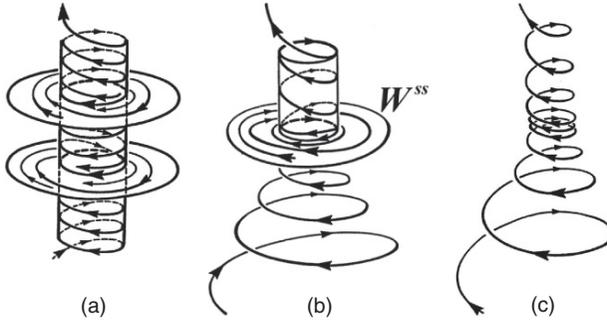
Fig. 1.100 Saddle-node bifurcation: (**a**) $\delta < 0$, (**b**) $\delta = 0$, (**c**) $\delta > 0$

In the first case, the single multiplier of the periodic orbit becomes equal to $(+1)$, this is the saddle-node bifurcation (fold). The second codimension-1 bifurcation is the flip or period-doubling, the single multiplier is equal to (-1), here the periodic orbit does not disappear after this bifurcation but loses its stability. In the third case, a pair of complex-conjugate multipliers crosses the unit circle and the periodic orbit remains but loses its skin, then a stable two-dimensional invariant torus is born.

Basically, such dynamics are constructed from dynamical systems which exhibit sustainable self-oscillations, i.e. a stable periodic orbit. Then, the behaviors of periodic orbits are analyzed when the parameters are modified.

Next, let us consider a one-parameter family X_δ (δ : parameter) of dynamical systems with an exponentially stable periodic orbit for δ. This periodic orbit remains stable within some interval of parameter values. Such a device is the starting point for the study of codimension-1 stability boundaries (for dynamical systems) whose types were enumerated above. At this stage, let us focus our attention on the *saddle-node bifurcation* and its internal behavior for various values of the parameter δ. The three emerging behaviors are as follows: (1) $\delta < 0$, in such a case there exist *two periodic orbits* (*stable and saddle*), (2) $\delta = 0$, here the *periodic orbits merge into a saddle-node* and its *Strong Stable Manifold* W^{ss} splits the neighborhood into the node region localized below W^{ss} and the saddle region localized above W^{ss} (see Fig. 1.100b). The unstable manifold is the part of the *center manifold* which is localized in the saddle region. (3) $\delta > 0$, in this case the saddle-node disappears (the drifting time throughout its neighborhood tends to $1/\sqrt{\delta}$). See Fig. 1.100 of the *saddle-node bifurcation*.

After this brief parenthesis about saddle-node bifurcation, we are going to geometrically depict a three-dimensional system in which there exists a blue sky catastrophe, as it was developed by Shilnikov and Turaev (see paper previously mentioned). It is assumed that there exists a *saddle-node periodic orbit* L^* whose unstable manifold $W^u_{L^*}$ comes back to L^* when t tends to $+\infty$ (see Fig. 1.101c).[68]

---

[68] *Remark:* the closure of the unstable manifold is not a Hausdorff manifold.
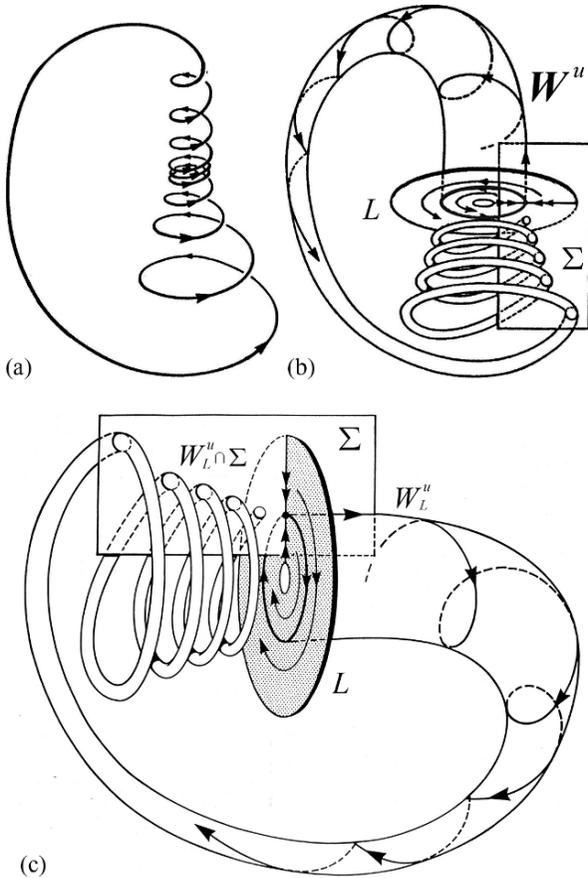
Fig. 1.101 Scenario of a "blue sky catastrophe" (ref. to Shilnikov A, Shilnikov L, Turaev D)

Figure 1.101b depicts the global shape of the manifold W^u which has been built to obtain *a blue sky catastrophe*. The intersection of W^u with the local section Σ in the node zone is a countable set of circles which accumulates on $\Sigma \cap L$.

Figure 1.101c depicts a scenario of a "blue sky catastrophe". The "blue sky" stability boundary is an "open subset" of a codimension-1 bifurcational surface corresponding to the *existence of a saddle-node periodic orbit*. This open subset is characterized by conditions defined by the geometry of the unstable manifold of the saddle-node (see Fig. 1.101a,b) and this open subset is also characterized by quantitative restrictions, i.e. the Poincaré map above (Fig. 1.101b,c) must be a contraction. *The stable periodic orbit L (noted L_δ before) whose period and length tend to the infinite when approaching the moment of bifurcation is born when the saddle-node orbit disappears.*
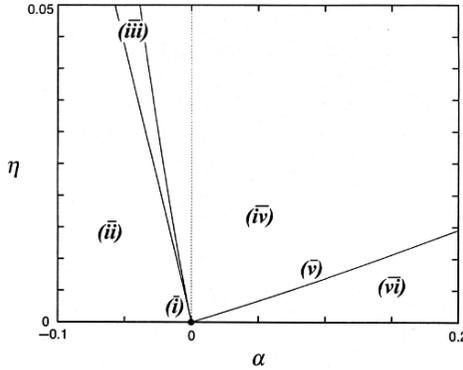
Fig. 1.102 Bifurcation diagram in the (α, η)-parameter plane

Example of a blue sky catastrophe (Gavrilov N. and Shilnikov A. 2000): Let us consider the following system:

$$\dot{x} = x(2 + \alpha - \beta(x^2 + y^2) + z^2 + y^2 + 2y \equiv A, \tag{1.369}$$
$$\dot{y} = -z^3 - (1 + y)(z^2 + y^2 + 2y) - 4x + \alpha y \equiv B, \tag{1.370}$$
$$\dot{z} = (1 + y)z^2 + x^2 - \eta \equiv C. \tag{1.371}$$

α, β, η are parameters and β is preset at 10 ($\beta = 10$). In the singular case $\alpha = \eta = 0$, the system describes a closed curve (see Fig. 1.103a) on which there are *two equilibrium states*; the first one $E'(0, -2, 0)$ is a saddle-node with one zero characteristic exponent $\lambda_1 = 0$ and two negative characteristic exponents extracted from $\lambda^2 + 40\lambda + 68 = 0$, the second equilibrium state $E(0, 0, 0)$ has also $\lambda_1 = 0$ and two imaginary characteristic exponents $\lambda_{2,3} = \pm 2i$ (see Fig. 1.103a). Beyond this preliminary singular case, let us briefly depict the *general bifurcating behavior of the system*, i.e. when the parameters (α, η) are modified in the clock-wise direction around the origin in the (α, η)-parameter plane. It is possible to show the *bifurcation diagram* for (α, η)-parameter as follows (for $\beta = 10$) (Fig. 1.102).

When η gets positive, the saddle-node E' disappears and the equilibrium state E splits in two equilibria E_1 and E_2. Inside the zone $(\bar{\bar{ii}})$, the point E_1 is stable and E_2 is a saddle-focus whose one-dimensional separatrices converge to E_1 when t tends to $+\infty$ (see Fig. 1.103b). On the frontier (i.e. bifurcation) between the zones $(\bar{\bar{ii}})$ and $(\bar{\bar{iii}})$, the point E_1 loses its stability by means of a supercritical Andronov–Hopf bifurcation and E_1 comes to be a saddle-focus. Moreover, the two unstable separatrices of the saddle-focus E_2 tend to a new stable periodic orbit L_1 (see Fig. 1.103c).

Figure 1.103 describe the occurrence of a blue-sky catastrophe. There is also an Andronov–Hopf bifurcation for the equilibrium state E_2 on the frontier between the zones $(\bar{\bar{iii}})$ and $(\bar{\bar{iv}})$ and E_2 comes to be a repelling point in the zone $(\bar{\bar{iv}})$. The unstable manifold of the new saddle periodic orbit L_2 continues to tend to L_1 (see
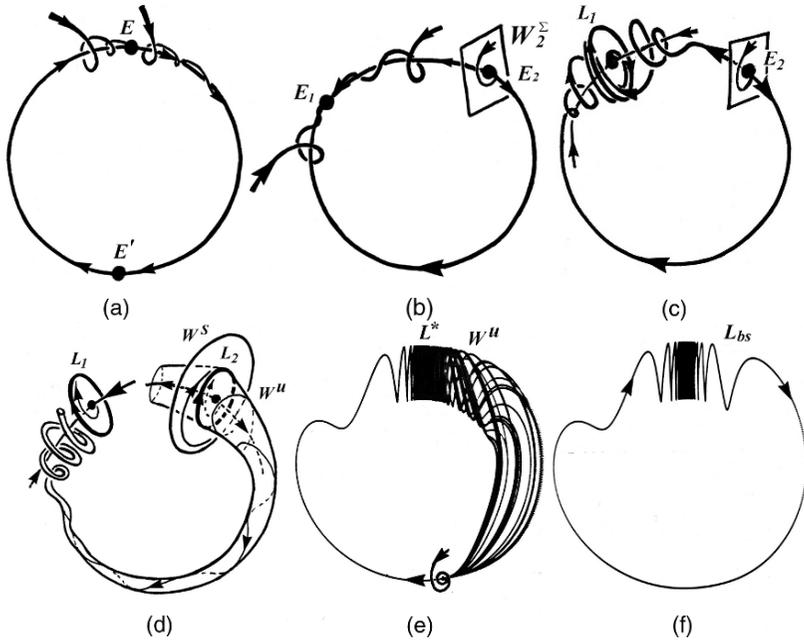
Fig. 1.103 Mechanizm of the blue-sky catastrophe (ref. to Gavrilov N. and Shilnikov A. 2000)

Fig. 1.103d). On the last frontier between the zones $(\bar{v})$ and $(\overline{vi})$, the two periodic orbits merge, constituting thus a saddle-node cycle L^* whose unstable manifold is bi-asymptotic to L^* when $t \to \pm\infty$ (see Fig. 1.103e). The cycle L disappears in the region $(\overline{vi})$, but the local stability of the system comes to be a new Unique Stable Large-Amplitude Periodic Orbit noted L_{bs} which is *not homotopic*[69] to one or the other of former cycles (see Fig. 1.103f).

Examples of the disappearance of the saddle-node (see Shilnikov et al. 2003): The topic here is to provide examples of global saddle-node bifurcations coming from (continuous-time) dynamical systems. Let us consider a one parameter family of C^2-smooth $(n+2)$-dimensional dynamical systems depending smoothly on $\gamma \in \gamma(-\gamma_0; \gamma_0)$; furthermore suppose that the following hypotheses are satisfied:

1. For $\gamma = 0$ there exists a *periodic orbit* L_0 of *saddle-node type* (i.e. multipliers $= (+1)$).

---

[69] *Homotopic*: Two mathematical objects are said to be homotopic if one can be continuously deformed into the other (e.g. the real line is homotopic to a single point, as is any tree. The *circle* is not contractible but is *homotopic to a solid torus*. The basic version of homotopy is between maps). *Homotopy:* Between two mappings of the same topological spaces, a continuous function representing how (in a step-by-step fashion) the image of one mapping can de continuously deformed onto the image of the other. *Homotopy group:* Associated to a topological space X, the groups appearing for each positive integer n, which reflect the number of different ways (up to homotopy) than an n-dimensional sphere may be mapped to X. *Homotopy theory:* The study of the topological structure of a space by examining the algebraic properties of its various homotopy groups.

2. Any trajectory in the unstable manifold W^u of the periodic orbit L_0 tends to L_0 when $t \to \infty$ and $W^u \cap W^{ss} = \emptyset$ (W^{ss} is a strong stable manifold).
3. The family of dynamical systems is *transverse* to the bifurcational set of systems with a periodic orbit of the saddle-node type. Then, when γ changes the saddle-node bifurcates into a saddle and a node when $\gamma < 0$ (but does not exist when $\gamma > 0$).

Then let us introduce in a small neighborhood of the orbit L_0 the following generic system:

$$\dot{x} = \gamma + x^2(1 + p(x, \theta, \gamma)), \tag{1.372}$$
$$\dot{y} = (A(\gamma) + q(x, \theta, \gamma))y, \tag{1.373}$$
$$\dot{\theta} = 1. \tag{1.374}$$

θ is an *angular variable*. The *eigenvalues* of the *matrix A* are localized in the left open half-plane. θ is defined modulo 1, this means that the points $(x, y, \theta) = 0$ and $(x, \sigma y, \theta = 1) = 0$ (with σ in $\mathbb{R}^n$). p is a one-periodic function in θ, and q is a biperiodic function. Furthermore, $p(0,0,0) = 0$ and $q(0,0,0) = 0$. The coordinates are used so that p becomes independent of θ for $\gamma = 0$. The saddle-node periodic orbit is L_0 is given by the equation $(x, y) = (0,0)$ for $\gamma = 0$. And its strong stable manifold W^{ss} is given locally by the equation $x = 0$. The manifold W^{ss} separates the saddle region, where $x > 0$ of L_0 from the node where $x < 0$. The invariant *center manifold is* $y = 0$. There exist two periodic orbits when $\gamma < 0$, one is stable, it is L_1, and the other is of the saddle type L_2, these two orbits merge for $\gamma < 0$ and become L_0. The are no periodic orbits when $\gamma > 0$ and a trajectory leaves the small neighborhood of the former saddle-node.

For $\gamma = 0$, the x-coordinate increases in a monotonous way. It slowly tends to zero in the zone $x < 0$ at rate $1/t$. Since the y-component decreases exponentially, any trajectory in the node zone tends to L_0 when $t \to +\infty$ in a tangent way to the cylinder given by $y = 0$. When $t \to -\infty$ in the saddle zone $x(t) \to 0$, and as y increases in an exponential way when t decreases, all the trajectories converge to the saddle node L_0 when $t \to -\infty$, this means that its unstable manifold W^u is the *cylinder* $\{y = 0, x \geqslant 0\}$. Any trajectory originating from $W^u \backslash \{L_0\}$ (when t increases) leaves a small neighborhood of the saddle-node. But due to the hypothesis 2, the trajectory must return to the node zone when $t \to +\infty$ (and thus converges to L_0 tangentially to $y = 0$). Then, it is possible to select a $\zeta > 0$ such that the unstable manifold W^u crosses the section $\Sigma_0 : \{x = -\zeta\}$, and the intersection between the unstable manifold and the section $W^u \cap \Sigma_0 = \bar{l}$ is a close curve. Moreover it is supposed that the median line $l_0 : \{y = 0\}$ in the section Σ_0 is oriented in direction of the increase of θ, and thus is the median line $l_1 : \{y = 0\}$ of the section $\Sigma_0 : \{x = +\zeta\}$. Due to $l_1 = W^u \cap \Sigma_1$, the curve $\bar{l}$ is an image of the curve l_1 through the map of trajectories, then the orientation on l_1 defines the orientation on $\bar{l}$. Considering the orientation, the curve $\bar{l}$ becomes homotopic to ml_0 where $m \in \mathbb{Z}$. In the case $m = 0$ (i.e. a blue-sky catastrophe), the manifold W^u is shown in Fig. 1.104a.
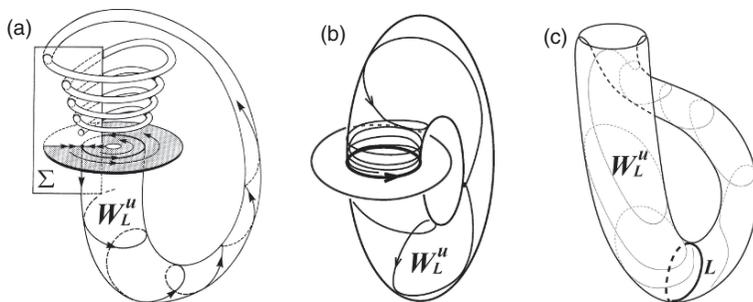
Fig. 1.104 (**a**) $m = 0$, (**b**) $m = 1$, (**c**) $m = -1$ (ref. to Shilnikov et al. 2003)

Fig. 1.105 $m = 2$
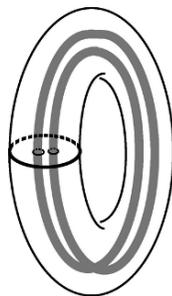
When $m = 1$, the manifold W^u is *homeomorphic to a two-dimensional torus* (see Fig. 1.104b $m = 1$). When $m = -1$, the manifold W^u is *homeomorphic to a "Klein bottle"* (see Fig. 1.104c $m = -1$). When $|m| \geqslant 2$, the manifold W^u is a $|m|$-branched manifold. And in the case $m = 2$, the solid torus is squeezed, doubly expanded, twisted and inserted back into the original and so on, giving thus a *solenoid* with a specific shape, which is sometimes called the Wietorus–van Danzig solenoid (see Fig. 1.105, $m = 2$).

1.36 Temporal Correlation: Periodicity, Quasiperiodicity, Aperiodicity

When the behavior of a system shows a chaotic regime, its *power spectrum* contains a *continuous part*, this means that the evolution of this variable over time is done in a disordered way. In order to measure this level of disorder, let us consider a function which measures the *rate of disorder* and which attempts to quantify the resemblance of the value taken by the variable at a given moment with the value taken by the variable at the next moment, separated from the previous moment by a *step* τ. This relatively simple function is obtained by making the average of the products of the variables at the moment t and the moment $t + \tau$. The temporal correlation is
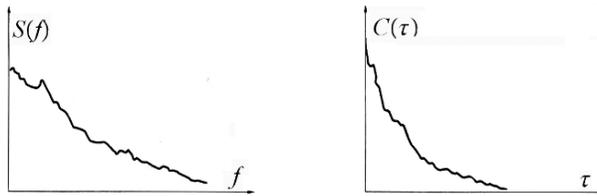
Fig. 1.106 Spectrum and autocorrelation function for an aperiodic regime

given by

$$C(\tau) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t)x(t + \tau)dt \tag{1.375}$$

or by using a scalar product: $C(\tau) = \langle x(t) \cdot x(t + \tau) \rangle$. This function is called the temporal autocorrelation function and expresses the *similarity* of the variable (with itself) in the course of time. Thus, the function is built by varying the interval τ (a delay variable). This construction can be related to the *power spectrum* whose construction is based on the *Fourier transform* and on the *Wiener–Kintchine theorem*. $C(\tau)$ is the Fourier transform of the power spectrum. Consequently, if $x(t)$ is *constant*, *periodic* or *quasi-periodic*, then the function $C(\tau)$ of temporal correlation will not tend to zero when τ will tend to ∞, because in such a case the Fourier spectrum is formed of distinct lines (expressing the power spectrum at each frequency). The periodic or quasi-periodic signals preserve their internal similarity in the course of time. This signifies that the behavior of the system is predictable, since its knowledge during a sufficient period of time allows to know (by simple comparison) what it will be at any ulterior instant. That means that it is possible to do forecasts. However, for a chaotic regime where the spectrum contains a continuous part, $C(\tau)$ generally tends towards 0 when τ increases (see Fig. 1.106).

Thus, the *temporal similarity of the signal (with itself) progressively decreases in the course of time* and even disappears completely for moments sufficiently distant in time. Thus, the knowledge of $x(t)$ during a period of time (as long as it is) does not allow to predict the future behavior of $x(t)$. Consequently, *the chaotic regimes are unpredictable by loss or by internal loss of similarity to their process*. The *loss of similarity* can be connected with the notion *loss of memory* of initial conditions of a system (although different). The loss of memory is not a total absence of sensitive dependence on initial conditions of a system. It is not either a progressive impoverishment of "information". It is in fact an *enrichment of information*, indeed the variety and the multiplicity of initial states of a system come to feed its attractor which generate a multiplicity of unpredictable final states. *The sensitive dependence on initial conditions* that expresses the increase of the spacing between two initially very close trajectories, *increases the uncertainty of the system over time*.

1.37 Power Spectral Density

1.37.1 Characterization of Dynamical Systems

An important objective is to try to characterize the type of subjacent dynamics that has engendered a time series that we study. It can be a more or less complex oscillation, but of defined period, or the superposition of several (different) oscillations, or another type of dynamics. Understanding a time series requires to know the basic period and amplitude. The periodicities are identified through spectral analysis, whether the subjacent model is known or not. This is a very important phenomenon that it is absolutely necessary to underline. We'll discuss the principle in the section concerning the Fourier transform in greater detail, however we can say that some dynamics result from a *superposition of oscillations* of different amplitudes and periods, but also from harmonics. In this case, the regime is described as quasi-periodic. Their associated *attractor*, of a higher order than the limit-cycle, is a torus. For *chaotic but deterministic regimes*, i.e. *for dynamics represented by a limited number of nonlinear differential equations*, the attractor identified in the phase-space is a *strange attractor*.

The approach by spectral analysis results from the theory of Waves; it uses the function called "power spectrum" that shows that the power of a signal (oscillation or wave) *is proportional to the square of its amplitude*. This concept can be connected with the notion of *Energy* of a signal mentioned in Part III.

Like the Poincaré section method that allows to identify the characteristics of dynamics by studying the shape of the section; similarly, spectral analysis also allows to identify by the graphic tool, the characteristics of a periodic, quasi-periodic or aperiodic dynamics. Spectral analysis is a principle resulting from two concepts, which are the Fourier transform and the autocorrelation function of a signal, that is an extraction of the Wiener–Khintchine theorem. Hereafter, we are going to describe the Wiener–Khintchine theorem. The autocorrelation function of the signal x_j is

$$\psi_m = \frac{1}{n} \sum_{j=1}^{n} x_j x_{j+m} \tag{1.376}$$

with Δt as "step", thus we can write

$$\psi_m = \psi(m \cdot \Delta t). \tag{1.377}$$

ψ_m corresponds to the average of the product of values of the signal at a moment t, then at a posterior moment $m \cdot \Delta t$. This technique makes it possible to know if and for how much time the value at time t of the signal depends on what it was before. ψ_m measures the resemblance of the signal to itself over time. If the studied series is periodic, the periodicity finds in the writing of ψ_m:

$$\psi_m = \psi_{m+n} \tag{1.378}$$

if we apply the inverse Fourier transform, we have

$$\psi_m = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k,k'=1}^{n} \hat{x}_k \cdot \hat{x}_{k'} \cdot \exp\left[i\frac{2\pi}{n}\left(jk+(j+m)k'\right)\right] \tag{1.379}$$

if we write $\hat{x}_{k'}^* = \hat{x}_{n-k}$ and make the sum on the indices j and k' we write

$$\psi_m = \frac{1}{n} \sum_{j=1}^{n} |\hat{x}_k|^2 \cos\left(\frac{2\pi mk}{n}\right). \tag{1.380}$$

This equality clearly tells us that the autocorrelation function corresponds to a Fourier transform of the variable $|\hat{x}_k|^2$. If we seek to go up on $|\hat{x}_k|^2$, we can write a function S_k as follows:

$$S_k = \sum_{m=1}^{n} \psi_m \cos\left(\frac{2\pi mk}{n}\right) \tag{1.381}$$

by replacing ψ_m:

$$S_k = \sum_{l=1}^{n} |\hat{x}_l|^2 \frac{1}{n} \sum_{m=1}^{n} \cos\left(\frac{2\pi mk}{n}\right) \cos\left(\frac{2\pi ml}{n}\right). \tag{1.382}$$

However, in connection with the Moivre formula, it is known that:

$$\cos\left(\frac{2\pi mp}{n}\right) = \frac{1}{2}\left[\exp\left(i\frac{2\pi mp}{n}\right) + \exp\left(-i\frac{2\pi mp}{n}\right)\right]. \tag{1.383}$$

If we carry out a summation on m, we obtain the following series with n terms:

$$\frac{1}{n} \sum_{m=1}^{n} \cos\left(\frac{2\pi mk}{n}\right) \cos\left(\frac{2\pi ml}{n}\right) = \frac{1}{4}\left[\delta_{k+l}^{(n)} + \delta_{k-l}^{(n)} + \delta_{-k+l}^{(n)} + \delta_{-k-l}^{(n)}\right] \tag{1.384}$$

$\delta_j^{(n)}$ being a function of the index j and by using $|\hat{x}_{n-l}|^2 = |\hat{x}_l|^2$, due to the periodicity, it comes

$$S_k = |\hat{x}_k|^2 = \sum_{m=1}^{n} \psi_m \cos\left(\frac{2\pi mk}{n}\right), \tag{1.385}$$

i.e. the relation of inversion that we looked for. The representative graph of $|\hat{x}_k|^2$ function of $f(f = k \cdot \triangle f)$ is called the *Power Spectral Density (PSD)*.
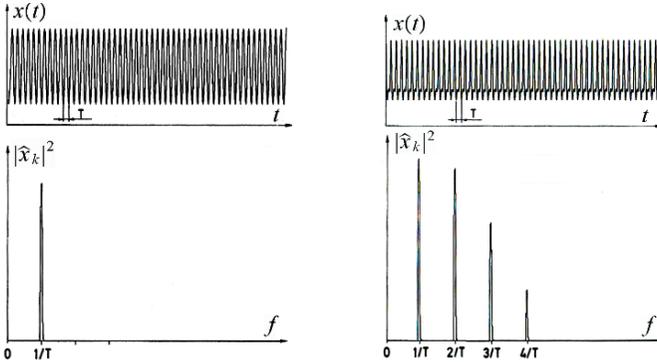
Fig. 1.107 Pure sinusoidal function (*left*), sinusoidal function with harmonics (*right*)

1.37.2 Different Types of Spectra

1.37.2.1 Periodic Signal

Given a periodic series written in a simple way $x(t) = x(t+T) = x(t+2\pi/\omega)$. This is a particular case where the period is exactly equal to the duration of the signal measurement. It can be written $T = t_{\max} = n \cdot \Delta t$. The Fourier components will be located exactly at the frequencies: $\frac{1}{T}, \frac{2}{T}, \frac{3}{T}, \ldots, \frac{n}{T}$. Consequently, in extreme cases if the signal is a sine, for example, its spectrum will have only one non-null component which will have as abscissa $1/T$ that will express only one frequency. For a slightly more complex periodic signal, the amplitude of harmonics of its frequencies $\left(\frac{2}{T}, \frac{3}{T}, \ldots\right)$ will not be zero. If there are harmonics in a spectrum, this reveals the non-sinusoidal character of its evolution. Hereafter, we provide two examples of periodic signals (Fig. 1.107).

1.37.2.2 Quasi-Periodic Signal and Coupling of Frequencies

Generally, the spectrum of a quasi-periodic function has a rather complicated representation. We call a quasiperiodic function of time, a periodic function whose variables are all directly proportional to time ($t_j = \omega_j t$, $j = 1, \ldots, r$). It is constituted of basic "peaks" located at the frequencies $f_1, f_2, \ldots, f_p$, but also of their harmonics denoted $a_1 f_1, a_2 f_2, \ldots, a_p f_p$, where $(a_1, a_2, \ldots, a_p)$ are positive integers. If (for example) the spectrum analyzes a signal made up of the product of two sinusoids: $x(t) = sin(\omega_i t) \cdot sin(\omega_j t)$, then the Fourier spectrum contains the basic frequencies $\left| f_i - f_j \right|$ and $\left| f_i + f_j \right|$ and their harmonics, because:

$$sin(\omega_i t) \cdot sin(\omega_j t) = \frac{1}{2}\cos(\left| f_i - f_j \right| 2\pi t) - \frac{1}{2}\cos(\left| f_i + f_j \right| 2\pi t) \qquad (1.386)$$

thus, the Fourier spectrum of a quasi-periodic signal, which depends (*in a non-linear way*) on periodic functions of variables $\omega_i t$, *contains components at all the frequencies:*

$$\left| a_1 f_1 + a_2 f_2 + \cdots + a_p f_p \right|. \tag{1.387}$$

If we choose a bi-periodic case $(p = 2)$, such as it is presented in the function $x(t)$ above, each component of the spectrum of the signal $x(\omega_1 t, \omega_2 t)$ is a *peak of abscissa*:

$$\left| a_1 f_1 + a_2 f_2 \right|. \tag{1.388}$$

The ratio f_1 / f_2 can be *rational or irrational*. In general, we characterizes a quasi-periodic spectrum by seeking the two basic frequencies f_1 and f_2 from which by combination $(|a_1 f_1 + a_2 f_2|)$, we can construct the other frequencies. If f_1 / f_2 is rational, the spectrum is not representable by a continuous function, such as $f_1 / f_2 = \sigma_1 / \sigma_2$, where σ_1 and σ_2 are integers. The quasi-periodic function is then regarded as periodic of period $T = \sigma_1 T_1 = \sigma_2 T_2$. According to the definition, we have

$$x(\omega_1 t, \omega_2 t) = x(\omega_1 t + 2\sigma_1 \pi, \omega_2 t + 2\sigma_2 \pi) \tag{1.389}$$

and

$$x(\omega_1 t, \omega_2 t) = x\left(\omega_1 \left(t + \frac{\sigma_1}{f_1} \right), \omega_1 \left(t + \frac{\sigma_2}{f_2} \right) \right). \tag{1.390}$$

There is *coupling of two frequencies* f_1 and f_2. The set of lines of the spectrum are harmonics of the lowest frequency:

$$f_0 = \frac{1}{T} = \frac{f_1}{\sigma_1} = \frac{f_2}{\sigma_2}. \tag{1.391}$$

Two consecutive lines of the spectrum are always separated by the same distance: $1/T$. Here are two examples of quasiperiodic functions:

(1) Figure 1.108 (left) shows a series for two basic frequencies f_1 and f_2 whose ratio f_1 / f_2 is irrational. The spectrum (beyond basic frequencies) exhibits main frequency peaks $f_p = m_1 f_1 + m_2 f_2$:

(a) $\rightarrow f_2 - f_1$, (b)$\rightarrow 3f_1 - f_2$, (c)$\rightarrow f_1 + f_2$, (d)$\rightarrow 3f_1$, (e)$\rightarrow 5f_1 - f_2$,
(g) $\rightarrow 3f_1 + f_2$, (h)$\rightarrow 5f_1$, (i)$\rightarrow 7f_1 - f_2$, (j)$\rightarrow 5f_1 + f_2$.

(2) Figure 1.108 (right) shows almost the same function as previously, but the frequency f_1 is changed so that the ratio $f_1 / f_2 = 2/3$. In such a case, all the peaks are harmonics of the frequency $f = f_2 - f_1 = (1/3)f_2$:

(a) $\rightarrow f_2 - f_1$, (b)$\rightarrow 3f(= f_2)$, (c)$\rightarrow 5f$, (d)$\rightarrow 6f$, (e)$\rightarrow 7f$,
(g) $\rightarrow 9f$, (h)$\rightarrow 10f$, (i)$\rightarrow 11f$, (j)$\rightarrow 12f$.

An experimental illustration is given by the Rayleigh–Bénard instability (R–B).[70] Here, the fluid is the water whose temperature is such that the Prandtl

---

[70] Ref. to J. Gollub and S. Benson.
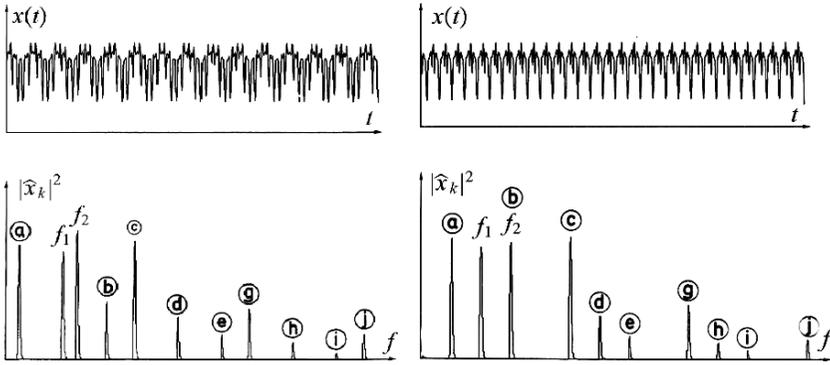
Fig. 1.108 Function with f_1/f_2 irrational (*left*), Function with f_1/f_2 rational (*right*)
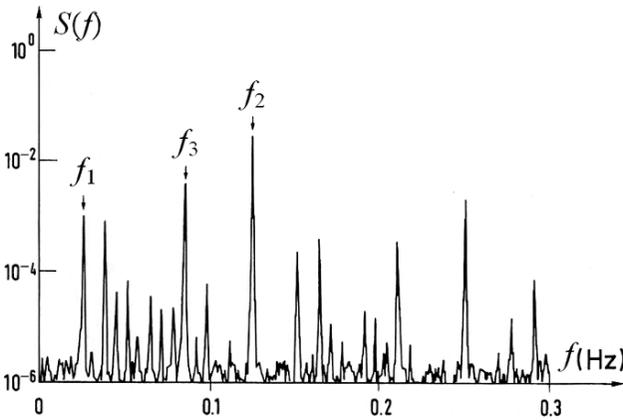
Fig. 1.109 Power spectrum of a quasiperiodic dynamic with three incommensurable frequencies (R–B instability, Ra/Ra$_c$ = 42.3). Ref. to Gollub J. and Benson S.

number is equal to 5. The power spectrum of a quasiperiodic dynamic with three incommensurable frequencies observed in the R–B instability is shown above (see Fig. 1.109).

The spectral scale is logarithmic. By increasing the ratio Ra/Ra$_c$ the stationary convection loses its stability to become periodic starting from Ra/Ra$_c$ = 30 (first Hopf bifurcation). For Ra/Ra$_c$ = 39.5 the previous periodic regime loses its stability to become quasiperiodic with two basic frequencies (second Hopf bifurcation). From Ra/Ra$_c$ = 41.5 a third bifurcation occurs with three basic frequencies (see spectrum above). The frequency of each ray is indexed as $f = m_1 f_1 + m_2 f_2 + m_3 f_3$ (with m_1, m_2, m_3 integers) and obviously cannot be indexed as $f = m'_1 f'_1 + m'_2 f'_2$. The incommensurability is verified because the ratios f_1/f_2, f_1/f_3, f_2/f_3 vary continuously with Ra/Ra$_c$.

1.37.2.3 Aperiodic Signal, Chaotic Signal

In contrast, if the function or the signal studied is neither periodic, nor quasi-periodic, it is *aperiodic or non-periodic*. It is generally acknowledged that the spectrum of such a signal is *continuous*, but the converse is not true, because this can be the result of the Fourier spectrum of a quasi-periodic function with a number of very high frequencies (even infinite). Suppose that we deal with an (true) aperiodic signal, then we have to distinguish two cases according to the degree of freedom of the system: either the degree of freedom is very restricted, or it is very high. In the first case, it is possible to develop a deterministic approach of dynamics, whereas in the second case, only a probabilistic approach could be implemented, introducing the notion of randomness (even if the underlying determinism is present).

(1) The extreme case of a random signal is the white noise (e.g. white light, source). In such a case, the signal can be regarded as new at each instant. Such a noise suggests the action of an almost infinite number of independent agents (e.g. molecular agitation, elastic vibrations of atoms). The spectrum of such a noise is flat, i.e. the amplitude is independent of the frequency and is thus devoid of harmonic structure. It is an infinite (stationary) random signal whose function of autocorrelation is proportional to a Dirac, i.e. a constant complex spectrum on the whole zone of frequencies.

(2) Hereafter, we show an aperiodic signal and its characteristic continuous spectrum (see Fig. 1.110).

(3) Ruelle–Takens theory (R–T) was introduced in 1971, then detailed in 1978 by Ruelle–Takens–Newhouse (R–T–N). This theory called into question the previous Landau mechanism in which an infinity of Hopf bifurcations would be necessary to generate the turbulence. On the contrary, Ruelle–Takens have considered that a small number of bifurcations was sufficient to produce chaotic
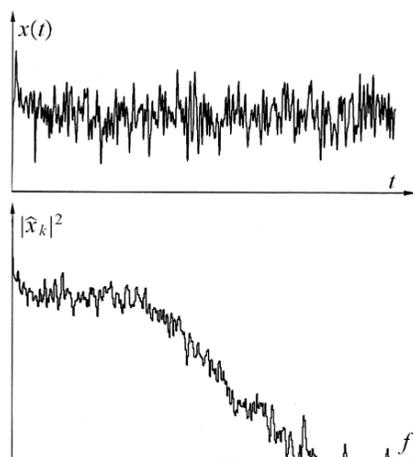
Fig. 1.110 Continuous spectrum of an aperiodic signal
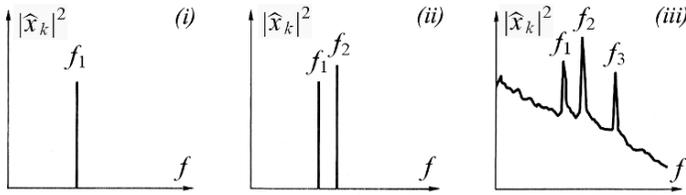
Fig. 1.111 Spectrum evolution according to Ruelle–Takens theory

behaviors. The experiment used was that of Reynolds. Given a laminar flow, then by increasing the Reynolds number, the system loses its stability and becomes oscillating at the frequency f_1. The same process is repeated twice, thus we successively obtain three Hopf bifurcations at the frequencies f_1, f_2, f_3. Then according to R–T–N, the corresponding torus T^3 can become (subject to conditions) unstable and be replaced by a strange attractor. The behavior is no longer quasiperiodic with three frequencies (tore T^3) but clearly chaotic. Here is the evolution of the spectrum according to Ruelle–Takens theory (see Fig. 1.111); the behavior is (1) periodic, (2) quasiperiodic with two frequencies, (3) chaotic.

1.37.2.4 ARMA Process Spectra

Some ARMA processes are generated (autoregression and moving average or combinations of both). x_t corresponds to the dependent variable and a_t are Gaussian risks. The graphs of eleven spectra are given below (see Fig. 1.112).

(1). AR(1): $x_t = 0.9x_{t-1} + a_t$

(2). AR(1): $x_t = -0.9x_{t-1} + a_t$

(3). AR(2): $x_t = 0.7x_{t-1} - 0.2x_{t-2} + a_t$

(4). AR(2): $x_t = -0.7x_{t-1} - 0.4x_{t-2} + a_t$

(5). MA(1): $x_t = a_t - 0.8a_{t-1}$

(6). MA(1): $x_t = a_t + 0.8a_{t-1}$

(7). MA(2): $x_t = a_t + 0.9a_{t-1} + 0.2a_{t-2}$

(8). MA(2): $x_t = a_t - 0.5a_{t-1} + 0.8a_{t-2}$

(9). ARMA(1,1): $x_t = -0.3x_{t-1} + a_t + 0.4a_{t-1}$

(10). AR12(1): $x_t = 0.6x_{t-12} + a_t$

(11). MA12(1): $x_t = 0.6a_{t-12} + a_t$

1.37.2.5 Spectra of French Stock Index (Cac40)

Cac40 and Its First-Differences

The signal is a sample of 2,847 days. The (rough) spectrum of the Cac40 decreases and shows a L-shape with an enormous localized peak. For the differences, the spectrum of the stationarized series is very different (Figs. 1.113 and 1.114).
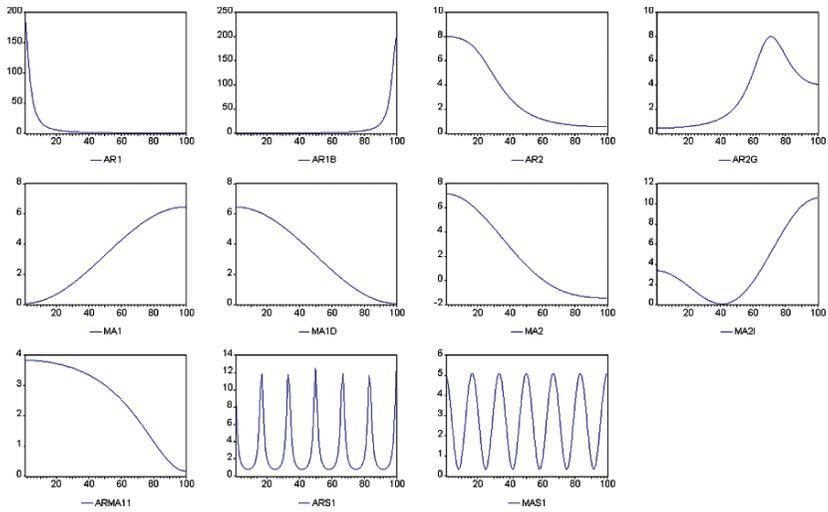
Fig. 1.112 Spectra of ARMA processes
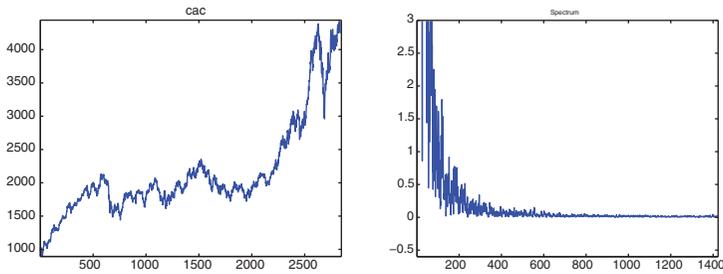
Fig. 1.113 *Left:* Cac (2,847 days). *Right:* Spectrum
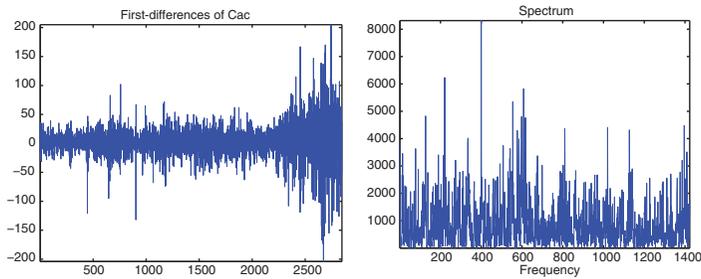
Fig. 1.114 *Left:* First-differences. *Right:* Spectrum

1.37.2.6 Spectra of the Logistic Equation

Spectrum of the Logistic Equation for $\alpha = 4$

The direct spectrum of the logistic equation is explosive. On the other hand, if we make it stationary by means of first-differences, the series provides a (rather) rich
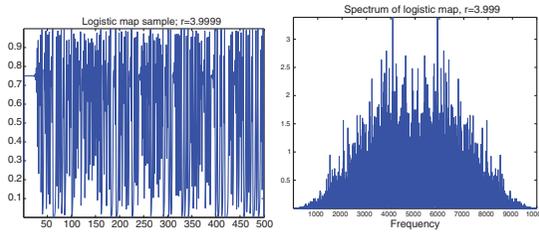
Fig. 1.115 Chaotic domain (*left*), spectrum of the chaotic domain (*right*)

representation in the frequency field of the energy of the signal (Fig. 1.115). (The symmetry of the spectrum ($|\hat{x}|^2$) was restored to vary the mode of representation.)

Set of Spectra of the Logistic Equation $\alpha \in [2.9, 4]$

Evolution of spectra according to the parameter:

One period and two fixed points for $\alpha \in [2.9, 3]$.
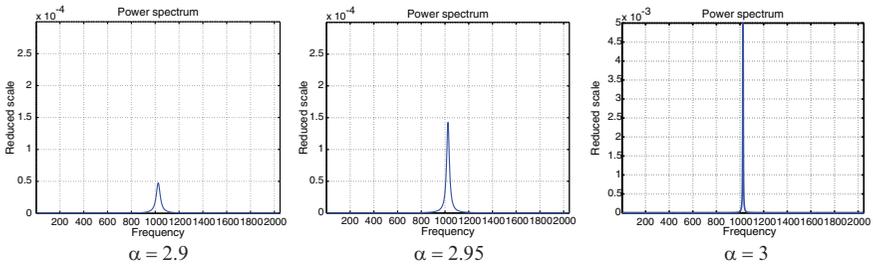
First bifurcation and period-doubling for $\alpha \subset [3, 3.4985]$.
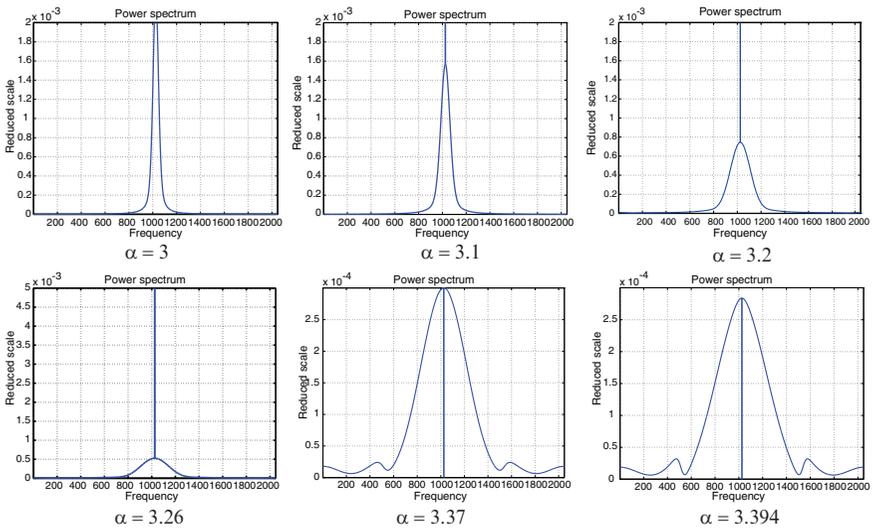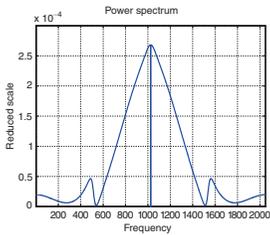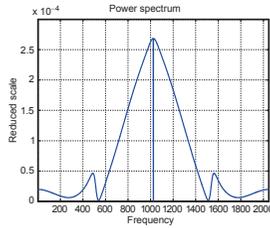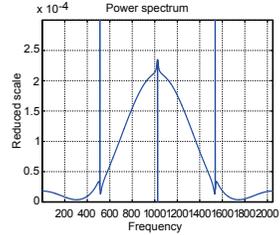
$\alpha = 3.41$ $\alpha = 3.434$ $\alpha = 3.454$
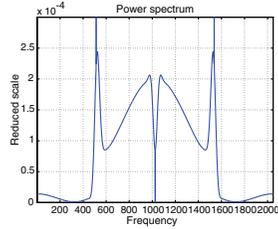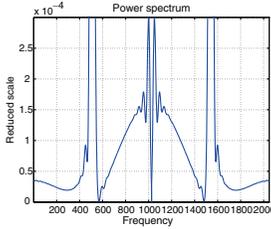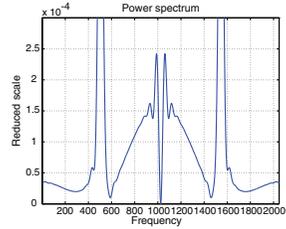
$\alpha = 3.474$ $\alpha = 3.494$ $\alpha = 3.496$

Second bifurcation and period-doubling for $\alpha \subset [3.498, 3.544]$.
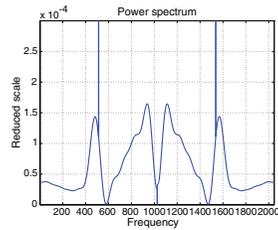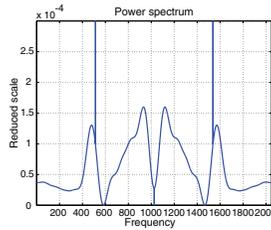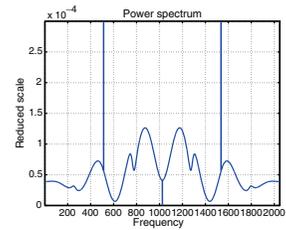
$\alpha = 3.512$ $\alpha = 3.514$ $\alpha = 3.532$

New bifurcations $\alpha \in [3.544, 3.5569]$. Chaotic regime $\alpha = [3.5699, 4]$.
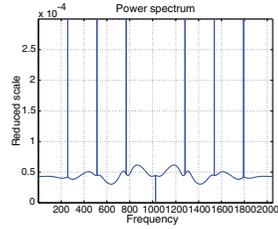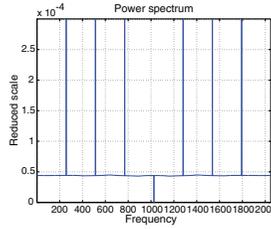
Six branches $\alpha = 3.558$ Seven branches

$\alpha = 3.552$ $\alpha = 3.558$

$\alpha = 3.572$ $\alpha = 3.576$ $\alpha = 3.599$

$\alpha = 3.632$ $\alpha = 3.84$ $\alpha = 3.854$
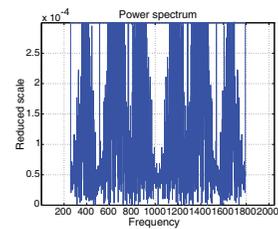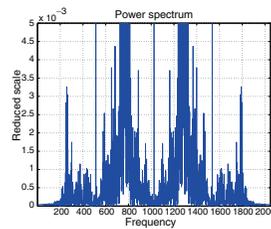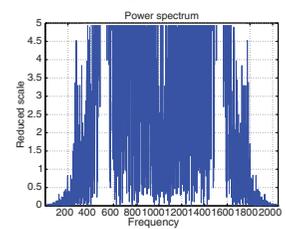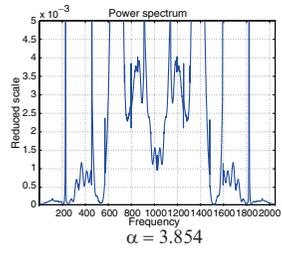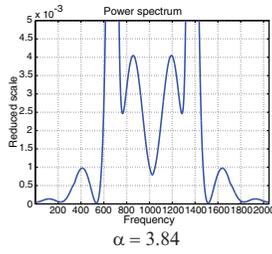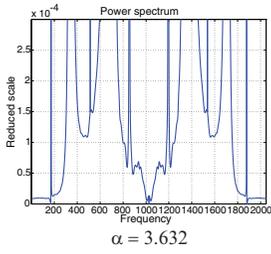
Extraction of symptomatic spectra

$\alpha = 2.95$

$\alpha = 3.2$

$\alpha = 3.434$

$\alpha = 3.494$

$\alpha = 3.532$

$\alpha = 3.558$

$$\alpha = 3.576 \qquad\qquad\qquad \alpha = 3.854$$
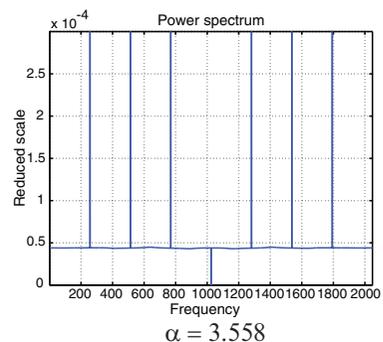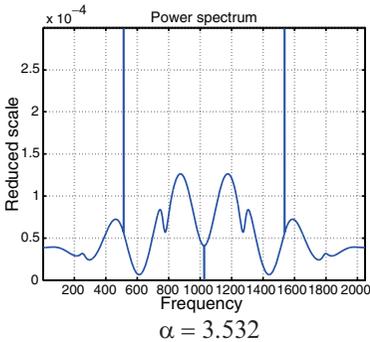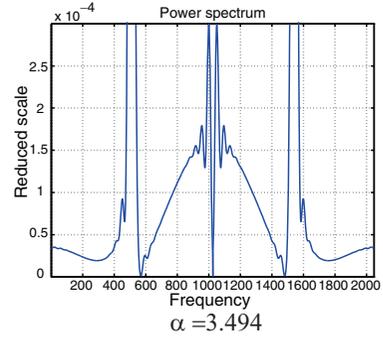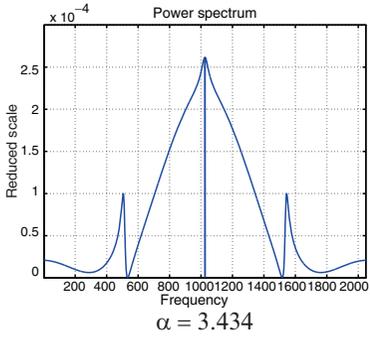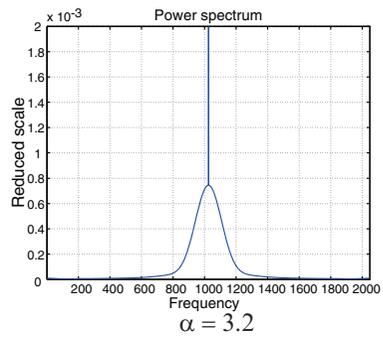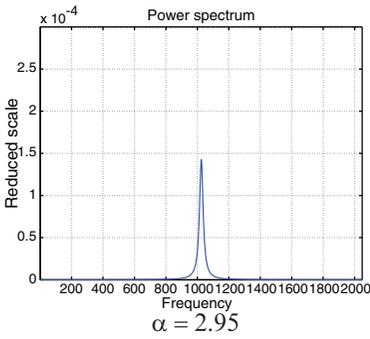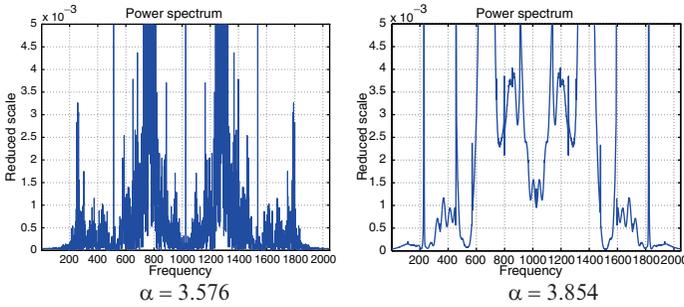
1.38 Van der Pol Oscillator and Spectra

In numerous scientific fields the study of a phenomenon is usually focused on its predominant effects (Bergé et al. 1987) Often this approach leads to *linearize* the characteristic phenomena of a system.[71] Indeed, the *nonlinear effects* are often regarded as *perturbative* and therefore neglected, however we know that they can lead to unexpected and surprising effects. It is possible to study such effects by means of the *Van der Pol oscillator*. As a preamble, let us consider the *simple pendulum*.

The simple pendulum: The figure hereafter describes the simple pendulum device where we can observe θ, g, l and m:
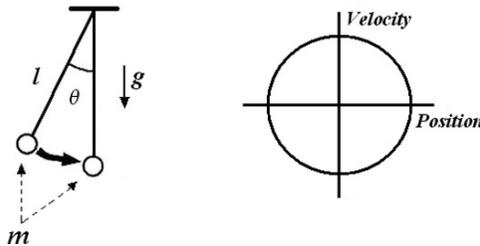
For such a device, the mechanics theory leads to the following equation, if there is *no friction* $\ddot{\theta} + \dfrac{g}{l} \sin\theta = 0$. By contrast, if there is *friction,* i.e. if dissipation is introduced, it is necessary to add a term dependent on $\dot{\theta}$ $\ddot{\theta} + 2\alpha\dot{\theta} + \dfrac{g}{l} \sin\theta = 0$. Such an *oscillator is nonlinear* except for the *small angles* where it is possible to make the approximation: $\sin\theta \approx \theta$ which is valid at the second order for θ. Then the differential equation is written

$$\ddot{\theta} + 2\alpha\dot{\theta} + \omega_0^2 \theta = 0, \qquad\qquad (1.392)$$

where $\omega_0 = \sqrt{g/l}$. Starting from this basic reminder, it is now possible to describe the van der Pol oscillator.

Van der Pol oscillator: For oscillations to occur it is known that in the differential equation the coefficient of the first order term must be negative. Furthermore, so that these oscillations have limited amplitudes, this coefficient must change its sign,

---

[71] Considering that "a priori" there is proportionality between cause and effect.

then the system evolves into a *limit circle*. This is the fundamental topic of the van der Pol oscillator. The differential equation of the van der Pol system can be written

$$\ddot{\theta} - \delta_0 \left(1 - \frac{\theta^2}{\theta_0^2}\right)\dot{\theta} + \omega^2\theta = 0, \tag{1.393}$$

where δ_0 and θ_0 are *constants* of the equation. In such an equation for *small amplitudes* $\theta \ll \theta_0$ the fraction is negligible, then the differential equation corresponds to the occurrence of oscillations whose amplitude increases exponentially since the coefficient of the first order term is $\simeq -\delta_0$. Conversely, when the amplitude of these oscillations is high, the fraction becomes higher than the unit, the sign of the associated term changes and the amplitude of oscillations decreases exponentially. The system then evolves between two limit infinitely close states, which define the amplitude and the shape of oscillations. (θ_0 influences the amplitude and δ_0 the sinusoid. Furthermore, when an external "forcing"[72] is taken into account, then a new quantity $\rho\cos\omega_\mu t$ is introduced and a new equation can be written

$$\ddot{\theta} - \left(\varepsilon - \theta^2\right)\dot{\theta} + \theta = \rho\cos\omega_\mu\, t \tag{1.394}$$

with $\varepsilon = \dfrac{\delta_0}{\omega}$. It is possible to rewrite such an equation in the form of an *autonomous differential system*:

$$\dot{x} = y, \tag{1.395}$$
$$\dot{y} = \rho\cos(z) - x + (\varepsilon - x^2)y, \tag{1.396}$$
$$\dot{z} = \omega_\mu, \tag{1.397}$$

with $x = \theta$. At this stage it is possible to carry out the numerical analysis successively of *Non-forced* and *Forced van-der-Pol oscillators*.

Non-forced van der Pol oscillator: Let us study the model by means of the following generic equation:
$$\ddot{\theta} - \left(\varepsilon - \theta^2\right)\dot{\theta} + \theta = 0 \tag{1.398}$$
and by means of the *autonomous differential system,* it is possible to write:

$$\dot{x} = y = \varphi_1(x,y), \tag{1.399}$$
$$\dot{y} = (\varepsilon - x^2)y - x = \varphi_2(x,y). \tag{1.400}$$

Consequently, it is possible to obtain the fixed-points:

$$\left(\dot{x} = 0 = y,\ \ \dot{y} = 0 = (\varepsilon - x^2)y - x\right). \tag{1.401}$$

---

[72] *Remark:* In mathematics the term "forcing" has a precise meaning: *Forcing* is a technique in set theory created by Cohen (63) used to prove that the axiom of choice and continuum hypothesis are independent of one another in Zermelo–Fraenkel set theory.

A single fixed-point results $(x,y) = (0,0)$. In order to study its stability, the following Jacobian matrix of the system is analyzed

$$J = \begin{pmatrix} \dfrac{\partial \varphi_1}{\partial x} & \dfrac{\partial \varphi_1}{\partial y} \\[2mm] \dfrac{\partial \varphi_2}{\partial x} & \dfrac{\partial \varphi_2}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 - 2xy & \varepsilon - x^2 \end{pmatrix} \qquad (1.402)$$

and for the fixed-point $(x,y) = (0,0)$, we have

$$J = \begin{pmatrix} 0 & 1 \\ -1 & \varepsilon \end{pmatrix}. \qquad (1.403)$$

The Jacobian matrix study is carried out by its trace. The results depend on the sign of ε. When $\varepsilon > 0$ or $\varepsilon < 0$ the system is dissipative, the fixed-point $(x,y) = (0,0)$ is repelling or attracting; conversely when $\varepsilon = 0$ the system is conservative and $(x,y) = (0,0)$ is a center. Here, ε corresponds to the *friction*, and for a *low friction* there will be oscillatory motions and therefore the eigenvalues of the Jacobian matrix will be complex and the stable point will be a *focus*. More precisely, from the equation of eigenvalues:

$$\lambda^2 - \varepsilon\lambda + 1 = 0, \quad \Delta = \varepsilon^2 - 4 \qquad (1.404)$$

thus the *general behavior* of the system is as follows:

$|\varepsilon| > 2$:	$|\varepsilon| < 2$:
$.\varepsilon < 0 \Rightarrow (0,0)$: Stable-node.	$\varepsilon < 0 \Rightarrow (0,0)$: Attracting focus
$.\varepsilon > 0 \Rightarrow (0,0)$: Unstable-node.	$\varepsilon = 0 \Rightarrow (0,0)$: Center
	$\varepsilon > 0 \Rightarrow (0,0)$: Repelling focus (source).

Since $\lambda_\pm = \varepsilon \pm i\omega$ for $\varepsilon = 0$ there exists a supercritical Hopf bifurcation, which means that at zero there is a transition from a stable focus to an unstable focus. Figures 1.116 and 1.117 give some examples of trajectories in the analytical framework described above (for initial conditions (x_0,y_0)):

For $0 < \varepsilon < 2$ a *stable limit cycle (here a limit circle)* occurs because of *amplitude limitation* consecutive to the *friction quadratic term*. Then, for $\varepsilon = 0$ there is a supercritical Hopf bifurcation.
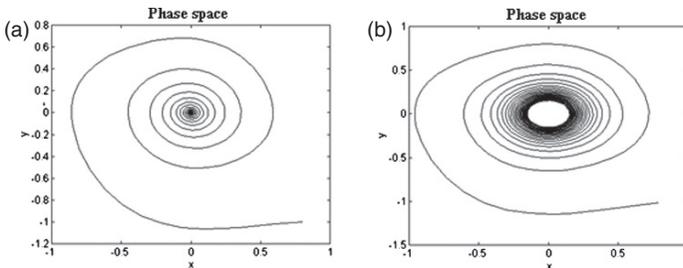
Fig. 1.116 (a) $-2 < \varepsilon < 0$ for $(x_0,y_0) = (0.8, -1)$; (b) $\varepsilon = 0$ for $(x_0,y_0) = (0.8, -1)$
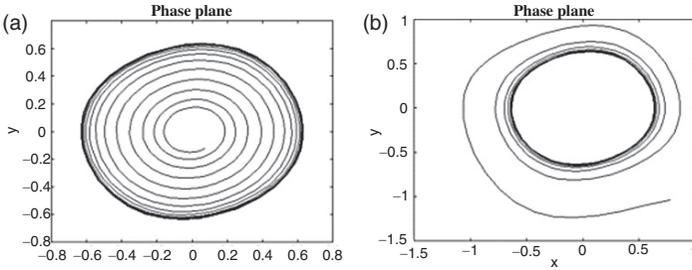
Fig. 1.117 (a) $0 < \varepsilon < -2$ for $(x_0, y_0) = (0, -0.1)$; (b) $0 < \varepsilon < 2$ for $(x_0, y_0) = (0.8, -1)$
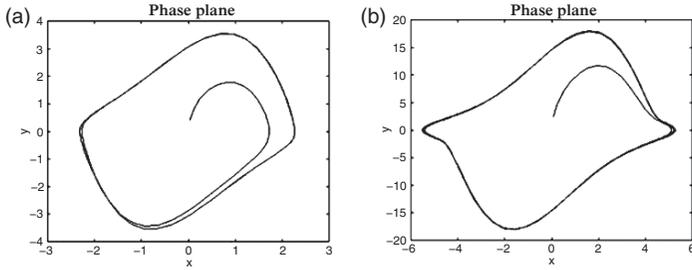
Fig. 1.118 (a) $\varepsilon = 0.1$, $\omega_\mu = 1$, $\rho = 3$; (b) $\varepsilon = 0.1$, $\omega_\mu = 1$, $\rho = 50$

Forced van der Pol oscillator: Let us study the model by means of the following generic equation:

$$\ddot{\theta} - (\varepsilon - \theta^2)\, \dot{\theta} + \theta = \rho \cos \omega_\mu\, t, \tag{1.405}$$

which can be rewritten in *autonomous form*:

$$\dot{x} = y, \tag{1.406}$$
$$\dot{y} = \rho \cos(z) + (\varepsilon - x^2)y - x, \tag{1.407}$$
$$\dot{z} = \omega_\mu. \tag{1.408}$$

For this system the fixed-point is $(\rho, 0, 0)$. When $\varepsilon = 0$ we return to the case of the *harmonic oscillator* with *circular trajectories*. In contrast, when ε is small and positive, $\varepsilon = 0.1$: near to the origin, the trajectories diverge and draw a spiral, and far from the origin, the trajectories tend to approach the fixed point.[73] Therefore, between these two extreme situations we understand that there exists a *limit cycle*. See hereafter the *almost rectangular form* of the *limit cycle* in Fig. 1.118a or the *distorted rectangular form* in Fig. 1.118b.

The almost rectangular course of the limit cycle in the Fig. 1.118a uses two different scales of time. If it is large enough the (forcing) parameter ρ constrains to a regularity on these scales in relation to its frequency. Within the framework of the van der Pol oscillator the parameters are defined as follows: ρ is the *strength of the*

---

[73] *Remark:* In this area θ is always very small and the term θ^2 remains small.
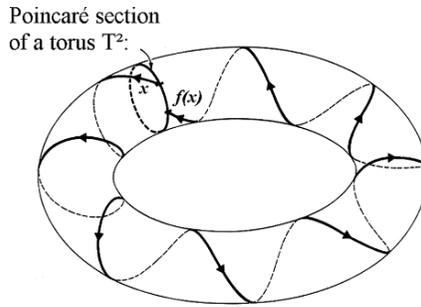
Fig. 1.119 Poincaré section of a flow on a torus T^2

driving (sometimes called *forcing parameter*), ε is the *coefficient of the negative linear damping* (also called *friction* parameter) and ω_μ is the *frequency of the external driving*.

As seen previously, when ε changes its sign, there exists a *Hopf bifurcation,* then (at least in the proximity of the bifurcation) the *attractor is a torus T^2*. Figure 1.119 shows a flow on a torus T^2 and the Poincaré section of this flow; thus we can see how to construct the Poincaré map for a circle drawn on a torus and see that the trajectory starting from the point x on this circle re-cuts the circle at the point $f(x)$.

At this stage, in the spectral analysis framework when Fourier transform is used to decompose the signal, we observe *two different pulses* (ω_1, ω_2) representing the periods of motions along the large and small circles. When these *frequencies* have a *"rational" ratio* we observe the *"coupling of frequencies" (phenomenon)* and we can *count the number of harmonics* which is *equal to this ratio*. Such observations are depicted in Fig. 1.120 in which it is almost possible to visually count the number of different frequencies.

In Fig. 1.120d, the number of frequencies is not visually countable, so we use the spectral analysis which highlights the peaks for each frequency in the spectrum corresponding to the frequencies which exist in dynamics. See Fig. 1.121 where it is possible to distinguish the peaks in the spectrum and, in particular, the basic frequencies (ω_1, ω_2) for the case $\omega_\mu = 12$.

In contrast, when the *ratio* of eigenfrequencies (and forcing) is *irrational*, theoretically the resulting (quasi-periodic) dynamics have spectra with an infinite number of dense lines for all the pulsations $|n_1\omega_1 \pm n_2\omega_2|$ where n_1 and n_2 are integers. However, in practice the sampling methods of signals imply that there is not an infinite number of dense lines (a better approximation of irrational numbers and infinite number of rotations would be necessary). Figures 1.122 and 1.123 show the phase-spaces and spectra for the cases $(\varepsilon, \omega_\mu, \rho) = (0.1, \pi, 3)$ and $(\varepsilon, \omega_\mu, \rho) = (0.1, \mu, 3)$. For $(0.1, \pi, 3)$ a periodic motion is always observed in the presence of both (basic) frequencies running on the torus; the spectrum exhibits always peaks although they are more complex and dense.

For $(0.1, \mu, 3)$ the result is clearer, thus the spectrum exhibits an infinite number of dense lines. The van der Pol oscillator is the most typical modeling of oscillatory
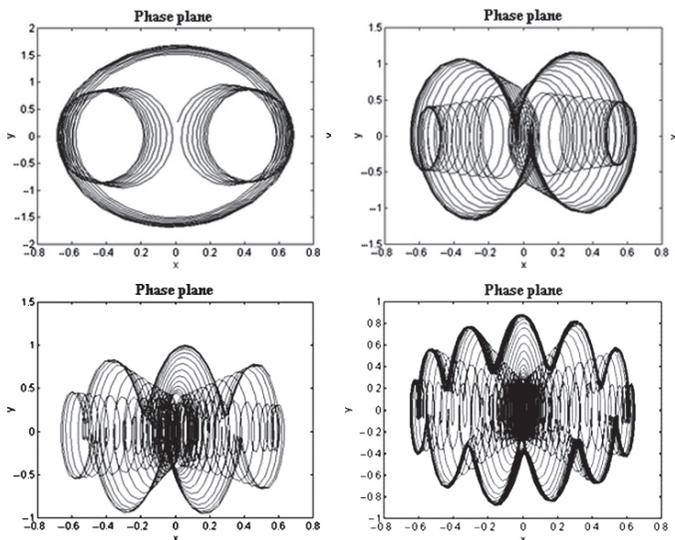
Fig. 1.120 (**a**) $\varepsilon = 0.1$, $\omega_\mu = 3$, $\rho = 3$, (**b**) $\varepsilon = 0.1$, $\omega_\mu = 5$, $\rho = 3$, (**c**) $\varepsilon = 0.1$, $\omega_\mu = 8$, $\rho = 3$, (**d**) $\varepsilon = 0.1$, $\omega_\mu = 12$, $\rho = 3$
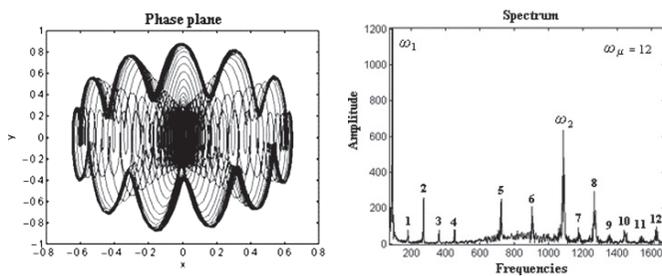
Fig. 1.121 $\varepsilon = 0.1$, $\omega_\mu = 12$, $\rho = 3$

motions and allows to highlight fundamental phenomena. In particular, this oscillator makes it possible to exhibit a *supercritical Hopf bifurcation* when the system passes from an attracting state to a repelling state while ε changes its sign, and in addition, it can exhibit the *coupling of frequencies*.

Remark 1.9 (Synchronization and van der Pol oscillator). It is interesting to highlight the synchronization phenomena through the van der Pol oscillator. The following paragraph provides (1) an introduction to this important topic, and (2) we will use this framework to show typical spectra of the van der Pol dynamics.

Synchronization phenomena and Van der Pol oscillator: It is well-known that interaction between *nonlinear oscillatory systems*, including chaotic behaviors, can result in their synchronization.
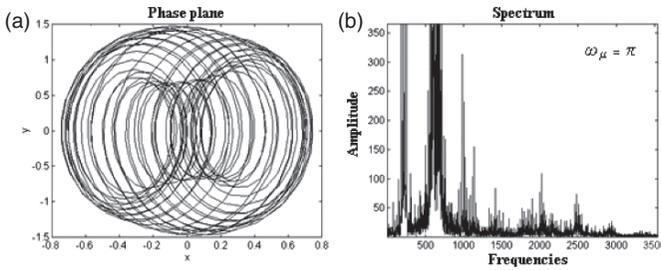
Fig. 1.122 (**a**) $\varepsilon = 0.1$, $\omega_\mu = \mu$, $\rho = 3$; (**b**) Spectrum for $\varepsilon = 0.1$, $\omega_\mu = \mu$, $\rho = 3$
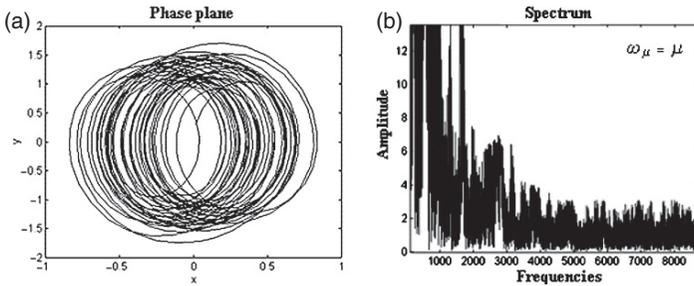
Fig. 1.123 (**a**) $\varepsilon = 0.1$, $\omega_\mu = \mu$, $\rho = 3$; (**b**) Spectrum for $\varepsilon = 0.1$, $\omega_\mu = \mu$, $\rho = 3$

(1) Various types of synchronization between oscillatory processes have been studied in physics, chemistry and biology (Pikovsky et al. 2001; Glass and Mackey 1988; Schäfer et al. 1999). Recently, the analyses of *synchronization in "living organisms" whose activity is induced by the interaction of a large number of complex rhythmic processes* has became an important field of research. Often the (underlying) sources of these oscillatory processes cannot be measured separately, but only superpositions of their signals are available (e.g. in the electroencephalo-grams recorded on the scalp the signals are the superpositions of various interacting sources). Therefore, it is possible to detect false or "spurious" synchronization between brain sources leading to wrong conclusions (Meinecke et al. 2005). For many multichannel measuring devices this fact is typical. Similar problems occur while studying synchronization between the rhythms of the *cardiovascular system* (CVS). The most significant oscillating processes characterizing cardiovascular dynamics, i.e. the main *heart rhythm*, *respiration*, and the *process of blood pressure slow regulation with the fundamental frequency close to 0.1 Hz*, appear in various signals, for example, *electrocardiogram* (ECG), blood pressure, blood flow, and *heart rate variability* (HRV). Such facts hinder the study of their synchronization. *The systems producing the main heart rhythm and the rhythm of slow regulation of blood pressure can be considered as "self-sustained oscillators" and the respiration can be taken as "an external forcing of these systems"*. But, as suggested above, for respiration frequencies near to 0.1 Hz, it is difficult to *distinguish* the case of *true synchronization between* the *respiration* and the process of *blood pressure regula-*

tion.[74] The existence of *external forcing* can result in *linear "mixing"* of the driving signal and the signal of the self-sustained oscillator without any synchronization. (Concomitant presence of mixing of signals and their synchronization can also be another cause.) Similar characteristics of *synchronization* can be observed in the case of *periodic driving of a van der Pol oscillator* and in the case of *respiratory forcing of the heartbeat* and the process with the basic frequency of about 0.1 Hz.

(2) In order to study the *interaction* between *respiration* and the process of *blood pressure* slow regulation, it is possible to use the (asymmetric) *van der Pol oscillator* under external forcing with linearly increasing frequency:

$$\ddot{x} - \nu(1 - \alpha x - x^2)\dot{x} + \Phi^2 x = A \sin \psi(t) \tag{1.409}$$

with $\nu = 1$ the *nonlinearity parameter*, $\Phi = 0.24\pi$ the *natural frequency*, and A, ψ the *amplitude* and *phase* of the *external force*. The phase equation

$$\psi(t) = (2\pi[(a + bt/T)]t) \tag{1.410}$$

expresses the linear dependence of the driving frequency $\omega_d(t)$ on time:

$$\omega_d(t) = \frac{d\psi(t)}{dt} = 2\pi(a + bt/T), \tag{1.411}$$

with $a = 0.03$, $b = 0.17$ and $T = 1800$ (maximum time of calculation). When $\alpha = 0$, the system *corresponds to the usual van der Pol oscillator with symmetric limit cycle*. In this case, the spectrum of oscillations has only *odd harmonics* $(2n + 1)f_0$ (with $n = 1, 2, \ldots$) of the *basic frequency* f_0. The second harmonic $2f_0$ with the basic frequency close to 0.1 Hz is prominent, then the van der Pol oscillator is modified by means of $\alpha = 1$. Due to *the effect of nonlinearity*, there is a *difference* between the *natural frequency* Φ and the *frequency* ω_0 of *self-sustained oscillations* ($\omega_0 = 2\pi f_0$,
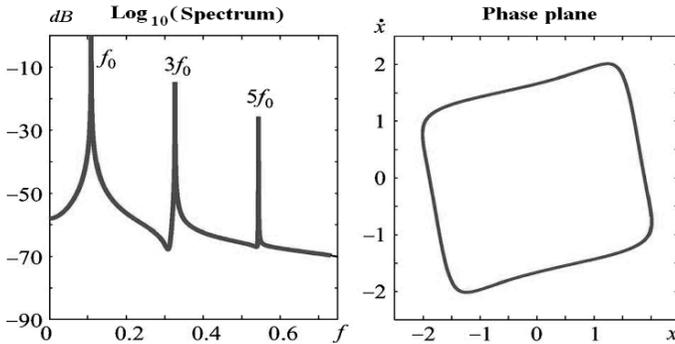
Fig. 1.124 Log_{10}(Spectrum) and Phase-plane for $A = 0, \Phi = 0.24\pi, \alpha = 0$

---

[74] Indeed, false synchronization can appear due to the presence of respiratory component in the HRV and blood pressure signals used for the analysis of the rhythm with the basic frequency of about 0.1 Hz.
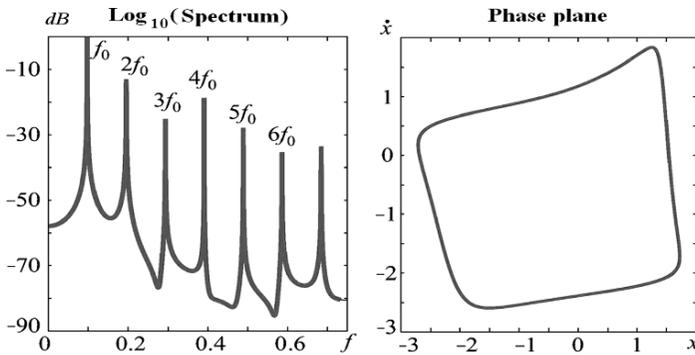
Fig. 1.125 Log_{10}(Spectrum) and Phase-plane for $A = 0, \Phi = 0.24\pi, \alpha = 1$

$\Phi = 0.24\pi$, $f_0 = \omega_0/2\pi = 0.106)$, (see Fig. 1.124). In the asymmetric van der Pol oscillator this difference between Φ and ω_0 is greater ($f_0 = 0.098$), Fig. 1.125.

It could be interesting to compare (1) the synchronization of oscillations by external driving and (2) the mixing of signals. This can be done considering

$$y(t) = x(t) + B\sin\psi(t), \tag{1.412}$$

where $x(t)$ is the signal of the autonomous asymmetric van der Pol oscillator and $B\sin\psi(t)$ is the additive signal with amplitude B, phase $\psi(t)$ and *varying frequency* $\omega_d(t) = d\psi(t)/dt = 2\pi(a + bt/T)$. The literature on such subjects is abundant, and we thus exhort the reader to consult the increasing number of publications on this topic (Shilnikov et al. 2003; Hramov et al. 2006).

1.39 Reconstruction Theorems

Before explaining the Taken's reconstruction theorem, we have to introduce the embedding notion and the general Withney's embedding theorem.

1.39.1 Embedding, Whitney Theorem (1936)

In a general way, an *embedding*[75] is a representation of a topological object, manifold, graph, field, etc., in a certain space in such a way that its connectivity or algebraic properties are preserved. A field embedding preserves the algebraic structure, an embedding of a topological space preserves open sets, and a graph embedding preserves connectivity. A space A is embedded in another space B when the properties of B restricted to A are the same as the properties of A. The rationals are embedded in the reals, and the integers are embedded in the rationals. In geom-

---

[75] Ref. to Riemannian geometry and topology.

etry, the sphere is embedded in $\mathbb{R}^3$ as the unit sphere. Let us specify the notion of "embedding":[76]

A *differentiable manifold* of class C^k (resp. C^∞, C^ω) is the pair of a *manifold* and an *equivalence class* of *atlas* C^k (resp. C^∞, C^ω) (see def. atlas in Appendix). (Ω, φ) will be a *local map* at P, $(x^1, \ldots, x^n)$ the associated coordinate system. A map f from a differentiable manifold $C^k : W_p$ to another V_n is differentiable C^r ($r \leq k$) at P if, $\widehat{\varphi} \circ f \circ \varphi^{-1}$ is differentiable C^r at $\varphi(P)$, and we define the *rank* of f at P as the rank of $\widehat{\varphi} \circ f \circ \varphi^{-1}$ at $\varphi(P)$. The map f is C^r is it is C^r at any point. This is an *immersion* if the rank of f is equal to p at any point of W. This is an *embedding* if f is an immersion such that f is a homeomorphism from W_p to $f(W_p)$ provided with a topology induced by that of V_n.

A *sub-manifold* of dimension p of a differentiable manifold V_n is a subset W of V_n such that, for any point $Q \in W$, there exists a local map (Ω, φ) of V_n where $\varphi(\Omega)$ is an open set of the form $\theta \times U$ with $\theta \subseteq \mathbb{R}^p$ and $U \subseteq \mathbb{R}^{n-p}$ such that $\varphi(\Omega \cap W) = \theta \times \{0\}$. Thus there exists a system of local coordinates $(x^1, \ldots, x^n)$ on a neighborhood of Q in V_n such that W_p is locally defined by $x^{p+1} = x^{p+2} = \cdots = x^n = 0^p$. Often the subset of $W \subset V_n$ will be defined by a system of $q = (n-p)$ equations $f_i(M) = 0$ ($1 \leq i \leq q$) where the f_i are functions C^1. If the map from V_n to $\mathbb{R}^q$ defined by $M \rightarrow (f_i(M), \ldots, f_q(M))$ is of rank q at any point M of W, then W is a sub-manifold of dimension p. We suppose that the manifolds are connected. Otherwise we study the connected components the one after the other. The definition of a differentiable manifold is abstract. But in fact, this is a hypersurface of dimension n in $\mathbb{R}^p$ with $p > n$. The manifold is supposed countable at infinity (there exists a sequence of compacts $K_i \subseteq V_n$ such that, for any i, $K_i \subset \mathring{K}_{i+1}$ and $\cup_{i \geq 1} K_i = V_n$), Whitney proved the theorem

Theorem 1.21 (Whitney). *A differentiable connected manifold C^1 of dimension n admits an embedding in $\mathbb{R}^{2n+1}$.*

This theorem (1936) states that every n-dimensional manifold admits an embedding into $\mathbb{R}^{2n+1}$ (and concerns a general class of maps). The theorem states that a *generic map* from an n-manifold to $2n + 1$-dimensional Euclidean space is an embedding: the image of the n-manifold is completely *unfolded* in the larger space. In particular, no two points in the n-dimensional manifold map to the same point in the $2n + 1$-dimensional space. As $2n + 1$ independent signals measured from a system can be considered as a map from the set of states to $2n + 1$-dimensional space, Whitney's theorem implies that each state can be identified uniquely by a vector of $2n + 1$ measurements, thereby reconstructing the phase space.

---

[76] **Definition (Embedding).** In short, a continuous map $f : X \rightarrow Y$, between two topological spaces is called an embedding, *if it is a homeomorphism on a subspace of Y.* (The notation $f : X \hookrightarrow Y$ is often used for embeddings.) The embeddings correspond to the subspaces. f and the inclusion map of the subspace $f[X]$ into X differ only up to a homeomorphism.

1.39.2 Takens Theorem (1981): A Delay Embedding Theorem

The Takens theorem (1981) corresponds to a delay embedding theorem that gives the conditions under which a chaotic dynamical system can be reconstructed from a sequence of observations of the state of a dynamical system. The reconstruction preserves the properties of the dynamical system that do not change under smooth coordinate changes, but it does not preserve the geometric shape of structures in phase space. The (delay embedding) Takens theorem provides the conditions under which a smooth attractor can be reconstructed from the observations made with a generic function. (Ulterior results replaced the smooth attractor with a set of arbitrary box counting dimension and the class of generic functions with other classes of functions.) The theorems published by Takens (1981) proved the link between dynamics of a true unknown system in the (high dimension) phase space and dynamics defined in a pseudo-phases space of a system reconstructed from observations. The equations of the movement (or system) are not necessarily known. Consider a series of observations $Y(t), Y(t + \tau_k), Y(t + 2\tau_k), \ldots$ taken as outputs of a dynamical system containing a certain number of variables which undergo certain dynamical laws. We want to determine its qualitative characteristics by only analyzing $Y(t)$. This aim seems impossible because we do not know the role of the other variables in the determination of the system; We do not know the dimension of the system that can be indefinitely large, nor the numbers of variables and equations involved in the dynamics. However, two arguments make possible this aim:

- The attractors of large (or infinite) systems can have low dimensions. Consequently, if we only take into account the asymptotic behavior of the system, the study of the geometrical objects of low dimensions can provide all information needed.
- The extraction from a reconstructed time series of the needed information about the original system is possible by applying the reconstruction theorems.

To this end, let us take n successive elements of the time series of Y, that we write $Y(t), Y(t + \tau_k), Y(t + 2\tau_k), \ldots, Y(t + (n-1)\tau_k)$. These elements can be regarded as a vector that defines a point in $\mathbb{R}^n$. Observe now the evolution in the course of time of this vector: an orbit is obtained in a space A included in $\mathbb{R}^n$. It can be shown that for almost any time delay denoted τ_k, the dynamics in the pseudo-phase space A, have the same properties as the original system (a condition concerning the size of n is to be sufficiently large compared to the original attractor).

1.39.2.1 Takens Theorem by Broomhead and King (1986)

Theorem 1.22 (Takens). *Let M be a compact manifold of dimension m and (F, v) a pair where F is a set of smooth vectors and v is a smooth function on M.*[77]

---

[77] (1) *The differentiability (possibly of a high order) of a function is regarded as a smoothing.* Thus, regarding a smoothed function, usually one considered a differentiable function (of a high order).

$$\Phi_{F,v}(y) : M \longrightarrow \mathbb{R}^{2m+1} \tag{1.413}$$

defined by: $\Phi_{F,v}(y) = (v(y), v(\varphi_1(y)), \ldots, v(\varphi_{2m}(y)))^T$. $\Phi_{F,v}(y)$ *is an embedding in which* φ_t *is a flow of F.*

Where ^{T} is the transpose, $v(y)$ is regarded as the measurement of the system at the point $y \in M$ (i.e. here $v(\varphi_i(y))$ corresponds to an observation of Y at time i). The space which contains the image of $\Phi_{F,v}$ is called *embedding space* and the dimension of this space is n. It is important to explain how the measurement of the system is made on $\Phi_{F,v}(M) : (v_1, v_2, \ldots, v_i, v_{i+1}, \ldots)$ with $v_i \equiv v(\varphi_i(y))$. n denotes the dimension of the embedding. We suppose $n \geq 2m + 1$ in order to satisfy the Embedding Whitney theorem. The Takens theorem insures that if n is sufficiently large, the n-dimensional image of the attractor provides a *faithful* (topological) representation to the origin, e.g. periodic orbits on an attractor correspond to periodic orbits in the reconstructed phase space, and chaotic orbits of the original system will appear chaotic in this space, etc. Note that a chaotic attractor obtained from the reconstructed series will have a positive Lyapunov exponent as the corresponding attractor in the initial system.

$y(t)$ is the state of system at the moment t.

$y(t)$ is a point in the phase-space S

$y(t) = (y1, y2, \ldots)$

Si $t = 0, \ldots, p$: $y = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(.) \end{bmatrix}$

y: portrait in the phase-space

S: is the phase-space.

F: is a smoothing vector field.

V: is a smoothing function on M.

M: is a manifold of dimension m.

(Source: D.S Broomhead & G.P King 1985).
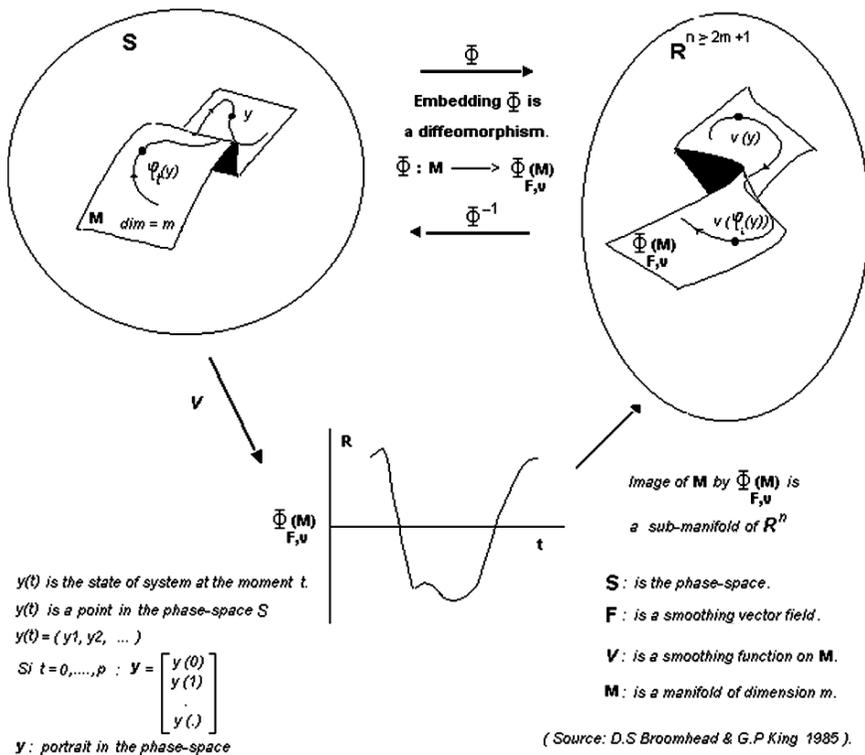
(2) A smoothed function could be also a selection, with regular intervals, of points of a function or a signal for example.

The dynamical system: The state of a system is denoted $y = (y_1, y_2, ..)$, where y represents a point in the phase space of the system. It is possible to indicate the phase space by S. Given *a linear operator* applied to the points belonging to $S : F(y) = dy/dt$. If we denote φ_t a family of one-parameter maps of the phase space in itself, we can write: $y(t) = \varphi_t\, y_0$ with y_0 as the initial point of the dynamics. Thus $\varphi_t\, S$ can be considered as a flow of points of the dynamics in the phase space. *In a dissipative system* the dimension of $\varphi_t\, S$, which at the beginning was that of S, can contract towards a lower dimension. Thus, if the flow contracts, we can thus represent the existence of the attractor. This attractor, which has a propensity to lower the dimension of the flow, exists and can be represented inside a field of a dimension lower than the phase space S. This field which is "smoothed" contains the attractor and is denoted M, thus we can write: $dimM < dimS$. The dynamics on M is easier to (analytically and numerically) study and handle, knowing that its dimension is lower and the number of degrees of freedom is lower. *The purpose is thus to consider that there is a topological equivalence between the initial attractor and the reconstructed attractor.*

Topological equivalence and phase-portrait, Poincaré map: The solution of the differential system $F(y) = dy/dt$ is equivalent knowing the family of maps φ_t. The *Poincaré map* allows to study the orbit to observe the phase portraits. Given two differentiable vector fields F and G of class C^r, i.e. r times differentiable, they are known as of C^k-*equivalent class* $(k \leq r)$ if there is a diffeomorphism $\Phi, (C^k)$, which allows to pass from the orbits $\varphi_t(y)$ of F to the orbits $\psi_{t'}(\Phi(y))$ of G and preserves their orientations. The aim is to describe a *topological equivalence* which preserves the properties of the original set.

Remark 1.10. The contribution of the Takens Embedding Theorem (1981) was to show that the same goal as the Whitney theorem (1936) could be reached with a single measured quantity. (The idea of using time delayed coordinates to represent a system state is reminiscent of the theory of ordinary differential equations, where existence theorems say that a unique solution exists for each $(y(t), \dot{y}(t), \ddot{y}(t), \ldots)$.)

Some references concerning reconstruction theorems: Abarbanel (1996), Sauer et al. (1991), Ott et al. (1994), Sauer (1994), Broomhead and King (1986), Aeyels (1981), Takens (1981), and Whitney H (1936).

1.39.2.2 Application to Nonlinear Signals

Nearly 20 years ago, new techniques have been developed to analyze and treat the signals corresponding to nonlinear dynamics of the chaotic type. These techniques are gathered under the name of "nonlinear signal processing". They are not built on the Fourier transform as it was the case in one of the previous sections. Here, we postulate that nonlinear dynamics are inserted within a "deterministic" framework as opposed to traditional methods which are placed within a purely "stochastic" framework that we try to characterize. The techniques of nonlinear signal processing are based mainly on reconstruction theorems. These theorems shows that from

a time series, we can built a unidimensional or multidimensional trajectory *topo-logically equivalent* to the initial trajectory extracted from the phase-space of the initial series. The trajectory reconstructed is represented in a pseudo-phase-space. Since we reconstruct the series according to the dimension of the system, with one or more delays, we posed the basis of a signal processing problem where we will handle the series in the pseudo-phase space. For a unidimensional series $X(t)$, the reconstruction can be done in a two-dimensional plane: $[X(t), X(t + \Delta t)]$.

Reconstruction of the logistic function for $\alpha = 3, 3.4, 4$.

$\alpha = 3$ $\alpha = 3.4$ $\alpha = 4$: chaotic domain
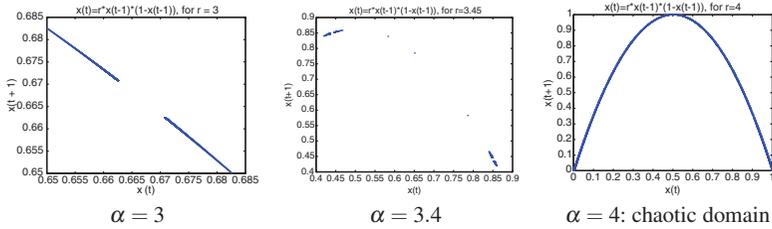
Reconstruction variants with an increasing delay for $\alpha = 4$.

$[x(t), x(t+2)]$ $[x(t), x(t+3)]$ $[x(t), x(t+3)]$

$[x(t), x(t+5)]$ $[x(t), x(t+6)]$ $[x(t), x(t+8)]$
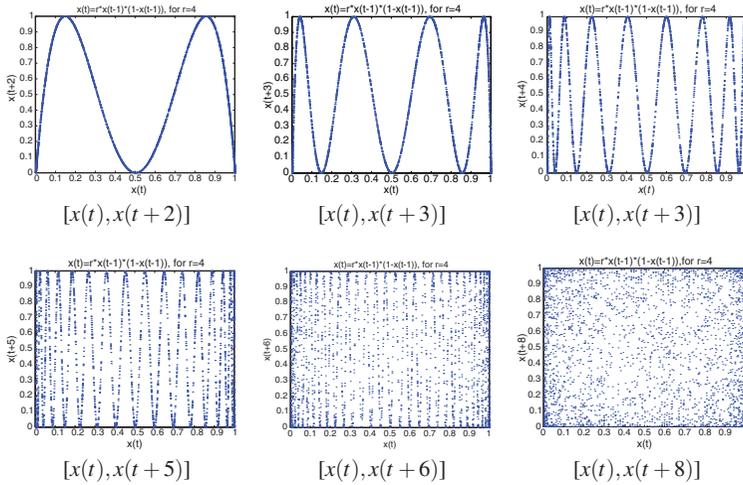
1.39.3 (n, J)-Window Concept

The window concept is directly derived from the principle mentioned in previous theorems. This concept makes visible n elements of the time series. We write: if $J = 1$, the elements are consecutive and if $J > 1$, then there is an interval between each visible element. Consequently, the $(n, 1)$-window corresponds to the (n)-window.

The elements thus selected in the window constitute the components of a vector in the *Embedding-space* $\mathbb{R}^n$ (With $\dim(\mathbb{R}^n) \geq \dim(\mathbb{R}^{2m+1})$). The vector made up in this space characterizes a trajectory. This trajectory is represented as follows: $x_i = \Phi_{F,v}(\varphi(y)) = (v_i, v_{i+J}, \ldots, v_{i+(m-1)J})^T$. Here is the example of the construction of a vector in an embedding space $\in \mathbb{R}^5$.

Example of a (5)-window: We denote: (a) the starting vector, (b) selection, (c) reconstructed vectors.

$(a): \{v_1, v_2, \ldots, v_{i-1}, v_i, v_{i+1}, v_{i+2}, v_{i+3}, v_{i+4}, v_{i+5}, \ldots\}$
$(b): v_1, v_2, \ldots, v_{i-1}, [v_i, v_{i+1}, v_{i+2}, v_{i+3}, v_{i+4}], v_{i+5}, \ldots$
$(c): \qquad\qquad x_i = (v_i, v_{i+1}, v_{i+2}, v_{i+3}, v_{i+4})^T$

As a consequence, we will be able to note X a vector trajectory constructed from a (5)-window in the following way:

$$X = N^{-\frac{1}{2}} \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \end{bmatrix} = N^{-\frac{1}{2}} \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_5 \\ v_2 & v_3 & v_4 & v_5 & v_6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{1.414}$$

This trajectory matrix will be used in the next chapters, in particular within the framework of the Singular Spectrum Analysis that we will apply to the logistic equation and to the French stock index.

Chapter 2
Delay Model, SSA and Brownian Motion

This chapter presents three other tools to approach complex, nonlinear and chaotic dynamics. We will consider the *Delay-model*, the *Singular Spectrum analysis* and the *Brownian motions* (fractional or non-fractional). Firstly, we present the delay-model which is applied to the logistic equation. According to the Medio's work, a discrete-delay is integrated into the construction of an economic model by means of a convolution. The lengths of lags are distributed in a random way in the population. The delay is in fact modelled by means of a random variable which is characterized by its probability distribution. We will notice that in this way, the system built rocks more tardily to the chaos. We will observe a shift of bifurcation points, but also an *unhooking* in the trajectory.

Delay-model applied to the logistic equation. We will use the equation with the first-order differences used by Robert May. The central element of the model is the concept of "delay". For a macroeconomic consumption model for example, if we postulate that there is an (unspecified) great number of agents, and that all these agents answer to a certain stimulation with given discrete-lags, the lengths of lags are different for various agents and are distributed in a random way in the population. In a global model, in the whole population, the reaction times are aggregate. In the described case, we can model the reaction-time by means of a random variable that will represent the global length of the lag.

The Singular Spectrum Analysis is the second investigation tool of complex dynamics presented in this chapter. The method associates the *Takens reconstruction* technique and a technique known in *matrix algebra* which is *the Singular Value Decomposition*. In such a framework, the purpose is to project a time series on a basis of *eigenvectors extracted* from this same time-series. We project a *trajectory matrix* on the space described by the *eigenvectors of the covariance matrix of the time series. The eigenvalues obtained* can be ordered and be the subject of a filtering with the aim to extract the *deterministic* part of the signal cleaned of its background noise. The SSA was used in signal theory, and its applications to dynamical system theory have been introduced by Broomhead and King in 1986, in connection with their version of the Takens theorem (see Chap. 1). The method is presented in the framework of the delay-model behavior applied to the logistic equation, but also on the French stock index.

The last concept described in this chapter is the Brownian motion, which is a formidable tool to study chaotic behaviors. We will describe their construction and experiment different types of Brownian motions, fractional ($H \neq 1/2$) or non-fractional ($H = 1/2$). The concepts associated with Brownian motions, such as *persistence*, *memory*, *Levy distribution*, *fractal dimension* and *Rescaled range* statistics are approached in this heading.

2.1 Delay Model Applied to Logistic Equation (Medio)

2.1.1 Nonlinearities and Lags

Here, the first-order difference equation of R.May is used and its general form is: $X_{t+T} = G(X_t)$, where T represents the length of the *lag* and G is a *smoothed one-hump function*. Such an equation can be understood as an aggregate system (i.e. a one-loop feedback system) with two components (1) a nonlinear relationship (here, single-hump functions) and (2) a lag (here, fixed delay). For (1): Nonlinearities are widely used in Economics, for example in rational consumption models, overlapping generations models, optimal growth models, we can mentioned: Sutzer (1980), Day (1981–1982), Benhabib and Day (1981), Grandmont (1985), Pelikan and Deneckere (1986), Baumol and Benhabib (1988), Lorenz (1989), Boldrin and Woodford (1990), Scheinkman (1990). For (2): The notion of lag is neglected in Economics, however it can be a source of important developments in connection with the notions of aggregates, agent behaviors, stochastic processes, aggregate models and chaotic systems.

2.1.1.1 Lag Distribution of Agents

This notion of "lag" has been revisited by A. Medio and can find a significance in certain economic models (e.g. rational consumption or overlapping generations models).[1] Suppose that there is an unspecified great number of agents and that all these agents *respond to a certain stimulation with given discrete lags. The lengths of lags are different for various agents, and are distributed in a random way in the population*. The subject here is the "reaction time" of agents. In a global model, in the whole population the reaction times are aggregate. In the described case, we can model the reaction time by means of a random variable, real, positive or null, that is called T (in accordance with the equation posed at the beginning of chapter) and will represent the Global length of the lag. A random variable is characterized by its probability distribution (if it is known).

---

[1] Ref. to Medio publications, in particular: Medio (1992).

2.1.1.2 Convolution of the Lag and Reaction of Agents

Let $X(t)$ be a variable that is function of another variable $Z(t)$ through a continuously distributed lag.[2] (Note that Z may indicate the same variable X at some time different from t.) m can be understood as a "weighting function" (i.e. a kind of "moment") that formalizes the strength of impact that values of Z in the more or less distant past have on the value of X (i.e. a kind of temporal correlation). The equation of the lag can given by:

$$X(t) = \int_0^\infty m(s)Z(t-s)ds \qquad (2.1)$$

with m continuous on $\mathbb{R}$, and $\int_0^\infty m(s)ds = 1$. In practice, s is bounded by t and the previous equation becomes $X(t) = \int_0^t m(s)Z(t-s)ds$. This equation can be taken as a (commutative) convolution.[3] Given the polynomial $v(p) = a_0 p^n + \cdots + a_n$, with $\mathscr{L}$ the Laplace transform that is written $\mathscr{L}[m(s)] = 1/v(p) = \mathscr{L}[\mathscr{L}^{-1}[1/v(p)]]$. Then $m(s)$ is defined as the inverse Laplace transform of $1/v(p)$, thus, $m(s) = \mathscr{L}^{-1}[1/v(p)]$, then $X = m \star Z$ is the solution of the differential equations $v(D)X(t) = Z(t)$ with $D \equiv d./dt$ and the initial conditions $X = X' = \cdots = X^{n-1} = 0$. If the lag is an *exponential lag* of order n, we have $\frac{1}{v(p)} = \left(\frac{\tau p}{n} + 1\right)^{-n}$. Thus the differential equation $v(D)X(t) = Z(t)$ is written $\left(\frac{\tau D}{n} + 1\right)^n X(t) = Z(t)$ or also

$$X(t) = \left(\frac{\tau D}{n} + 1\right)^{-n} Z(t), \qquad (2.2)$$

where $n \in \mathbb{Z}^+$ and τ is the time constant of the lag. For multiple exponential lags, $m(t)$ can be calculated through the inverse Laplace transform, we have: $m(t) = \left(\frac{n}{\tau}\right)^n \frac{t^{n-1}}{(n-1)!} e^{-nt/\tau}$. For $n \geq 2$, $m(t)$ has a one-hump shape:

(a) When $n = 1$, we deal with the ordinary differential system which can be written in the following way: $\dot{X}(t) = \gamma(Z(t) - X(t))$ where γ represents the speed of adjustment of the model (note that $1/\gamma = \tau$).
(b) When n *becomes large*, the weighting function tends to a Dirac delta function and the exponential lag tends to a fixed delay of length τ. Then we have: $\lim_{n\to\infty} \left(\frac{\tau D}{n} + 1\right)^{-n} = e^{-\tau D}$ (Fig. 2.1).

At this stage, we have to combine the delay and the model. If we place the lag in the generic initial model $X_{t+T} = G(X_t)$, this can be written:

$$X_n = \left(\frac{D}{n} + 1\right)^{-n} G(X_n). \qquad (2.3)$$

---

[2] The lag can be understood as a shift in the reaction time of agents.
[3] See Appendix.
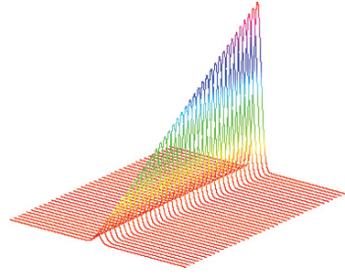
After the application of the factor $(\frac{D}{n}+1)^{-n}$ to $G(X_n)$ we encounter the value of
X_n again. We can turn over the equation, and move the lag term on the other side
of the equality and change the sign of the exponent. Then we write the following
equivalent relations that describe a differential equation system:

$$((D/n)-1)X_i = X_{i-1}, \quad \text{with } i = 2\dots n, \tag{2.4}$$
$$((D/n)-1)X_1 = G(X_n), \tag{2.5}$$

where $G(X_n)$ is a one-hump function. The system can also be written:

$$((D/n)+1)X_2 = X_1,$$
$$\vdots \tag{2.6}$$
$$((D/n)+1)X_1 = G(X_n).$$

2.1.2 Application to the Logistic Equation

The logistic equation is written $X_{n+1} = \alpha X_n(1 - X_n)$. Consequently: $G(X_n) = \alpha X_n(1 - X_n)$ and the lag takes the form $(\frac{D}{n}+1)^{-n}$, $X_n = ((D/n)+1)^{-n}(\alpha X_n(1 - X_n))$. We face a *nonlinear functional relation* and a *delay-function*. We will see that
the application of the lag to the logistic model shifts the chaotic zone.

2.1.2.1 Solution of the Logistic Model

At the equilibrium we have $X_1 = X_2 = X_3 = \dots = \bar{X}$, thus for $X_{n+1} = \alpha X_n(1 - X_n)$,
there are the equilibrium solutions: $\bar{X} = 0$ and $\bar{X} = 1 - (1/\alpha)$. The stability of
these equilibrium points was previously studied in Chap. 1. Finally, we write the
differential system with the delay function as follows:

2.1.2.2 Differential System Combining Logistic Model and Delay Funct

The system of ordinary differential equations of dimension n is:

$$((D/n)+1)X_2 = X_1,$$
$$\vdots \tag{2.7}$$
$$((D/n)+1)X_1 = \alpha X_n(1-X_n).$$

For a system of dimension 10 and of length N, we write:

$$((D/10)+1)X_2 = X_1,$$
$$\vdots \tag{2.8}$$
$$((D/10)+1)X_2 = \alpha X_{10}(1-X_{10}).$$

By simple multiplication, the system is written:

$$((\dot{X}_2/10)+X_2) = X_1,$$
$$\vdots \tag{2.9}$$
$$((\dot{X}_{10}/10)+X_{10}) = X_9,$$
$$((\dot{X}_1/10)+X_1) = \alpha X_{10}(1-X_{10}).$$

The differential system can be written in matrix form:

$$[\dot{X}] = [A]*[X]. \tag{2.10}$$

2.1.2.3 Figures of Various Simulations for $\alpha \in [3,\dots,5]$

Pictures of differential system solutions for different α values (Figs. 2.2–2.9).

For each value of alpha, we have to face a rectangular matrix made with vectors X of dimension $10 \times N$. A weight $(\frac{D}{10}+1)$ is applied to each vector of the matrix, and this weight changes the trajectory.

2.1.2.4 Shift of Bifurcation Points, and Periodicity

For the logistic model the ultimate bifurcation before chaos occurs for $\alpha \simeq 3.56$. In our case, for a system dimension equal to $10 \times N$, this occurs for $\alpha \simeq 5$. The period-doublings visible in the graphs of the preceding section are more tardy under the effect of delay function. The periodic behaviors and their periods are identifiable with the numbers of distinct orbits in the figures. Starting from the value equal to 5, the trajectory seems to describe different and distinct orbits which do not overlap, contrary to what occurred for $\alpha = 4$, for example, where if we eliminate the

Fig. 2.2 $\alpha = 4$. Transitory behavior was preserved before convergence
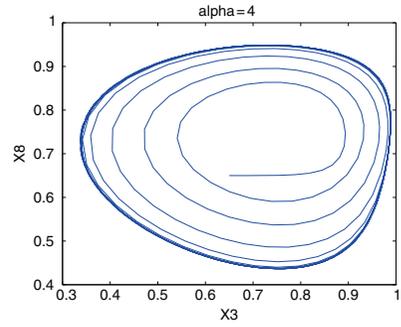
Fig. 2.3 $\alpha = 4.35$. Transitory behavior suppressed to make the orbit visible
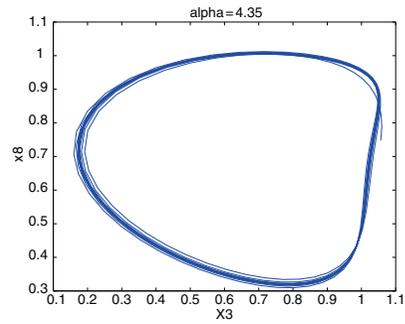
Fig. 2.4 Period-doubling. $\alpha = 4.5$. Asymptotic behavior
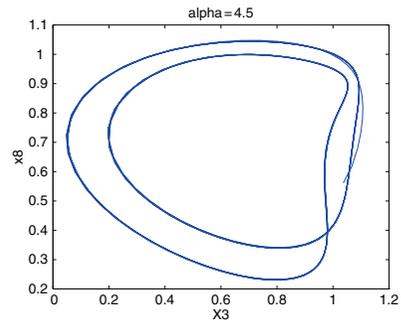
transitory behaviors before convergence, we have a *"dense closed" orbit* without *unhooking* and *period-doubling*. Moreover, in order to underline the periodicities, we can analyse the spectrum of these trajectories and calculate the Lyapunov exponent to highlight the "moment" during which the system rocks towards a pure chaotic behavior, as we could do it in the first chapter for the logistic equation itself. We did not represent the *sensitive dependence on initial conditions of the system* which is the characteristic of chaos nor the value of the *Lyapunov exponent* during the evolution of the system on the route towards chaotic zone. The capacity dimension can also be measured rather easily. The set of these elements converges

Fig. 2.5 $\alpha = 4.75$
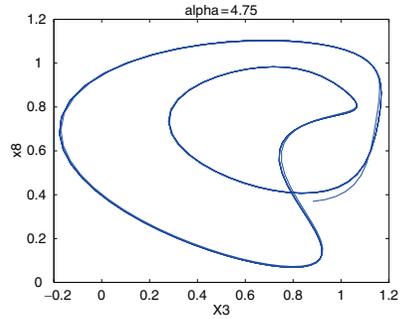
Fig. 2.6 $\alpha = 4.85$
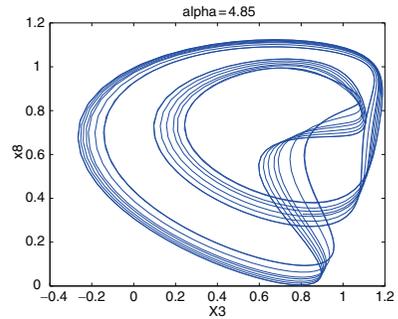
Fig. 2.7 $\alpha = 4.95$
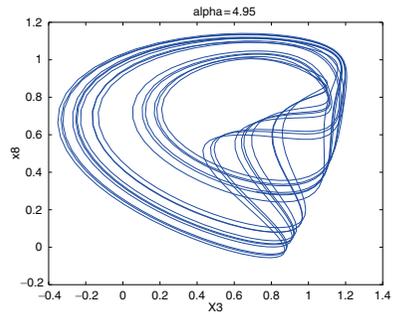
towards a value of alpha equal to 5 to characterize the critical point. *The value of the capacity dimension for* $\alpha = 5$ *is close to 2.15, which is a non-integer value and characterizes the presence of a simple attractor.* The principle of the exponential lag makes it possible to model a large variety of economic situations by controlling the parameters T and n. The nonlinearities of one-hump functions coupled to a delay function produce chaos. One of the characteristics of this exponential lag is that the variance which was used to build it is low. Thus we are close to the aspect of a Dirac function. In conclusion, it is possible to say that a high order lag, for example $n = 10$ is a condition for chaos occurrence. The type of model that we have just seen can find applications in macroeconomics, in particular in the models of

Fig. 2.8 $\alpha = 5$. Orbit and chaotic attractor, with unhooking
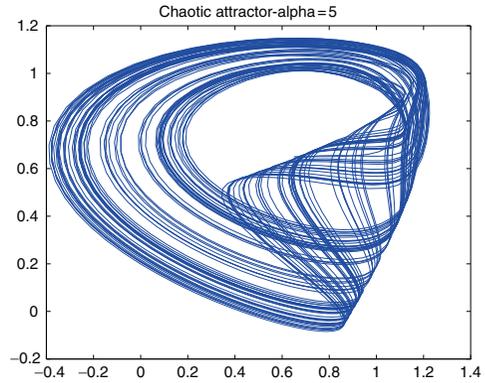
Fig. 2.9 Behavior of $X(8)$ with $n = 10$
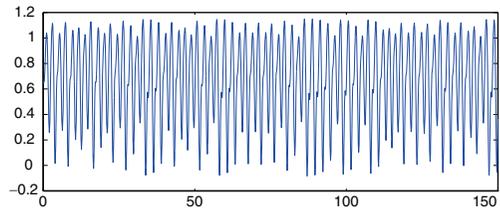
production–consumption, for example the developments of Grandmont models on the equilibrium cycles ("Overlapping-generations models").

2.2 Singular Spectrum Analysis

2.2.1 Singular Spectrum Analysis Principle: "Windowing", Eigenvector and Projection

The SSA method is built from a technique known in matrix algebra as the "singular value decomposition". This method consists in projecting a time series on a basis of eigenvectors extracted from this same time-series. Or more exactly, it is the projection of a matrix trajectory, built from the initial time series, on the eigenvectors of an intermediate matrix, itself built from the studied experimental series. The SSA was used in signal theory, and its applications to the dynamical system theory have been introduced by Broomhead and King in 1986 (see Chap. 1). Their version of the Takens theorem states that the space containing the image of the map $\Phi_{F,v}$ is called the embedding-space and its dimension n is called embedding dimension.[4]

---

[4] See Takens theorem. Recall: $\Phi_{F,v} : A \longrightarrow R^{2m+1}$.
    $\Phi_{F,v}(y) = (v(y), v(\phi_1(y)), \ldots, v(\phi_{2m}(y)))^T$.

This is the reconstruction of phase space of solutions of an arbitrary dynamical system. And the dimension of this reconstructed phase space is also the dimension of the embedding space. It is pointed out that the conclusions of the Takens theorem impose the following constraint $n \geqslant 2m + 1$ with:

- m: Dimension of the attractor
- n: Dimension of the embedding space

Consider for example the values taken by a variable $Y(t)$ with regular time intervals denoted τ, then we can write this without dimensioning the number of observations in the following way $Y(t), Y(t + \tau), Y(t + 2\tau), \ldots$. If the set is dimensioned, we can write a group of observations of the variable $Y_p = Y(p\tau)$ as follows with $p = 1, \ldots, n$ and where τ is a "step" corresponding to a "periodic measurement of the variable". We will use the concept of (n, τ)-window presented in the first chapter (see Takens theorem). Here, it is pointed out that a window is the combination of two criteria: the number of selected measurements that becomes, in fact, the length of the series and the step τ which is the periodicity of the measurement. Thus, the periodicity of the measurement with the number of measurements makes visible n elements of the initial time series, which is therefore sampled on intervals of length τ. An optimal approach in a discrete case makes us take τ equal to 1, i.e. each measurement of the value of the variable, without descending to the value of the step below the unit. Thus for a (n, τ)-window, we make visible n elements of the initial time series. In the continuous case, things are different since we can construct windows with the length we select and show exactly what we want about the model. This makes much more easier the handling of trajectories. Our limits in this construction are rather those fixed by what we wish to exhibit of the trajectory, i.e. in general the attractor and in particular to operate its reconstruction whose principal constraint is $n \geqslant 2m + 1$. The subject is to manage scale problems that are important in this type of construction. For the choice of a (n, τ)-window which obviously is of the size n, we must generate, by means of the model equation, a discrete time series of a sufficient length so that the window can "exist" and in a significant way.

(a) *Classical method of reconstruction and trajectory matrix.* It is possible to construct a set that is sequence of vectors in the embedding space of dimension n, which is written: $\{x_i \in \mathbb{R}^n \mid i = 1, \ldots, N\}$. Each vector x_i of this set is a point in $\mathbb{R}^n$. The set contains thus N points, each one of the dimension in $\mathbb{R}^n$. Consequently, we represent the set called "n-history" by a rectangular matrix of dimension $N \times n$. We can write the constraint on the set:

$$N = N_o - n + 1, \tag{2.11}$$

with N_o: number of observations (i.e. the length of the series), N: dimension of the reconstructed series, and n: dimension of the embedding space. These sequences are

---

Knowing that $n \geq 2m + 1$. ϕ_t is a flow of F. ϕ is a map representing the dynamics such that: $y_{j+1} = \phi(y_j)$. And the function $v(y)$ can be taken as a measurement made on the system at the point $y \in A$, i.e. $v(\phi_i(y))$ would be equal to an observation of Y at time i.

used to build a matrix X, which is the trajectory matrix of dimension $N \times n$. If we pose the trajectory matrix in the following way:

$$X = N^{-\frac{1}{2}} \begin{pmatrix} X_1^T \\ \vdots \\ X_N^T \end{pmatrix}, \quad \text{or } X = N^{-\frac{1}{2}} \begin{pmatrix} (x_1^1, \dots, x_1^n) \\ \vdots \\ (x_N^1, \dots, x_N^n) \end{pmatrix}, \quad (2.12)$$

with $\dim(X) = (N, n)$, and the factor $N^{-\frac{1}{2}}$ being introduced by convenience. "By plotting the columns of X against the principal directions of the embedding space while respecting $n \geqslant 2m + 1$, we obtain the reconstructed attractor" according to Takens method.

(b) *The SSA method.* The previous method has evolved. A better reconstruction technique has replaced the Takens method. We will describe it hereafter. In spite of a rather long presentation, the resulting system is of a quite simple handling. Consider a set of vectors

$$\left\{ s_i \in \mathbb{R}^N / i = 1, \dots, n \right\} \quad (2.13)$$

such as by their action on X, they generate a new set of vectors "linearly independent and orthogonal"[5] which is

$$\left\{ c_i \in \mathbb{R}^n / i = 1, \dots, n \right\}. \quad (2.14)$$

It is possible to assume that the vectors $\{c_i\}$ are also "normal" and provide a basis in $\mathbb{R}^n$

$$s_i^T X = \sigma_i c_i^T, \quad (2.15)$$

where $\{\sigma_i\}$ is a set of real constants used to normalize the vectors. From the algebra of matrices we know that:

$$s_i^T X X^T s_j = \sigma_i \sigma_j \delta_{ij}, \quad (2.16)$$

where δ is the Kronecker symbol.[6] In addition, the matrix $H = X X^T$ is real and symmetrical and, moreover, its eigenvalues form an orthogonal basis for $\mathbb{R}^N$. More precisely the eigenvectors of H satisfy the preceding equation $s_i^T X = \sigma_i c_i^T$. The extraction of eigenvectors and eigenvalues is made from the square matrix H whose

---

[5] We will refer to the following different notions: basis of vectors, linearly independent vectors, singular values, and "rank" of a group of vectors. The rank of a family of p vectors $(V_1, V_2, \dots, V_p)$ is the greatest number of linearly independent vectors among them. It is also said: row of the p vectors. It is also said: Rank of the p vectors $(V_1, V_2, \dots, V_p)$. The rank is lower or equal to p. If A is a matrix of vectors, we note by $k = rank(A)$ the number of singular values of A (which are larger than the $[max(size(A)) \cdot norm(A)]$).

[6] δ is the Kronecker symbol: *A family of vectors* $(v_1, v_2, \dots, v_s)$ *of* $\mathbb{R}^n$ *is said orthogonal system if for all* $(i, j), i \neq j, v_i \cdot v_j = \delta_{ij}$ *(where* δ *is the symbol of Kronecker:* $\delta_{ij} = 0$ *if* $i \neq j$, $\delta_{ij} = 1$, *if* $i = j$). *Any orthogonal system of non-null vectors is free.*

dimensions are very large, since they are those of the length of X. Then, we will see that it is easier to work with a matrix that is a variant of H and of which the dimension is much lower. This new square matrix will note V and its dimension will be equal to the width of matrix X. We can write:[7]

$$H s_i = \sigma_i^2 s_i. \tag{2.17}$$

Then we have the set $\{s_i\}$ of the corresponding eigenvectors of H, and the set $\{\sigma_i^2\}$ of the corresponding eigenvalues of H all real and non-negative (H being positive-semidefinite). H can be written as follows:

$$H = N^{-1} \begin{pmatrix} X_1^T X_1 & \cdots & X_1^T X_N \\ \vdots & \ddots & \vdots \\ X_N^T X_1 & \cdots & X_N^T X_N \end{pmatrix} \tag{2.18}$$

with $dim(H) : (N,n) * (n,N) = (N,N)$. The writing in the simplified form is:

$$H = X^T X. \tag{2.19}$$

H can be understood as a correlation matrix between pairs of vectors (generated by the window of size n). With this technique, we have to face an important difficulty: indeed, H has a dimension $N \times N$ usually very large and consequently its diagonalization is often impossible to practice.[8] To remove this serious constraint in particular, a new technique was developed.

(c) *SSA 2nd method.* A more effective method to obtain the desired result is the following: it is to take the "transpose" of the equation $s_i^T X = \sigma_i c_i^T$, thus:

$$X^T s_i = \sigma_i c_i. \tag{2.20}$$

If we pre-multiply by X: $XX^T s_i = \sigma_i X c_i$. By using the equation $H s_i = \sigma_i^2 s_i$ and by simplifying:

$$X c_i = \sigma_i s_i. \tag{2.21}$$

And if we pre-multiply the preceding equation by X^T and we introduce it into the equation $s_i^T X = \sigma_i c_i^T$ the following equation is obtained:

$$V c_i = \sigma_i^2 c_i, \tag{2.22}$$

---

[7] Let us note that:
  $f(V) - \lambda V = 0$; $AV - \lambda V = 0$. $H s_i - \sigma_i^2 s_i = 0$; $(H - \sigma_i^2) s_i = 0$
  Collinearity:
  $AV - \lambda V$
  $\downarrow \qquad \downarrow$
  $H s_i - \sigma_i^2 s_i.$
[8] See appendix about the diagonalization of matrices.

where $V \equiv X^T X \in \mathbb{R}^{n \times n}$, the extraction of eigenvectors and eigenvalues is carried out on the square matrix V

$$V = \begin{pmatrix} X_1^T X_1 & \cdots & X_n^T X_n \\ \vdots & \ddots & \vdots \\ X_n^T X_1 & \cdots & X_n^T X_n \end{pmatrix}. \tag{2.23}$$

The equation $V c_i = \sigma_i^2 c_i$ allows to deduce $\{c_i\}$ as the set of eigenvectors of V and $\{\sigma_i^2\}$ as the set of the corresponding eigenvalues of V. As said for the matrix H, note that V can be taken as a covariance matrix of observations. The number of "steps" (τ) will be equal to the dimension of the "embedding" n being appreciably smaller than N, the equation $V c_i = \sigma_i^2 c_i$ is much easier to treat than the equation $H s_i = \sigma_i^2 s_i$. The numerical calculation of eigenvectors and eigenvalues of H was often very long, sometimes even impossible, due to the high dimensions of the matrix, directly connected with the length of observed vectors. For example, for an experimental series constructed from 1,000 measurements, the matrix H becomes a square matrix of dimensions $1,000 \times 1,000$ (i.e. millions of values) which has to be diagonalized to obtain the eigenvectors and eigenvalues. Often the calculations by H fail. On the other hand, the use of the matrix V (which is of a considerably reduced size) facilitates the calculation by its simplification. Let C be the matrix whose columns are composed by the c_i and $\Sigma^2 = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right)$, where the σ_i^2 are ordered from the largest to the smallest, i.e. $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2 \geq 0$. Consequently, the equation $V c_i = \sigma_i^2 c_i$ can be written:

$$VC = C\Sigma^2 \tag{2.24}$$

$X^T X \, c_i \equiv c_i \, diag\left(\sigma_1^2, \ldots, \sigma_n^2\right)$. Using the definition of V, consider $V \equiv X^T X \in \mathbb{R}^{n \times n}$, we have:

$$(XC)^T (XC) = \Sigma^2. \tag{2.25}$$

The matrix XC represents the trajectory matrix projected on the basis $\{c_i\}$. The subject is the choice of $\{c_i\}$ as a basis for the projection, i.e. to project the trajectory matrix onto the space spanned by the eigenvectors of the covariance matrix of the time series. This projection is optimal because the columns of the trajectory matrix are independent $(XC)^T (XC) = \Sigma^2$ and minimize the mean square error of the projection. ("Thus the plots are not squeezed any more onto the diagonal and the projections on the planes *(I, J) and (i + p, j + p)* are not equal any more, like with the Takens method" Medio 1992.) We will find an application of the method above in the section that follows. There are other developments of this method which approached the time-series with a background noise that disturbs the analysis. The statistical approaches that aim at sorting the eigenvalues of the matrix V, make it possible to "denoise" the reconstruction, but they are not depicted in the present work.

2.2.2 SSA Applied to the Logistic Equation with Delay Function

2.2.2.1 Projections and Reconstructions of the Initial Series

Consider the delay model applied to the logistic equation, whose system is:

$$X_{t+1} = \alpha X_t (1 - X_t), \quad X_n = ((D/n) + 1)^{-n} G(X_n) \quad \text{with } n = 10, \ \alpha = 5 \quad (2.26)$$

then we have $((D/n) + 1)^n + 1^n X = \alpha X (1 - X)$. Consider the trajectory of X_1 for $\alpha = 5$ to be positioned at the beginning of the chaotic regime, and then we will compute outputs for $\alpha = 3$, $\alpha = 4$, $\alpha = 20$. (Recall: Takens constraint for the attractor reconstruction is $n \geqslant 2m + 1$.)

(a) Observe the logistic system attractor for a function delay $\alpha = 5$ (Fig. 2.10).

X (above) is the solution of the delay model applied to the logistic equation for a parameter equal to 5. X has the shape of a rectangular matrix. Starting from the first vector x_1 of X, we will apply the SSA. Let us visualize x_1 (Fig. 2.11).

(b) Then, let us calculate the trajectory matrix by the first method with $N = N_o - n + 1$, $N = 1000$, $n = 10$, $N_o = 1,009$, $\dim(X) = (N, n) = (1000, 10)$:

$$X = 1000^{-\frac{1}{2}} \begin{pmatrix} X_1^T \\ \vdots \\ X_{1000}^T \end{pmatrix} = 1000^{-\frac{1}{2}} \begin{pmatrix} x_1^T(1) & \cdots & x_1^T(10) \\ \vdots & \ddots & \vdots \\ x_{1000}^T(1) & \cdots & x_1^T(10) \end{pmatrix}, \quad (2.27)$$

$$H = X^T X, \quad (2.28)$$

$\dim(H) = (N, n) * (n, N) = (N, N)$. $\dim(H) = (1000, 10) * (10, 1000) = (1000, 1000)$

$$H = 1000^{-1} \begin{pmatrix} X_1^T X_1 & \cdots & X_1^T X_{1000} \\ \vdots & \ddots & \vdots \\ X_{1000}^T X_1 & \cdots & X_{1000}^T X_{1000} \end{pmatrix}. \quad (2.29)$$
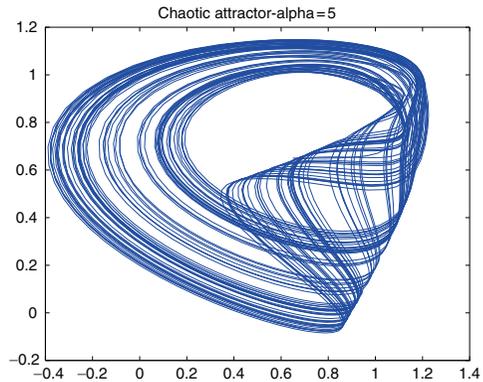
Fig. 2.10 Logistic system attractor for a function delay with $\alpha = 5$

Fig. 2.11 Behavior of x1
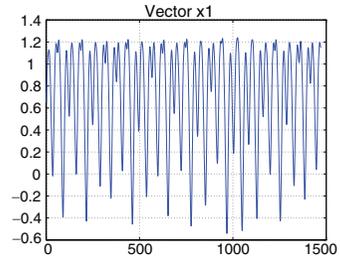
(c) Let us compute the new trajectory matrix V. According to what precedes, note that as the matrix H is heavy to handle, we use the following more efficient way: Given $\{s_i \in \mathbb{R}^{1000}/i = 1,\ldots,10\}$ and $\{c_i \in \mathbb{R}^{10}/i = 1,\ldots,10\}$ with $\dim(V) = (n,N) * (N,n) = (n,n)$, $\dim(V) = (1000,10) * (10,1000) = (10,10)$

$$V = X^T X = \begin{pmatrix} X_1^T X_1 & \cdots & X_1^T X_{10} \\ \vdots & \ddots & \vdots \\ X_{10}^T X_1 & \cdots & X_{10}^T X_{10} \end{pmatrix}. \tag{2.30}$$

(d) *Results.* The vectors and matrices contributing to the decomposition are:

  X which is the initial matrix of vectors
  X^T: the transpose of X
  $V = X \cdot X^T$: the product of both preceding vectors
  c: eigenvectors of V
  σ^2: eigenvalues of V
  $V \cdot c$: product of the matrix V with its eigenvectors
  $\sigma^2 \cdot c$: product of the eigenvalues with eigenvectors
  $X \cdot c$: product of the initial matrix of vectors with vectors of matrix V
  $(X \cdot c)^T \cdot (X \cdot c)$: product of transpose of previous projection with itself

If we extract the eigenvalues from V, by taking the diagonal of the matrix of eigenvalues σ^2 composed of $\sigma_1^2,\ldots,\sigma_n^2$, the values of the diagonal are: [0 0 0 0 0 0 0.0004 0.0148 0.3185 6.9033]. These eigenvalues are naturally ordered in an ascending order. And the construction of Σ^2 which is carried out in a descending order of eigenvalues does not give more information. But such a case is obviously not frequent in practice, especially if we increase the size of n. When the matrix $X \cdot c$ is constructed, several types of Poincaré sections are shown from the different components of $X \cdot c$ and with various steps. These projections of the attractor reconstructed by the SSA method are depicted in Fig. 2.12 ($\alpha = 5$). The results of this SSA construction could be compared with the projections resulting from the Takens method – not shown here –: in the Takens method, the figures would show orbit projections more or less *squeezed* on the diagonal according to selected "steps". Figure 2.12 of pairwise components of the matrix $X \cdot c$, we observe very different results. Indeed, in such a case, we do not observe any more squeezing on the diagonal. Note that each plane constructed with two different vectors from the matrix $X \cdot c$ provides different plots, whereas the Takens method always showed similar orbits.
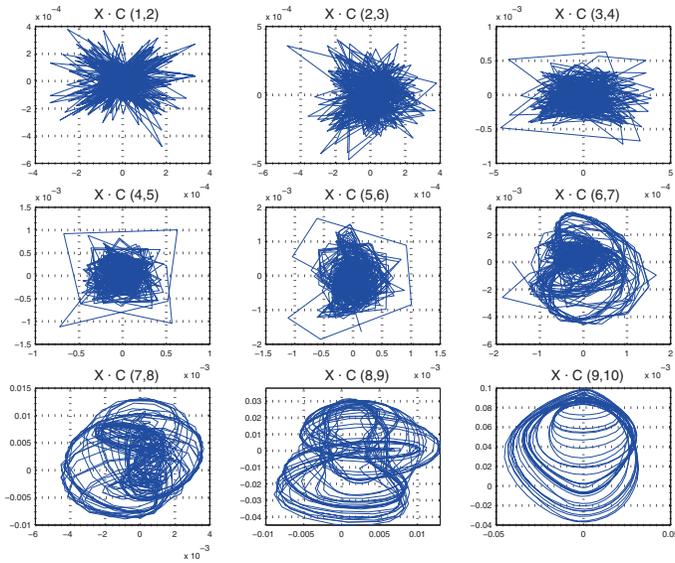
Fig. 2.12 Pairwise components of $X \cdot c$ by disregarding time for $\alpha = 5$

Thus, the projection of X on the eigenvectors c_i of C allows to isolate different structures of the signal, which correspond to the many aspects of the signal. This is due to the eigenvectors of V that are individually different. Periodicities and stationarities can be observed in the vectors $X \cdot c$. Note that the more the sequence number of the vector increases, the more its amplitude above and below zero increases. Then, we construct a sample of Poincaré sections similar to the previous construction and we *observe eigenvectors of V which constitute the basis on which X is projected*.

(f) Case $\alpha = 3.5$. The behavior of the system for this value of alpha is not chaotic but periodic. The orbit exhibits a spiral form for $X \cdot c(9, 10)$. Figure 2.13 shows different components of XC by disregarding time.

(g) *Case* $\alpha = 4$. Same remarks as previously, the orbit shows periodicities at the same time on the $V \cdot c$ vectors and also on the eigenvectors. The matrix XC exhibits all the periodicities in course of time (not shown here). Figure 2.14 show the various components of XC by disregarding time.

2.2.3 SSA Applied to a Financial Series (Cac40)

The method is based on the decomposition of a time series by using a basis generated by the initial time series itself. In Economics and Finance, it is usual to observe time series whose terms are autocorrelated. To study the series, it is often necessary to suppress the trend.

In order to stationarize the time series, a regression on time can be used or the nth-differences. Here we stationarize by the first-differences. As previously, we compute
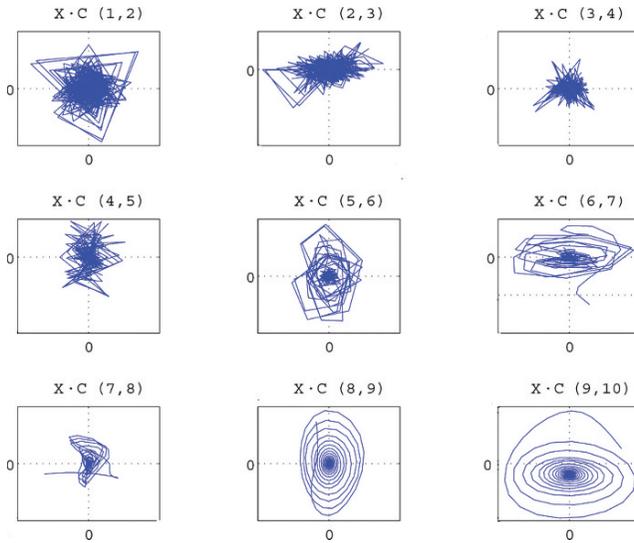
Fig. 2.13 Pairwise components of XC for $\alpha = 3.5$
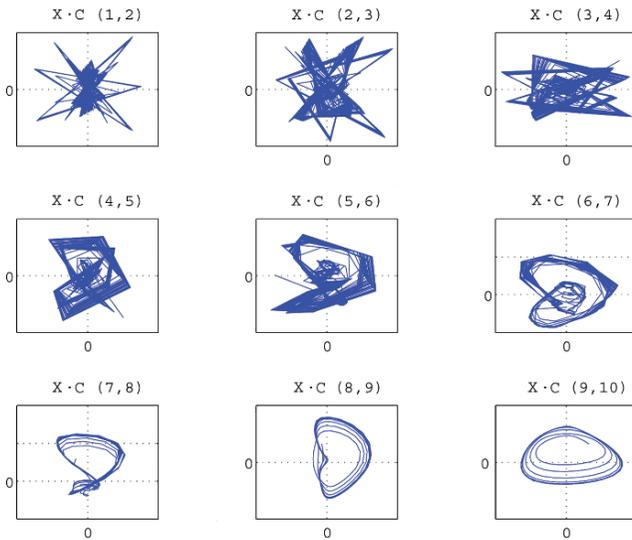
Fig. 2.14 Pairwise components of XC for $\alpha = 4$

the following vectors: $X, X^T, (V = X * X^T), c, \sigma_i^2, (V * c), (\sigma_i^2 * c), (X * c)$. If the diagonal is extracted $(\sigma_1^2, \ldots, \sigma_n^2)$ from matrix of eigenvalues of V, we obtain: [770.5 761.4 826.2 850.2 965.3 952.8 952.8 1,005.4 1018.5 693.5 647.1]. Note that the eigenvalues are not naturally ordered. In Fig.2.15 we project matrix X on the matrix of eigenvectors of V.
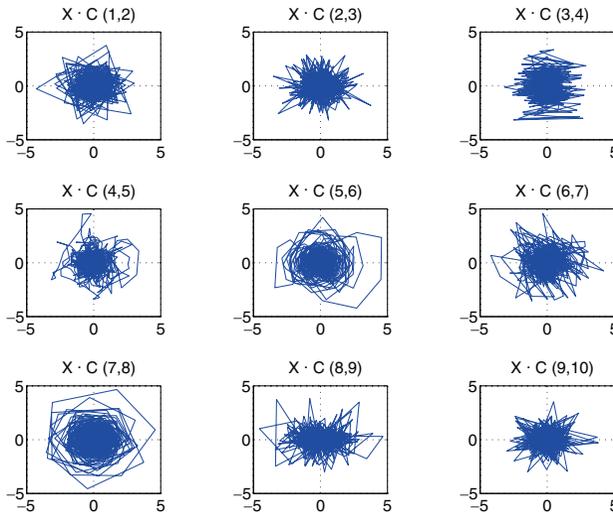
Fig. 2.15 Nine main components of $X \cdot C$ for SSA applied to a stock index (Cac40)

2.2.3.1 Role of Eigenvalues, and Filtering of Background Noises

One of the elements contributing to the SSA is $\Sigma^2 = diag(\sigma_1^2, \ldots, \sigma_n^2)$ (i.e. the vector containing the eigenvalues of the matrix V), whose eigenvalues σ_i^2 are classified in descending order from the largest to the smallest, i.e. $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_n^2 \geq 0$. After their classification, the elements σ_i^2 of Σ^2 are: [1,018.5 1,005.4 965.3 952.8 850.2 826.2 770.5 761.4 693.5 647.1]. From these ordered eigenvalues, we obtain a diagonal matrix: Σ^2. Here, the size of the window or the dimension of the embedding space is 10. One of the properties of the SSA method is the filtering of *background noise* of a time series. According to the words of A. Medio (1992, p. 186), this method allows "the identification of directions along of which the deterministic component of motion takes place, which we shall henceforth call significant (or deterministic) directions", whereas the infinite rest will be denoted by "stochastic directions". These assertions were applied to unidimensional systems in which deterministic chaos appear (e.g. logistic equation). Here, we will not compare the eigenvalues and the dimension of phase-space orbits of a dynamics, which has been largely presented by A. Medio in 1992 in "Chaotic dynamics". We will just say that we can split the embedding space n into a space in which the attractor is immersed, i.e. a subspace called d where the orbits exist without background noise and a stochastic subspace $(n - d)$ in which the only involved motion is the noise. We also mention[9] that the rank[10] of the matrix V gives the higher limit of the dimension

---

[9] By considering the Broomhead and King assumptions.

[10] The rank of a matrix V is the number of linearly independent columns in the matrix V. For a square matrix, this number is always equal to the number of linearly independent lines. If the matrix is rectangular $m \times n$, then the rank is lower or equal to the $min(m, n)$.

of the subspace explored by the deterministic component of the trajectory. Thus, the dimension d of the deterministic subspace is obtained by computing the *rank of the matrix V*. In the case of a series without background noise, a hypothesis has been enunciated saying that d (with $d \leq n$) is equal to the number of strictly positive eigenvalues of V, and the rest of eigenvalues is equal to the number of $(n - d)$ whose value is equal to zero. In this particular case (i.e. differentiated Cac40), we observe that there is no eigenvalue equal to zero. It was different for the vector x_1 resulting from the delay model applied to the logistic equation (for $\alpha = 5$), for which six values among ten were equal to zero. Observe below the eigenvalues (after classification in descending order), (1) for the delay model, (2) then for the stock index:

Logistic. Eq/Delay model:
[6.9033 0.3185 0.0148 0.0004 0 0 0 0 0 0]
First-difference of Cac40:
[1,018.5 1,005.4 965.3 952.8 850.2 826.2 770.5 761.4 693.5 647.1]

However, the previous assumptions about the eigenvalues proved to be incomplete and non-exhaustive. In addition, the noise level of the background noise is not dependent on the dimension n of the embedding space. Thus, when we increase the size of the embedding space (by increasing the size of the window-(n, τ) or by reducing the size of the interval τ which is used to sample), then the level of background noise is lowered and new eigenvalues appear. From these observations, we can work on the eigenvalues σ_i^2 and treat them to distinguish (in a noisy signal) the *deterministic* part corresponding to the cleaned signal, and the part corresponding to the background noise. $\sigma_i^2 = (\sigma_i^D)^2 + (\sigma^B)^2$ (*D*: determinist, *B*: noise). The SSA is an important source of research for complex dynamics in Economics and Finance.

2.3 Fractional Brownian Motions

2.3.1 Brownian Motion and Random Walk

A Brownian motion can be defined as *a random series $x(t)$* with Gaussian increases and whose variance

$$var[x(t_2) - x(t_1)] \text{ is proportional to } |t_2 - t_1|^{2H}$$

with $2H = 1/2$. Although the internal structure of Brownian motion is different according to the value of H ($0 < H < 1$), in a generic way, we speak of a fractional Brownian motion, whatsoever H. Figure 2.16 shows examples of three Brownian motions respectively for $H = 0.1, 0.5$ and 0.72.

Without presenting its genesis, it is possible to say that H is a statistical indicator known under the name of the Hurst exponent. In particular, for an experimental series, the goal of this indicator was to dissociate random walk from non-random walk. The Hurst subject was that the experimental dynamics in Nature do generally
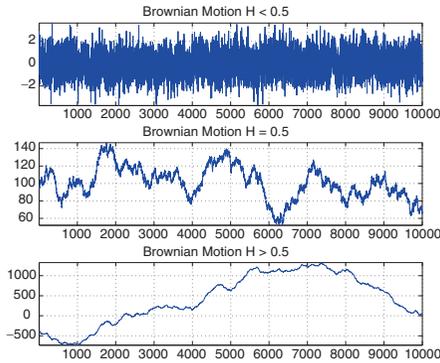
Fig. 2.16 Examples of Brownian motions respectively for H $= 0.1, 0.5$ and 0.72

not follow random walks. The Hurst-Test was used by analogy on financial markets in order to test the concept of random walk, which in econometrics has been used for a long time to characterize financial series. We will see later on how the Hurst statistic also plays a role in the following concepts: Persistence, Long memory of the series (part II), ARFIMA models (part II), Self-similarities (part III, and fractal series).

2.3.1.1 Rescaled Range Statistic and Fractional Brownian Motion

The long-term temporal dependence was approached by the statistic called "Rescaled Range" that corresponds to a ratio: the ratio of the *extent* of a series to a standard deviation. This is the extent of partial sums of variations to the average (of a time series) divided by its standard deviation. Given a series $(x_1, x_2, \ldots, x_n)$ and the average of sample $\bar{x}_n = \frac{1}{n} \sum_{j=1}^{k} x_j$, we write R and S:

$$R = \max_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}_n), \qquad (2.31)$$

$$S = \left(\frac{1}{n} \sum_j (x_j - \bar{x}_n)^2 \right)^{1/2}. \qquad (2.32)$$

(k $=$ shifts). Thus the R/S statistic is written:

$$R/S = \frac{\max_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^{k} (x_j - \bar{x}_n)}{\left[\frac{1}{n} \sum_j (x_j - \bar{x}_n)^2 \right]^{1/2}}.$$

The relation between these statistics and the Hurst exponent can be written:

$$R/S = a \cdot n^H, \qquad (2.33)$$
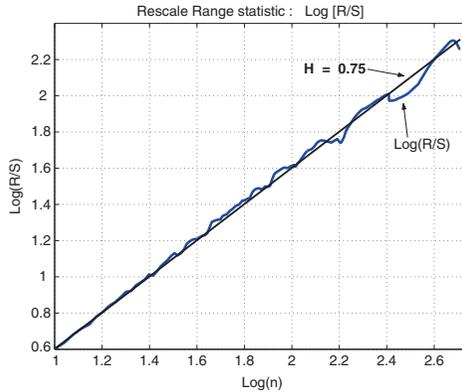
Fig. 2.17 Log (R/S) of a Brownian motion $H = 0.75$

a is a constant. If the series is a random walk, then $H = 0.5$. On the other hand, when $H > 0.5$, it is not a random walk. This statistical criterion, which implicitly rests on the autocorrelation concept, would allow to identify the stochastic aspect of a series. This test is also expressed in a logarithmic form:

$$log(R/S) = H \cdot \log(n) + b. \tag{2.34}$$

By using a log-log scale, we show the graph of the statistic (in relation to the number of observations of the time series). From the preceding equation and for $b = 0$, we deduce $H = \log(R/S)/\log(n)$. Thus, we can read the estimate of H compared with the value chosen to construct the series. Plane $[\log_{10}(n), \log_{10}(R/S)]$ (Fig. 2.17).

We mentioned above that when $H = 0.5$, we have to face a random walk, and not for $H > 0.5$. This assertion means that for $H = 0.5$, the variable of the series is not autocorrelated, as we could see it in a statistical analysis. On the other hand for $H > 0.5$, there are autocorrelation or dependence of terms. It is said that: "Each observation carries a memory of events which precedes it, this is a long term memory: the most recent events have an impact larger than those which are prior to them. What happens today has an influence on the future; the present is a consequence of the past. The time plays an important role" (Abraham-Frois and Berrebi 1995).[11]

$$C = 2^{2H-1} - 1. \tag{2.35}$$

C is an autocorrelation of long period or a correlation of future values with the past values. If $H = 0.5$ then $C = 0$, there is not a "temporal correlation" between the terms of the series, we are thus faced with a random process of a random walk. And it is noted that the characterization of the process is done without using the probability law of the series.

---

[11] *Remark:* These remarks will be able to echo in econometrics concerning concepts of *process of the DS type* (Difference Stationary) where the method used to make the time series stationary is done by differentiation, and of *TS type* (Trend Stationary) where the method used to make the time series stationary is done by a regression over time.
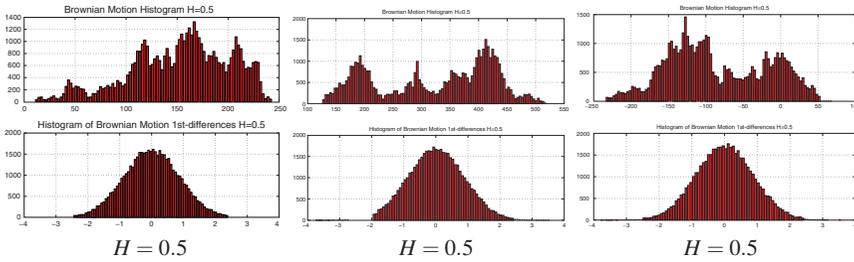
Fig. 2.18 Empirical distributions (histograms) of three arbitrary Brownian motions for H = 0.5. Below histograms of their 1st-differences

If H is not equal to 0.5 and in particular when it is higher than 0.5, the correlation is an increasing function of H. It is said that there is *persistence*, it is also said that the Brownian motion is fractional and for this reason, *there are global and even local tendencies* that emerge from the series. The series has frequent "exits" of the tunnel described around the average of values which precede. These exits are kinds of fractures ("unhookings") in relation to the prior walk of the process which does not have a periodicity. This random component of a process, which was stated as non-stochastic, requires to use the probability laws to describe it, and the statistic H is understood as the probability so that two consecutive events occur.[12] On the other hand, if $H > 0.5$, there is an occurrence of non-periodic cycles, and the more H is high, the more the aperiodic oscillations frequently deviate from the average of values that the time series took prior to each fracture. In Fig.2.18, we show for $H = 0.5$ the histograms (empirical distributions) of three Brownian motions (50,000 steps) and below, the histograms of their first-differences (increases). Note that signal distributions have particular forms and the distributions of their first-differences seemingly tend towards a structure of the Gaussian type (to verify with normality test).

In Fig. 2.19, (1) we show for $H > 0.5$ the histograms of three Brownian motions, (2) the histograms of their first-differences and (3) the histograms of their second-differences.

In this case, on the other hand, for $H = 0.8 > 0.5$, only the distributions of second-differences appear to tend towards a structure of the Gaussian type.

2.3.2 Capacity Dimension of a Fractional Brownian Motion

Although the characterization of a Brownian motion is complex, even if we have given a definition above, we can depict a Brownian motion in the following way. Let us imagine an object represented by a point that moves and that at every moment

---

[12] Ref: Abraham-Frois and Peters: "For $H = 0.6$, there is a probability of 60% so that if the last change were positive the following movement will be also" (cf. Abraham-Frois and Berrebi 1995, p. 330).
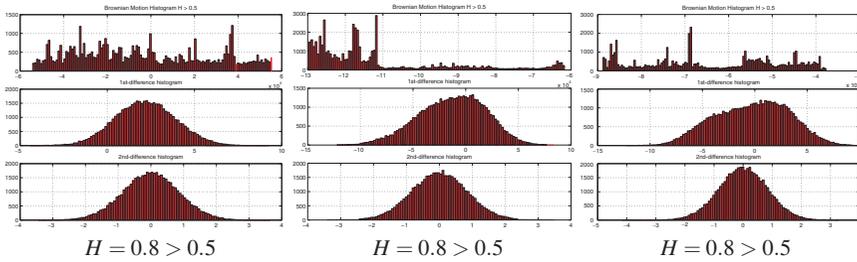
Fig. 2.19 Histograms of three arbitrary Brownian motions for H > 0.5. Below histograms of their 1st-differences. Below histograms of their 2nd-differences

makes a jump (or a step) in an unspecified direction (knowing that here, it is a two-dimensional or three-dimensional Brownian motion). More exactly, we imagine that this displacement is that of a drunk walker as it is usual to symbolize it. Viewed from a big distance, one does not know if it is a continuous curve where each step is represented by a point, or if it is discontinuous, i.e. if each point corresponds to 100 or 1,000 steps for example. If we compare the trajectories taken step by step, all the 100 steps or all the 1,000 steps, we will observe that the trajectories resemble each other. It is usual to say that the trajectory is a fractal curve and *its capacity dimension (or fractal dimension) is equal to 2* ($D = 2$). This type of trajectory has been highlighted by Jean Perrin at the beginning of the twentieth century, he observed the movements of atoms under the microscope. One of the consequences of *this capacity dimension equal to 2 is that the trajectory tends to "blacken" the plane* while advancing in the course of time. And if despite everything the length of the time series is not sufficient, we will be able to observe portions of *the plane "filled" uniformly by the trajectory*. A "singular geometry" can be associated with this type of trajectory which is (certainly too quickly) defined as a random walk. The selected example of a walker, lets precise that the step taken by the walker excludes all the "long distance jumps". We can choose another way of symbolizing this type of movement borrowed from the physical science by observing the trajectory of an atom or a molecule belonging to a gas from which the temperature is different from the absolute zero. The random movements of atoms consecutive to the presence of a non-zero ambient temperature are called *phenomena of diffusion* which are related to Physics but also the biology for example. Generally, any matter undergoes this diffusion phenomenon through an increase in the ambient temperature. Above, we evoked the notion of *the long distance jump*, but the length of the walker's steps is always close for two arbitrary steps, which may be slightly different if we measured them with precision. But this way of representing the movement excludes *the long distance jump*. This type of long distance jump has been approached by Paul Levy who in the 1930s studied fundamental variants of the Brownian motion where these long jumps are allowed, in opposition to what is stated above. The probability of these jumps is weak, however they have the very important capacity to modify the general structure of a movement. This type of event is called *Levy jump* or *Levy flight*.

2.3.2.1 Capacity Dimension of Trajectories, Gauss and Levy Laws

This concept of Levy jump can find transpositions in Nature. The most usual case is the displacement of an albatross above the ocean, seeking the fish shoals to nourish itself. The displacement of the bird is done in three dimensions. However if we observe the trajectory in a plane (i.e. two-dimensions), we cannot observe the movements (elevation) in the vertical dimension any more. The characteristic is that the bird that found a fish shoal will describe displacements of weak length and width during a time, the majority of its displacements at this place being vertical (not represented). Then, having finished at this place, it will move much further in the plane until finding another fish shoal. This aspect of the set in the plane corresponds to concentrations of movements in the form of very localized loops which will be separated by long curves without loops. The distance between two fish shoals is much larger than the one inside the same fish shoal. In short, there will be many short flights and from time to time long distance flights. This observation was done by two teams, one from the British Antarctic Survey from Cambridge and the other from the Boston University. Unlike the Brownian motions described by the displacement of a walker, where the steps have close lengths, the Levy jumps produce immense steps (in the trajectory) even if they have a weak probability to occur. This type of Levy jump offers trajectories whose *capacity dimension or fractal dimension is lower than 2 ($D < 2$). A Levy jump will thus be defined as a movement where the probability $P(s > S)$ of a jump s of distance higher than S varies like a negative power of S: $P(s > S) = S^{-D}$, with $D < 2$.*[13]

This principle of the Levy distribution is used in astronomy in the study of the distribution of "celestial objects". The trajectories of the stock exchange indexes do obviously not follow the Gauss law. The strong fluctuations of the stock exchange, the shocks and the explosions over long periods could be integrated into the probability law. Sapoval (2001) gives a recent example of the Frankfurt Stock Exchange index, for which we notice *brutal explosions* and *self-similarities* on different scales. He notices that many stock market indexes have representations of this type and have the statistical property of the *self-affinity*. (The stake is to consider that the stock-exchange trajectories result from a subjacent deterministic dynamics, like the trajectories with Levy jumps.)

2.3.2.2 Capacity Dimension and Fractional Integration

The fractional integration parameter d is defined as follows, in connection with the Hurst exponent H evoked in the preceding sections: $H = d + 1/2$. In connection with the Levy law, we can indicate the functional relation between the fractal dimension and the Hurst exponent: $D = 2 - H$ or, the relation between the fractal dimension and fractional integration: $D = 3/2 - d$. There exists a relation between the Levy law and fractional integration: $D = 3/2 - d < 2, -d < -3/2 + 2,$

---

[13] The central limit theorem is impugned when the expectation value: $E(s^2) \rightarrow \infty$.

$d < 3/2 - 2, d < -1/2$. This case corresponds to the lower bound of the parameter of fractional integration d of a process ARFIMA(0,d,0) for which we have to face "anti-persistence" phenomena. On the contrary, the relation between the Levy law and the Hurst exponent is immediate: $D < 2, 2 - H < 2, H > 0$. H will have to be strictly higher than 0, what appears to be a too large sample to be significant. The relation between the fractal dimension and the exponent is also written in certain works (Mandelbrot 1975, p. 114) and under some conditions no longer described as a difference but as a ratio: $D = 1/H$.

2.3.2.3 Definition of a Brownian Motion by Mandelbrot

Definition 2.1 (Mandelbrot, Brownian motion). If x is a point of the plane, $x(t)$ is called a Brownian motion it is a succession of small displacements which are mutually independent and "isotropic".[14] The last characteristic means that all the directions for the displacement of the point in the plane are possible. For any couple of moments t and $t' > t$, one dissociates two points of vector x by $x(t)$ to $x(t')$ and considers that (Mandelbrot 1975, p. 45):

(1) The direction as well as the length of the trajectory are independent of the initial position and of the position of each previous point.
(2) The vector must be isotropic.
(3) The length of the vector is such that its projection on an unspecified axis obeys the Gaussian distribution of density: $\frac{1}{\sqrt{2\pi|t'-t|}} \exp \frac{-x^2}{2|t'-t|}$.

2.3.3 Introduction to Persistence and Loops Concepts

We implicitly evoked in different previous sections the fact that the Brownian motion is spread in a natural space, i.e. in a plane or in a volume. This is why in this section we chose to represent some Brownian motions in dimension higher than 1. Brownian motions have been generated by the same simulator (simulator seeds influence the results).

2.3.3.1 Brownian Motions: Highly Persistent, Fairly Persistent and Weakly Persistent

Highly persistent. The Hurst exponent chosen for the simulation hereafter is close to 0.85. We are beyond the value 0.5 of H for which we define a random walk

---

[14] *An isotropic line:* an isotropic line passes through the circular points at infinity. Isotropic lines are perpendicular to themselves.
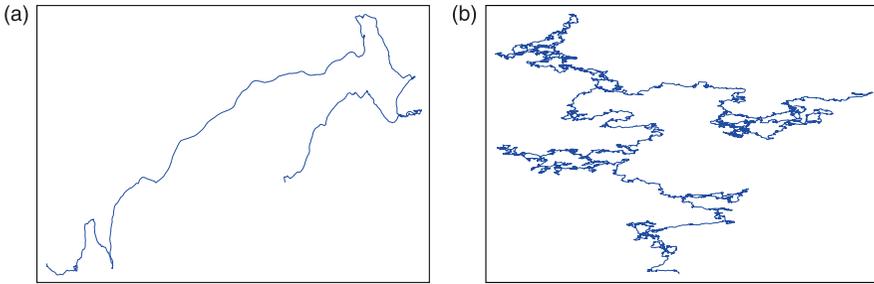
Fig. 2.20 (**a**) "Highly" persistent, $H = 0.85$; (**b**) "Fairly" persistent, $H = 0.6$
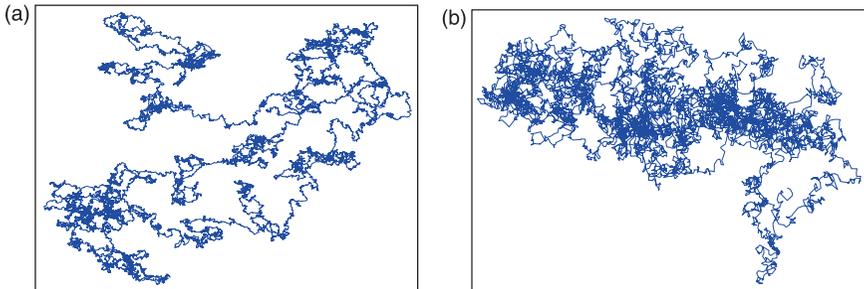
Fig. 2.21 (**a**) Brownian motion, $H = 0.6$; (**b**) Brownian motion for $H = 0.51$

in the sense of Hurst. The value $H = 0.85$ provides a fractal dimension near $D = 2 - 0.85 = 1.15$. The second approach also provides a dimension lower than $2, D = 1/0.85 = 1.17$. The construction of such a time series, according to the terms of Mandelbrot *"discourages very strongly, without forbidding them", the formation of loops, because we "forced this trajectory to be very persistent"*. We can observe above in Fig. 2.20a that *the loops are rare*. The low frequency "drift" is very high.

Fairly persistent. The Hurst exponent chosen in this case is around 0.6. We are beyond the value 0.5 of H for which we define a random walk in the sense of Hurst. The value $H = 0.6$ provides a fractal dimension near $D = 2 - 0.6 = 1.4$. The second approach also provides a dimension lower than $2, D = 1/0.6 = 1.66$. We observe the proliferation of many small loops that the trajectory describes on itself. *The more the Hurst exponent approaches* 0.5, or *the more the fractal dimension approaches* 2, *the more the curve forms convolutions*. The time series becomes *dense* (see Fig. 2.20b built by means 1,000 steps). Figures 2.21a and 2.22 represent fairly persistent Brownian motions built by means of 10,000 steps, the first one is a two-dimensional motion (see Fig. 2.21a) and the second one is a three-dimensional Brownian motion (see Fig. 2.22):

Weakly persistent. The Hurst exponent chosen in this case is around 0.51. We are very close to the value 0.5 of H for which we define a random walk in the sense

Fig. 2.22 Three-dimensional
Brownian motion, $H = 0.6$
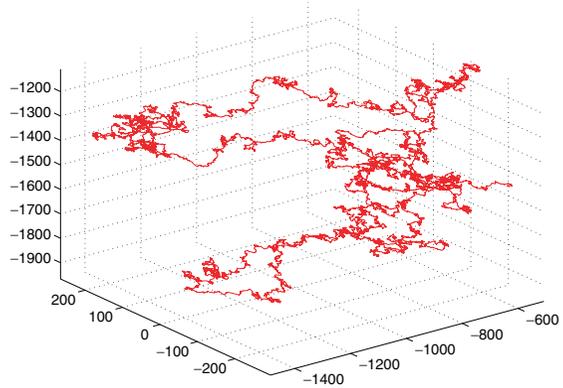
Fig. 2.23 Brownian motion
for $H = 0.5$
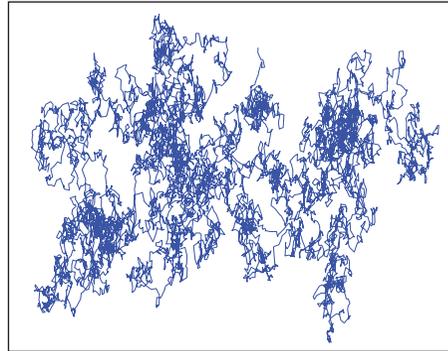
of Hurst. The value $H = 0.51$ provides a fractal dimension about $D = 2 - 0.51 =$ 1.49. the second approach also provides a dimension lower than 2 but very close, $D = 1/0.51 = 1.96$. *The number of loops* that the trajectory describes on itself is growing. *The more the fractal dimension approaches* 2*, the more the curve forms convolutions*, the time series becomes more and more *dense*. The low frequency "drift" becomes very variable and is almost invisible in our case (see Fig. 2.21b).

Brownian motion for $H = 0.5$. The different convolutions are similar. *Such a situation is similar to the statistical analysis where the spectrum is white*, we have to face a white noise, by the average or after the differentiation of the time series. Low frequency drift is invisible. We will be able to visualize the density of the trajectory that blackens and uniformly fills the plane (see Fig. 2.23).

2.3.4 Comment on DS/TS Process and Brownian Motions

In this section, we open a parenthesis that has the form of a question. The "memory" notion of observations mentioned above (which appears when $H > 0.5$, i.e. when

"the present is a consequence of the past") interrogates about the "tendency" notion in a time series. *The analysis of the trend in econometrics*, or more exactly *the analysis of the stationarity* by means of the Dickey and Fuller (Unit root) statistic, leads to choose for an arbitrary series between the two tendency types. The series are identified by these tests as, either (1) or (2):

(1) DS type (Difference stationary), i.e. stationary by difference. The "stationariza-tion" is done by differentiation.[15] We speak of the *stochastic stationarity* that concerns *the non stationary random processes*.
(2) TS type (Trend stationary), i.e. with "stationary trend". The "stationarization" (or detrend) is done by a regression on time. We say that there is a *deterministic non-stationarity*.

The stock markets and economic time series are processes that are rarely anal-ysed as being stationary (or even Gaussian). The non-stationarity (non-stationariness) can result from moments of the first order (i.e. Expectation) or of the second order (i.e. Variance). Before the end of the 1970s, there was no analytical method to study non-stationariness. The Box–Jenkins method of graphic analysis, that makes it possible to visualize the tendencies or the cycles and the saisonali-ties, is interesting but not sufficient. In 1984, works of Nelson and Plosser analysed the non-stationarity (non-stationariness) by means of two processes: TS and DS processes.

2.3.4.1 TS Processes: Non-Stationarity of Deterministic Type

Such a process is written $x_t = f_t + \varepsilon_t$, where f_t is a polynomial function of time, linear or nonlinear and ε_t is a stationary process. An example of an elementary process TS of order 1 is written: $x_t = a_0 + a_1 t + \varepsilon_t$. If ε_t is a white noise (Gaussian or not) $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, the process is determined by:

$$E(x_t) = a_0 + a_1 t + E(\varepsilon_t) = a_0 + a_1 t, \tag{2.36}$$
$$V(x_t) = E(a_0 + a_1 t + \varepsilon_t - (a_0 + a_1 t))^2 = \sigma_\varepsilon^2, \tag{2.37}$$
$$cov(x_t, x_{t'}) = 0 \text{ for } t \neq t'. \tag{2.38}$$

Such a process is non-stationary because the expectation depends on time (in a linear way), as one can see above, t is provided with a coefficient. The expectation is calcu-lated at every moment and obviously depends on t. We speak of (non-stationariness) non-stationary of the deterministic type. The parameters of the tendency has a_0 and a_1, we can estimate them by the method of least squares. The estimators obtained are the "Best Linear Unbiased Estimators (BLUE)" which make forecasts possible. The detrend is done by removing from the value of x_t the estimation $\hat{a}_0 + \hat{a}_1 t$ at each moment t. In this type of process, it is said that after a *random shock*, *known as transitory*, the series *re-takes* its walk around its tendency curve.

---

[15] *Stationarization* (mathematical barbarism or neologism): To make a time series stationary.

2.3.4.2 DS Processes: Random Walk

Such a process is called "random walk" (and often used concerning stock markets). Here, the process becomes stationary by a filter using the differences: $(1 - B)^d = \beta + \varepsilon_t$, usually the filter is of order 1 $(d = 1)$. ε_t is a stationary process or a white noise and β is a constant symbolizing the possible drift of the process. The process is written: $x_t = x_{t-1} + \beta + \varepsilon_t$, ε_t is a white noise (Gaussian or not). If the process is without drift $(\beta = 0)$, then: $x_t = x_0 + \sum_{i=1}^{t} \varepsilon_i$, thus $E(x_t) = x_0$, $V(x_t) = t\sigma_\varepsilon^2$, $cov = \sigma_\varepsilon^2 \times Min(t,t')$ for $t \neq t$. If the process has a drift $(\beta \neq 0)$, then we write: $x_t = x_0 + \beta t + \sum_{i=1}^{t} \varepsilon_i$, thus $E(x_t) = x_0 + \beta t$, $V(x_t) = t\sigma_\varepsilon^2$, $cov = \sigma_\varepsilon^2 \times Min(t,t')$ for $t \neq t'$.

2.3.4.3 Fractional Brownian Motion and DS Process

The fractional Brownian motions are mainly non-stationary. For a Hurst exponent $(H > 0.5)$, the process is not analysed as a random walk but as a more or less persisting process, and the correlation C is an "increasing function of H for $0.5 < H < 1$". It is known that in this case, the movement has frequent exits of the tunnel described by its average and its variance. These exits are fractures in relation to its prior walk. It was also said that these "exits" have a *random periodicity*. (It is the random component of a process which for $H > 0.5$ is considered as *non-stochastic* by Hurst.) The more H approaches 1, the more the non-periodic cycles deviate frequently from the average of values that the time series took before. For $H = 0.5$, we have pure randomness as in this case the correlations $C = 0$. The present does not influence the future and is not influenced by the past. This is known as a random walk.

Traditional Statistics analyse DS processes as a random walk. We can study the stationarity by means of the Dickey and Fuller (Unit root) Test of some samples of fractional Brownian motions for $H = 0.5$ and $H > 0.5$. In the cases where the series exhibits a non-stationarity, usually the conclusion is that we are faced with a DS process, i.e. a random walk that becomes stationary by means of a filter using the differences (test not presented here). Later we will present the traditional statistical tests of stationarity, in Part II and they will be applied to a stock index.

Chapter 3
Nonlinear Processes and Discrimination

3.1 Reminders: Statistics and Probability

3.1.1 Random Experiment and Measurement

Definition 3.1 (Random experiment). A random experiment is represented by a triplet (Ω, a, P) where the following conditions are verified:

(1) Ω is the set of possible results of the experiment,
(2) a is a σ-algebra, i.e. a set of parts of Ω (called events), containing the parts ϕ and Ω, stable by complementation and by denumerable union,
(3) P is a set function,[1] i.e. a map from a to $\mathbb{R}^+$, which satisfies $P(\Omega) = 1$ and the condition of σ-additivity: if (A_n) is a sequence of disjoint events and if $\sum_n A_n$ indicates their union:

$$P\left(\sum_n A_n\right) = \sum_n P(A_n).\tag{3.1}$$

The map P is a probability.[2]

3.1.1.1 Measurement

More generally, a given couple (Ω, a) can be provided with a measure, if it satisfied the first and the second conditions of a random experiment above. Such a measurement μ is defined as a set function, map from a to $\mathbb{R}^+$, such that $\mu : a \longrightarrow \mathbb{R}^+$, satisfying the condition of σ-additivity. Thus a probability is a measurement whose total *mass* is

$$\mu(\Omega) = 1.\tag{3.2}$$

---

[1] *Set function:* A relation that assigns a value to each member of a collection of sets.
[2] $P(\Omega) = 1$ and $P(\sum_n A_n) = \sum_n P(A_n)$ are the Kolmogorov axioms.

3.1.2 Reduction Principles of Estimators: Invariance Principle, Unbias Principle, Asymptotic Principle

The choice of estimators is fundamental (Gourieroux and Monfort 1989). In reality, some "intuitive" or "natural" presuppositions impose properties to the estimators. These properties are gathered under three principles: the invariance principle, the Unbias principle (without bias) and the asymptotic principle.

3.1.2.1 Invariance Principle

If we wish to estimate parameters of the *mean* type, one will choose *estimators that are linear functions* of observations:

$$\delta(x_1 \cdots x_n) = \sum_{i=1}^{n} a_i x_i. \qquad (3.3)$$

If we wish to estimate parameters of the *variance* type, one will choose *estimators that are quadratic functions* of observations:

$$\delta(x_1 \cdots x_n) = \sum_{i,\,j}^{n} a_{ij} x_i x_j. \qquad (3.4)$$

These choices, which are *intuitive and natural* presuppositions (i.e. a priori), imposed on the nature of estimators, are understood as *a property of invariance* or *a principle of invariance*. If the linear estimators are considered, they satisfy a property of invariance on linear combinations. Indeed such a estimator satisfies ($\forall$ α, α^* and $\forall x, x^*$ of $\mathbb{R}^n$):

$$\delta[\alpha x + \alpha^* x^*] = \alpha \delta(x) + \alpha^* \delta(x^*). \qquad (3.5)$$

3.1.2.2 Unbias Principle

An estimator δ is an unbiased estimator of $g(\theta)$ if and only if

$$E_\theta \delta(x) = g(\theta), \quad \forall : \theta \in \Theta. \qquad (3.6)$$

An estimator is without bias if, in the *mean*, the value proposed is equal to the searched value and this for any unknown value of the parameter.

3.1.2.3 Asymptotic Principles and Convergence

When the number of observations is large, $n \to \infty$, it is possible to impose only asymptotic constraints to the estimators. To estimate a function $g(\theta)$ of the

parameter, we choose a sequence of the estimator $\{\delta_n, n \in \mathbb{N}\}$: The sequence of the estimator is *asymptotically* unbiased if:

$$\lim_{n \to \infty} E_\theta \delta_n(x) = g(\theta), \quad \forall \theta \in \Theta. \tag{3.7}$$

A sequence of estimators is:
- *Weakly convergent* if $\delta_n(x)$ converges in probability towards $g(\theta)$, $\forall \theta$:

$$\forall \varepsilon > 0, P_{n,\theta}[\|\delta_n(x) - g(\theta)\| > \varepsilon] \to 0, \quad \forall \theta \in \Theta \tag{3.8}$$

- *Convergent in the quadratic mean* $\delta_n \underset{m.q}{\to} g(\theta)$ if:

$$E_\theta \|\delta_n(x) - g(\theta)\|^2 \to 0, \quad \forall \theta \in \Theta \tag{3.9}$$

- *Strongly convergent*, if δ_n converges "almost surely" towards:

$$g(\theta), \quad \forall \theta \in \Theta \tag{3.10}$$

3.1.3 Definition of a Process

A process is a sequence $(X_n, n \in I)$ of random vectors defined on the same space (Ω, a, P) and taking values in $\mathbb{R}^p$. The set of indexes I is generally $\mathbb{N}, \mathbb{N}^*$ or $\mathbb{Z}$. To each state of Nature ω corresponds thus an element $(X_n(\omega), n \in I)$ of $[\mathbb{R}^{pI}]$; such an element is called trajectory of the process. A process admits a law, we can show that this one is characterized by the knowledge of laws of all finite sub-family $(X_{n1}, \ldots, X_{nk})$ extracted from the sequence $(X_n, n \in I)$. From this characterization, we can define particular process classes.

3.1.4 Probability Law, Cumulative Distribution Function, and Lebesgue Measure on $\mathbb{R}$

3.1.4.1 Discrete Law

When the set Ω of possible results is finite or denumerable and when a coincides with the set of parts Ω, a measure on (Ω, a) is said discrete. A discrete measure is characterized by the values which it takes for the *sets reduced to a point*. A discrete probability law has elementary values:

$$P_\omega = P([\omega]), \quad \omega \in \Omega \text{ and } \sum_{\omega \in \Omega} p_\omega = 1. \tag{3.11}$$

Conversely, let $(p_\omega, \omega \in \Omega)$ be a denumerable family of positive real numbers whose sum is 1. There exists a single law of probability on Ω, whose elementary probabilities are $p_\omega \in \Omega$.

3.1.4.2 Probability Law on $\mathbb{R}$ and Cumulative Distribution Function

If $\Omega = \mathbb{R}$ is posed, we call a cumulative distribution function associated with the law P the map F of $\mathbb{R}$ in $[0, 1]$, defined by:

$$F(x) = P(]-\infty, x]), \quad \forall x \in \mathbb{R}. \tag{3.12}$$

A cumulative distribution function is increasing and continuous to the right, such that:

$$\lim_{x \to +\infty} F(x) = 0, \ \lim_{x \to +\infty} F(x) = 1. \tag{3.13}$$

Conversely, any increasing numerical function continuous to the right such that $\lim F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$, can be regarded as the cumulative distribution function of a probability law on $\mathbb{R}$. This law is unique.

3.1.4.3 Lebesgue Measure on $\mathbb{R}$

The measurements on $\mathbb{R}$ which take finite values on any limited interval are characterized by the set of the values $\mu(]a, b])$, $a < b$, $a, b \in \mathbb{R}$. It is possible to show that it exists a measure noted λ, such that:

$$\lambda(]a, b]) = b - a, \tag{3.14}$$

this is called *the Lebesgue measure on* $\mathbb{R}$.

3.1.5 Integral with Respect to a Measure

A measure is a map defined on the set of characteristic functions (called indicators)[3] of events, which is in fact an integral with respect to μ, noted $\int_\Omega \cdot d\mu$ [or under a discrete form $\int_\Omega \cdot \mu(d\omega)$] and defined such as:

$$\forall A \in a, \quad \int_\Omega d\mu : \mathbf{1}_A \longrightarrow \int_\Omega \mathbf{1}_A d\mu = \mu(A). \tag{3.15}$$

---

[3] *Indicator:* $\mathbf{1}_A$ is the characteristic function (called indicator) of the set A, such that:
$$\mathbf{1}_A = \begin{cases} 1 : & \text{if } y \in A, \\ 0 : & \text{otherwise.} \end{cases}$$

If μ is a discrete measure, it is written:

$$\int_\Omega d\mu = \int_{\omega \in \Omega} f(\omega)\mu(\{\omega\}). \tag{3.16}$$

3.1.6 Density and Lebesgue Measure Zero

Consider a positive function f integrable with respect to a measure μ, i.e. such that $\int_\Omega f d\mu < \infty$. The map v that associates the event A with $v(A) = \int_\Omega \mathbf{1}_A f d\mu$ defines a measure on (Ω, a). This measure is called the *density f with respect to μ*. It is noted $v = f \cdot \mu$ or $f = dv/d\mu$.

A probability law on $\mathbb{R}^n$ is continuous if and only if the cumulative distribution function is continuous and if the set of points, whose derivative

$$\frac{\partial^n F(x_1 \cdots x_n)}{\partial x_1 \cdots \partial x_n} \tag{3.17}$$

does not exist, is of zero Lebesgue measure. A density f can then be selected equal to

$$f(x_1 \cdots x_n) = \frac{\partial^n F(x_1 \cdots x_n)}{\partial x_1 \cdots \partial x_n}. \tag{3.18}$$

And on $\mathbb{R}$, a continuous law is given by $f(x) = \dfrac{dF(x)}{dx}$ at any point where the derivative exists.

3.1.7 Random Variables and Transfer Formula

Given two sets provided with the events (Ω, a) and $(X, \mathbf{B})$. A random variable is a map X of Ω such that: $\forall B \in \mathbf{B}, X^{-1}(B) \in a$. A real random variable is a variable in $\mathbb{R}$ provided with the smallest family of events containing the intervals. Such maps are characterized by the condition:

$$\forall b \in R, X^{-1}(]-\infty, b]) = \{X \leq b\} \in a. \tag{3.19}$$

When the starting space (Ω, a) is provided with a probability law P, we can use a random variable X in order to provide $(X, \mathbf{B})$ with an adapted law. Let $(\Omega, a, \mathbf{P})$ be a "probabilized" space and X a random variable of (Ω, a) in $(X, \mathbf{B})$. The function P^X defined by $\forall B \in \mathbf{B}, P^X(B) = P[X^{-1}(B)]$ is a probability law on $(X, \mathbf{B})$. It is the image law P by X or the law of X.

3.1.8 Some Laws of Probabilities

Discrete law	Set	Probabilities	Mean	Variance
Dirac measure	$\{x\}$	$p_x = 1$	x	0
Geometric	$\mathbb{N}^*$	$p_x = p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$

Continuous law	Density	Cumulative.D.F	Mean	Variance
Normal law N(0,1)	$\frac{1}{\sqrt{2\pi}}\exp\frac{-x^2}{2}$	$\Phi(x) = \frac{1}{\sqrt{2\pi}}$ $\int_{-\infty}^{x}\exp-\frac{t^2}{2}dt$	0	1
Pareto law	$\frac{\alpha A^{\alpha}}{x^{\alpha+1}}1_{x\geq A}$	$1-\left(\frac{A}{x}\right)^{\alpha}$	$\alpha > 1$ $\frac{\alpha A}{\alpha-1}$	$\alpha > 2$ $\frac{\alpha A^2}{(\alpha-1)(\alpha-1)^2}$

3.1.9 Autocovariance and Autocorrelation Functions

Definition 3.2 (Autocovariance). An *autocovariance function* of a random process x_t with a finite variance is written:

$$\gamma_k = Cov(x_t, x_{t+k}) = E([x_t - E(x_t)][x_{t+k} - E(x_{t+k})]). \tag{3.20}$$

Definition 3.3 (Autocovariance). An *autocovariance function* of a stationary process x_t verifies at the same time:

$$\gamma_0 = Cov(x_t, x_t) = E([x_t - E(x_t)]^2) = Var(x_t) = \sigma_x^2 \geqslant 0, \tag{3.21}$$

$$|\gamma_k| \leqslant \gamma_0, \tag{3.22}$$

$$\gamma_k = \gamma_{-k}. \tag{3.23}$$

Definition 3.4 (Autocorrelation). An *autocovariance function* of a stationary process x_t is written:

$$\rho_k = \frac{Cov(x_t, x_{t+k})}{\sigma_{x_t}\sigma_{x_{t+k}}} = \frac{\gamma_k}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0}. \tag{3.24}$$

3.2 The ARMA Processes: Stock Markets and Random Walk

For a long time, traditional statistical analysis had used the *ARMA* processes, which are representative of linear modelling. In this section, we will recall the structure of these processes in conjunction with the construction of stationarity tests. Another objective is to point out how traditional statistical analysis by means of the *ARMA* models characterized the stock exchange index as "random walk". This reminder is done from a series of more than 8 years of daily values of the French stock market (Cac40) starting 1988. After having studied stationarity, the Box–Jenkins statistical

method will be used. As a preliminary, we proceed to a phase of identification of characteristics of the stationarity of the series in order to define then the model adapted within the framework of the $ARIMA(p,d,q)$ process. With this intention, we will use the Dichey–Fuller method and we will analyze the correlograms. Following this analysis, the tendency will be eliminated (in this case by differentiation). We will then proceed to the determination of the orders p and q of $AR(p)$ and the $MA(q)$ processes. After the estimation of the model, we will test the adequacy of the model by analyzing the residuals.

3.2.1 Reminders: ARMA Processes and Stationarity

3.2.1.1 ARMA(p,q) Process

In an autoregressive process of order p, the variable y_t is generated by a weighted average of observations until the *p-th* period:

$$AR(p) : y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + \varepsilon_t, \qquad (3.25)$$

the coefficients θ_i are to be estimated and ε_t are uncorrelated random errors, also called "white noise errors". The equation can be also written with a shift operator D: $(1 - \theta_1 D - \theta_2 D^2 - \cdots - \theta_p D^p) y_t = \varepsilon_t$. The autocorrelations of a $AR(p)$ process are defined by a geometrical decrease of terms. The partial correlogram has its p first terms different from 0. A moving average process is written:

$$MA(q) : y_t = \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} - \cdots - \alpha_q \varepsilon_{t-q}, \qquad (3.26)$$

the coefficients α_i are to be estimated and ε_t being the white noise errors. The equation can be also written with a shift operator D: $(1 - \alpha_1 D - \alpha_2 D^2 - \cdots - \alpha_q D^q) \varepsilon_t = y_t$. There is an equivalence between the AR and MA processes: $MA(1) = AR(\infty)$.

3.2.1.2 Normality Tests

The normality tests of the distribution of a time series or distribution of residuals resulting from the estimation of a model use the following statistic tests:
    Skewness Test (asymmetry):

$$S = \frac{\frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^3}{\sigma^3}. \qquad (3.27)$$

    Kurtosis Test (flattening):

$$K = \frac{\frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^4}{\sigma^4}. \qquad (3.28)$$

Jarque–Bera Statistic: this statistic brings together the results of the Skewness and Kurtosis and is built by means of the centred moment of order-k:

$$m_k = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^k, \tag{3.29}$$

where the Skewness coefficient is $\beta_1^{1/2} = m_3 / m_2^{3/2}$ and the Kurtosis coefficient is $\beta_2 = m_3 / m_2^2$. If the distribution is normal and the number of observations large enough, i.e. if it is higher than 30 $(T > 30)$: $\beta_1^{1/2} \to N(0; \sqrt{6/T})$ and $\beta_2 \to N(3; \sqrt{24/T})$. It is necessary to extract the statistic:

$$v_1 = \left| \beta_1^{1/2} - 0 \right| / \sqrt{6/n} \quad \text{and} \quad v_2 = |\beta_2 - 3| / \sqrt{24/n}, \tag{3.30}$$

where each one of the statistics is to be compared with the value of the normal law for the threshold of 5% which is: 1.96. If $v_1 < 1.96$ and $v_2 < 1.96$, then the symmetry and flatness assumptions (i.e. Normality) are checked. Thus the Jarque–Bera statistic takes again the preceding tests, and if the Skewness and Kurtosis coefficients follow Normal laws, then $JB = \frac{n}{6}\beta_1 + \frac{n}{24}(\beta_2 - 3)^2$ follows a χ^2 with two degrees of freedom. Thus, if $JB > \chi_{1-\alpha}^2(2)$ the Normality hypothesis is rejected (for a time series or the residuals of an estimation):

$$JB = \frac{T-k}{6} \left[S^2 + \frac{1}{4}(K-3)^2 \right]. \tag{3.31}$$

3.2.1.3 Stationarity and Dickey–Fuller Tests (Unit Root): TS or DS Process

If for a given process, its expectation value and its variance (as well as its covariance) are modified over time, then the process is regarded as non-stationary. Thus, a stochastic process or a white noise process is regarded as stationary, subject to the independence of its expectation and its variance with time.

Dickey–Fuller-DF and Augmented Dickey–Fuller-ADF Tests (Unit Root Tests)

Before practising the analysis of correlograms, a thorough study of the stationarity and tendency is necessary in the majority of cases. This is why we present the unit root test, using the Dickey–Fuller procedure.

The Simple DF Test

It must estimate the three following models:

1. Model without trend nor constant term. Autoregressive model of order 1:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \tag{3.32}$$

(or $\Delta y_t = (\phi_1 - 1)y_{t-1} + \varepsilon_t$).

2. Model without trend and with constant term. Autoregressive model with constant:

$$y_t = \phi_1 y_{t-1} + \beta + \varepsilon_t, \tag{3.33}$$

(or $\Delta y_t = (\phi_1 - 1)y_{t-1} + \beta + \varepsilon_t$).

3. Model with trend and constant term. Autoregressive model with trend:

$$y_t = \phi_1 y_{t-1} + bt + c + \varepsilon_t, \tag{3.34}$$

(or $\Delta y_t = (\phi_1 - 1)y_{t-1} + bt + \beta + \varepsilon_t$).

If the studied series is of the *TS* type (*Trend Stationary*), then it is necessary to "stationarize" by a regression on time and the residual of the estimation must be studied to check the quality of the estimated model $ARMA(p,q)$, according to the Box–Jenkins method. A TS process is written $x_t = f_t + \varepsilon_t$, where f_t is a polynomial function of time and ε_t always a stationary process. If the series is of the *DS* type (*Difference Stationary*), it is necessary to "stationarize" by means of the differences of order (d)-*th*. Here also the differentiated series must be studied through the Box–Jenkins method in order to determine the order p and q of the $ARIMA(p,d,q)$ model.

The ADF Test

It is possible that the error is autocorrelated in a model. This is a hypothesis that is not considered in the simple DF tests, since the residual is regarded as a *white noise*. The ADF test (by the OLS, Ordinary Least Squares) must consider the three models corresponding to the following hypotheses:

4. Model without trend nor constant term. Autoregressive model:

$$\Delta y_t = \rho y_{t-1} - \sum_{j=2}^{p} \phi_j \Delta y_{t-j+1} + \varepsilon_t. \tag{3.35}$$

5. Model without trend and with constant term. Autoregressive model with constant:

$$\Delta y_t = \rho y_{t-1} - \sum_{j=2}^{p} \phi_j \Delta y_{t-j+1} + c + \varepsilon_t. \tag{3.36}$$

6. Model with trend and constant term. Autoregressive model with trend:

$$\Delta y_t = \rho y_{t-1} - \sum_{j=2}^{p} \phi_j \Delta y_{t-j+1} + c + bt + \varepsilon_t. \tag{3.37}$$

The DFA test must lead to similar results than the results obtained with the simple DF test.

Unit Root, Random Walk and Stationarity

Let ε_t be a white noise and the following $AR(1)$ model: $y_t = \phi y_{t-1} + c + \varepsilon_t$. If the autoregression coefficient $\phi = 1$, then the model has *a unit root* and follows *a random walk* with *non-stationary* characteristics. This walk is with *a drift* if $c \neq 0$ and *without drift* if $c = 0$. If $\phi > 1$, then the model is an *explosive process*. If $\phi < 1$, then the model is *stationary*.

Impact of Shocks on the Series and Correlation of Errors of the Model

At a given moment of time, the impact of a "shock" on the series is different if we consider a TS process or a DS process. Indeed, in a process of the TS type (where ε_t is a stationary process), the shock is transitory, and the series will re-find its deterministic walk around its tendency curve. In a DS process (where ε_t is a white noise) the impact of a shock at a given moment, of the time is prolonged infinitely on the series, even if it is decreasing.

3.2.2 Dickey–Fuller Tests Applied to French Stock Index (Cac40)

The sample of the Cac40 and its histogram are represented on the following graphs. The sample concerns almost 8 years of daily values from 1988 to 1996:

Histogram
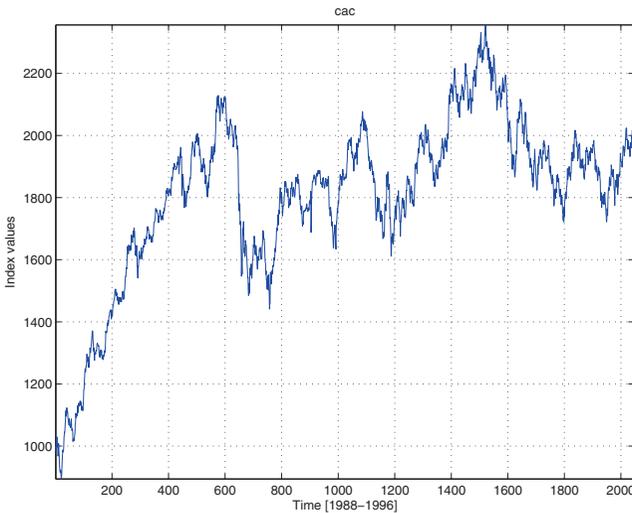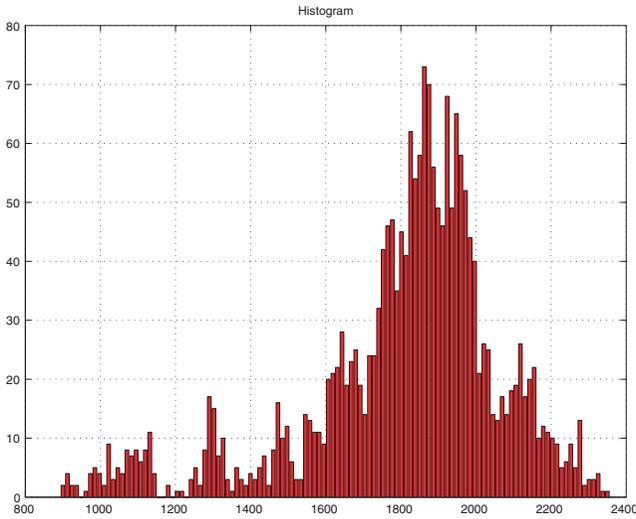
3.2.2.1 The Simple DF Test

The model [3] *to be estimated is:*

$$D(CAC) = C(1) * CAC(-1) + C(2) + C(3) * Time$$

	Coefficient	Std. error	T-statistic	Prob.
C(1)	−0.007729	0.002284	−3.384428	0.0007
C(2)	13.47671	3.596888	3.746770	0.0002
C(3)	0.001063	0.001016	1.046725	0.2954

R-squared	0.993884	Mean dependent var	1811.885
Adjusted R-squared	0.993877	S.D. dependent var	255.2393
S.E. of regression	19.97176	Akaike info criterion	5.990142
Sum squared resid	794152.7	Schwartz criterion	5.998564
Log likelihood	−8798.535	F-statistic	161761.4
Durbin–Watson stat	1.917135	Prob(F-statistic)	0.000000

The estimated model [3] *is:*

$$D(CAC) = -0.007729 * CAC(-1) + 13.47671 + 0.0011063 * Time$$
$$(-3.384) \cdots\cdots\cdots\cdots\cdots (3.74) \cdots\cdots (1.04)$$

$n = 1994 −$ (T-Statistic) = t empirical. The coefficient of the trend (0.0011063) is not significantly different from 0 ($t = 1.04$), thus we reject the assumption of a TS

process. Furthermore, the empirical value of $t_{\widehat{\phi_1}}$ is -3.384, whereas in the table, the value of t_{ϕ_1} for the 5% threshold is -3.41. $(t_{\widehat{\phi_1}} > t_{\phi_1})$. Although the empirical value is very close to the tabulated value, it is nevertheless higher. Thus we accept the null hypothesis H_0, the process is not stationary. Or we can analyze on $n(\widehat{\phi_1} - 1) = 1994(-0.007729) = -15.4116$ that is higher than the value of the table for a 5% threshold which is: -21.8. Here we accept the null hypothesis H_0, the process is not stationary.

The estimated model [2] *is:*

$$D(CAC) = -0.006190 * CAC(-1) + 11.75$$
$$(-3.54) \cdots\cdots\cdots\cdots\cdots (3.67)$$

The constant term is somewhat difficult to analyze, it seems significantly different from 0 ($t = 3.67$). Thus we tend to accept the hypothesis of a DS process with drift. In addition, the empirical value of $t_{\widehat{\phi_1}}$ is -3.54, whereas in the table the value of t_{ϕ_1} for a 5% threshold is -2.86. $(t_{\widehat{\phi_1}} < t_{\phi_1})$. Thus, we reject the null hypothesis H_0, the process is stationary. Or we can analyze on $n(\widehat{\phi_1} - 1) = 1994(-0.00619) = -12.3429$ that is higher than the value of the table for a 5% threshold which is: -14.1. The tabulated value for a 10% threshold is -11.3. H_0 is rejected with a 5% threshold and is accepted with a 10% threshold according to the used method. For this model, the conclusion about the stationarity and the trend of the process are more difficult.

The estimated model [1] *is:*

$$D(CAC) = 0.000170 * CAC(-1)$$
$$(0.693411)$$

The empirical value of $t_{\widehat{\phi_1}}$ is 0.693, whereas in the table, the value of t_{ϕ_1} for the 5% threshold is -1.95. $(t_{\widehat{\phi_1}} > t_{\phi_1})$. Thus, we accept the null hypothesis H_0, the process is not stationary. And for $n(\widehat{\phi_1} - 1) = 1994(0.00017) = 0.339 > -8.1$, the conclusion is identical, the null hypothesis H_0 is accepted, the *process is non-stationary*.

The conclusion of the analysis through the simple DF test is that the *Cac40* sample is an non-stationary process of the DS type (and not the TS type) without drift. Let us check these results by means of the DF test of joint hypotheses and by means of the DFA test.

DF Test of Joint Hypotheses

The matter here is to test the following joint hypotheses (see models for b, c): H_0^1: $(c, \phi_1) = (0, 1)$, H_0^2: $(c, b, \phi_1) = (0, 0, 1)$, H_0^3: $(c, b, \phi_1) = (c, 0, 1)$. We can write a fast program to calculate the F_1, F_2, F_3 statistics. They will be compared with the values of the Dickey–Fuller table, which come from the Fisher table:

$$F_1 = \frac{(SSR_c - SSR_2)/2}{SSR_2/(n-2)}, \quad F_2 = \frac{(SSR_c - SSR_2)/3}{SSR_3/(n-3)}, \quad F_3 = \frac{(SSR_c^2 - SSR_3)/2}{SSR_3/(n-3)}.$$

(With: SSR_c = sum of the squares of the residuals of the model [1] constrained under the hypothesis H_0^1, SSR_2 = sum of the squares of the residuals of the model [2] non-constrained estimated with the OLS, SSR_c^2 = sum of the squares of the residuals of the model [2] constrained under the hypothesis H_0^3, SSR_3 = sum of the squares of the residuals of the model [3] non constrained estimated with the OLS, $SSR_c^2 = \sum_t (x_t - x_{-1} - \hat{c})^2$, $SSR_c = \sum_t (x_t - x_{-1})^2$, and finally the degree of freedom is related to the numerator of the F-statistic.) The empirical values of the statistics F are: $F_1 = 6.998$, $F_2 = 5.031$, $F_3 = 6.822$. The hypothesis which concerns us more is H_0^3. It tests the possibility of a process with a unit root that follows a random walk, of the DS type with drift, on which we have a doubt. The empirical value is higher than the tabulated value. The hypothesis is rejected. It is not a DS process with drift.

The ADF Test Estimates the Model of the Type [4]

This model without trend nor constant term is written as explained previously (table of results not presented):

$$\Delta y_t = \rho y_{t-1} - \sum_{j=2}^{p} \phi_j \Delta y_{t-j+1} + \varepsilon_t. \tag{3.38}$$

The test t $(t_{\hat{\phi}_1} = 0.563870)$ is higher than the critical values for all the given thresholds. H_0 is accepted. Thus the traditional statistical methods lead to conclude that our sample of the *Cac40* is a "random walk", i.e. a DS process without drift. This observation is checked by means of the correlograms of first differences, which according to the traditional analysis highlight a "white noise" process. Indeed, it will be observed that the Q-statistic has a critical probability of 0.482 (for $k = 15$) largely higher than 0.05. The *Cac40* series is thus a DS process, i.e. a random process (or a random walk) that we can make stationary by means of a filter using the differences: $(1 - D)^d x_t = \beta + \varepsilon_t$, where ε_t is a stationary process, β a real constant, D the shift operator and d the order of differences (filter). Here $\beta = 0$, then the process is known as without drift. (to confirm the analysis it would be necessary to test the ADF model n° 5 which is not presented here. It will simply be said that it introduces a constant term c, and that the test T $(t_{\hat{\phi}_1} = -3.29)$ is higher only for the 1% critical value and is lower for the other 5% and 10% critical values. Thus we tend to reject the hypothesis H_0, except for the critical values lower than 1%. The ADF tests of Joint Hypotheses are not implemented within the framework of this work.

Conclusion of the Stationarity Tests

The set of DF and ADF tests leads to conclude that the French stock exchange sample *(Cac40)* is a DS process without drift, even if certain components from this relatively elementary group of tests (presented or not) somewhat impugn this conclusion.

3.2.3 Correlogram Analysis of the Cac40 Sample

The autocorrelation function written: $\rho_k = \sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})/\sum_{t=1}^{T}(y_t - \bar{y})^2$ measures the correlation of the shifted series of k periods (with $\bar{y}$ average of the series on $n - k$ periods). For a lag k equal to 15, we obtain the results presented in the following table with the correlogram graphs:

	AC	PAC	Q-Stat	Prob
1	0.994	0.994	1972.6	0.000
2	0.987	-0.011	3920.6	0.000
3	0.981	0.017	5845.0	0.000
4	0.975	0.009	7746.7	0.000
5	0.969	-0.022	9624.8	0.000
6	0.963	0.020	11481.	0.000
7	0.957	-0.003	13314.	0.000
8	0.951	0.004	15126.	0.000
9	0.945	0.010	16917.	0.000
10	0.939	0.017	18688.	0.000
11	0.934	-0.014	20439.	0.000
12	0.928	0.001	22170.	0.000
13	0.923	-0.003	23881.	0.000
14	0.917	0.008	25572.	0.000
15	0.912	0.003	27245.	0.000
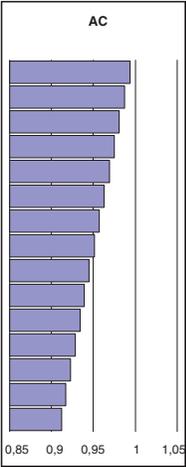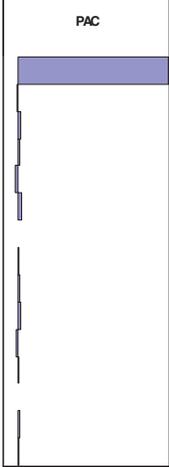
If the Cac40 correlograms are observed, it is noted, first of all, that the autocorrelation (simple correlogram) decreases in a regular way. Thus the signal has a trend that we will have to eliminate (by first differences) to better identify the behavior of the data. Let us recall that the conclusion of the DF and ADF tests was a DS process without drift. Furthermore, in the partial correlogram, only its first term is significantly different from zero. Indeed, the partial correlation of order 1 is the only significant one, since there is only an influence of cac_t on cac_{t-1}, and the effect of the other variables is removed.

We also study the hypothesis of normality of the series distribution (see histogram) by means of the empirical values of the Skewness, Kurtosis and Jarque Bera statistics $v_1 = |\beta_1^{1/2} - 0|/\sqrt{6/n} = |-1.02|/\sqrt{6/1995} = 18.61 > 1.96$ and $v_2 = |\beta_2 - 3|/\sqrt{24/n} = |4.33 - 3|/\sqrt{24/1995} = 12.12 > 1.96$.

Thus we are led to reject the normality hypothesis with regard to the symmetry and the flatness of the distribution, which is confirmed by the Jarque Bera statistics. Here, $JB = 494.32 > \chi_{0.05}^2(2) = 5.99$, the distribution of the *Cac40* is *non Gaussian*. Thus, the *Cac40* has *a trend* and is a *non Gaussian* process (*the experimentation mainly shows that complex signals seem non Gaussian*). A priori, here, the study of correlograms leads to the conclusion that the series is an *AR(1)* process.

First-Differences

To make the series stationary, it is necessary to use the first-differences: $(1 - D)cac_t = cac_t - cac_{t-1}$ where D is a shift operator. Once the signal is corrected of its trend by means of first differences (table and graph not presented here), the observation of both correlograms of the series shows that all the terms of the simple and partial correlograms are inside the confidence interval, beyond which it is considered that the values are significantly different from zero. Thus, the terms are not significantly different from zero. Traditional approaches usually said that such an observation is characteristic of a "white noise" process. The Q-statistic of the Ljung–Box (calculated by Matlab or Eviews) confirms this observation: Q-stat $= 14.58$ (with a delay $k = 15) < \chi^2_{0.05;15} = 25$. Thus, we would be led to accept the hypothesis of the nullity of coefficients ρ_k (the critical probability of this test is indicated $\alpha_c = 0.482 > 0.05$, therefore one accepts the null hypothesis H_0). The Ljung–Box Q-Statistic evoked above is given by:

$$Q_{LB} = T(T+2) \sum_{j=1}^{P} \frac{r_j^2}{T-j}, \tag{3.39}$$

where r_j is the j-th autocorrelation and T the number of observations. Below, the first-differences of the series:
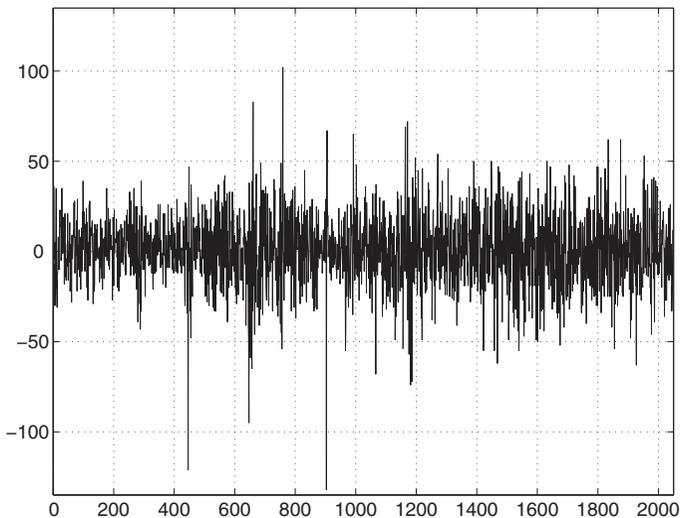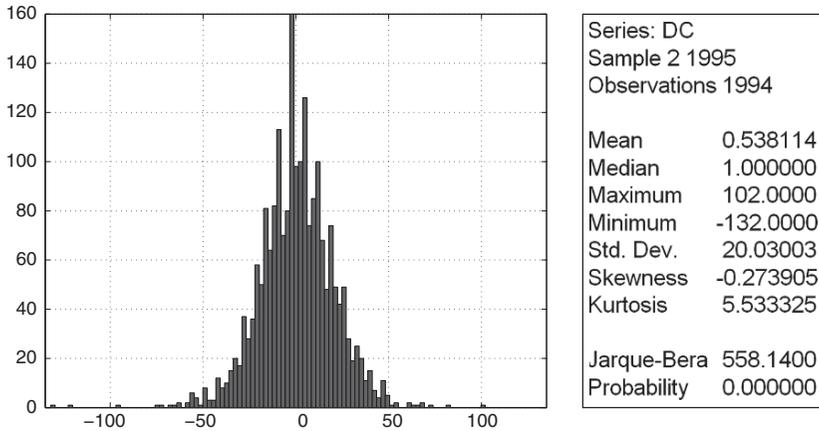
The normality tests highlight a *non Gaussian* distribution. The analyzed series is thus a non Gaussian "white noise". As we have observed it at the end of the treatment of the series by the first differences, the correlograms observed are thus satisfactory, and the series is *detrended*. The cac40 is an *autoregressive process of an order 1,* written *AR(1) or ARMA(1,0).*

3.2.4 Estimation of the Model

Within the framework of the analysis of both correlograms, we observed that the first differences give good results and thus, our series is taken as an $AR(1)$ process (although we could consider it as an $ARMA(1,1)$ process). The estimated model $AR(1)$ gives the following equation:

$$CAC = 11.750002 + 0.99381019 * CAC(-1)$$

LS // Dependent Variable is CAC
Sample: 2 1995. Included observations: 1994

Variable	Coefficient	Std. error	T-statistic	Prob.
C	11.75000	3.196375	3.676040	0.0002
CAC(−1)	0.993810	0.001747	568.7761	0.0000

Above on the right, the distribution of the residual does not seem Gaussian.

Series: Residual	
Sample 2 1995	
Observations 1994	
Mean	7.61E-14
Median	-0.276168
Maximum	99.27474
Minimum	-132.4846
Std. Dev.	19.96723
Skewness	-0.248110
Kurtosis	5.511868
Jarque-Bera	544.6708
Probability	0.000000
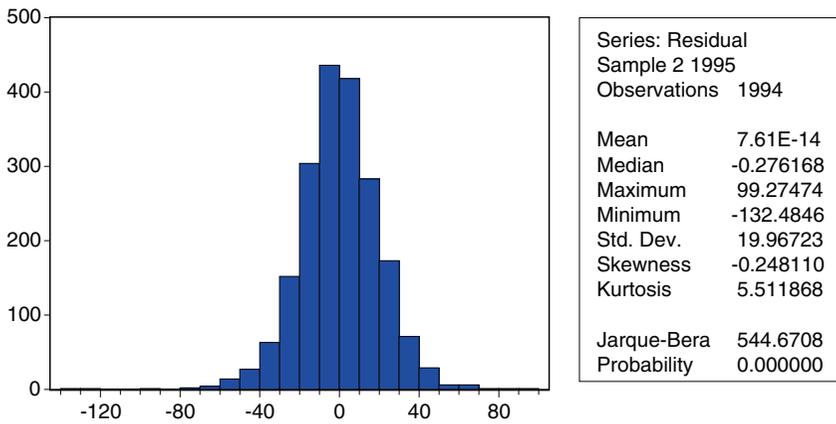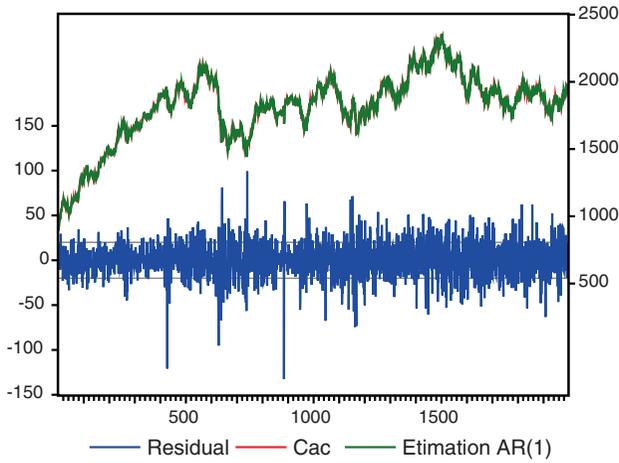
3.2.4.1 Analysis of the Residuals of the Model

Analysis of the Residuals of the ARMA(1,0) Model

The graphic analysis (not presented here) of residuals leads to predict an absence of autocorrelation. The correlograms of residuals of the $AR(1)$ model on the Cac40 are of good quality, i.e. they are clearly inside the confidence interval. But the visual analysis does not always allow to conclude, thus it is sometimes necessary to use the Durbin Watson test.

Durbin–Watson Test

The conditions of use of the Durbin–Watson test must be respected, i.e. the model must be specified in time series, the number of observations must be higher than 15

and the estimated model must have a constant term, which is the case here. The test statistic calculated on the residuals is:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}. \tag{3.40}$$

The e_t obviously are the residuals of the model. (The test statistic is to be studied in the Durbin–Watson table.) The test of DW can identify an autocorrelation of errors of order 1^2 of the type: $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$, with $v_t \to N(0, \sigma_v^2)$. The following hypotheses are tested, H_0: $\rho = 0$ and H_1: $\rho \neq 0$. The purpose is to accept or reject the hypothesis of the autocorrelation of residuals. Here the test statistic of DW is equal to 1.92. This value is to be compared with the value of the table of DW. ($k =$ variables, $T = n =$ size of the sample, here $k = 1$ and $T = 1995$). The value of the empirical test statistic DW is in the zone: $d_2 < DW < 4 - d_2$, (i.e. $1.69 < DW < 4 - 1.69$). Thus, we are led to accept the null hypothesis H_0 with $\rho = 0$. This allows us to predict a good adjustment. The conclusion is the absence of an autocorrelation of the residuals, i.e. with their *independence*.

Q-Test of Autocorrelation of Residuals

If one studies the statistic Q of the correlograms of residuals, the empirical value Q of the Ljung–Box statistic is equal to 20.13 with the delay $k = 20$. It is lower than the value with the 5% threshold given by the table $\chi_{0.05}^2$ which is 31.41. Thus, we are led to accept the hypothesis of the nullity of the coefficients ρ_k (the critical probability of this test is given $\alpha_c = 0.45 > 0.05$, thus one accepts the H_0 assumption). Then, the *residual can be assimilated to a "white noise" process*. (We could compare these results with the results of an *ARMA*(1,1) model.)

3.3 Econometrics of Nonlinear Processes

3.3.1 Stochastic Processes: Evolution of Linear Modeling Towards Nonlinear Modeling

The ARMA modeling is representative of linear modeling. But the hypothesis of linearity is in numerous domains improper to represent the real, economic or financial phenomena.

In Economics, the models used during the twentieth century were primarily of a linear nature, including those aiming at representing business cycles, in particular the Frisch–Slutsky linear model (1930). According to these models, cycles resulted from the propagation of external shocks in our economies and were not generated in an endogenous way by the model. All the models derived from these principles quickly showed their limits. The exploitation of nonlinear models proved to be

necessary, Hicks (1950) and Goodwin (1951) had already introduced them, but they had little resonance at that time, in particular due to the lack of statistical and computational tools. The ARMA processes thus remained dominant in econometrics for a long time. The linearity hypothesis is an enormous constraint. It cannot depict, for example, the relationship between the values of a stock exchange index on the financial markets. The latter are characterized by nonlinearities and a volatility which vary in the course of time.

The ARMA processes do not depict asymmetrical phenomena that exist, for example, during business cycles between the length of the tops and that of the bottoms of cycles. Moreover, these ARMA processes do not render or depict the ruptures with great variances. Lastly, an interesting characteristic of these processes is that they are not able to formalize the moments of a higher order than two for the autocovariance, only nonlinear models are able to formalize them.

Usually, it is said that there are two types of nonlinear models in Statistics (or Econometrics), the nonlinear models on average and the nonlinear models on variance. The nonlinear models on average are either built on an evolution of ARMA models or rest on the analysis of frequencies, and in particular on the bispectral analysis, on the path of *Polyspectra*. The nonlinear models on variance concern the models of ARCH type.

The purpose of this part is not to depict the statistics of nonlinear stochastic processes, the literature on the subject is currently quite abundant. The purpose is rather to depict the recent and relevant developments of the statistical analysis relating to the chaotic nonlinear processes: Non-parametric analysis, Nonlinearity tests, Density Estimators, Estimators of dynamical systems, Long memory processes, ARFIMA process, etc.

3.3.2 Non-Parametric Test of Nonlinearity: BDS Test of the Linearity Hypothesis Against an Unspecified Hypothesis

Beyond the tests of linearity, i.e. Keenan (1985) or Granger and Newbold (1976) tests (for which the alternative hypothesis of nonlinearity does not assign a particular form to the nonlinearity), there exists the more recent and particularly interesting BDS test. The non-parametric test-statistic of Brock, Dechert and Scheinkman (BDS) in 1987 rests on the correlation integral of a series. This test is a test of nonlinearity which attempts to observe if the elements of a series are independent and are identically distributed: *iid*. First, let us consider an observed time series $x_1, \ldots, x_T$, with $t = 1, \ldots, T$. Then we construct a vector from this time series that deploys it inside an embedding-space (see Takens theorem) of dimension m:

$$x_t^m = [x_t, x_{t+1}, \ldots, x_{t+m-1}], \tag{3.41}$$

where $t = 1, \ldots, T - m + 1$. Then, we construct and calculate the correlation integral $C_m(m, \varepsilon, T_m)$ that measures the number of couples (x_i, x_j) at a distance lower than ε, considering ε as a limiting length. C_m is defined by:

$$C_m(m, \varepsilon, T_m) = \frac{1}{T_m(T_m - 1)} \sum_{i,j=1}^{T_m} H(\varepsilon - |x_i^m - x_j^m|), \qquad (3.42)$$

$H(\cdot)$ is a Heaviside function where $H = 1$, if $(\cdot) > 0$. The test considers that:

$$T \to \infty, \text{ then } C_m(\varepsilon) \to (C_1(\varepsilon))^m,$$

and

$$\sqrt{T}[C_m(m, \varepsilon, T_m) - (C_1(1, \varepsilon, T_m))^m] \to N(0, \sigma_m^2), \qquad (3.43)$$

what means that when the length of the time series tends towards the infinite, $T \to \infty$, then $C_m(\varepsilon)$ tends towards $(C_1(\varepsilon))^m$ and, moreover, the numerator of the BDS statistic: $\sqrt{T}[C_m(m, \varepsilon, T_m) - (C_1(1, \varepsilon, T_m))^m] \to N(0, \sigma_m^2)$, where the variance σ_m^2 has a heavy expression that we do not present here (see Hsieh 1991). [$\sigma_m(\varepsilon)$ is the standard error of $C_m(m, \varepsilon, T_m) - (C_1(1, \varepsilon, T_m))^m$ which can be estimated by using the Monté Carlo method]. The BDS statistic can be written:

$$BDS(\varepsilon) = \sqrt{T} \frac{[C_m(m, \varepsilon, T_m) - (C_1(1, \varepsilon, T_m))^m]}{\sigma_m(\varepsilon)}. \qquad (3.44)$$

It can be shown that asymptotically, as $T \to \infty$, the distribution of the BDS is a (reduced and centered) Normal law $N(0, 1)$, when the x_t^m are independent and are identically distributed: *iid*. Opposed to an independent and identically distributed behavior, we can find: linear dependence, non-stationary, dynamical chaos or nonlinear stochastic processes.

This BDS test aims at the acceptance or the rejection of the independence hypothesis of the "terms" of a time series but is not used for the measurement or the evaluation of a "deterministic chaos". The test makes it possible to know if we are faced with a white noise with "independent terms". *We can use the BDS as a test of nonlinearity.* Indeed, if our objective is the estimation of a linear model, the residuals resulting from this estimation are evaluated as non independent, then one can conclude that the analyzed time series is of nonlinear nature. On the other hand, we cannot skip this step and announce that the time series is chaotic. The test is not a measurement of the chaotic nature of a time series but can help to give information about its nonlinearity. Let us precise that before using this test, we must "stationarize" (make stationary) the initial time series and remove all the linear dependences that the time series could contain.

3.4 The Non-Parametric Analysis of Nonlinear Models

3.4.1 Parametric Analysis: Identification and Estimation of Parametric Models

The major problem is the observation of time series of which we do not know if there is a theoretical model able to explain it, i.e. we do not know the nature of subjacent equations to these observations. Statistical analysis tries to reconstitute the subjacent deterministic model of a time series (see Gouriéroux and Monfort 1989, Tome I, Chaps. I, II). The try of reconstitution of the model is done by the construction of a stochastic model, with the implicit idea of the existence of a subjacent deterministic structure:

$$x_{t+1} = \varphi(x_t). \tag{3.45}$$

The traditional statistical (or econometric) approach consists in the construction of parametric models by means of linear or non linear functions. The objective is to adapt the stochastic model to the observed time series as well as possible, so that it is its best possible theoretical representation.

A standard stochastic model can be written: $x_{t+1} = \varphi(x_t) + \varepsilon_t$, where ε_t is a white noise.[4] If the observed time series is of linear nature, the relation $\rho(x_t)$ which links the variables will be of linear nature. If the observed time series is of nonlinear nature the relation $\rho(x_t)$, which links the variables, will be of non linear nature. For nonlinear processes, there arises in the field, at least three main categories of models:

1. Asymmetrical heteroscedastic processes: ARCH, GARCH, EGARCH, TARCH, β-ARCH,
2. Long memory processes : GIGARCH, FARMA, ARFIMA, GARMA,
3. Processes with changes of states: SETAR (self-exciting Threshold Autoregressive), process with changes of regime.

The construction of a stochastic model initially passes through an identification of the process, then through *an estimation of parameters of the identified model*. Then, it is necessary to test the quality of the estimation in relation to the initial time series. The quality of the estimation about the past of the chronicle will make it possible or not to be projected towards the future by making forecasts by means of the estimated model. The identification of processes is quite essential, it allows a better adjustment of the (stochastic) model with the observations.

The stochastic modeling of nonlinear processes unfortunately does not allow, for a great number of processes, to produce robust estimations by experimentation. The models are badly specified and forecasts prove to be impossible. We have to face a specification deficit about complex or chaotic processes, thus our attention is drawn to the *non-parametric analysis*. These methods, which are a part of this work, do not necessarily seek to specify a model; they often aim at reconstructing

---

[4] See Stochastic model: $x_{t+1} = \varphi(x_t) + \varepsilon_t$, where ε_t is a white noise.

the orbits or trajectories of an observed phenomenon, without a deterministic or stochastic model. The objective is always to make forecasts. In fact, *the "Dynamical Relation" $\rho(x_t)$ which links the variables can be considered or approached in a non-parametric way,* i.e. without seeking to specify a stochastic parametric model, with its estimators. The non-parametric analysis, usually practiced for robustness reasons, on long time series, seduces by its capacity to extract the properties from a observed phenomenon. The extraction of properties of the time series and its estimators, allows the reconstruction of the dynamical relation ρ. Thus, the nonlinear, complex or chaotic Dynamics which escape from the stochastic-modeling could be reconstructed and if required used for forecasts.

3.4.2 Non-Parametric Analysis

3.4.2.1 The Periodogram is the First Non-Parametric Estimator: An Introduction to "Windowing" and Kernel

The concept of periodogram has been written for more than one century and was the first known non-parametric estimator (Shuster 1896; Prestley 1981). It estimates spectral density, and the most recent evolution of this concept integrated "windows" in its construction. This window concept constituted a true technical contribution, it is applied for example to:

- Histograms
- Polynomial regressions
- Weighting functions
- Approximations by moving average
- Fourier Transformations, with variable or slipping windows
- Smoothing functions
- Estimators of the density
- Or any other type of estimator, as the wavelet transform with variable or slipping windows for example

In this evocation, the purpose is to highlight the concept of window, which largely exceeds the simple idea of segmentation of a time series, because it is associated with a function of various forms, which has vocation to analyze the signal. In part III, time-frequency analysis will highlight the interest of the windowing concept and (core) kernel concept which are of a great utility in the construction of Fourier (or wavelet) transforms with slipping (or variable) windows.

3.4.3 Construction of a Non-Parametric Estimator of Density: From Windowing to Kernel Concept

There are numerous types of windows, sometimes described differently according to their source, from signal analysis, to Statistics, and "Engineering", they are called:

window, filters, or kernel. Statistics seem to offer a unity to this concept with the "kernel" term, while defining its own conditions for applications (stationarity of the process...) and by attributing widened qualities to it (Robinson 1986). It is not only an interval on its fields of analysis, but it is also a weighting function where the kernel can adopt different forms. We will further note the symmetrical forms of these kernels, which in addition could be compared with densities. We consider a long series of observations on $\mathbb{R}$:

$$x_1, \ldots, x_n, \tag{3.46}$$

and we seek to estimate its probability law μ and its associated density f by means of their estimators $\widehat{\mu}$ and $\widehat{f}$. We introduce a constant length interval around the values taken by $x : [x - \frac{\tau}{2}, x + \frac{\tau}{2}]$, this interval segments the set of values of x; the purpose is to work out a histogram. The choice of a narrow τ allows a finer approach of the distribution of values of the time series and also to diminish the loss in relation to the initial data. For $i = 1, \ldots, n$, and for an event:

$$A = \left\{ x_i \in \left[x - \frac{\tau}{2}, x + \frac{\tau}{2} \right] \right\}, \tag{3.47}$$

by using the indicator function of the event A:

$$\mathbf{1}_{x_i \in [x - \frac{\tau}{2}, x + \frac{\tau}{2}]} = \begin{cases} 1 & \text{if } x_i \in [x - \frac{\tau}{2}, x + \frac{\tau}{2}], \\ 0 & \text{otherwise.} \end{cases} \tag{3.48}$$

We simply proceed to the enumeration of values of the time series included in the interval $[x - \frac{\tau}{2}, x + \frac{\tau}{2}]$. We order the possible values of x and the enumeration obtained for each value of x is the subject of a representation on a histogram. We have just introduced an interval on the distribution of x, it constitutes the work window of values of the time series. It is pointed out that the objective is to estimate the density f associated with the law μ by means of an estimator $\widehat{f}$. One of the usual estimators of the density can be written in the following way:

$$\widehat{f}(x) = \frac{1}{n\tau} \sum_{i=1}^{n} K \left(\frac{x - x_i}{\tau} \right), \tag{3.49}$$

τ is the size of the interval or window and K is the kernel, which is a weighting function on x (with $\tau \to 0$, when $n \to \infty$). We present some examples of kernels (Epanechnikov, biquadratic, cubic, Gaussian) below:

Epanechnikov kernel: $K_e(x) = (3/4)(1 - x^2)$,
Biquadratic kernel: $K_b(x) = (15/16)(1 - 2x^2 + x^4)$,
Cubic kernel: $K_c(x) = (35/32)(1 - 3x^2 + 3x^4 - x^6)$,
Gaussian kernel: $K_g(x) = (1/\sqrt{2\pi}) e^{-x^2/2}$.

The kernel can work on the variable on different length intervals, i.e. the weighting effect around x is more or less great and depends on the construction of each

kernel. The kernels are represented on the graph below and we will observe the length of the *compact support* on which each kernel exists. Except the Gaussian kernel, we note that the supports of these kernels are $[-1,1]$: the window of these kernels is written: $\mathbf{1}_{[-1,1]}(x)$.
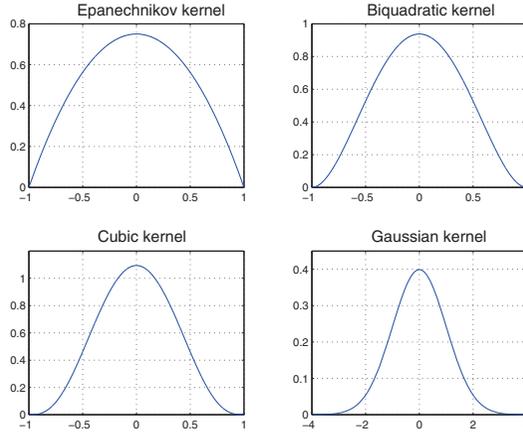
Generalization of the density estimator: It is interesting to give a generalized formulation of the estimator of the density for the spaces whose dimension is higher than one. This means that for example a vector of size equal to D, whose elements are: $x = (x^{(1)}, x^{(2)}, \ldots, x^{(m)}, \ldots, x^{(D)})$, we can note the generalized estimator of the density in the following way:

$$\widehat{f_n}(x) = \frac{1}{n\tau^{(1)}\cdots\tau^{(D)}} \sum_{i=1}^{n} K\left(\frac{x^{(1)} - x_i^{(1)}}{\tau^{(1)}}, \ldots, \frac{x^{(D)} - x_i^{(D)}}{\tau^{(D)}}\right), \qquad (3.50)$$

or:

$$\widehat{f_n}(x) = \frac{1}{n} \sum_{i=1}^{n} \prod_{m=1}^{D} \frac{1}{\tau^{(m)}} K\left(\frac{x^{(m)} - x_i^{(m)}}{\tau^{(m)}}\right), \qquad (3.51)$$

with the kernel (core): $K(x) = \prod_{m=1}^{D} K_m\left(\frac{x^{(m)} - x_i^{(m)}}{\tau^{(m)}}\right)$.

A simple example: By means of a pseudo-random generator of series, we create a Gaussian distribution (for example) of dimension 1, then we represent on a graph the normalized histogram of the distribution (Fig. 3.1). Then, we estimate the density with a Gaussian kernel and we superimpose the obtained result (the curve below) on the normalized histogram of the distribution.

Fig. 3.1 Histogram
and estimated density
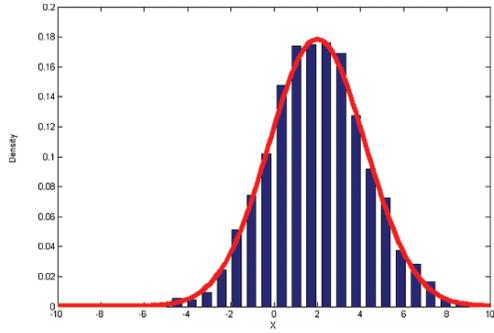
3.4.4 Estimator of Density and Conditional Expectation of Regression Between Two Variables

If we imagine two random variables (y,z), we can estimate the conditional density of the variable y given z. The density is written:

$$\widehat{f_n}(y,z) = \frac{1}{n\tau^2} \sum_{i=1}^{n} K\left(\frac{y-y_i}{\tau}\right) K\left(\frac{z-z_i}{\tau}\right). \tag{3.52}$$

The generic form of the conditional expectation $E(y/z)$ is written:

$$E(y/z) = \frac{\int yf(y,z)dz}{\int f(y,z)dz}. \tag{3.53}$$

This conditional expectation of y given z, can be estimated from a sample length n of variables (y,z), and we write the estimator of the conditional expectation by combining the two preceding equations:

$$\widehat{E}_{y/z} = \frac{\sum_{i=1}^{n} y_i K\left(\frac{z-z_i}{\tau}\right)}{\sum_{i=1}^{n} K\left(\frac{z-z_i}{\tau}\right)}. \tag{3.54}$$

Such an estimator of the conditional expectation gives a method to reconstruct the dynamics of y given z.

Example of conditional densities and "a posteriori" probabilities (i.e. posterior probabilities): The chosen illustration initially consists in producing data of dimension two, calculated from a mixture of three Gaussian random series, centered at three centers: $[(0, -0.1), (1, 1), (1, -1)]$. One of these series is assigned to a class and the two others are assigned to the second class. (The first series is of *class 1: [red points]* and the two others are of *class 2: [blue stars]*.) In the figure below on the left, the first series of class 1, centered close to the origin $(0, -0.1)$ is represented by the *(red) points* and the two other series gathered under class 2, are represented by

(blue) stars. We distinguish the centers marked by the crosses. Below on the right, the density of the non-conditional probability of the set of data gathered under the variable x.
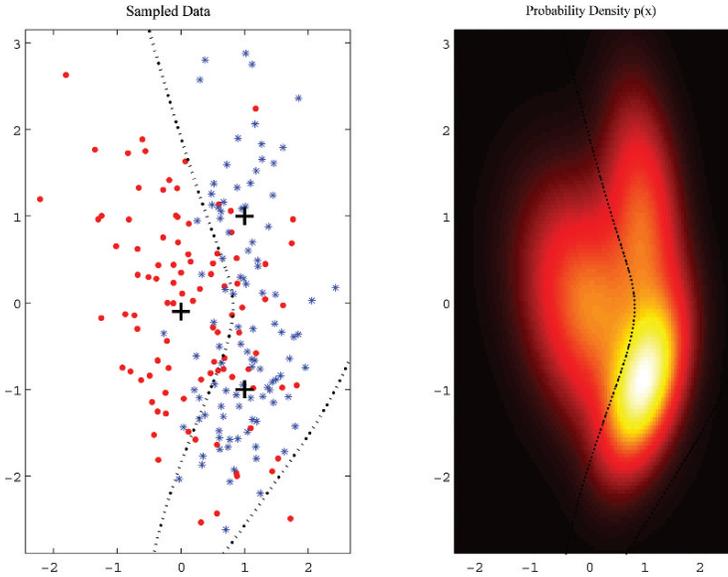
In order to better visualize the contours of the non-conditional probability density, we show it below with the perspective:
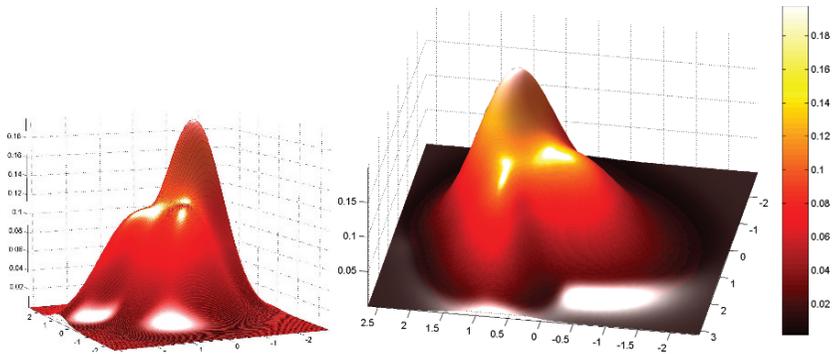
We distinguish the three poles around which the density is concentrated. They correspond to the centers of the Gaussian series, and the strongest concentration of the density is around $(1, -1)$. Below, the conditional densities and the posterior probabilities are represented successively. The top-left image below shows the conditional density of x which is function of the series (of class 1 represented by) the "(red) points", i.e. $p(x/red\ points)$. The top-right image below shows the conditional density of x, which is function of the other series (of class 2 represented by) the "(blue) stars", i.e. $p(x/blue\ stars)$. The bottom-left image shows the posterior probability of the series "(red) points" which is function of x, i.e. $p(red\ points/x)$. The bottom-right image shows the posterior probability of the series "(blue) stars"

which is function of x, i.e. $p(blue\ stars/x)$. The scales of colors at the right of the images assign a probability to a color (or nuance of gray). One can thus see the zones with high and weak probabilities immediately.
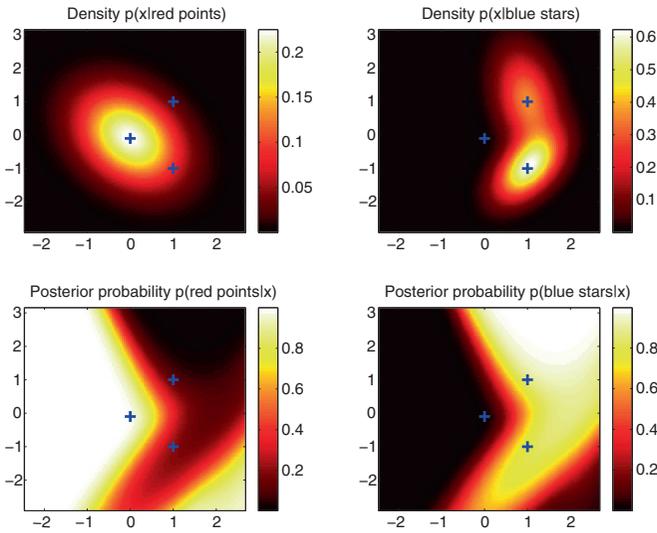
3.4.5 Estimator of the Conditional Mode of a Dynamics

It is from the conditional law of x_t and x_{t-1} that we can construct such an estimator, which is efficient for the dynamics of a deterministic nature. Then we estimate the density associated with conditional law, it is written:

$$\widehat{f_n}(x, y) = \frac{1}{n\tau^2} \sum_{t=1}^{n} K\left(\frac{x - x_t}{\tau}\right) K\left(\frac{y - x_{t-1}}{\tau}\right), \qquad (3.55)$$

with $\tau \to 0$, when $n \to \infty$. An estimator of the conditional mode $\widehat{\varphi}_n^{[mc]}$ of the dynamics φ verifies the following inequality:

$$\widehat{f_n}(x, \widehat{\varphi}_n^{[mc]}(x)) \geq \sup_{y} \widehat{f_n}(x, y) - \Theta_n, \qquad (3.56)$$

with $\Theta_n \to o$, when $n \to \infty$.

3.4.6 A First Estimator of Dynamics by Regression

To estimate the dynamics of a variable, we can of course proceed to a regression known as Nadaraya–Watson regression. It will be noticed that the conditional

expectation of y given z, presented previously:

$$\widehat{E}_{y/z} = \frac{\sum_{i=1}^{n} y_i K\left(\frac{z - z_i}{\tau}\right)}{\sum_{i=1}^{n} K\left(\frac{z - z_i}{\tau}\right)}, \tag{3.57}$$

is constructed on a similar mode with that of the regression which is written:

$$\widehat{\varphi}_n^{[reg:N.W]}(x) = \frac{\sum_{i=1}^{n} x_i K\left(\frac{x - x_i}{\tau}\right)}{\sum_{i=1}^{n} K\left(\frac{x - x_i}{\tau}\right)}. \tag{3.58}$$

It is known that such an estimator is efficient for the dynamics which is of a deterministic nature or which has components with a deterministic dominant.

3.4.7 Estimator by Polynomial Regression

Such an estimator is written in the following way:

$$\widehat{\varphi}_n^{[reg.p]}(x) = \underset{\{\alpha_0, \alpha_1, \dots, \alpha_p\}}{\arg\min} \sum_{i=1}^{n} [(x_i - \alpha_0 - \alpha_1(x_i - x) - \dots - \alpha_p(x_i - x)^p]^2 K\left(\frac{x - x_i}{\tau}\right)$$

with $\tau \to 0$, when $n \to \infty$ and $\{\alpha_0, \alpha_1, \dots, \alpha_p\}$ are the parameters of the regression. The degree p of the polynomial is to be chosen as a preliminary to do the estimation, whose quality increases when the degree increases. Generally this method is coupled with the least squares technique.

3.4.8 Estimator by the k-Nearest Neighbors Method: KNN

To establish the link with the section concerning the construction of the estimator of the density, we know that a density estimator can be constructed by the method of the k-nearest neighbors. Indeed, it is necessary to replace the width of the window τ in the expression of the estimator of the density presented previously $\widehat{f}(x) = \frac{1}{n\tau}\sum_{i=1}^{n} K((x - x_i)/\tau)$, by a random variable which measures the distance between a point x and the k-nearest neighbors taken among the observed values of the initial series $x_1, x_2, \dots, x_n$. The technique thus consists in adjusting τ, which is an interval or a step, in the density estimator with an usual kernel, in order to always use only the same number k of observations to construct an estimation of the density. The technique of nearest-neighbors exploits the idea that the number of observations

locally used is always identical, whatsoever the stability of the trajectory around the considered point x. The estimator construction of nearest neighbors of the dynamics φ is built on the same principle.

The method of nearest neighbors to reconstruct the original dynamical system consists in finding in our time series $x_1,\ldots,x_n$ the nearest observed value from x_n that we note $x_i^* \in x_1,\ldots,x_n$ then in taking the following value x_{i+1} of x_i^* and using it to reconstruct x_{n+1} the future of x_n (which is the last historical value of the time series). We can give an example for $k=1$, i.e. for the estimator of the dynamics φ with 1-nearest neighbors, we simply write:

$$\widehat{\varphi}_n^{[knn]}(x_n) = x_{i+1}^*. \tag{3.59}$$

It is also possible to choose the k nearest neighbors to construct an estimation of x_{n+1}. Thus, the estimator of φ is written then:

$$\widehat{\varphi}_n^{[knn]}(x_n) = \widehat{x}_{n+1} = \sum_{i=1}^{k} w(\|x_n - x_i^*\|)x_{i+1}^*. \tag{3.60}$$

The weighting function $w(\cdot)$ can take different forms such as for example:

$$w(\|x - x_i^*\|) = \frac{e^{-\|x-x_i\|^2}}{\sum_{i=1}^{k} e^{-\|x-x_i\|^2}} \, or : w(\|x - x_i^*\|) = \frac{1}{k}. \tag{3.61}$$

The major stake is the choice of the number k of k-nearest neighbors. The experiment shows that the method of nearest neighbors is efficient enough to estimate and reconstruct a chaotic dynamics.

3.4.9 Estimator by the Radial Basis Function Method: RBF

This method is not so different from the precedent. It consists in taking C points in our time series $x_1,\ldots,x_n$ which is a discrete expression of the dynamics. These C points, generally resulting from a classification technique, are noted y_c. The estimator is written:

$$\widehat{\varphi}_n^{[rbf]}(x_n) = \widehat{x}_{n+1} = \sum_{c=1}^{C} \lambda_c w_c(\|x_n - y_c\|), \tag{3.62}$$

where $w_c(\cdot)$ are the radial functions, r_c is the radius of the radial function w_c and λ_c is generally estimated by least squares. To build this estimator it is necessary for us thus to select: a radius r_c, the radial function $w_c(x)$ and the points y_c. Some models of radial functions are given below:

Linear:	$w_c = x/r_c,$	Spline:	$w_c = (x/r_c)^2 \log(x/r_c),$
Gaussian:	$w_c = e^{-(x/r_c)^2},$	Multi-quadratic:	$w_c = \sqrt{(x^2 + r_c^2)}.$
Cubic:	$w_c = (x/r_c)^3,$		

The experiment shows that the method of the radial basis function is very powerful to estimate and reconstruct a chaotic dynamics. One gives a diagrammatic representation of the regression by radial basis functions.
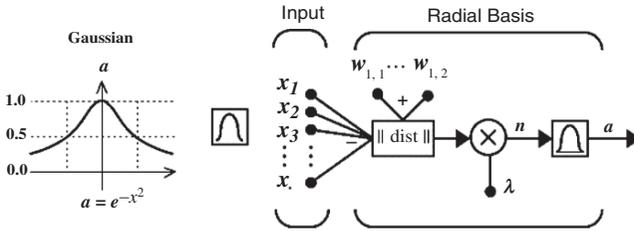
3.4.10 A Neural Network Model: Multi-Layer Perceptron and Limit of Decision

Explain the multi-layer perceptron model in a relevant and concise manner seems difficult, especially because it comprises many variants since the origin (Rosenblatt 1962). To place the perceptron within a history of the set of neuronal models (Networks of Hopfield) could have appeared very interesting, as well as to explain its construction from the point of view of algorithms. However in spite of the interest of these methods, it seemed to us that a detailed presentation would be too heavy within the framework of this book.[5]

The perceptron is a "Classifier", a "linear separator". The perceptron proceeds to a discrimination for a decision. It is pointed out that a *classification* consists in examining the characteristics of an "object" or a "model" and assigning a class to it, the class being a particular field with discrete values. The "training" by the perceptron (as well as by the method of least squares) is a technique of "linear separation" (comparable with certain techniques used in statistics). It is a *non-parametric method,* which does not require to make any other hypothesis on the data, except on the *separability*.

The perceptron has been the subject of a very strong interest at its beginnings from the first work of Rosenblatt in 1962, criticized by Minsky and Papert in 1969.[6] Minsky and Papert rejected in particular the possibility of building "training" algorithms for the networks with several layers. The perceptron was the center of different stakes and in particular that of the transition from the analogic to the digital. And it is by means of the simulation possibilities offered by the digital systems that the perceptron had a strong resurgence of interest since 1980. Today the

[5] It is possible to consult the work of Marc Tommasi (1997 MET DST) who gives an excellent analytical presentation of the multi-layer networks and in particular of the perceptron. It is possible to also consult the Matlab-Neural Network tools which clearly present the construction of the algorithms.

[6] Ref: see about the limits of the perceptron the following work which has been republished: Minsky and Papert (1988).

multi-layer networks are a good answer to overhaul the limitations of the simple perceptron and Hopfield models of that time. The separator or classifier of the perceptron can be applied to a dynamical system, either on its flow or on an *embedding space* of the dynamics.

While leaving the precise framework of the perceptron, we can say about the Hopfield networks (which are equivalent networks) that they classify the configurations according to their *basin of attraction*, and two configurations are equivalent if they have same *attractor*. Such classifications generally rest on a *criterion of neighborhood, i.e. a distance*, usually it is a distance of Hamming. This criterion is not always adapted, in particular within the framework of the "Form Recognition", for which the neural networks are interested in the *invariances*, and for which we can say that two forms are equivalent if they result one from the other by *translation*. And in such a case, the distance of Hamming of both models (or patterns) can be important.

When we work on a rather significant number of data and when the problem allows it, we observe that the perceptron has a good predictive capacity. However, the majority of the "training" problems cannot be solved by so simple methods, the possibility that natural phenomena are distributed easily on both sides of a hyperplane is not very probable. A way of solving this problem would be either to develop *nonlinear separators* or in an equivalent way to "complexify" the *space of representation* in order to *linearize* the initial problem. *The multi-layer networks* make this possible. It is necessary to understand choosing a too simple architecture *limits the nature of correlations* that the network treats, and on the opposite, choosing a too complex architecture (too big number of layers and cells) generates a too long "training" time and prevents the generalization (for the decision). The architecture of the network depends on the studied problem. In its simplest, architecture the perceptron (one hidden layer) is made up of:

1. A first layer which is that of the input cells on which we apply the models (or "patterns") to classify.
2. A second layer made up of associative cells (neurons), whose inputs are taken among the input cells. (Generally in the networks with layers, the role of hidden cells is of course very important, it is related to the regularities which the cells detect in the models or patterns presented. We say that there is an internal representation of data.)
3. A final layer that includes one or more cells of decision which are usually functions with threshold whose weights are modified during the "training" procedure.

We give below the example of a multi-layer perceptron with one input layer, then five cells, then two hidden layers having respectively three and two neurons, then one output layer (one neuron):
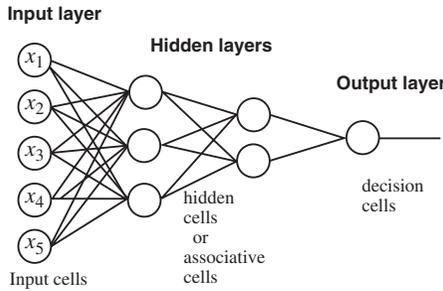
To choose an architecture, we generally start with "training" tries with a minimum number of layers, i.e. initially without intermediate layer (association layer or hidden layers), then with only one layer, then we gradually increase the number of layers according to the results obtained (the convergence theorem of the perceptron makes it possible to say that if there is a function which carries out the chosen classification, then the algorithm satisfactorily converges in a finite time). Below an example of possible architectures of a perceptron with four inputs and one output:
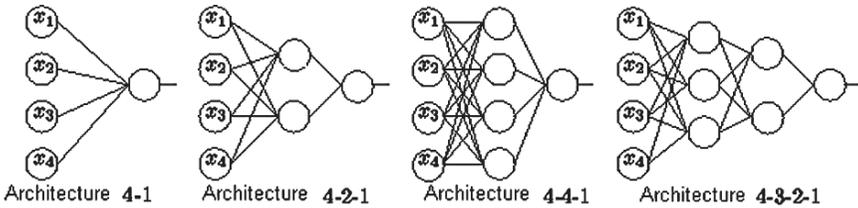
The mathematical model of the multi-layer perceptron: Given a quantitative variable or a qualitative variable y, with q modalities that we must predict using the p "predictives" variables $(x_1, x_2, \ldots, x_p)$. Furthermore, we have n individuals or observations, which are often called the *training sample*, described by the p variables $(x_1, x_2, \ldots, x_p)$ and for which we know the values of y. Then we suppose that the entry layer is made of p-entries (i.e. inputs), to which coefficients will be applied that are called the *synaptic weights* w_{jm}. Moreover, there exists a constant term in the input device (called input constant) which, for practical reasons, takes the value 1. The "hidden layer" includes c *neurons* which each one will be activated by an integration (usually monotonic function of the sum) of the p signals coming from the "input layer". There is a similar operation for the q elements of the "output layer" implying the "synaptic" weights W_{mk}. There is also a direct connection from the constant input to the output.

The introduction of the unitary input constant, connected to each neuron located in the hidden layer and connected to each output, avoids separately introducing what

the computer specialists call a "bias" for each unit. The "bias" become simply parts of the series of weights. Thus, the model will be written for an observation (i):

$$y_k = f\left\{ b_k + \sum_{m=1}^{c} W_{mk}\, \Phi\left(b_m + \sum_{j=1}^{p} w_{jm} x_j \right) \right\}. \tag{3.63}$$

The formula enunciated above corresponds to an observation (i). In the expression which precedes, the function Φ is the *logistic* function, which can be written:

$$\Phi(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}. \tag{3.64}$$

The function f called a transfer function can be: Threshold (Heaviside, Hard-limit, Indicator function), linear or logistic (log-sigmoid):

| Hard-limit Transfer Function | Linear Transfer Function | Log-Sigmoid Transfer Function |
| *(Hardlim)* | *(Purelin)* | *(Logisig)* |
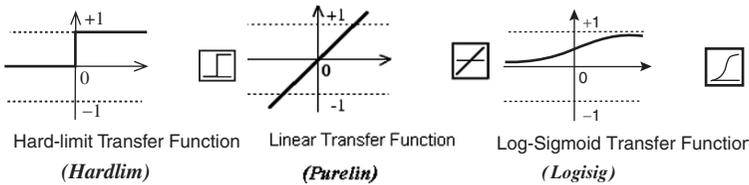
The expression $y_k = f\left\{ b_k + \sum_{m=1}^{c} W_{mk}\Phi(b_m + \sum_{j=1}^{p} w_{jm}x_j) \right\}$ as presented previously, corresponds to an observation (i). To conclude the process we are faced with n equations of this type, each one utilizing q values $y_k^{(i)}$ and p values $x_j^{(i)}$. The estimation of parameters is done by minimizing a *loss function*, which can be simply the sum of the squares of the gap between the computed values $\widehat{y}_k^{(i)}$ and the actual values $y_k^{(i)}$ in the training sample.

Diagrammatic representation of the perceptron: To better understand such a model, it is sometimes necessary to use a diagrammatic representation that helps to perceive the mechanism of the perceptron. Later we will show in the diagram of the multi-layer perceptron with S neurons, but below we will observe the simple perceptron with p-inputs and one neuron:

logsig $(\mathbf{W}x+b)$
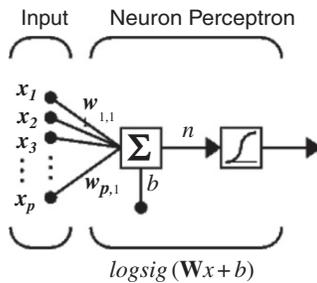
Limits of decision and classification (case of a perceptron with two inputs and one neuron): To give an example, we take the simple case of two inputs with a threshold

transfer function of the "Hard-limit" type (if $n \geqslant 0$, then we have 1, and 0 otherwise). It is known that each external input is weighted by an adapted weight w_j and the sum of weighted inputs is sent towards the hard-limit transfer function which produces 0 or 1. To simplify, we note here the result of the hard-limit transfer function by: $a = hardlim(\mathbf{w}x + b)$. Thus, the perceptron will answer 1, if in the transfer function there is "the input network" higher than 0, i.e. $\mathbf{w}x + b > 0$, otherwise the perceptron will answer 0. The transfer function (hard-limit) gives a perceptron which is able to classify vectors inputs by dividing the space of inputs into two areas. More precisely, the outputs will be equal to 0 if the input network n is lower than 0, or the outputs will be equal to 1 if the input network n is $\geqslant 0$.

$$\text{If} \quad \mathbf{w}x + b \leqslant 0, \quad \text{then the perceptron } a = 0,$$
$$\text{If} \quad \mathbf{w}x + b > 0, \quad \text{then the perceptron } a = 1.$$

The two-inputs "hard-limit" space with the weights $w_{1,1} = -1$ and $w_{2,1} = 1$ and a bias $b = 1$, is represented below, in the right figure:
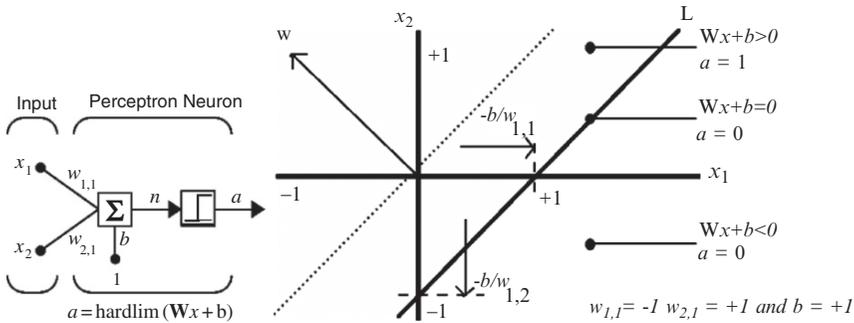
Thus, we have two areas of classification on both sides of a *limit of decision* which is noted $\mathbf{L}$ in $\mathbf{w}x + b = 0$. This line $\mathbf{L}$ is perpendicular to the matrix w and can be moved by the bias b. The vectors inputs above and on the left of the line $\mathbf{L}$ will have the $\mathbf{w}x + b > 0$ and moreover the neuron (hard-limit) will have an output 1. In the same way, the vectors inputs below and on the right of the line $\mathbf{L}$ will have an output 0.

The *ligne* $\mathbf{L}$ dividing the plane can be oriented or *moved everywhere* to *classify* the space input as we wish it *by choosing the weights w* and the *bias b*. The neurons without bias b will always have a line of classification not passing through the origin. By adding a bias b, it allows the neurons to solve problems where the two sets of inputs are not localized on the different sides in relation to the origin. The bias allows the limit of decision to be moved from the origin, as it is possible to observe it easily in the figure above.

Some numerical computation softwares and in particular MATLAB®-Neural Network make it possible to try out the classification method, in particular by the displacement of the limit of decision, by picking new inputs to classify and by observing how the repeated application of the rule of training produces a network which classifies the vectors inputs in a suitable way.
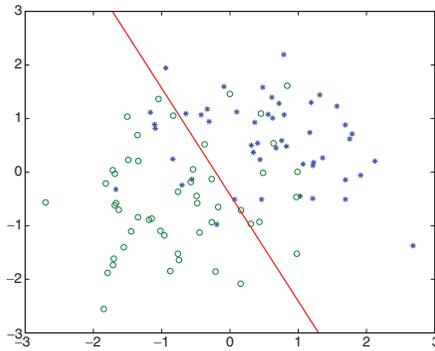
Fig. 3.2 Limit of decision **L** in $\mathbf{w}x + b = 0$ during the training

Graphic illustration of the simple perceptron: The inputs here are two random series, each one assigned to a class of symbols: the "circle" or the "star". Figure 3.2 shows the separator, i.e. the limit of decision (here a straight line) during the training. *Figure of a one-layer perceptron including S neurons:* As presented previously, the figure of the perceptron with one neuron, we depict in the following figure the case of the perceptron with one layer including S neurons and with a Hard-limit transfer function (Heaviside):
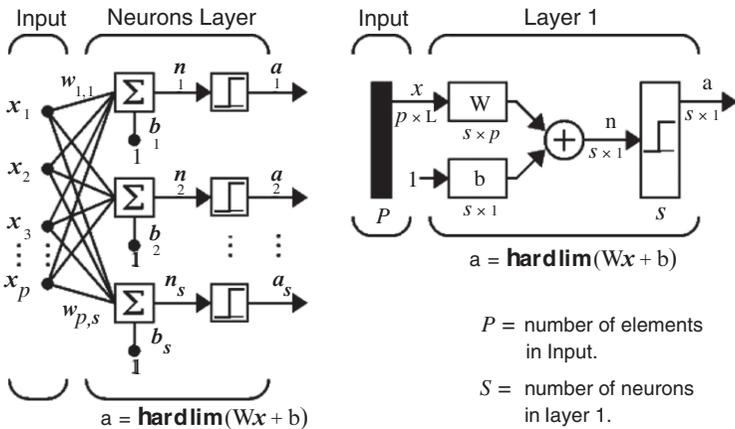
Diagram of a three-layers perceptron: Now remains to show a diagram of a perceptron with several layers in order to help to represent the mechanism of the treatment of the input within the framework of the multi-layer perceptron, and we limit it to three layers for obvious reasons. The layer 1 contains S^1-neurons, the layer 2 has S^2-neurons and the layer 3 has S^3-neurons.

In this diagram, the nature of the transfer function f is not specified. All the elements of the multi-layer perceptron, i.e. the transfer function, the weights, the bias, are indexed according to the layer to which they belong. (**IW** is a

matrix-weight-layer and **LW** is a matrix weight-input. Moreover matrices are noted **W** (bold) and the vectors are noted: *w*.)

$$a^1 = f^1(IW_{1,1}x + b^1) \qquad a^2 = f^2(LW_{2,1}\,a^1 + b^1) \qquad a^3 = f^3(LW_{3,2}\,a^2 + b^3)$$

$$a^3 = f^3(LW_{3,2}\,f^2\,(LW_{2,1}\,f^1\,(IW_{1,1}x + b^1) + b^2) + b^3 = Y$$

$$a^1 = f^1\,(IW_{1,1}\,x + b^1) \qquad a^2 = f^2\,(LW_{2,1}\,a^1 + b^2) \qquad a^3 = f^3\,(LW_{3,1}\,a^2 + b^3)$$

$$a^3 = f^3\,(LW_{3,2}\,f^2(LW_{2,1}\,f^1(IW_{1,1}\,x + b^1) + b^2 + b^3)$$
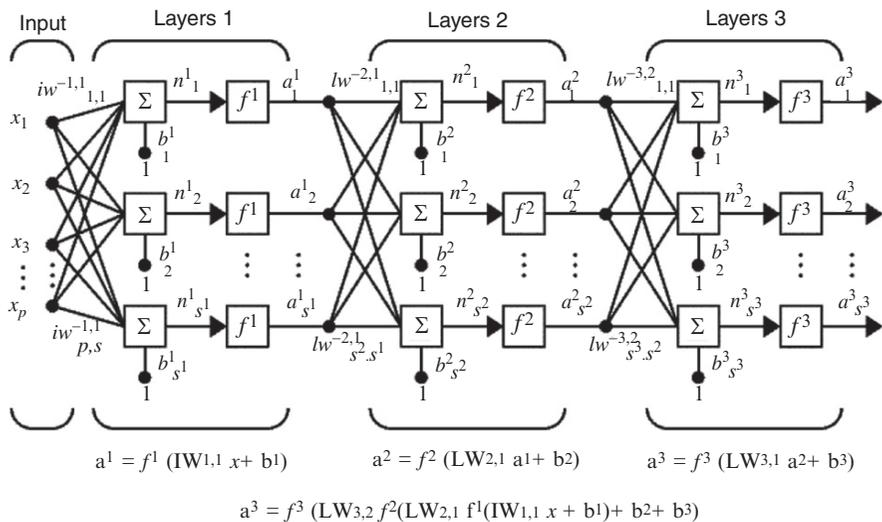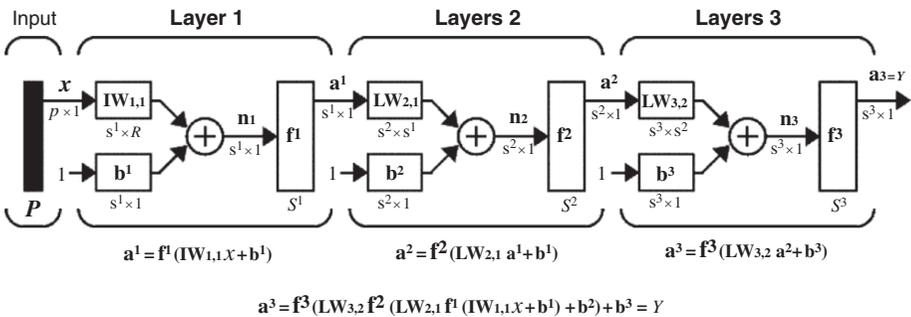
Industrial applications of neural networks:[7] The different methods are used in many spheres of activities:

1. Telecommunications for image and data compression
2. Aerospace industry for flight control systems and simulators
3. Automatic for guidance systems
4. Banks for customer analysis or control of the credit card activity
5. Defense for trajectory guidance, facial recognition, recognition of radar objects or risk analysis
6. Electronics for process control, voice synthesis or nonlinear modeling

---

[7] For a general presentation of the neural networks, it is possible to consult the work of Marc Tommasi (1997): analytical presentation of the multi-layer networks and the perceptron. It is also very interesting to consult the MATLAB® tutorial.

7. Finance for exchange rate and market forecasts and evaluation of companies
8. Robotics for trajectory controls and vision systems
9. Signal analysis for voice recognition, ...

Graphic illustration of classification by the multi-layer perceptron compared with the rule of the optimal Bayesian decision: The illustration that is selected here, retakes the example used in the section concerning the conditional density. It initially consists in producing data of dimension two, calculated from a mixture of three Gaussian random series, centered at three centers: $[(0, -0.1), (1, 1), (1, -1)]$. One of these series is assigned to a class and the two others are assigned to the second class. (The first series is of *class 1: (red points)* and the two others are of *class 2: (blue stars)*.) The multi-layer perceptron with logistic transfer function is applied to these data (we use six hidden layers and the quasi-Newton algorithm of optimization). This example shows the results obtained by means of the multi-layer perceptron and, highlights, in particular the limit of decision that is furthermore compared to the limit of decision obtained by means of the Bayesian method of conditional probabilities. In the figure which follows below on the left, the first series of class 1, centered close to the origin $(0, -0.1)$ is represented by the "(red) points" and the two other series gathered under class 2, are represented by "(blue) stars".

Hereafter on the right the density of the non conditional probability of the set of data gathered under the variable x (moreover, through superimposition in the figure we indicate by means of dotted lines the limit of the optimal Bayesian decision):
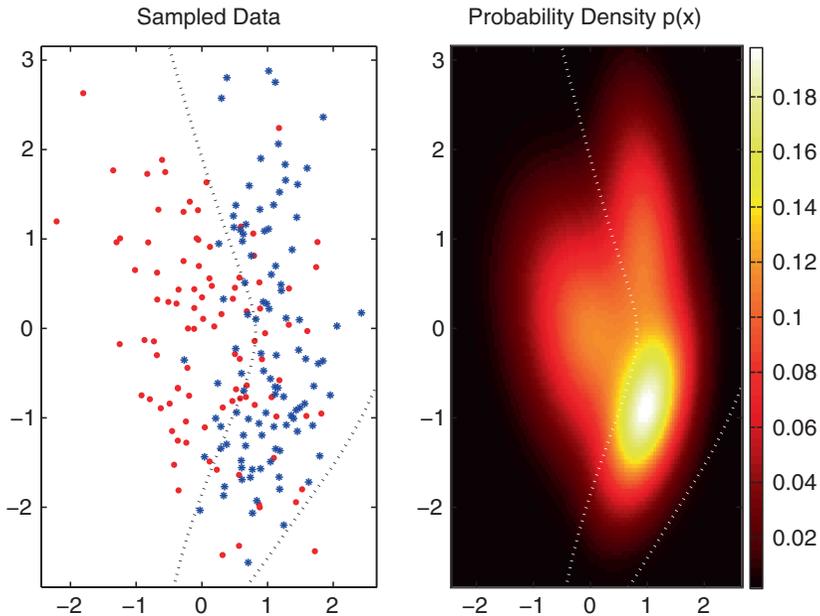
The figure above shows the sampled data of class 1 (red points) and of class 2 (blue stars) as well as the density of probability $p(x)$ of the set of the sample. By
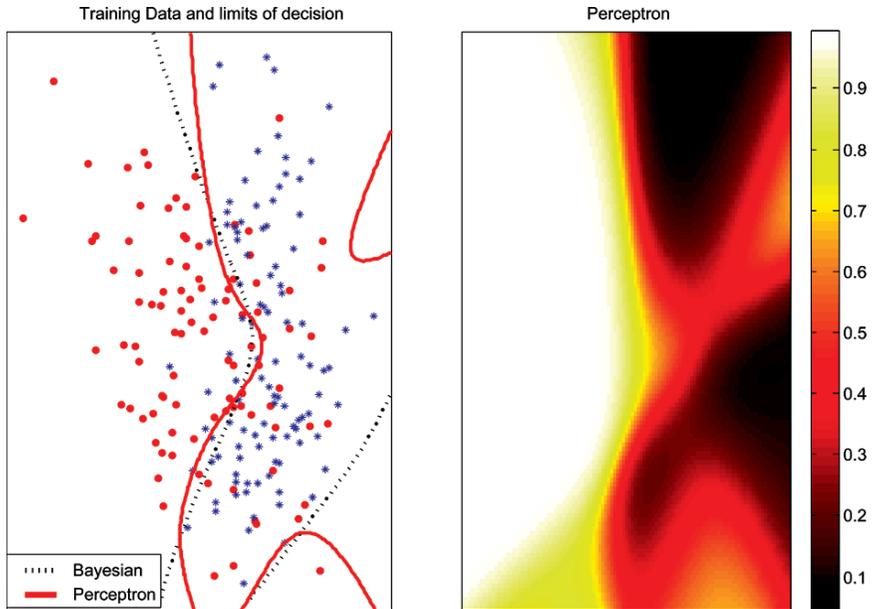
Fig. 3.3 Data with assignments of classes and limit of decision of the multi-layer perceptron. (The bayesian limit of decision corresponds to the *dotted lines*.) It is easy to find the perceptron decision limit on the *right-hand side image*

means of dotted lines, we superimpose the Bayesian decision limit presented before in the section concerning the estimator of the density for the same samples (ref. to Netlab–Matlab, Bishop and Nabney 1995).

Hereafter on the left, the (red) curve shows the limit of decision obtained by means of the multi-layer perceptron, whereas the (black) dotted lines shows the limit of decision resulting from the Bayesian method (Fig. 3.3). The image hereafter on the right shows the output of the perceptron with the limit of decision and the network forecast of the posterior probability of the class 1 (red). The scale on the right of the image assigns a color (or nuances of gray) to the probability of the zone, the black color is close to zero, the white color is close to 1 and the nuances of red (or gray) the intermediate probabilities.

3.5 First Statistical Tests of Validation of Chaotic Process Detection: Brock Test and LeBaron and Scheinkman Random Mixture Test

The book of Lardic and Mignon (2002) underlines that the *tests of chaos detection*, as well as the Lyapunov test of trajectories divergence, or the correlation dimension, *are not statistical tests in a strict sense*. Thus, it was necessary to create statistical

validation tests of these detection tests. It is true also that, according to the field of application of these detection tests, very different problems can be posed, those related to the measurement, which are not neutral on the nature of the time series, those related to the noises which exist inside a series, those related to the aggregation of data are also crucial, or the length of the studied series which is a serious problem in particular in Economics. Beyond this type of difficulties which appear upstream but which is certainly much more structural, it is fundamental for the statisticians to build a theory core of the chaotic dynamics which remains still today to be worked out, in spite of numerous current contributions. The tests indicated in the title are in relation with the Lyapunov exponent and the correlation dimension.

3.5.1 Residual Test of Brock (1986)

The linear transformation of a deterministic system preserves its properties, it is on this concept that Brock constructed its test. The quality of the statistical estimation of a model is measured in particular through its residuals. An autoregression model (AR or ARMA), which has the vocation to provide a deterministic explanatory structure, must thus have a Lyapunov exponent and a correlation dimension of its residuals similar to those of the initial series. The matter is to postulate that if:

$$\lambda_{f(x)}^{Brock} \sim \lambda_x^{Brock} \tag{3.65}$$

and

$$D_{f(x)}^{Brock} \sim D_x^{Brock}, \tag{3.66}$$

then the initial series is deterministic (where x is the initial series and $f(x)$ its linear transformation). In the opposite case, the time series that we analyze, will be regarded as stochastic. If the test is applied to *profitabilities of stock exchange indexes* of the American or European financial markets and obviously to the *residuals* resulting from ARMA models, we note two remarks from the tests applied by S.Lardic and V.Mignon to which we will be able to refer on the indexes evoked above:

- The values of the Lapunov exponent are positive.
- The values of the Lyapunov exponent between initial profitabilities and residuals are similar.

In the same way, the results about the correlation dimension show very similar values between initial profitabilities and residuals.

Two contradictory conclusions are underlined. On one hand, the positivity of Lyapunov coefficients would plead in favor of the presence of a (deterministic) chaos, but on the other hand, the similarities of values between initial profitabilities and residuals would lead to conclude to a deterministic structure. It thus appeared necessary to specify another type of test to try to resolve this contradiction. The "random mixture" test deals with this objective.

3.5.2 Scheinkman and LeBaron Random Mixture Test (1989): The Test Weakly Rejects the Hypothesis of Deterministic Chaos and Always Regards Financial Markets as Stochastic Processes

This test proceeds to a random mixture of values of the initial time series. The method is surprising and the idea is that the mixture destroyed the possible deterministic structure of the initial time series. If the "mixed" series loses its structure, the correlation dimension of the mixture will be higher than the initial series; moreover the Lyapunov exponent of the mixed series will be lower than the initial series. On the other hand, if the initial series x is not deterministic but stochastic, the $m(x)$ mixture will not significantly change the values of the correlation dimension and of the exponent. If we have:

$$\lambda_{m(x)} < \lambda_x \quad \text{and} \quad D_{m(x)} > D_x, \tag{3.67}$$

then the initial series is deterministic, in contrast if: $\lambda_{m(x)} \sim \lambda_x$ and $D_{m(x)} \sim D_x$, then the initial series is a stochastic process. The conclusions of this test are interesting, indeed if the comparison of the Lyapunov exponent is obvious since the exponents of the mixed series are not lower than those of the initial series, in contrast the dimensions from the mixed series is higher than the initial series and could lead to the acceptance of the hypothesis of a deterministic structure (and in particular the hypothesis of deterministic chaos).

In spite of this encouraging result, the fact that the differences between the dimensions are weak, coupled with unchanged Lyapunov exponents, lead to conclude to the rejection of the hypothesis of a subjacent determinism, and *to accept the hypothesis of a stochastic process for the financial markets again.*

3.6 Long Memory Processes

Long memory processes are observed in many fields and in particular in Finance. The identification of such processes can be done by the observation of the autocorrelation function of the time series. If there is long memory, the *autocorrelation decreases hyperbolically towards zero,* shown in Fig. 3.4.

Another manner of identifying such processes is to observe their spectra, i.e. their frequencies. If there is long memory, the spectrum exhibits only very rare peaks of frequency, i.e. "monster" peaks. Lastly, these processes can also be detected by their persistent or anti-persistent behaviors.

Fig. 3.4 Hyperbolic shape of the autocorrelation function

The hypothesis of the independence of elements of a time series usually is only an approximation of the true subjacent correlation of the dynamics. Weak delays and strong correlations are sometimes identified by short memory modelings as the processes of the ARMA type. Frequently many series have weak correlations and their periodogram shows a peak at the central frequency into zero. This observation is to be connected with the hyperbolic decrease of the autocorrelation function mentioned above, which can either express a non-stationarity of the series or a more interesting behavior: long term dependence, in other words a long-memory phenomenon. This will lead to define the ARFIMA processes which are representative of this type of phenomena in the following paragraphs.

As evoked above, the long memory processes are defined both in the temporal field and in the spectral field. In the course of time, these processes are characterized by a autocorrelation function decreasing in a hyperbolic way, whereas the processes with short memory decrease exponentially. In the space of frequencies, the processes with long memory are characterized by a spectral concentration that increases when we approach the *zero central frequency*. We will use again the two definitions corresponding to these assertions and relating to *the asymptotic behavior* of long memory processes, first over time and second on the frequencies; by looking at the way in which *the processes converge*. It is commonly said that the more a process is persistent, the more the convergence is slow, the more the sum of autocorrelations is high. For a short memory process, the sum of autocorrelations is low. In short, *we are focused on the hyperbolic and geometrical speed of convergence towards zero.*

Definition 3.5 (Long memory process defined by the autocorrelation). A stationary process x_t is a long memory process if $\exists$ a real $\omega : 0 < \omega < 1$ and if $\exists$ a positive constant c_r, such that

$$\lim_{k \to \infty} \frac{\rho_k}{c_r \, k^{-\omega}} = 1, \tag{3.68}$$

k is a delay and ρ is the autocorrelation function. Thus, asymptotically for $k \to \infty$ we verify that $\rho_k \approx c_r \, k^{-\omega}$.

Definition 3.6 (Long memory process defined by the spectral density). A stationary process x_t is a long memory process if $\exists$ a real $\theta : 0 < \theta < 1$ and if $\exists$ c_f: a positive constant, such that

$$\lim_{\lambda \to 0} \frac{f(\lambda)}{c_f \, |\lambda|^{-\theta}} = 1, \tag{3.69}$$

λ is the frequency and $f(\lambda)$ is the spectral density for this frequency. Thus, it is verified that $f(\lambda) \approx c_f \, |\lambda|^{-\theta}$ for $\lambda \to 0$; which justifies the peak of frequency into zero mentioned above.

The ARFIMA processes (autoregressive fractionally integrated moving average processes) correspond to a type of long memory process. They are a development of ARIMA processes, and we will try to quickly describe them in the following section.

3.6.1 ARFIMA Process

The unit root test (used in another part of this work) indicates that if there is indeed a unit root, then the time series is non-stationary. In such a case, these time series can be modelled by processes of $ARIMA(p,d,q)$ type, where the unit root highlights a long-memory phenomenon and even infinite ($d = 1$: non-stationary with the difference). If there is not unit root the time series is stationary ($d = 0$). Between $d = 1$ and $d = 0$, the main point is to be interested in the *non-integer fractional differentiation exponents d*. These $ARFIMA(p,d,q)$ processes have been developed by Hosking in 1981 and Granger, Joyeux in 1980.

The generalized $ARFIMA(p,d,q)$ processes can be presented in the following way:

$$\Phi(L)x_t = \Theta(L)\varepsilon_t \qquad (3.70)$$

with $\varepsilon_t = (1-L)^{-d}\upsilon_t,$[8] where υ_t corresponds to a white noise with variance σ^2. Φ and Θ are lag polynomials of degrees p for the first and q for the second. We obtain:

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \qquad (3.71)$$

where

$$\varepsilon_t = \upsilon_t + d\upsilon_{t-1} + \frac{d(d+1)}{2}\upsilon_{t-2} + \frac{d(d+1)(d+2)}{6}\upsilon_{t-3} + \cdots \qquad (3.72)$$

The $ARFIMA(p,d,q)$ processes are stationary long-memory processes and invertible when $d \in \left]-\frac{1}{2},\frac{1}{2}\right[$ and $d \neq 0$.

3.6.1.1 Particular Case of ARFIMA(0,d,0) Processes

The $ARFIMA(0,d,0)$ process is written: $(1-L)^d x_t = \upsilon_t$: where υ_t is a white noise.

If we select the particular case: $ARFIMA(0,d,0)$ with $d > -\frac{1}{2}$, the process can be written as an *AR* process:

$$\sum_{k=0}^{\infty} \beta_k x_{t-k} = \upsilon_t \qquad (3.73)$$

with

$$\beta_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} \qquad (3.74)$$

---

[8] L is a polynomial and $(1-L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 + \frac{d(1-d)(2-d)}{3!}L^3 - \cdots = \sum_{i=0}^{\infty} p_i L^i$.

$$\sum \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)} = \prod_{0 \leq k \leq i} \left(\frac{k-1-d}{k}\right), i = 1,2,\ldots,$$ where Γ is an Eulerian function (or Gamma) of second type. Γ is defined by the integral: $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt$, with x real. (According to the property $\Gamma(n) = (n-1)!$ for n integer strictly higher than 1, we can use it to calculate the factorial of a natural integer n, by : $n! = \Gamma(n+1)$. Example: $\Gamma(0.5) = -0.5! = \sqrt{\pi}$.)

then:

$$\lim_{k\to\infty} \beta_k = \frac{k^{-d-1}}{\Gamma(-d)}. \tag{3.75}$$

If we select the particular case: $ARFIMA(0,d,0)$ with $d < \frac{1}{2}$, the process can be written as a MA process:

$$x_t = \sum_{k=0}^{\infty} \psi_k v_{t-k} \tag{3.76}$$

with

$$\psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \tag{3.77}$$

then:

$$\lim_{k\to\infty} \psi_k = \frac{k^{d-1}}{\Gamma(d)}. \tag{3.78}$$

If we select the particular case: $ARFIMA(0,d,0)$ with $-\frac{1}{2} \leq d \leq \frac{1}{2}$, and v_t a white noise of unit variance, then the autocovariance is written:

$$\gamma_k \sim \frac{\Gamma(1-2d)\Gamma(k+d)}{\Gamma(d)\Gamma(1-d)\Gamma(k+1-d)} \tag{3.79}$$

with

$$\gamma_k \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{2d-1} \quad \text{when } k \to \infty. \tag{3.80}$$

Lastly, let us write the *Spectral Density:*

$$f_d(\lambda) = (2\sin(\lambda/2))^{-2d} \tag{3.81}$$

and

$$\lim_{\lambda\to 0} f_d(\lambda) = \lambda^{-2d}. \tag{3.82}$$

The process is stationary and invertible for $-1/2 < d < 1/2$. The ψ_k and β_k decrease hyperbolically as well as the autocorrelations, i.e. less quickly than for the ARMA processes. We will not present in this present section estimations from the ARFIMA processes, the literature about the subject is numerous today.

3.6.1.2 R/S Analysis: Hurst Exponent and Memory of a Series

The Hurst statistic, called R/S analysis, already evoked in different other sections of this work, allows to identify the long memory phenomena. This statistic proposes a coefficient that is the Hurst exponent which characterizes a time series according

to the nature of its "memory". Statistics R/S are written as a ratio of a length R to a standard deviation S. If we consider a time series x_t, $t = 1, \ldots, T$ with average $\bar{x}_T$, the statistic R/S is written:

$$R/S = \frac{\left[\max_{1 \le k \le T} \sum_{j=1}^{k}(x_j - \bar{x}_T) - \min_{1 \le k \le T} \sum_{j=1}^{k}(x_j - \bar{x}_T) \right]}{\left[\frac{1}{T} \sum_{j=1}^{T}(x_j - \bar{x}_T)^2 \right]^{1/2}}. \tag{3.83}$$

Asymptotically this statistic is proportional to T^H, where H is the Hurst exponent and is included between 0 and 1:

$$H \sim \frac{\log(R/S)}{\log T}. \tag{3.84}$$

H makes possible an ordering of the time series according to the level of dependence of the variable; it is said that H is *a measurement of the dependence level*.

Fractional Parameter of Integration d and Hurst Exponent $H : d = \phi(H)$

We will give again a definition of a *standard Brownian motion $B(t)$*, also called *Wiener process $W(t)$*, then we will develop the expression of a *fractional Brownian motion*.

Definition 3.7 (Brownian motion). A Brownian motion $B(t)$ is a stochastic process in continuous time such that: $B(0) = 0$ and such that: $[B(t) - B(t-1)]$ are stationary and independent.[9] Consequently, for $0 \le t_1 \le \cdots \le t_n$, $[B(t_i) - B(t_{i-1})]$ are independent random variables such that:

$$E[B(t_i) - B(t_{i-1})] = 0, \tag{3.85}$$
$$var[B(t_i) - B(t_{i-1})] = \sigma^2(t_i - t_{i-1}). \tag{3.86}$$

$[B(t_i) - B(t_{i-1})]$ have a normal distribution of null average and variance $\sigma^2(t_i - t_{i-1})$.

If $\sigma^2 = 1$, then the Brownian motion is called *standard*. The expression of a fractional Brownian motion $B_H(t)$ (where H is the Hurst exponent), is written in the following way:

$$B_H(t) = \frac{1}{\Gamma(H + 1/2)} \left[\int_0^t (t-s)^{H - \frac{1}{2}} dB_H^1(t) \right]. \tag{3.87}$$

Γ: Eulerian function (or Gamma function[10]) of the second type. H: Hurst exponent: $0 \le H \le 1$. $B_H^1(t)$: standard Brownian motion of variance equal to 1. The

---

[9] The increments of $B(t) : B(t) - B(t-1)$.

[10] $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, with x real. Such Gamma or Eulerian functions are shown in Fig. 3.5.

Fig. 3.5 Gamma function
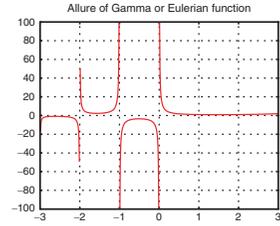
differences $[B_H(t) - B_H(t-1)]$ of the fractional Brownian motion, constitute a *stationary Gaussian process* (or *fractional Gaussian noise*) whose autocovariance is written:

$$\gamma_k = \frac{1}{2}[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}] \tag{3.88}$$

with

$$\gamma_k \sim H \cdot (2H-1) \cdot k^{2H-2} \quad \text{when } k \to \infty. \tag{3.89}$$

Although one of the formulas is at continuous time and the other is discrete, if we bring closer, with their limits, the autocovariance of generalized ARFIMA processes and the autocovariance of the fractional Gaussian process, it is possible to establish a functional relation between the Hurst exponent H and the fractional integration parameter d, which are both characterized by a hyperbolic decrease. The relation between d and H is read in the following way:

$$2H - 2 = 2d - 1 \tag{3.90}$$

thus, it is obtained:

$$d = H - \frac{1}{2}. \tag{3.91}$$

1. If $d = 0$, the ARFIMA series becomes an ARMA and does not contain long term dependence.
2. If $0 < d < 1/2$, the ARFIMA process is stationary and with long memory. We progressively find for these values the well-known hyperbolic decrease towards zero of autocorrelations during the increase of lags. And we find again the peaks of frequencies near frequency zero. The long memory of this process is expressed as an effect of the persistence of this memory.
3. If $-1/2 < d < 0$, in such a case there is an alternation of the sign of correlations and the process is anti-persistent. As an example an illustration of this last case will be given. Figure 3.6 shows for the Hurst exponent equal to $0.1, (d = -0.4)$, the trajectory of the motion, the histogram, the autocorrelation, the partial autocorrelation, the spectrum and incidentally the Poincaré section $[t, t+1]$. (In order to do a comparison, we present the same plots for the Logistic equation in chaotic zone.)
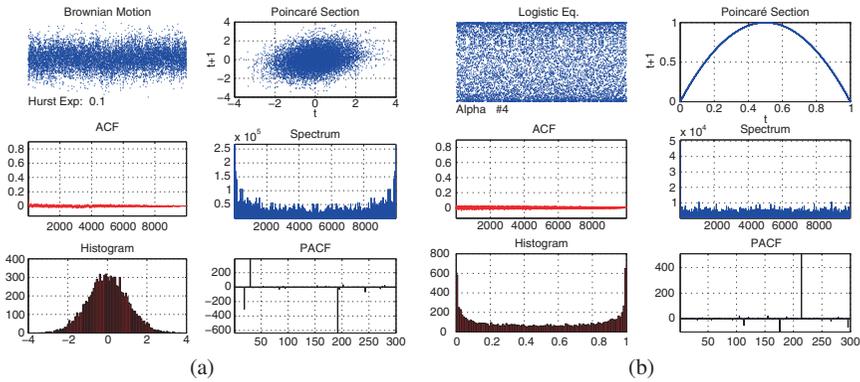
Fig. 3.6 (a) Brownian motion: $H = 0.1$. (b) Logistic equation $\alpha \simeq 4$

3.7 Processes Developed from ARFIMA Process

3.7.1 GARMA Processes: To Integrate the Persistent Periodic Behaviors of Long Memory

The subject is to *integrate* the *periodic or cyclic persistent behavior* of a process and its long memory; Hosking, Gray, Zhang, Woodward and Porter-Hudak contributed to elaborate these methods. To involve the spectrum in the analysis in order to model the cyclic behaviors is one of the methods. Let us pose a Gegenbauer[11] process with a real parameter $\lambda \neq 0$, with ε_t a white noise

$$(1 - 2uL + L^2)^\lambda x_t = \varepsilon_t, \tag{3.92}$$

where $(1 - 2uL + L^2)^\lambda$ is a filter. Then, by introducing the $\Theta(L)$ terms of the moving average and $\Phi(L)$ the autoregression, we obtain a $GARMA(p, u, \lambda, q)$ process, that is written:

$$\Phi(L)(1 - 2uL + L^2)^\lambda x_t = \Theta(L)\varepsilon_t \tag{3.93}$$

(where $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\Theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$). This is a long memory process and it is stationary if $|u| < 1$ and $0 < \lambda < 1/2$ or if $|u| = 1$ and $0 < \lambda < 1/4$. Let us observe the behavior of the autocorrelation functions of u and λ:

1. If $|u| = 1$ and $0 < \lambda < 1/4 : \rho_k \sim k^{4\lambda - 1}$.
2. If $|u| = -1$ and $0 < \lambda < 1/4 : \rho_k \sim (-1)^k k^{4\lambda - 1}$.

---

[11] This process is constructed from the Gegenbauer polynomial:

$$C_n^\lambda(u) = \sum_{k=0}^{Integer[n/2]} (-1)^k \frac{\Gamma(\lambda + n - k)}{\Gamma(\lambda)} (2u)^{n-2k} [k!(n-2k)!]^{-1}, (\Gamma : Gamma).$$

3. If $|u| < 1$ and $0 < \lambda < 1/2 : \rho_k \sim c\, k^{2\lambda-1} \cos(k\omega_0)$, where c is an independent constant and ω_0 is the Gegenbauer frequency.
4. If $|u| = 1$, the process is equivalent to an $ARFIMA(p,d,q)$ with $d = 2\lambda$.

3.7.2 ARCH Processes Towards FIGARCH Processes: To Integrate the Persistence of Shocks in the Volatility of Long Memory Processes

A widespread approach of financial time-series consists in studying the conditional variance to highlight the persistence of shocks, by using an extension of IGARCH processes (integrated GARCH), the FIGARCH processes (Fractionally Integrated GARCH). The GARCH processes are obviously an extension of ARCH processes (Autoregression Conditional Heteroscedasticity); the last processes were conceived *so that the variance depends* on all the information available about a series and in particular concerning *time*. The ARCH processes supplanted the ARMA processes conceived before, the latter were not adapted to *financial series* that have in particular, *asymmetrical structures* and a *strong volatility of their variances*, and that are more and more studied. The ARCH processes integrate in an *endogenous* way the parameter of conditional variance (Lardic and Mignon 2002, Chaps. 7 and 8). Such models were often used for the optimization of financial portfolio choices and equities portfolio. In an ARCH model, the process ε_t has a conditional expectation and variance which are written:

$$E(\varepsilon_t \mid (\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots)) = 0 \ \text{ and } \ V(\varepsilon_t \mid (\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots)) = \sigma_t^2 \quad (3.94)$$

(The conception of ARCH processes, usually understand ε_t as a "residual", therefore, as directly non observable but function of an explanatory variable.) The characteristic of such a process ε_t is that its conditional variance can change over time, unlike the characteristic of ARMA models. The writing of the ARCH process, stated by Engle in 1982, is done from a quadratic form on the conditional variance and coefficients.

Definition 3.8 (ARCH process). ARCH(q) process is defined by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \alpha_0 + \alpha(L)\varepsilon_t^2 \quad (3.95)$$

with $\alpha_0 > 0$, $\alpha_i \geq 0$, for all i.

GARCH process extension results from introduction of a lag in the variance.

Definition 3.9 (GARCH process). A $GARCH(p,q)$ process is defined by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 = \alpha_0 + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2 \quad (3.96)$$

with $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, for all i, j; the positivity of coefficients ensures the positivity of the conditional variance. When $p = 0$,this $GARCH(p,q)$ process is an $ARCH(q)$ process.

There is an essential link between $GARCH$ and $ARMA$ processes by writing:

$$a_t = \varepsilon_t^2 - \sigma_t^2 \Leftrightarrow \sigma_t^2 = \varepsilon_t^2 - a_t \tag{3.97}$$

by replacing σ_t^2 in the GARCH process definition, $\sigma_t^2 = \alpha_0 + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2$:

$$\varepsilon_t^2 - a_t = \alpha_0 + \alpha(L)\varepsilon_t^2 + \beta(L)(\varepsilon_t^2 - a_t) \tag{3.98}$$

thus:

$$(I - \alpha(L) - \beta(L))\varepsilon_t^2 = \alpha_0 + (I - \beta(L))a_t, \tag{3.99}$$

which can be understood as an ARMA process. The writing $\alpha_0 + (I - \beta(L))a_t = (I - \alpha(L) - \beta(L))\varepsilon_t^2$, thus allows to highlight the asymptotic independence of the variance according to the time for this $ARMA(\max(p,q),q)$ process on the series of the *squares of residuals* (or *squares of innovations*). When the lag polynomial $(1 - \alpha(L) - \beta(L))$ has a unit root, the $GARCH(p,q)$ process is an "integrated" process of the type $IGARCH(p,q)$:

$$\Phi(L)(1-L)\varepsilon_t^2 = \alpha_0 + (1 - \beta(L))a_t, \tag{3.100}$$

where $\Phi(L) = (1 - \alpha(L) - \beta(L))(1-L)^{-1}$. The $FIGARCH(p,d,q)$ process, that is between $GARCH(p,q)$ process and $IGARCH(p,q)$ process, is written by replacing the expression $(1 - L)$ by the expression $(1 - L)^d$ that contains the parameter d of fractional intégration. The following definition is given:

Definition 3.10 (FIGARCH process). A $FIGARCH(p,d,q)$ process is written:

$$\Phi(L)(1-L)^d \varepsilon_t^2 = \alpha_0 + (1 - \beta(L))a_t \tag{3.101}$$

- If $d = 1$: $FIGARCH(p,1,q)$ is an integrated $GARCH$.
- If $d = 0$: $FIGARCH(p,0,q)$ is an integrated $GARCH(p,q)$.

Then, by substitution of $a_t = \varepsilon_t^2 - \sigma_t^2$ in this definition, we obtain: $(1 - \beta(L))\sigma_t^2 = \alpha_0 + [1 - \beta(L) - \Phi(L)(1-L)^d]\varepsilon_t^2$. And since the subject of this type of process is to study the behavior of the variance, we extract this variance: $\sigma_t^2 = \alpha_0(1 - \beta(1))^{-1} + [1 - (1 - \beta(L))^{-1}\Phi(L)(1-L)^d]\varepsilon_t^2$. And by posing for the expression in between square brackets on the right $\lambda(L) = [1 - (1 - \beta(L))^{-1}\Phi(L)(1-L)^d] = \lambda_1 L + \lambda_2 L^2 + \lambda_3 L^3 + \cdots$, with λ_k positive or zero, we finally obtain:

$$\sigma_t^2 = \alpha_0(1 - \beta(1))^{-1} + \lambda(L)\varepsilon_t^2. \tag{3.102}$$

The *FIGARCH* processes that integrate the fractional parameter d *are conceived to highlight the persistence of shocks on the conditional variance*;[12] they show that *it decreases hyperbolically* for $0 < d < 1$.

---

[12] The asymmetries of the shocks on the conditional variance are expressed by another type of extension which are the "exponential" processes: FIEGARCH.

3.8 Rejection of the "Random Walk" Hypothesis for Financial Markets: Lo and MacKinlay Test on the Variance of the NYSE (1988)

Lo and MacKinlay have revisited the random walk hypothesis concerning financial markets (Lo and MacKinlay 1988, 1989; Campbell et al. 1997). Prior studies have not been able to reject this hypothesis, and Lo and MacKinlay have introduced more sensitive test statistics able to detect a small but significant abandonment of the random walk. In other words, they developed a more powerful test with a high probability to reject the random walk hypothesis. In particular, they used the work of Merton (1980) that proposed to estimate with more precision the variances when the data are sampled on finer intervals. The suggested Test consists in comparing the variances on different intervals and then by testing them by means of a specification test of the Hausman type (1978). As we evoked in the introduction of this book, the random walk hypothesis has been associated for a long time with the postulate of the efficiency of the financial markets ("Efficient Markets Hypothesis") (Lo and MacKinlay 1999, p. 17).

Closely related to the birth of the probability theory, the random walk hypothesis has a famous history, with remarkable "actors" such as Bachelier, Lévy, Kolmogorov and Wiener. More recently, one of the first serious applications of the random walk hypothesis to financial markets has been the Paul Samuelson contribution in 1965, stated in an article entitled "Proof that Properly Anticipated Prices Fluctuate Randomly". It explains why in an efficient market regarding information, the price changes are unpredictable if they are properly anticipated, i.e. if they fully incorporate the expectations, the information and the forecasts of all the market participants. In 1970, Fama summarizes what precedes in a rather explicit formula: "prices fully reflect all available information". In contrast with many applications of the random walk hypothesis in natural phenomena, for which randomness is supposed almost by default because of the absence of any natural alternative, Samuelson argues that the randomness is achieved through the active participation of many investors who seek the increase their wealth. They try to take advantage from the least information at their disposal, as tiny as it may be. And while doing so, they incorporate their information in the prices of the market and quickly eliminate capital-gain and profit opportunities. If we imagine an "ideal" market without friction and trading cost, then the prices must always reflect all the available information, and no profits can be garnered from trading based on information because the profits have already been captured. Thus, in a contradictory way, we are led to say that, the more efficient the market, the more the time series of prices generated by such a market is random, and the most efficient of all markets is the market in which price changes are completely random and unpredictable. Thus the random-walk hypothesis and the efficient market hypothesis became emblematic in Economics and Finance, although more recently, in 1980, Grossman and Stiglitz considered that the hypothesis of the efficiency of markets is an economically unrealizable idealization. Moreover, some recent works of the last 15 years initiated an

approach aiming to reject this random walk hypothesis.[13] Lo and MacKinlay econo-
metric studies (since 1988) relating to the American stock exchange have rejected
the random walk hypothesis for weekly values of New York Stock Exchange, using
a simple test based on the volatility of courses (Lo and MacKinlay 1988); by speci-
fying however that the rejection of the random walk hypothesis does not necessarily
imply the inefficiency of the generation (or formation) of prices for the markets.

3.8.1 Specification of the Test: Variances of the Increments for Their Ratio and Difference

The general subject of the Test is to verify that if a process follows a random walk,
the variance of increments (or differences) is a linear function of observations (Lo
and MacKinlay 1999).[14] It is said that the variance of the nth-differences is equal to
n times the variance of the first-differences:

$$Var[n\text{th-}differences] = n \cdot Var[first\text{-}differences]. \qquad (3.103)$$

If we name x_t an index of stock exchange prices, and $X_t = \log x_t$ a transformation
by the logarithm of this index, the following relation is posed:

$$X_t = \mu + X_{t-1} + \varepsilon_t \qquad (3.104)$$

μ is an *arbitrary drift parameter* and ε_t is the *random disturbance term*, with
$\forall t$, $E[\varepsilon_t] = 0$. The random walk hypothesis usually requires that the ε_t are "inde-
pendent and identically distributed"[15] Gaussian variables (*iid*), but the economists
agree more and more to say that financial time series generally have variances with
variable volatilities over time (Merton 1980; French et al. 1987). Beyond (*idd*) prop-
erties and gaussianity, this test of the variances is based on the non-autocorrelation
postulate and on the fact that the law is not obligatorily a Normal law and *variances
can change in the course of time*. In short, the postulates are as follows:

- Homoscedasticity: Idd and gaussianity.
- Heteroscedasticity: Non-autocorrelation, the law is not obligatorily a normal law,
  and the variance of the law can vary in the course of time.

3.8.1.1 H_0-Hypothesis of Homoscedasticity of Increments

The null hypothesis H_0 says that the disturbance terms ε_t are independent normal
random variables and identically distributed (*iid*), with variance σ_0^2:

---

[13] Lo and MacKinlay in "Newspaper of Econometrics", volume 40, 1989; volume 45, 1990.
Econometrica, volume 59, 1991.

[14] Example of an increment of order 1 of X_t is written : $[X_t - X_{t-1}]$.

[15] *iid*: independent and identically distributed.

$$H_0 : \varepsilon_t \ iid \ N(0, \sigma_0^2). \tag{3.105}$$

Moreover with the homoscedasticity, the hypothesis of a Gaussian law of increments was posed. Taking the initial subject of the test again, the random-walk hypothesis implies that the variance of differences is linear. In this sense, this supposes that the variance of $X_t - X_{t-2}$ is the double of $X_t - X_{t-1}$. It is on this basic idea that the test is built, which is to compare the variances of $X_t - X_{t-1}$ to that of $X_t - X_{t-2}$. With this intention and by extension of the principle, we select a series of observations $X_0, \ldots, X_{2n}$ of length $2n+1$ ($2n$ in order to handle the time series on multiples of 2, $+1$ to complete the calculation with the differences). Then we build the *Maximum Likelihood Estimators* of first-differences, for the unknown average and variance μ and σ_o^2:

$$\widehat{\mu} = \frac{1}{2n} \sum_{k=1}^{2n} (X_k - X_{k-1}) \tag{3.106}$$

and if we carry out the theoretical sum of the expression above, it obviously follows:

$$\widehat{\mu} = \frac{1}{2n}(X_{2n} - X_0)$$

$$\widehat{\sigma}_a^2 = \frac{1}{2n} \sum_{k=1}^{2n} (X_k - X_{k-1} - \widehat{\mu})^2. \tag{3.107}$$

Then as previously considered, we carry out a variant of the estimator of the variance, i.e. an estimator with the second-differences that corresponds to:

$$\widehat{\sigma}_b^2 = \frac{1}{2n} \sum_{k=1}^{n} (X_{2k} - X_{2k-2} - 2\widehat{\mu})^2. \tag{3.108}$$

If we refer to the standard asymptotic theory, these three estimators are strongly consistent. Moreover, we say that the estimators $\widehat{\sigma}_a^2$ and $\widehat{\sigma}_b^2$ asymptotically follow a Gaussian law. The asymptotic equivalences are written as follows:

$$\sqrt{2n}(\widehat{\sigma}_a^2 - \widehat{\sigma}_0^2) \sim N(0, 2\sigma_o^4), \tag{3.109}$$

$$\sqrt{2n}(\widehat{\sigma}_b^2 - \widehat{\sigma}_0^2) \sim N(0, 4\sigma_o^4). \tag{3.110}$$

The subject here is the limit-distribution of the difference of variances. If we refer to Hausman 1978, who states that any estimator asymptotically efficient $\widehat{\theta}_{efficient}$ of the parameter θ can have the property to be asymptotically non-correlated with the difference: $[\widehat{\theta}_{other} - \widehat{\theta}_{efficient}]$ (*where* $\widehat{\theta}_{other}$ is simply another estimator of θ). Otherwise, i.e. if there is a correlation, we can find a *linear combination* between $\widehat{\theta}_{efficient}$ and $[\widehat{\theta}_{other} - \widehat{\theta}_{efficient}]$, which is *more efficient* than $\widehat{\theta}_{efficient}$, which is obviously contrary to the assertion presenting $\widehat{\theta}_{efficient}$ as an efficient estimator. Thus, in connection with the asymptotic variances $Var_\infty(\cdot)$, we can write:

$$Var_\infty(\widehat{\theta}_{other}) = Var_\infty(\widehat{\theta}_{efficient} + \widehat{\theta}_{other} - \widehat{\theta}_{efficient})$$
$$= Var_\infty(\widehat{\theta}_{efficient}) + Var_\infty(\widehat{\theta}_{other} - \widehat{\theta}_{efficient}), \tag{3.111}$$

$$Var_\infty(\widehat{\theta}_{other} - \widehat{\theta}_{efficient}) = Var_\infty(\widehat{\theta}_{other}) - Var_\infty(\widehat{\theta}_{efficient}). \tag{3.112}$$

At this stage, it is possible to write a test-statistic Ψ_{diff} of the *difference* of the variances of increments:

$$\Psi_{diff} = \widehat{\sigma}_b^2 - \widehat{\sigma}_a^2, \quad \text{with} \quad \sqrt{2n}\Psi_{diff} \sim N(0, 2\sigma_o^4). \tag{3.113}$$

Or it is possible to choose as test-statistic Ψ_{ratio} of the *ratio* of the variances of increments. [By knowing that if we take $(\widehat{\sigma}_a^2)^2$ as an etimator of σ_o^4, then the standard t-test of $\Psi_{diff} = 0$ will give a similar result to $\Psi_{ratio} = \frac{\widehat{\sigma}_b^2}{\widehat{\sigma}_a^2} - 1 = 0$, thus: $\frac{\Psi_{diff}}{\sqrt{2\widehat{\sigma}_a^4}} = \frac{\widehat{\sigma}_b^2 - \widehat{\sigma}_a^2}{\sqrt{2\widehat{\sigma}_a^2}} = \frac{\Psi_{ratio}}{\sqrt{2}} \sim N(0,1)$]. It is written:

$$\Psi_{ratio} = \frac{\widehat{\sigma}_b^2}{\widehat{\sigma}_a^2} - 1, \quad \text{with} \quad \sqrt{2n}\,\Psi_{ratio} \sim N(0,2). \tag{3.114}$$

Although the variance estimator $\widehat{\sigma}_b^2$ is based on the differences of every other observations, it is possible to create a new variance estimator by means of differences of every qth observation (i.e. the qth-differences). Then we obtain a series of observations $X_0, \ldots, X_{nq}$ of length $nq + 1$ (for any integer $q > 1$). The new estimators are written:

$$\widehat{\mu} = \frac{1}{nq} \sum_{k=1}^{nq} (X_k - X_{k-1}) \tag{3.115}$$

and if we carry out the theoretical sum of the expression above, we obviously have: $\widehat{\mu} = \frac{1}{nq}(X_{nq} - X_0).$

$$\widehat{\sigma}_a^2 = \frac{1}{nq} \sum_{k=1}^{nq} (X_k - X_{k-1} - \widehat{\mu})^2, \tag{3.116}$$

$$\widehat{\sigma}_b^2(q) = \frac{1}{nq} \sum_{k=1}^{n} (X_{kq} - X_{qk-q} - q\widehat{\mu})^2, \tag{3.117}$$

thus:

$$\Psi_{diff}(q) = \widehat{\sigma}_b^2(q) - \widehat{\sigma}_a^2 \quad \text{and} \quad \Psi_{ratio}(q) = \frac{\widehat{\sigma}_b^2(q)}{\widehat{\sigma}_a^2} - 1. \tag{3.118}$$

The test can be refined by using the fact that under the null hypothesis the asymptotic distributions are written:

$$\sqrt{nq}\,\Psi_{diff}(q) \sim N(0, 2(q-1)\sigma_o^4), \tag{3.119}$$

$$\sqrt{nq}\,\Psi_{ratio}(q) \sim N(0, 2(q-1)). \tag{3.120}$$

It is possible to create an alternative of the estimator $\widehat{\sigma}_b^2(q)$ of σ_o^2 which is written:

$$\widehat{\sigma}_c^2(q) = \frac{1}{nq^2} \sum_{k=1 \cdot q}^{n \cdot q} (X_k - X_{k-q} - q\widehat{\mu})^2 \tag{3.121}$$

and from this estimator, to recreate the test-statistics for the difference and the ratio of variances:

$$M_{diff}(q) = \widehat{\sigma}_c^2(q) - \widehat{\sigma}_a^2, \tag{3.122}$$

$$M_{ratio}(q) = \frac{\widehat{\sigma}_c^2(q)}{\widehat{\sigma}_a^2} - 1. \tag{3.123}$$

Then we can give the unbiased version of estimators $\widehat{\sigma}_c^2(q)$ and $\widehat{\sigma}_a^2$, that are denoted $\bar{\sigma}_c^2(q)$ and $\bar{\sigma}_a^2$:

$$\bar{\sigma}_a^2 = \frac{1}{nq-1} \sum_{k=1}^{nq} (X_k - X_{k-1} - \widehat{\mu})^2, \tag{3.124}$$

$$\bar{\sigma}_c^2(q) = \frac{1}{q(nq-q+1)\left(1 - \dfrac{q}{nq}\right)} \sum_{k=1 \cdot q}^{n \cdot q} (X_k - X_{k-q} - q\widehat{\mu})^2, \tag{3.125}$$

the statistics were refined $M_{(\cdot)}(q)$:

$$\bar{M}_{diff}(q) = \bar{\sigma}_c^2(q) - \bar{\sigma}_a^2, \tag{3.126}$$

$$\bar{M}_{ratio}(q) = \frac{\bar{\sigma}_c^2(q)}{\bar{\sigma}_a^2} - 1. \tag{3.127}$$

Under the initial hypothesis of homoscedasticity H_0 : asymptotically the statistics $M_{diff}(q)$, $M_{ratio}(q)$, $\bar{M}_{diff}(q)$, $\bar{M}_{ratio}(q)$ behave in the following way:

$$\sqrt{nq}M_{diff}(q) \sim \sqrt{nq}\bar{M}_{diff}(q) \sim N\left(0, \frac{2(2q-1)(q-1)}{3q}\sigma_o^4\right), \tag{3.128}$$

$$\sqrt{nq}M_{ratio}(q) \sim \sqrt{nq}\bar{M}_{ratio}(q) \sim N\left(0, \frac{2(2q-1)(q-1)}{3q}\right). \tag{3.129}$$

They can be normalized:

$$z_1(q) = \sqrt{nq} \cdot \bar{M}_{ratio}(q) \cdot \left[\frac{2(2q-1)(q-1)}{3q}\right]^{-1/2} \sim N(0,1). \tag{3.130}$$

Indeed, under the homoscedasticity of the variance of increments, asymptotically the ratio of the variances $\left[\bar{\sigma}_c^2(q)/\bar{\sigma}_a^2\right] \to 1$, and thus the statistic $\bar{M}_{ratio}(q) = \left[(\bar{\sigma}_c^2(q)/\bar{\sigma}_a^2) - 1\right] \to 0$. Thus, the random walk hypothesis of a time series can be tested by means of the statistics $z_1(q)$ above.

Lo and MacKinlay highlight the fact that for $q = 2$, the statistic $M_{ratio}(q)$ can be written:

$$M_{ratio}(2) = \widehat{\rho}_1 - \frac{1}{4n\widehat{\sigma}_a^2}\left[(X_1 - X_0 - \widehat{\mu})^2 + (X_{2n} - X_{2n-1} - \widehat{\mu})^2\right] \simeq \widehat{\rho}_1, \quad (3.131)$$

i.e. for $q = 2$, the statistic M_{ratio} corresponds to $\widehat{\rho}_1$, which is the estimate of auto-correlation coefficients of order 1 of first-differences of X_t. Thus, they underline by generalizing that there is an equivalence between $M_{ratio}(q)$ and the linear combination of estimators of autocorrelation coefficients of the differences of X_t, which decrease arithmetically for $i = 1, \ldots, q - 1$. Thus, the $M_{ratio}(q)$ statistic are rewritten as a function of the estimates of autocorrelation coefficients and q ($q =$ integer higher than 1) in the following way:

$$M_{ratio}(q) = \frac{2(q-1)}{q}\widehat{\rho}_1 + \frac{2(q-2)}{q}\widehat{\rho}_2 + \cdots + \frac{2}{q}\widehat{\rho}_{q-1}. \quad (3.132)$$

Remark 3.1. The generic writing of autocorrelation coefficients of order i of first-differences is the following:

$$\widehat{\rho}_i = \frac{\displaystyle\sum_{k=i+1}^{nq}(X_k - X_{k-1} - \widehat{\mu})\,(X_{k-i} - X_{k-i-1} - \widehat{\mu})}{\displaystyle\sum_{k=1}^{nq}(X_k - X_{k-1} - \widehat{\mu})^2}.$$

3.8.1.2 H_0-Hypothesis of Heteroscedasticity of Increments

Lo and MacKinlay state that under the hypothesis of heteroscedasticity, firstly, there is non-correlation and secondly, the law is not obligatorily a normal law and its variance can change in the course of time.[16] The $\bar{M}_{ratio}(q)$ statistic still tends towards zero under the present hypothesis, and it is only necessary to calculate its asymptotic variance $\theta(q)$ to reach the required result. We highlight the following expression of $\bar{M}_{ratio}(q)$ obtained asymptotically:

$$\bar{M}_{ratio}(q) = \frac{2}{q}\sum_{i=1}^{q-1}(q-i)\widehat{\rho}_i \quad (3.133)$$

under the present hypothesis, the correlation coefficients $\widehat{\rho}_i$ are supposed to be asymptotically non-correlated. If we can obtain the asymptotic variances δ_i of each $\bar{\rho}_i$ under the hypothesis H_0, it is possible to calculate the asymptotic variances $\theta(q)$ and calculate $\bar{M}_{ratio}(q)$ as a weighted sum of the δ_i. Thus, if we formalize the subject, and if we note: $\delta_i = Var[\widehat{\rho}_i]$ and $\theta(q) = Var[\bar{M}_{ratio}(q)]$, under the Hypothesis H_0 of the heteroscedasticity, we obtain:

- The test statistics $\Psi_{diff}(q)$, $\Psi_{ratio}(q)$, $M_{diff}(q)$, $M_{ratio}(q)$, $\bar{M}_{diff}(q)$, $\bar{M}_{ratio}(q)$ converge almost surely towards zero.

---

[16] For a discussion about the construction of the Hypothesis, see Lo and MacKinlay 1999, p. 24 and 25.

- A consistent estimator of the heteroscedasticity of the variance δ_i:

$$\widehat{\delta}_i = \frac{nq \sum\limits_{k=i+1}^{nq} (X_k - X_{k-1} - \widehat{\mu})^2 (X_{k-i} - X_{k-i-1} - \widehat{\mu})^2}{\left[\sum\limits_{k=1}^{nq} (X_k - X_{k-1} - \widehat{\mu})\right]^2}. \tag{3.134}$$

- A consistent estimator of the heteroscedasticity of the variance $\theta(q)$:

$$\widehat{\theta}(q) = \sum_{i=1}^{q-1} \left[\frac{2(q-i)}{q}\right]^2 \widehat{\rho}_i. \tag{3.135}$$

In spite of the presence of heteroscedasticities, the following statistic still follows a normal law:

$$z_2(q) = \sqrt{nq} \cdot \frac{\bar{M}_{ratio}(q)}{\left[\widehat{\theta}(q)\right]^{1/2}} \sim N(0,1). \tag{3.136}$$

This statistic $z_2(q)$ is used by Lo and MacKinlay to test the random-walk hypothesis on the weekly values of the NYSE index. The tables of the obtained results are a bit heavy and not presented here, it will be necessary to consult their reference works. However we present a summarized table of the variance-ratio test based on weekly observations for nearly 13 years of the index. Under the null hypothesis, the variance ratios: $1 + M_{ratio}(q)$ is 1 and the statistical test has asymptotically a standard normal distribution. On the table, the asterisk "*" indicates that the variance ratios are significantly different from 1, for a 5% threshold, thus for these values there is rejection of the null hypothesis of the random walk:

Index[17]		Number				
CRSP NYSE-AMEX	nq		$q = 2$	$q = 4$	$q = 8$	$q = 16$
Sept 1962 to Dec 1985	1216					
Variance ratios			1.30	1.64	1.94	2.05
Stat z(q)*			(7.51)*	(8.87)*	(8.48)*	(6.59)*

* Rejection of the null hypothesis of the random walk

As a second example, we carried out the test on data of a French stock exchange index, i.e. 2,847 daily values of the cac40 index, between Jan 1, 1988 and June 30, 1999. The results are as follows, subject to the test conditions (i.e. length and nature of the series, and used algorithm):

---

[17] Detailed tables: Lo and MacKinlay (1999, p. 28, 29, 31, 33).

Index Cac40	$q = 2$	4	8	16	32	512
Jan-1988 to Jun-1999						
Variance ratios	1.0553	1.0575	1.0561	1.0587	1.0526	1.191
Psi*	(2.27)*	(1.255)	(0.78)	(0.56)	(0.36)	(0.37)

The results are contradictory here, the rejection for $q = 2$ is limited and for the other values of q, the null hypothesis is not rejected.

3.9 Estimation of the Fractional Integration Parameter d or the Hurst Exponent H of an ARFIMA(p,d,q) Process

3.9.1 General Information About Long Memory (LRD) Estimations and Self-Similarity

In particular, it is in the fields of telecommunication, network traffic information flow that certain concepts have found matter to be used. *Long memory phenomena*, *self-similarities* and *power-laws* but also, more difficult, fractals have been often observed. The estimation methods of the Hurst exponent are varied enough now, but it is not always easy to know which one of these methods offers the most precision or robustness. Moreover sometimes, the multidisciplinary nature of all the existing analysis methods of time series, due to the diversity of sources, can make it difficult to choose and use these tools. The fundamentals of this work are based on the Hurst exponent and on the "long memory" notion, of which the definition, let us recall it, can be written as follows: *A stationary process x_t has a long memory* or *a long range dependence* (LRD), if $\exists$ a real $\alpha \in [0, 1]$ and if $\exists$ a constant $c_a > 0$ such that:

$$\lim_{k \to \infty} \frac{\rho_k}{c_r k^{-\alpha}} = 1, \tag{3.137}$$

where ρ_k : is the autocorrelation function. And the relation between α et the Hurst exponent H is $H = 1 - \alpha/2$, knowing that the long memory processes appear when $1/2 < H < 1$. The autocorrelations decrease slowly towards zero, whereas on the contrary the short memory phenomena appear when the autocorrelations decrease very quickly towards zero: in such cases they can be ARMA process or Markov chains, for example.

Many methods are used to estimate the Hurst exponent.[18] In the lines which follow we will briefly evoke the most common ones, starting from with those that use the periodogram, the power spectrum, the Fourier transform and finishing with the Abry–Veitch estimator constructed by means of the wavelet transform:

1. *Method of the aggregated absolute value,* where an aggregated vector is defined $x^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{k\,m} x_i, \ k = 1, 2, \ldots, \frac{N}{m}, (x_i$ is a time series) using different

---

[18] Some of them use the concept of embedding space resulting from the Takens theorem.

sizes for *m* "blocks". If there is long memory phenomenon or long range dependence (LRD), in a log–log plane, the graph of the aggregation level (abscissa) and of the absolute value of the first-moment of the aggregated series $x^{(m)}$ (in ordinate) is a *straight line of slope H* − 1.

2. *Method of the aggregated variance:* where the log–log graph of the variance (abscissa) and of the level of aggregation (in ordinate) is a *straight line of slope* $\beta > -1$, with $H = 1 - \frac{\beta}{2}$.

3. *R/S Method,* the *oldest* (presented in Part I) where the log–log graph of the *R/S* statistic and number of points of the aggregated series is a *straight line* with a slope that is an estimate of the Hurst exponent.

4. *Periodogram method.* This technique provides a graph of the logarithm of the spectral density of the time series studied on the logarithm of frequencies, and the slope of this graph gives an estimate of H. The periodogram is written: $I(\lambda) = \frac{1}{2\pi T} \left| \sum_{j=1}^{N} x(j) e^{ij\lambda} \right|^2$, where *x* is the series, *T* is the size of the series, λ is the frequency.

5. *"Whittle" estimator.* This estimator is built on the function of likelihood applied to the periodogram.

6. *Variance of residuals*, where the log–log graphs of the aggregation level and of the average of the variance of the residuals of the time series is a *straight line* with a slope equal to $H/2$.

7. *Abry–Veitch method by the wavelet,* where the wavelets are used to estimate the Hurst exponent, we will explain it in one of the following sections.

Except for the R/S method previously presented in Part I, we will not present all these methods which are now well-known and developed in the specialized literature. However, we will present the Abry–Veitch method that uses the wavelets properties. Also in the sections that follow, we will present the two spectral methods of the Hurst-exponent estimation or of the fractional integration parameter *d*.

3.10 Estimation of the Parameter *d* by the Spectral Methods of an ARFIMA Process

3.10.1 Estimation of *d* Based on the Form of the Spectral Density: Regression Method of the Geweke and Porter-Hudak Estimator (GPH: 1983)

The test studies in particular the behavior of the *spectrum at low frequencies of the series*. It is based on the form of the spectral density that is written:

$$f(\lambda) = \left| 1 - e^{-i\lambda} \right|^{-2d} \cdot f_{\text{ARMA}}(\lambda) \tag{3.138}$$

with $f_{\mathrm{ARMA}(p,q)}(\lambda)$ which is the spectral density of an $ARMA(p,q)$ process and that is written:

$$f_{\mathrm{ARMA}}(\lambda) = \frac{\sigma^2 \left| \Theta(e^{-i\lambda}) \right|^2}{2\pi \left| \Phi(e^{-i\lambda}) \right|^2}. \tag{3.139}$$

By applying the logarithm to the initial expression, we obtain:

$$\log f(\lambda) = \log f_{\mathrm{ARMA}}(0) - d \cdot \log \left| 1 - e^{-i\lambda} \right|^2 + \log \left[\frac{f_{\mathrm{ARMA}}(\lambda)}{f_{\mathrm{ARMA}}(0)} \right]. \tag{3.140}$$

Let us recall the periodogram:

$$I(\lambda) = \frac{1}{2\pi T} \left| \sum_{j=1}^{T} x_j e^{ij\lambda} \right|^2 \tag{3.141}$$

and by introducing the logarithm of the periodogram $\log(I(\lambda_j))$ and $\lambda_j = \frac{2\pi j}{T}$, $j = 1, \ldots, \frac{T}{2}$, (T is the length of the series, i.e. the observations number), we have:

$$\log I(\lambda_j) = \log f_{\mathrm{ARMA}}(0) - d \cdot \log \left| 1 - e^{-i\lambda_j} \right|^2 + \log \left[\frac{f_{\mathrm{ARMA}}(\lambda_j)}{f_{\mathrm{ARMA}}(0)} \right] + \log \left[\frac{I(\lambda_j)}{f(\lambda_j)} \right]. \tag{3.142}$$

The idea of the test is to notice that around the zero value of the frequency λ_j, i.e. around the null-frequency for a sampling of frequencies on $2\pi_j/T$, the term $\log \left[\frac{f_{\mathrm{ARMA}}(\lambda_j)}{f_{\mathrm{ARMA}}(0)} \right]$ tends to zero. *Thus, the initial expression can be written as a linear regression whose coefficients can be the subject of a least squares estimation.* Even if the writing is heavy, it is possible to observe how the estimator of the parameter d emerges: $\widehat{d}$. If we substitute $Y_j = \log I(\lambda_j)$, $a = \log(f_{\mathrm{ARMA}}(0))$, $b = -d$, $Z_j = \log \left| 1 - e^{-i\lambda_j} \right|^2$, $\xi_j = \log \left[\frac{I(\lambda_j)}{f(\lambda_j)} \right]$, with $j = 1, 2, \ldots, m$ (m is equal to ordinate of the periodogram). Then, we write:

$$Y_j = bZ_j + \xi_j + a. \tag{3.143}$$

The estimation of d is written as the ratio of the covariance of Z_j and Y_j to the variance of the Z_j:

$$\widehat{d} = -\frac{\sum_{j=1}^{m} (Z_j - \bar{Z})(Y_j - \bar{Y})}{\sum_{j=1}^{m} (Z_j - \bar{Z})^2}. \tag{3.144}$$

The result of this test expresses that when $T \rightarrow \infty$ and $-1/2 < d < 1/2$, then the law of $\widehat{d}$ is:

$$\widehat{d} \sim N\left(d,\ \pi^2/6 \sum_{j=1}^{m}(Z_j - \overline{Z})^2\right).$$ (3.145)

The objective of the test is also to capture any *fractal structure* in the low frequencies, i.e. the correlation structures, which are neither $I(\lambda) = I(0)$ nor $I(1)$, but $I(d)$. (The choice of the number of ordinates of the interval from which the periodogram is expressed is important, and has in particular been discussed and evaluated by Hurvich 1998.) For example, we carried out the test on data of the French stock exchange index: Cac40 (2,847 daily values, between January 1, 1988 to June 30, 1999). Subject to the test conditions, the results are as follows:

	$\widehat{d}$
Cac40 Index Jan 1988 to June 1999.	
Estimation of d :	0.046721

3.10.2 Estimation of d by the Logarithm of the Power Spectrum: Estimator of Janacek (1982)

The test studies in particular Lardic and Mignon (2002) the logarithm of the power spectrum. We select the dth-differences of a series $x(t)$, noted $z(t)$ which is thus "stationarized" and which has the spectrum $f_z(\lambda)$. The spectrum of $x(t)$ is written:

$$\log f_x(\lambda) = -d \cdot \log[2(1 - \cos \lambda)] + \log f_z(\lambda).$$ (3.146)

We introduce in this expression a judiciously chosen weighting, that we note $w(\lambda) = -\frac{1}{2}\log[2(1 - \cos\lambda)]$, which becomes:

$$\log f_x(\lambda) = 2d \cdot w(\lambda) + \log f_z(\lambda).$$ (3.147)

$w(\lambda)$ is also written $w(\lambda) = \sum_{K=1}^{\infty} \frac{1}{K}\cos K\lambda$, by exploiting the writing of the Fourier series and its coefficient:

$$a_K = \frac{1}{2\pi}\int_0^{\pi}\log(2\pi f_x(\lambda))\cdot\cos(K\lambda)d\lambda.$$ (3.148)

It is written:

$$\frac{1}{\pi}\int_0^{\pi} w(\lambda)\log f_x(\lambda)d\lambda = d\sum_{K=1}^{\infty}\frac{1}{K^2} + \sum_{K=1}^{\infty}\frac{a_K}{2K}$$ (3.149)

and

$$f_z(\lambda) = a_0 + a_1\cos\lambda + a_2\cos 2\lambda + \cdots.$$ (3.150)

The minimal forecasting $log(\sigma^2)$ of the mean quadratic error σ^2 is written:

$$log(\sigma^2) = \frac{1}{\pi}\int_0^{\pi}\log(2\pi)f(\lambda)d\lambda.$$ (3.151)

The "step" of the quadratic error is written $\sigma_K^2 = \sigma^2(1 + b_1^2 + b_2^2 + \cdots + b_K^2)$. The relation between the a_K and the b_K is written:

$$exp\left[\frac{a_1 z + a_2 z^2 + \cdots}{2}\right] = 1 + b_1 z + b_2 z^2 + \cdots. \tag{3.152}$$

There is also a relation chain:

$$b_1 = c_1,$$
$$b_2 = c_2 + c_1^2/2!,$$
$$b_3 = c_3 + c_1 c_2 + c_1^3/3!,$$
$$\vdots$$

with $c_K = a_K/2$. The *estimation of the Fourier coefficients* is carried out through the estimate of the c_K, then we extract the estimator $\hat{d}$ from the equation initially enunciated: $\frac{1}{\pi}\int_0^\pi w(\lambda)\log f_x(\lambda)d\lambda = d\sum_{K=1}^\infty \frac{1}{K^2} + \sum_{K=1}^\infty (a_K/2K)$. The $\hat{c}_K$ is written:

$$\hat{c}_K = \frac{1}{\tau}\sum_{p=1}^{\tau-1}\log I(p)\cos K\lambda_p + \frac{1}{2\tau}[\log I(0) - \mathbf{1}_T[\log I(0)]], \tag{3.153}$$

where $\mathbf{1}_T$ is an indicator function: $\mathbf{1}_T = \begin{cases} 1 : & \text{if } T \text{ is even} \\ 0 : & \text{otherwise} \end{cases}$, T being the number of observations of the series. $I(\lambda)$ is the periodogram of $x(t)$ and τ is the integer resulting from the expression $(T-2)/2$. The estimators $\hat{c}_K$ can be written:

$$\hat{c}_K = (d + c_K)/K + e_K, \tag{3.154}$$

where e_K is a sequence of independent random variables, with $d = 0$ if the time series is stationary. At this stage and due to the *asymptotic normality of the coefficients* $\hat{c}_K$, the left member of the initial equation $\frac{1}{\pi}\int_0^\pi w(\lambda)\log f_x(\lambda)d\lambda = d\sum_{K=1}^\infty \frac{1}{K^2} + \sum_{K=1}^\infty (a_K/2K)$ follows a normal law with the following average

$$d\sum_{K=1}^\infty \frac{1}{K^2} + \sum_{K=1}^\infty \frac{c_K}{K}. \tag{3.155}$$

We write that

$$S = \frac{1}{\pi}\int_0^\pi w(\lambda)\log f_x(\lambda)d\lambda \sim N\left(d\sum_{K=1}^\infty \frac{1}{K^2} + \sum_{K=1}^\infty \frac{c_K}{K}, \cdot\right). \tag{3.156}$$

We extract the estimator of d which is noted $\widehat{d}_{Truncation:M}$:

$$\widehat{d}_{Truncation:M} = \frac{S - \sum\limits_{K=1}^{M} \dfrac{\widehat{c}_K}{K}}{\sum\limits_{K=1}^{M} \dfrac{1}{K^2}}, \tag{3.157}$$

where M is a *truncation parameter* aiming to the approximation of the infinite sum. The estimator follows a normal law:

$$\widehat{d}_{Truncation:\,M} \sim N \left(\frac{d + \sum\limits_{K=1}^{M} \dfrac{c_K}{K}}{\sum\limits_{K=1}^{M} \dfrac{1}{K^2}}, \frac{\Psi'(1)}{T \sum\limits_{K=1}^{M} \dfrac{1}{K^2}} \right), \tag{3.158}$$

where $\Psi(x) = \frac{d}{dx} \log \Gamma_x$, ($\Gamma_x$ is a gamma function) and $\Psi^{(K)}(x) = \frac{d^K}{dx^K} \Psi(x)$. The estimator of d becomes an unbiased estimate if we choose the truncation parameter M rather large, so that the coefficients c_K are negligible.

3.11 Abry–Veitch Estimator (1998) of the Hurst Exponent – Wavelet Analysis of Long Memory Processes: An Effective Approach of Scale Phenomena

This test conceived by Abry and Veitch in 1998,[19] has the same objective than the tests that seek to estimate the Hurst exponent H (or d the fractional integration parameter), but it uses the wavelet transform. One of the difficulties of this test resides in the wield of the dyadic scales 2^j. The wavelet technique is built on a Multi-Resolution Analysis (MRA) which is largely used in signal analysis today. In this test, "the wavelet analysis is used to detect the presence and to locate the long memory phenomena (LRD)". The wavelet transform is of a major interest, indeed during this transformation, the information about time is not lost, whereas in the Fourier transform the information is lost. Thus, the wavelets make it possible to have simultaneously information about time and frequencies. The wavelet transform allows a reading of frequencies over time, which is not the case with the Fourier analysis. The sinusoidal functions of the Fourier theory are replaced by wavelet functions that are written

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi_0 \left(\frac{t - u}{s} \right), \tag{3.159}$$

which are generated by the *translation* and *dilation* of a *mother wavelet* ψ_0, and by a *bandpass function* which *filters* at the same time on the scales of frequencies and on time. The theory of multi-resolution analysis explains why there is no loss of

---

[19] Abry and Veitch (1998a,b); Abry et al. (2000); Veitch et al. 2003; Abry et al. (2002a).

information if we sample the continuous wavelet coefficients a Time-Scale plane on a dyadic grid defined by $(u,s) = (2^j, 2^j k)$, where j and k are integers (relative). This practice makes it possible to pass from a Continuous Wavelet Transform (CWT) to a Discrete Wavelet Transform (DWT) and to obtain the discrete coefficients $d_{j,k}$, also called *detail* coefficients.

"The approach by wavelets for the statistical analysis of scale phenomena[20] such as the long memory", is interesting because "the wavelet functions have also a scale property"; "they constitute a system with optimal co-ordinates to observe such phenomena".[21]

Reminder: Long memory or Long Range Dependence (LRD) and self-similarity notions:

- A stationary process x_t is a long memory process if its spectral density $f(\lambda)$ satisfies:

$$f(\lambda) \sim c_f |\lambda|^{-2d}, \quad \text{for } \lambda \to 0, \tag{3.160}$$

  in order to harmonize the notations with the presentation of the Abry–Veitch estimator, we pose $\alpha = 2d$ and $f(\lambda) = \Gamma_x(\lambda)$, then the condition is written:

$$\Gamma_x(\lambda) \sim c_f |\lambda|^{-\alpha}, \quad \text{for } \lambda \to 0, \tag{3.161}$$

  with $0 < \alpha < 1$ and the constant $c_f > 0$, and λ the frequency. This implies that the autocovariance function $r_k = E[(x(t) - E(x))(x(t+k) - E(x))]$ satisfies:

$$r_k \sim c_r k^{\alpha-1}, \quad \text{for } k \to \infty. \tag{3.162}$$

  with moreover the relation $c_r = c_f 2\mathbf{G}(1 - \alpha)\sin(\pi\alpha/2)$, with $\mathbf{G}$ a gamma function.
  It is well known that the relation $d = H - \frac{1}{2}$ links the fractional integration parameter d and the Hurst exponent H, and since we posed $\alpha = 2d$, then $\alpha = 2H - 1$.
- The self-similarity, or more exactly the self-affinity (which is observed for example in the high frequencies of financial series or more recently in Internet traffic) can be presented by means of the Hurst exponent in the following way, for a stochastic process $x(t), t \in R^+$, $\exists H > 0$:

$$x(t) \stackrel{d}{=} a^{-H} x(at) \quad \text{for all tout } a > 0, \tag{3.163}$$

where $\stackrel{d}{=}$ is the equality by the distribution.

The core of this test is built, in particular, around the multi-resolution analysis whose rudiments are presented in the following lines (refer to Mallat 1998 or to the article of Abry and Veitch 1998a).

---

[20] We can evoke the case of the scales of the frequencies λ, that are more or less length, composing the spectrum *highlighting the periodicities*, or the case of the lags k of the autocorrelation function in the construction of the tests of the long memory process.

[21] Ref: Real-time estimation of the parameters of long-range dependence (Roughan et al. 2000).

Presentations of the Multi-Resolution Analysis (MRA) and context of the Abry–Veitch estimator:

At this stage, a presentation of the Multi-Resolution Analysis (MRA) and its corollary, the discrete wavelet transform (DWT), is necessary (Abry and Veitch 1998a).

Definition 3.11 (Multi-Resolution Analysis). A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $\mathbf{L}^2(\mathbb{R})$ is a multiresolution approximation, if the six following properties are satisfied:[22]

$$\forall (j,k) \in \mathbb{Z}^2, f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j. \tag{3.164}$$

$$\forall j \in \mathbb{Z}, \mathbf{V}_{j+1} \subset \mathbf{V}_j, \tag{3.165}$$

$$\forall j \in \mathbb{Z}, f(t) \in \mathbf{V}_j \Leftrightarrow f\left(\frac{t}{2}\right) \in \mathbf{V}_{j+1}, \tag{3.166}$$

$$\underset{j\to+\infty}{Lim}\, \mathbf{V}_j = \bigcap_{j=-\infty}^{+\infty} \mathbf{V}_j = \{0\}, \tag{3.167}$$

$$\underset{j\to-\infty}{Lim}\, \mathbf{V}_j = Closure\left(\bigcup_{j=-\infty}^{+\infty} \mathbf{V}_j \right) = \mathbf{L}^2(\mathbb{R}), \tag{3.168}$$

There exists θ such that $\{\theta(t - n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathbf{V}_0$. (3.169)

The interpretation of these six properties is as follows:

- $\mathbf{V}_{j+1}$ is the image of $\mathbf{V}_j$ by a dilation of a factor 2: There is a subjacent frequential grid in a geometric progression.
- For any j, $\mathbf{V}_{j+1}$ is a subspace of $\mathbf{V}_j$.
- $\mathbf{V}_j$ is invariant by translation of 2^j: There is a subjacent temporal grid by steps of 2^j.
- The intersection of $\mathbf{V}_j$ is reduced to 0 in $\mathbf{L}^2$: with a minimal resolution, one loses all the signal.
- The union of $\mathbf{V}_j$ is dense in $\mathbf{L}^2$: with infinite resolution, one reproduces all the signals perfectly.
- There is a function θ such as the whole translations of θ are a Riesz basis[23] of $\mathbf{V}_0$: Each resolution is generated by a basis of translated atoms of 2^j.

---

[22] $\mathbb{Z}$: denotes the ring of the integers and corresponds to the following space $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, +1, +2, +3, \ldots\}$ (positive and negative integer including 0), which is an extension of the space of the natural numbers $\mathbb{N}$.

[23] *Riesz basis:* A family of vectors noted $\{e_n\}_{n\in\mathbb{N}}$ is a Riesz basis of H the Hilbert space, if this basis is *linearly independent* and if there exists $A > 0$ and $B > 0$, such that for all $f \in \mathbf{H}$, we can find a discrete sequence $\lambda[n]$ with $f = \sum_{n=0}^{+\infty} \lambda[n] e_n$, which satisfies $(1/B)\|f\|^2 \leq \sum_n |\lambda[n]|^2 \leq (1/A)\|f\|^2$. A Riesz basis is a "lightened" concept of the orthonormal basis concept.

Hilbert Space: A Hilbert space is a space on $\mathbb{R}$ or $\mathbb{C}$, provided with a *scalar product* whose associated normed space is complete. The elements of these spaces historically were functions coming from the formalization of oscillatory phenomena and of the calculation of the variations where the required solutions (integral) seem to be the sum of a series of functions, often trigonometrical, which one approaches by orthogonal polynomials for a scalar product.

In the context of the Abry–Veitch estimator, this definition is specified as follows: the multiresolution analysis (MRA) consists in a collection of embedded subspaces $\{V_j\}_{j \in \mathbb{Z}}$ which have the following properties:

- $\bigcap\limits_{j \in \mathbb{Z}} V_j = \{0\}$, $\bigcup\limits_{j \in \mathbb{Z}} V_j$ is *dense* on $\mathbf{L}^2(\mathbb{R})$.[24]
- $V_j \subset V_{j-1}$.
- $x(t) \in V_j \Leftrightarrow x(2^j t) \in V_0$.
- There is a function $\phi_0(t)$ in V_0, called the *scaling function*, such that the set $\{\phi_0(t-k)\}_{k \in \mathbf{Z}}$ is a *Riesz basis* for V_0.

In a same way, the dilated and translated functions:[25]

$$\{\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k)\}_{k \in \mathbf{Z}} \tag{3.170}$$

also constitute a Riesz basis for V_0. *It is fundamental to understand that to carry out a multiresolution analysis of a time series $x(t)$, this is successively to project this time series on each subspace of approximation V_j:*

$$Approx_j(t) = \mathrm{Proj}_{V_j} x(t) = \sum_k a_{j,k}\phi_{j,k}(t). \tag{3.171}$$

As presented above $V_j \subset V_{j-1}$, consequently it is known that the approximation: $approx_j(t)$ of the series $x(t)$ is an approximation coarser than $approx_{j-1}(t)$. Thus, another fundamental idea of the multiresolution analysis (MRA) resides in the analysis of the information loss that is indicated by the term: "*detail$_j$*"; this term thus expresses the loss of information related to a more and more rough approximation, as follows:

$$Detail_j(t) = approx_j(t) - approx_{j-1}(t). \tag{3.172}$$

The multiresolution analysis (MRA) calculates the "*detail$_j$*" by the projections of $x(t)$ on a collection of subspaces, noted W_j, called the *subspaces of wavelets*. Moreover, the multiresolution analysis shows that there is a function ψ_0 called mother wavelet, which as what is made from the scaling function ϕ_0, is used to build the functions:

$$\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\} \tag{3.173}$$

that also constitute a Riesz basis for W_j:

$$Detail_j(t) = \mathrm{Proj}_{W_j} x(t) = \sum_k d_{j,k}\psi_{j,k}(t). \tag{3.174}$$

Banach Space: A complete normed vectorial space (i.e. a normed space in which any *Cauchy sequence* converges) and called *Banach space*.

[24] $\mathbf{L}^2(\mathbf{R})$: Functions of finite energy, such that: $\int |f(t)|^2 \, dt < +\infty$. And the spaces of integrable square functions.

$\mathbf{L}^\mathbf{P}(\mathbf{R})$: Functions such that $\int |f(t)|^p \, dt < +\infty$.

$\mathbf{l}^2(\mathbf{Z})$: Discrete signals of finite energy, such that: $\sum_{n=-\infty}^{+\infty} |f(t)|^2 \, dt < +\infty$.

[25] It is also said modulating and translating or scaling.

The principle of the multiresolution analysis is to transform the information contained in the time series $x(t)$, in a collection, which on one hand gathers an approximation with low resolution J, and on the other hand, the set of "$detail_j$" for different $j = 1,\ldots,J$ resolution $\left(\text{noted} \sum_{j=1}^{J} detail_j\right)$. This is written:

$$x(t) = approx_J + \sum_{j=1}^{J} detail_j(t) = \sum_k a_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_{j,k}\psi_{j,k}(t). \qquad (3.175)$$

To construct $approx_J$ which is a coarse representation with low resolution of the series, it is necessary that the *scaling function* ϕ_0, from which one calculates $\phi_{J,k}$ and $\sum_k a_{J,k}\phi_{J,k}(t)$, are a *low-pass filter*.

About the "$detail_j$" terms, which correspond to the differences $approx_j(t) - approx_{j-1}(t)$, the function ψ_0, which enters in their construction, is *a bandpass filter. The bandpass filter* despite of its source (because historically it belongs to the Fourier analysis) *also corresponds in its form to a wavelet.* The theory of the multiresolution analysis describes a wavelet as a function of integral zero and that satisfies the following condition on its Fourier transform $F_{\psi_0}(\lambda) = \int e^{-2\pi\lambda t}\psi_0(t)dt$:

$$\left|F_{\psi_0}(\lambda)\right| \sim \lambda^N, \quad \lambda \to 0, \qquad (3.176)$$

where λ obviously is the frequency and N a positive integer which is the number of vanishing moments of the wavelet. (*It is interesting to link the convergence above:* $\left|F_{\psi_0}(\lambda)\right| \sim \lambda^N$, $\lambda \to 0$ *with the condition about the spectral distribution in the definition of the long memory processes* $\Gamma_x(\lambda) \sim c_f |\lambda|^{-\alpha}$ *for* $\lambda \to 0$.)

Definition 3.12 (A wavelet). A wavelet ψ is a function that is written

$$\psi : \mathbb{R} \to \mathbb{R} : \int \psi_0(t)dt = 0, \qquad (3.177)$$

where ψ_0 is a mother wavelet. One can select any type of mother wavelet, i.e. any basic form: Morlet, Daubechies, Haar, sombrero, etc.

Vanishing moments of a Wavelet: Vanishing moments are fundamental to measure the local regularity of a signal. The wavelet transform can be interpreted as a multiscale differential operator. If the wavelet has N vanishing moments, it could be possible to show that the wavelet transform can be understood as a multiscale differential of order N (the proof is not presented here). The wavelet ψ_0 has $N > k$ vanishing moments if:

$$\int t^k \psi_0(t)dt = 0,\ 0 \leq k < N, \ \text{ and } \ \int t^N \psi_0(t)dt \neq 0. \qquad (3.178)$$

Scaling function: Within the general framework of signal analysis, a scaling function is interpreted as the impulse response of a *low-pass filter:*[26] If we note $\phi_s(t) = \frac{1}{\sqrt{s}}\phi\left(\frac{t}{s}\right)$ (and with $\overline{\phi}_s(t) = \phi_s^*(-t)$), then the low frequency approximation of a time series $x(t)$ for the scale s is written:

$$Lx_{u,s} = \left\langle x(t), \frac{1}{\sqrt{s}}\phi\left(\frac{t-u}{s}\right)\right\rangle = x \star \phi_s^*(u).$$

Within the framework of the description of the Abry–Veitch estimator, $\phi_s(t)$ becomes $\phi_j(t) = \frac{1}{\sqrt{2^j}}\phi_0\left(\frac{t}{2^j}\right)$, then the low frequency approximation of a time series $x(t)$ at the (dyadic) scale 2^j is written: $Lx_{j,k} = \left\langle x(t), \frac{1}{\sqrt{2^j}}\phi_0\left(\frac{t-k}{2^j}\right)\right\rangle$ and $\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k)$. The family $\{\phi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_j$ for all $j \in \mathbb{Z}$.

Definition of a scaling function: As presented previously, the approximation of a time series $x(t)$ at the resolution 2^j is defined as an orthogonal projection of the time series $x(t)$ over $\mathbf{V}_j$:$\text{Proj}_{V_j}x(t)$. To calculate this projection, we must be provided with a basis of $\mathbf{V}_j$. The theory of the Multiresolution says that it is possible to build an orthonormal basis by means of each space $\mathbf{V}_j$ by *dilation* and *translation* of a simple function ϕ_0 called the *scaling function*. A scaling function can be coarsely defined as an aggregation of wavelets at scales higher than 1.

Approximation of coefficients: The projection of $x(t)$ over $\mathbf{V}_j$ is obtained by a progressive dilation of the scale:

$$\text{Proj}_{V_j}x = \sum_{k=-\infty}^{+\infty} <x, \phi_{j,k}> \phi_{j,k}. \tag{3.179}$$

The *inner products* $a_j[k] = <x, \phi_{j,k}>$ provide a discrete approximation at the scale 2^j, from which we can rewrite them as the result of a convolution: $a_j[k] = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{2^j}}\phi\left(\frac{t-k}{2^j}\right)dt$, or $a_j[k] = \int_{-\infty}^{+\infty} x(t)\ 2^{-j/2}\phi_0(2^{-j}t - k)dt$. The discrete approximation $a_j[k]$ is a low-pass filtering of $x(t)$ sampled on the interval 2^j. We retain as expression of the $a_{J,k}$:

$$a_{J,k} = \int x(t)2^{-j/2}\phi_0(2^{-j}t - k)dt. \tag{3.180}$$

By analogy, we obtain the coefficients $d_{j,k}$ not by means of a scaling function of the low-pass filter type (as previously), but by a bandpass filter, which also corresponds to a wavelet constructed from a mother wavelet ψ_0 (note this correspondence between the bandpass filter resulting from the Fourier analysis and the wavelet). The projection of $x(t)$ on $\mathbf{W}_j$ is obtained by the filter of the wavelet function: $2^{-j/2}\psi_0(2^{-j}t - k)$. The inner products $\left\langle x, \psi_{j,k}\right\rangle$ provide an approximation at the

---

[26] From the gauge of a generic bandpass filter, one makes filters of the low-pass and high-pass types, which filter (by means of their specific forms) the Fourier coefficients, either from the high to the low frequencies, or from the low to the high frequencies.

scale 2^j. The coefficient at the scale 2^{-j} is written:

$$d_{j,k} = \int x(t)2^{-j/2}\psi_0(2^{-j}t - k)dt. \tag{3.181}$$

The index j corresponds to the resolution level of the wavelet analysis, j is also called octave. The octave j is the logarithm in base 2 of the scale 2^j, and k plays the role of "time". n is the length of the chronicle. The number of coefficients available at the octave j is noted $n_j = 2^{-j}n$.

Discrete wavelet transform: The non-redundant discrete transform allows to pass from a Hilbert space $\mathbf{L}^2(\mathbb{R})$ constructed on real numbers to the Hilbert space $\mathbf{l}^2(\mathbb{Z})$ constructed on whole numbers and in particular on integers $\mathbb{Z}$.[27] Provided with a scaling function ϕ_0 and with a mother wavelet ψ_0, the discrete transform produces the set $\{\{a_{J,k}\}_{k\in\mathbb{Z}}, \{d_{j,k}, \ j = 1,\ldots,J\}_{k\in\mathbb{Z}}\}$ from the time series $x(t)$. As previously enunciated, these coefficients are the result of the inner products $<x,\phi_{j,k}>$ and $<x,\psi_{j,k}>$ where $\phi_{j,k}$ and $\psi_{j,k}$ are respectively obtained by dilation and translation of the gauge of a scaling function ϕ_0 and a mother wavelet ψ_0. The algorithm used to produce the discrete wavelet transform is the recursive pyramidal algorithm (RPA).

Principle of the Abry–Veitch estimator: Patrice Abry and Darryl Veitch write that $|d_{j,k}|^2$ measures the quantity of energy of the analyzed signal at the instant $2^j k$ and at the frequency $2^j \lambda_0$, where λ_0 is an arbitrary reference-frequency resulting from the choice of a mother wavelet ψ_0.[28] And their intention is to consider "that a spectral estimator can be calculated from the temporal average of $|d_{j,k}|^2$", i.e. $\frac{1}{n_j}\sum_{k=1}^{n_j}|d_{j,k}|^2$ (or $\mathbf{E}|d_{j,k}|^2$). Abry and Veitch choose to designate the *spectral estimator* by $\widehat{\Gamma}_x$, thus on a given scale and for an arbitrary frequency λ_0, we have:

$$\widehat{\Gamma}_x(2^{-j}\lambda_0) = \frac{1}{n_j}\sum_k |d_{j,k}|^2. \tag{3.182}$$

$\widehat{\Gamma}_x(\lambda)$ is thus:

- A quantity of energy located in the neighborhood of the frequency λ
- A statistical estimator for the spectrum $\Gamma_x(\lambda)$ of $x(t)$

In fact, they show that for a stationary signal, the Expectation of $\widehat{\Gamma}_x$ is written:

$$\mathbf{E}\widehat{\Gamma}_x(2^{-j}\lambda_0) = \int \Gamma_x(\lambda)2^j \left|F_{\psi_0}(2^j\lambda)\right|^2 d\lambda \tag{3.183}$$

---

[27] *The generic writing of the DWT is:* $W f[n,a^j] = \sum_{n=0}^{N-1} f[m]\frac{1}{\sqrt{a^j}}\psi_j^*[m-n]$ *with a wavelet* $\psi_j[n] = \frac{1}{\sqrt{a^j}}\psi_j(\frac{n}{a^j})$.

[28] Abry and Veitch write: "$|d_{j,k}|^2$ measures the quantity of energy of the analyzed signal at the instant $2^j k$ and at the frequency $2^j \lambda_0$, where λ_0 is an arbitrary reference-frequency resulting from the choice of the mother wavelet ψ_0".

as explained previously, where F_{ψ_0} is *the Fourier transform of the wavelet* that analyzes the signal: $F_{\psi_0}(\lambda) = \int e^{-2\pi\lambda t}\psi_0(t)dt : \left|F_{\psi_0}(\lambda)\right| \sim \lambda^N, \lambda \to 0$. And let us recall that a stationary process $x(t)$ is known as a long memory process if its spectral density $\Gamma_x(\lambda) \sim c_f \left|\lambda\right|^{-\alpha}$, for $\lambda \to 0$. Thus, as $\alpha = 2H - 1$, the preceding equation is rewritten:

$$\mathbf{E}\widehat{\Gamma}_x(2^{-j}\lambda_0) = c_f \left|2^{-j}\right|^{1-2H} \int \left|\lambda\right|^{1-2H} \left|F_{\psi_0}(\lambda)\right|^2 d\lambda \qquad (3.184)$$

$$= \Gamma_x(2^{-j}\lambda_0)\left|\lambda_0\right|^{2H-1} \int \left|\lambda\right|^{1-2H} \left|F_{\psi_0}(\lambda)\right|^2 d\lambda. \qquad (3.185)$$

or, in an equivalent way, via α:

$$\mathbf{E}\widehat{\Gamma}_x(2^{-j}\lambda_0) = c_f \left|2^{-j}\right|^{-\alpha} \int \left|\lambda\right|^{-\alpha} \left|F_{\psi_0}(\lambda)\right|^2 d\lambda \qquad (3.186)$$

$$= \Gamma_x(2^{-j}\lambda_0)\left|\lambda_0\right|^{\alpha} \int \left|\lambda\right|^{-\alpha} \left|F_{\psi_0}(\lambda)\right|^2 d\lambda. \qquad (3.187)$$

Moreover, there is an estimator $\widehat{H}$ of the parameter H based on a linear regression of $\log_2 \widehat{\Gamma}_x(2^{-j}\lambda_0)$ at the octaves j, for large j:

$$\log_2 \widehat{\Gamma}_x(2^{-j}\lambda_0) = \log_2 \left(\frac{1}{n_j}\sum_k \left|d_{j,k}\right|^2\right) \approx (2\widehat{H} - 1)j + \widehat{c}, \qquad (3.188)$$

where $\widehat{c}$ is a constant independent of j. $\widehat{c}$ is a constant because the expression which corresponds (to it) converges: $\log_2[c_f \int \left|\lambda\right|^{1-2H} \left|F_{\psi_0}(\lambda)\right|^2 d\lambda]$.

Or, in an equivalent way, there is an estimator $\widehat{\alpha}$ of α, and since $\alpha = 2H - 1$, we write:

$$\log_2 \left(\frac{1}{n_j}\sum_k \left|d_{j,k}\right|^2\right) \approx \widehat{\alpha}j + \widehat{c} \qquad (3.189)$$

or $\log_2 \mathbf{E}\left|d_{j,k}\right|^2 \approx \widehat{\alpha} j + \widehat{c}$. The value n_j is a number of coefficients at the scale 2^j, for practical purposes: $n_j \sim n/2^j$. The estimator $\widehat{\alpha}$ of α is defined as an estimator of the *weighted least squares* of the slope, and in a field of resolution ranging between the octaves: $[j_1, j_2]$.[29] In order to reduce the writings, one poses $y_j = \log_2((1/n_j)\sum_k \left|d_{j,k}\right|^2)$. Moreover, the *weighting S_j* in the calculation of the least squares estimator is taken as being *the inverse of the asymptotic variance of*

---

[29] *Ordinary Least Squares (OLS):* Let be a model $y_t = \alpha x_t + \beta + \varepsilon_t$, the analytic resolution is written: $Min\sum_{t=1}^n \varepsilon_t^2 = Min\sum_{t=1}^n (y_t - \alpha x_t - \beta)^2 = Min\sum_{t=1}^n S^2$. To find the minimum of the function, we derive with respect to α and to β, i.e. : $\frac{dS}{d\beta} = 0$ and $\frac{dS}{d\alpha} = 0$, and while doing the sum on t, we obtain the normal equations: $\sum x_t y_t - \widehat{\alpha}\sum x_t^2 - \widehat{\beta}\sum x_t = 0$ and $\sum y_t - \widehat{\alpha}\sum x_t - n\widehat{\beta} = 0$, which involve: $\widehat{\alpha} = \frac{\sum_{t=1}^n (x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^n (x_t - \overline{x})^2} = \frac{\sum_{t=1}^n x_t y_t - \overline{xy}}{\sum_{t=1}^n x_t^2 - n\overline{x}^2}$ and $\widehat{\beta} = \overline{y} - \widehat{\alpha}\overline{x}$.

$y_j : S_j = [Var(y_j)]^{-1}$, $(Var(y_j) \sim -2/n_j(\log 2)^2$ and $n_j \sim n/2^j : S_j = [Var(y_j)]^{-1} = (n\log^2 2)/2^{j+1})$. Then, the $\widehat{\alpha}$ estimator is written:

$$\widehat{\alpha}_{[j_1,j_2]} = \frac{\displaystyle\sum_{j=j_1}^{j_2} S_j \, j \, y_j - \sum_{j=j_1}^{j_2} S_j \, j \sum_{j=j_1}^{j_2} S_j \, y_j}{\displaystyle\sum_{j=j_1}^{j_2} S_j \sum_{j=j_1}^{j_2} S_j \, j^2 - \left[\sum_{j=j_1}^{j_2} S_j \, j\right]^2} \tag{3.190}$$

and in an equivalent way the expression of $\widehat{H}$ can be provided by a simple substitution of $\widehat{\alpha}$ in the expression $\widehat{H} = \dfrac{1}{2}(\widehat{\alpha}+1)$:

$$\widehat{H}_{[j_1,j_2]} = \frac{1}{2}\left[\frac{\displaystyle\sum_{j=j_1}^{j_2} S_j \, j \, y_j - \sum_{j=j_1}^{j_2} S_j \, j \sum_{j=j_1}^{j_2} S_j \, y_j}{\displaystyle\sum_{j=j_1}^{j_2} S_j \sum_{j=j_1}^{j_2} S_j \, j^2 - \left[\sum_{j=j_1}^{j_2} S_j \, j\right]^2} + 1\right]. \tag{3.191}$$

The Abry and Veitch articles can be consulted to learn more about the *absence of bias* and about the *efficiency* of H. The articles can be also consulted about the "additional trends", the *wavelet estimator* proves to be *robust* with *additional trends*, and on the other hand, this robustness depends on the selected subjacent wavelet. Abry and Veitch used the Daubechies wavelet (D#) at different (#) *vanishing moments*.[30] Let us recall that the wavelet ψ_0 has N vanishing moments if $\int t^k \psi_0(t)dt = 0$, $0 \le k < N$ and $\int t^N \psi_0(t)dt \ne 0$. Moreover, the question of the optimal choice of the number of vanishing moments arises in the construction of the estimator. Elements of convergences appear around a value such that: $N \simeq H + 1$.

A *variant of the $\widehat{\alpha}$ presentation is written:*

$$y_j = \log_2\left(\frac{1}{n_j}\sum_{k=1}^{n_j} |d_{j,k}|^2\right) - g(j), \tag{3.192}$$

where the $g(j)$ are such that $\mathbf{E}y_j = \mathbf{E}\log_2 \mathbf{E}|d_{j,0}|^2$ for a *gaussian process* ($g_j \sim -1/(n_j\log 2)$). The graph on the logarithmic scale represents y_j and j with a 95% confidence interval for the $\mathbf{E}y_j$ ($Var(y_j) \sim -2/n_j(\log 2)^2$). The estimator $\widehat{\alpha}$ is written:

$$\widehat{\alpha} = \sum_{j=j_1}^{j_2} w_j y_j, \quad \text{where } w_j = \frac{S_0 j - S_1}{\sigma_j^2(S_0 S_2 - S_1^2)},$$

where $\sigma_j^2 = var(y_j)$ and $S_k = \sum_{j=j_1}^{j_2} \dfrac{j^k}{\sigma_j^2}$, with $k = 0,1,2$. Remember that N is the number of vanishing moments $N = 3$ and k is an index (playing the role of time) which varies from zero to $N-1 : k = 0,1,\ldots,N-1$.

---

[30] And also Haar wavelets, noting that D1 is also a "Haar".

The semi-parametric estimator $\widehat{\alpha}$ is an unbiased estimator, robust with the non-stationarities and does not depend on the conception of an "a priori" model. The estimator can be tested for example on stochastic processes or fractional Brownian motions, fractional white noise, ARIMA process, ARFIMA process, etc. Indeed, when we possess the Hurst exponent that has been used for to construct ARFIMA processes, we can test the quality of the estimation $\widehat{\alpha}$ on these processes. It is possible to illustrate this point by applying the test to a fractional Gaussian noise with $H = 0.8,\ N = 3,\ n = 2^{15} = 32768 : \widehat{\alpha} = 0.6$.
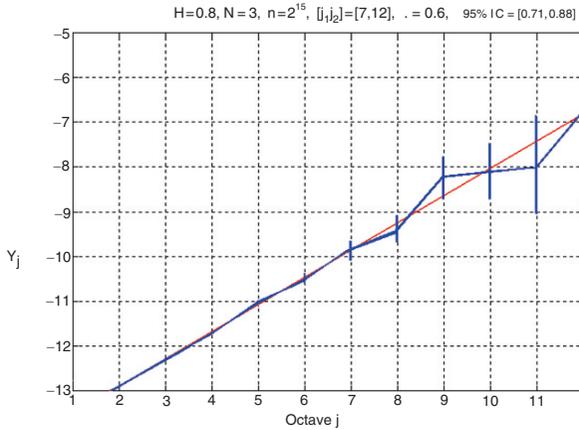
Estimations of H: (Some illustrations). An example of experimentation (Sibbertsen 2002) of the estimator by wavelets is given based on the German stock exchange price, the wavelet used is a Daubechies D4 (D#):

	$\widehat{\alpha}$	$\widehat{H}$
BASF	0.486	0.743
Daimler	0.526	0.763
Deutsche Bank	0.506	0.753
Dresdner Bank	0.498	0.749
Hoechst	0.488	0.744

The conclusion of these estimates is obvious, we are faced with long memory phenomena, and the use of the other methods (aggregate variance, whittle, R/S, etc.) provides results very close to the previous ones. Another illustration of the Abry–Veitch estimator $\widehat{AV}$, compared to the other usual estimators, is carried out on a sample of experimental signals with different H: Fractional Gaussian noises (fGn) and Markov processes (Clegg 2003).

	H	R/S	Var	Whit	AV
fGn	0.625	0.62	0.63	0.61	0.63
fGn	0.75	0.71	0.73	0.74	0.77
fGn	0.875	0.80	0.81	0.86	0.90
Markov C.	0.625	0.64	0.58	0.63	0.69
Markov C.	0.75	0.64	0.70	0.76	0.80
Markov C.	0.875	0.73	0.74	0.84	0.88

Note that the estimator by wavelets provides results which are often at least as efficient as the majority of the other estimators.

Comments about the $d_{j,k}$ coefficients properties:

- For an octave j fixed, the set $\{d_{j,k}\}_{k\in\mathbb{Z}}$ is stationary.
- In the case of a process with self-similarities for j fixed, $d_{j,k} = 2^{j\alpha/2}d_{0,k}$ and in particular $E\left|d_{j,k}\right|^2 = 2^{j\alpha}E\left|d_{0,0}\right|^2$.
- In the case of a process with long memory, when $j \to \infty$: $E\left|d_{j,k}\right|^2 \sim 2^{j\alpha}c_f$ $\int_{-\infty}^{+\infty}|\lambda|^{-\alpha}|\widetilde{\psi}(\lambda)|^2\,d\lambda, (= 2^{j\alpha}c_f C_{\psi,\alpha}$: ref. Abry–Veitch).
- In the case of a process with long memory, for j fixed, when $|k-k'| \to +\infty$, $\left|Ed_{j,k}d_{j,k'}\right| \leq C(j)|k-k'|^{\alpha-1-2N}$, where N is the number of vanishing moment.
- For $N > \alpha/2$, $d_{j,k}$ is stationary and does not show any more a long memory but a short memory.
- For $j \neq j'$ and $k \neq k'$, when $\left|2^j k - 2^{j'}k'\right| \to \infty$, then $Ed_{j,k}d_{j',k'} \approx \left|2^j k - 2^{j'}k'\right|^{\alpha-1-2N}$.

Chapter 4
Statistical and Topological Invariants and Ergodicity

4.1 The Measurement of a Deterministic Chaos Is Invariant in Time

4.1.1 Ergodic Theory and Invariant Measurement Associated with a Dynamics

At the end of the nineteenth century, first, the works of Poincaré and Boltzmann (1885) came first, followed by those of G.D. Birkhoff and J. Von Neumann in 1931. *"The Ergodic hypothesis" is equivalent supposing that the "major part" of the trajectories of a dynamical system is "equally distributed"* (on surfaces of constant energy of the phase space) *and makes it possible asymptotically to "replace the temporal averages by the spatial averages"*.[1] In 1931, the Birkhoff theorem established a rigorous general framework from which the Ergodic theory has been developed with the purpose to study the asymptotic behavior of a dynamical system by means of its invariant measurements (iteration of a transformation, one-parameter flow). The ergodic theory applies to the *deterministic case*, i.e. *dynamical systems* defined by differential equations and coupled with the martingale theory, or applied to the *probabilistic case* of *stochastic processes*, in particular of the Markovian type.

A way of approaching the study of the statistical properties of chaotic processes results from the works of Birkhoff and Von Neumann about "the absolutely continuous invariant distributions" which have a positive Lebesgue measure. The dynamics of aperiodic nature have sometimes distributions of this type, which indicate the frequency with which they take values inside a given interval. The most studied invariant distributions are those which can be represented by a function of density. Today, there are techniques which make it possible to build such functions for chaotic processes. In certain cases, it is possible to associate invariant

---

[1] *Ergodic theory:* The study of measure-preserving transformations.

measurements with a dynamical system, by using the Frobénius–Perron opera-tor, if these measurements are absolutely continuous with respect to the Lebesgue measure. First, let us recall the concepts of invariant measurement and Ergodicity.[2]

Definition 4.1 (Random experiment). A random experiment is represented by a triplet (Ω, a, P) where the following conditions are satisfied:

(1) Ω is the set of possible results of the experiment.
(2) a is a σ-algebra, i.e. a set of parts of Ω (called events), containing the parts ϕ and Ω, stable by complementation and by denumerable union.
(3) P is a set function,[3] i.e. a map from a to $\mathbb{R}^+$, which satisfies $P(\Omega) = 1$ and the condition of σ-additivity: if (A_n) is a sequence of disjoint events and if $\sum_n A_n$ indicates their union: $P(\sum_n A_n) = \sum_n P(A_n)$, the map P is a probability.

Definition 4.2 (Measure). A given couple (Ω, a) can be provided with a mea-sure, if this couple satisfies the first and the second conditions above of a random experiment. Such a measurement μ is defined as a set function, map from a to $\mathbb{R}^+$, such that $\mu : a \longrightarrow \mathbb{R}^+$, satisfying the condition of σ-additivity. (A probability is a measurement of which the total mass is equal to 1: $\mu(\Omega) = 1$.)

Thus, a measure associates a non-negative number with a set and this measure has the additivity and complementarity properties. One can also rewrite the definition of a measure: Consider a given couple (Ω, a), where Ω is the set of possible results, and a is the σ-algebra. If (A_n) is a sequence of disjoint events, a measure μ is a map with an image in $\mathbb{R}^+$, such that:

$$\mu(\phi) = 0, \tag{4.1}$$
$$\forall A_i \in a, A_i \cap_{i \neq j} A_j, \quad \mu(\cup_i A_i) = \sum_i \mu(A_i). \tag{4.2}$$

Given a dynamical system $\varphi(X_t) = X_{t+1}$, where φ is a continuous function, then the probability space is reintroduced (Ω, a, μ), Ω is the set of possible results, where a is the σ-algebra and μ is a probability measure. The measure μ is called *invariant* if we have:

$$\mu[\varphi^{-1}(\delta)] = \mu(\delta), \quad \forall \delta \in a. \tag{4.3}$$

Example. Empirical distribution for three standard series: (1) French stock index daily growth rate (Cac40), (2) a white noise, (3) the logistic equation for $\alpha = 3.99$.

---

[2] *Lebesgue Measure:* The measures on $\mathbb{R}$ which take finite values on any bound interval are characterized by the set of the values $\mu(]a, b])$, $a < b$, $a, b \in \mathbb{R}$. One can also show that there exists a measure noted λ (corresponding to the intuitive representation of the length), such that $\lambda(]a, b]) = b - a$, this is *the Lebesgue measure on* $\mathbb{R}$.

[3] *Set function:* A relation that assigns a value to each member of a collection of sets.
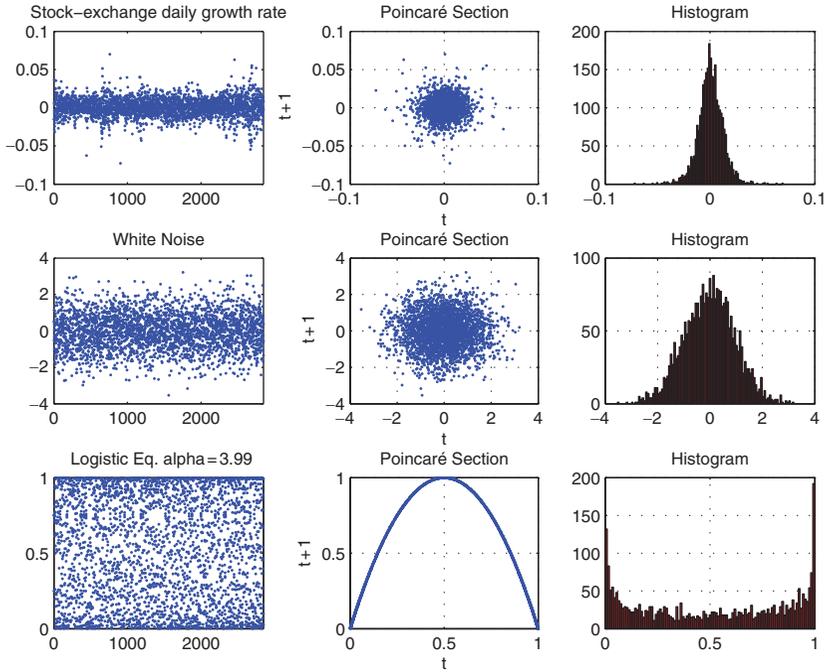
Above, we will distinguish at the center the Gaussian distribution from the white noise, below the "U"-shape of the distribution of the logistic equation in the chaotic field. We also point out the Poincaré sections of each trajectory of these three time series.

4.1.2 The Measure of Probability of a Deterministic Chaotic System Is Invariant in Time

If we study the behavior over time of a deterministic chaotic dynamical system, we note that there is an invariant probability measure on an given time interval. An interesting approach of the measurement concept is to note that the *measure* characterizing a dynamical system *describes with which frequency all the parts of an attractor are visited*. One of the fundamental properties of a dynamical system is the *spatial distribution of points of the (possible) attractor* (Urbach 2000). The spatial distribution is a kind of geometrical relation between the points of the attractor, such as can be the density of points in a small neighborhood of the attractor. We can express this concept of density of points as a probability measure. With this intention, first, we define the indicator function:

$$\mathbf{1}_B(s) = \begin{cases} 1 & \text{if } s \in B, \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Let us suppose that a dynamical system has an attractor $Q \subset M$ and that $s(0)$ is the set of points in the basin of attraction of the attractor, with as trajectory

$$s(t) = \varphi^t[s(0)], \tag{4.5}$$

then for any open set $B \subset Q$, we define μ a natural measure of B by:

$$\mu(B) = \mathop{Lim}_{T \to \infty} \frac{1}{T} \int_0^T \mathbf{1}_B[s(t)]dt. \tag{4.6}$$

If μ exists, then for any continuous function $\phi : M \longrightarrow R$, the quantity:

$$E(\phi) = \int_M \phi(x)\mu(dx) \tag{4.7}$$

is the expectation or the spatial average of ϕ and

$$\bar{\phi} = \mathop{Lim}_{T \to \infty} \frac{1}{T} \int_0^T \phi[s(t)]dt \tag{4.8}$$

is called the temporal average of ϕ, which respects the dynamics on the basin of attraction. The conditions necessary for the existence of both preceding limits $\mu(B) = \lim_{T \to \infty} 1/T \int_0^T \mathbf{1}_B[s(t)]dt$ and $\bar{\phi} = \lim_{T \to \infty} 1/T \int_0^T \phi[s(t)]dt$, are the subject of *the ergodic theory*. This theory says that if a natural measure exists and if this measure is ergodic, then the average in space is equal to the average over time and inversely. (The limit $\bar{\phi} = \lim_{T \to \infty} 1/T \int_0^T \phi[s(t)]dt$ provides an invariant measure of probability $\mu(\phi)$. It is often also said that the average value of ϕ is independent of time.)

Chaotic systems are considered ergodic, but the respect of the conditions of the ergodic theorem is sometimes difficult to satisfy in practice. Fortunately, the majority of physical systems has stable averages over time which suggests the existence of natural ergodic measurements. The temporal average can be written as a probability measure for a neighborhood $B_\varepsilon[x]$ of x:

$$P_\varepsilon(x) = \frac{1}{T} \int_0^T \mathbf{1}_{B_\varepsilon[x]}[s(t)]dt, \tag{4.9}$$

this supposes its ergodicity and leads to the equality of the temporal and spatial averages. The probability measure $P_\varepsilon(x)$ is fundamental for the invariants of dynamical systems, and in particular as regards concepts so important that are the correlation dimension, the Lyapunov exponent and the entropy. A discrete form of this probability can be written: $P_\varepsilon(x) = 1/N \cdot \sum_{k=1}^N \Theta(\varepsilon - \|s(k) - x\|)$, where Θ is the Heaviside function defined by:

$$\Theta(h) = \begin{cases} 1 & \text{if } h > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.10}$$

To write that a set A is invariant with respect to a dynamics $\varphi(\cdot)$ is equivalent to writing that:

$$\varphi(A) = A. \tag{4.11}$$

Intuitively, it is written that a set A is invariant if the trajectories that start from A do not leave A. And write that the natural probability measure μ is ergodic, comes down to write that A is invariant, so:

$$\mu(A) = 0 \quad \text{or } \mu(A) = 1. \tag{4.12}$$

A chaotic attractor is an invariant set, and the *ergodicity implies* that the *trajectories* of a system are almost always *recurrent in the attractor*. It follows that for the natural probability measure μ *almost all the trajectories will cut a dense curve through the attractor* and "the quantity of time that these trajectories use to traverse all the areas of the attractor is proportional to the size of the area". The starting states that would not have a *dense orbit* on the *attractor* belong to a set of measurement zero $\mu(\cdot) = 0$ and include unstable orbits and periodic orbits of the saddle type and also of fixed points. A *dynamical system* is known as *ergodic if its natural probability measurement is ergodic*. A fundamental point will be noted: *a chaotic dynamics implies more than the ergodicity since the periodic movements of a torus are considered as ergodic and not as chaotic*.

4.1.2.1 Frobénius–Perron Operator: Extraction of Functions of Invariant (Limit) Densities

The Frobénius–Perron operator allows to find the density functions representing *the statistical properties of attractors* that result from the iteration of the dynamical function. To simplify, the presentations of the construction method of the operator have frequently avoided several stages, consequently, they are not always very explicit. Thus here we choose a way of presenting the subject that seems more detailed to us and that we hope to be clearer.[4]

Given the map φ on $[0, 1]$ and a set $X_1^o, X_2^o, \ldots, X_N^o$ of N initial states. The map φ transforms these N initial states into N new states:

$$X_1^1 = \varphi(X_1^o), X_2^1 = \varphi(X_2^o), \ldots, X_N^1 = \varphi(X_N^o). \tag{4.13}$$

Given the indicator[5] function $\mathbf{1}_E(X)$ which is defined on E, and the density of the initial states is $f_o(X)$. Consequently, $\forall E_o \subset [0, 1]$,

$$\int_{E_0} f_o(u) \simeq \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{Eo}(X_j^o), \tag{4.14}$$

---

[4] Ref. to Guegan (2003, p. 118): Perron–Frobénius operator.

[5] Which is also written $\mathbf{1}_E$ the indicator function of the set E : $\mathbf{1}_E = \begin{cases} 1 \text{ if } X^1 \in E \\ 0 \text{ otherwise} \end{cases}$.

The integration domain of f_0 is E_o. Moreover, the density of new states is $f_1(X)$. We write: $\forall E_o \subset [0,1]$,

$$\int_{E_0} f_1(u) \simeq \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{Eo}(X_j^1). \tag{4.15}$$

The integration domain of f_1 is E_o. The objective of the approach is to study the density of initial states and the density of new states. The density of the initial states E_o is denoted $f_o(X)$ and the density of new states E is denoted $f_1(X)$. The map φ transforms $X^o \in E_o$ into $X^1 \in E$, i.e. more generally, initial states E_0 by φ are transformed into new states E: $\varphi(E_0) = E$. However, remember that we are interested in the *invariance*[6] *of dynamical system flow solutions* and in the *invariance of densities* for various states of the flow. *We are thus led to study the functional relation that links the densities of the initial states and the densities of new states.* That leads to study the reciprocal image by φ^{-1} of an interval E included in $[0,1]$, i.e. $\varphi^{-1}(E) = \{x, \varphi(x) \in E\}$. We have $X_j^1 \in E$ if $X_j^o \in \varphi^{-1}(E)$ and $\mathbf{1}_E(\varphi(x)) = \mathbf{1}_{\varphi^{-1}(E)}(x)$, then we have:

$$\int_E f_1(u)du \simeq \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{\varphi^{-1}(E)}(X_j^o). \tag{4.16}$$

By definition, we knew that $\varphi(E_0) = E$, then, we are going to *impose* a constraint on φ, E and E_o, by writing that:

$$E_o = \varphi^{-1}(E), \tag{4.17}$$

We intuitively understand its significance, it is equivalent to supposing a kind of bijective form for the map φ or a bijective morphism, of which the reciprocal map φ^{-1} is also a morphism; i.e. a kind of isomorphism. Provided with this constraint, we can write that:

$$\int_E f_1(u)du = \int_{\varphi^{-1}(E)} f_o(u)du. \tag{4.18}$$

The integration domain of f_1 is E and the integration domain of f_0 is $\varphi^{-1}(E)$. If $E = [a,x]$, we have:

$$\int_a^x f_1(u)du = \int_{\varphi^{-1}([a,x])} f_o(u)du. \tag{4.19}$$

This equation is fundamental. The left part expresses a measurement of the density of the dynamics, for the new states, on the domain of integration $E = [a,x] \subset [0,1]$. The right part expresses a measurement of the density of the dynamics,

---

[6] **Definition (An invariant set).** For a flow ϕ_t defined on $U \subset \mathbb{R}^n$, a subset $S \subset U$ is said invariant if $\phi_t(S) \subset S$, $\forall t \in \mathbb{R}$. [In our case: For a flow φ_t defined on $[0,1] \subset R^n$, a subset $E \subset [0,1]$ is said invariant if $\varphi_t(E) \subset E$, $\forall t \in R$].

for the initial states, on an integration domain corresponding to the image by φ^{-1} of $E = [a,x] : \varphi^{-1}([a,x])$. *This equality shows clearly what we are looking for, i.e. an invariant measurement of the density between the initial states and the new states.* This measurement is an invariant measurement associated with the dynamical system. We can also write the equation above in the following way:

$$f_1(x) = \frac{d}{dx} \int_{\varphi^{-1}([a,x])} f_o(u)du \tag{4.20}$$

or in a simplified form:

$$f_1 = Pf_o, \tag{4.21}$$

with:

$$Pf(x) = \frac{d}{dx} \int_{\varphi^{-1}([a,x])} f(u)du. \tag{4.22}$$

P is the Frobénius–Perron operator.

Invariant Density of the Logistic Equation $x_{t+1} = \alpha x_t(1 - x_t)$ with α Close to 4

When a dynamical system admits a density f (with respect to the Lebesgue measure) and an associated measurement μ, the Frobénius Perron operator can be used to determine the aforementioned measurement μ.

The responses of the logistic equation are distributed on an interval $[0, 1]$. When we measure the empirical distribution of the values taken by the variable, we choose a window (or a "subinterval") $[\frac{i-1}{n}, \frac{i}{n}]$ with $i = 1, \ldots, n$. And inside this interval, we count the values that the system takes and which belong to the interval. This *enumeration* on the interval $[\frac{i-1}{n}, \frac{i}{n}]$ is noted f_i. The enumeration gives a "summarized situation" of the "sub-density" inside the interval. The number of values taken by the system inside the interval is noted j. The totality of j is equal to the total number of values taken by the system, i.e. N. The total number of values N is necessarily higher than the total number of intervals noted n, that means: $\Sigma j > \Sigma i$, $(N > n)$. For $i = 1$ the interval is equal to $[0, \frac{1}{n}]$, for $i = 2$ the interval is equal to $[\frac{1}{n}, \frac{2}{n}]$, for $i = 3$ the interval is equal to $[\frac{2}{n}, \frac{3}{n}]$ (where n is the maximum number for i), etc. It is a histogram of responses of the dynamical system, $\varphi(x) : x_{t+1} = \alpha x_t(1 - x_t)$. (For this system with $\alpha = 4$ we know that the output on $[0, 1]$ theoretically behaves as a random sequence distributed according to the arc-sine law having as density $(1/\pi)\sqrt{x(1-x)}$, this is the subject of the following paragraphs). The shape of our empirical distribution, or histogram, is specific. We are not faced with the shape of the normal law. The enumeration is highest near the values 1 and 0, the famous "U" shape of the histogram expresses this observation.

Histogram
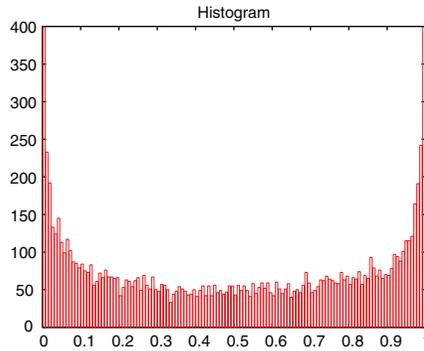
From this empirical distribution, it is possible to build the density of the dynami-cal system, by means of the Frobénius–Perron operator. Remember that the extrac-tion equation of the operator P is written:

$$Pf(x) = \frac{d}{dx} \int_{\varphi^{-1}([a,x])} f(u)du \quad \text{and} \quad f_1 = Pf_0. \tag{4.23}$$

We explain the approach that *consists in associating an invariant measurement with a dynamical system by means of the Frobénius–Perron operator* in the following way:

1. Select the initial state of a dynamical system with its density
2. Calculate by iterations, via the Frobénius–Perron operator, the successive densi-ties until obtaining the limit-density
3. At the end of the iteration process, extract *the invariant density* associated with the dynamical system

The limit density which can be noted f^l, results from the calculation of the limit:

$$Lim_{n \to \infty} P^n f = f^l, \tag{4.24}$$

and respects the equation:

$$Pf^l = f^l. \tag{4.25}$$

Remember that the measure μ (which is not necessarily single), associated with the dynamical system, must have a density f with respect to the Lebesgue measure $[0,x] \subset [0,1]$. Let us observe what occurs with the Frobénius–Perron equation on the reciprocal image (i.e. inverse) of the interval $[0,x]$ belonging to E, that means $\varphi^{-1}([0,x])$.

If we consider the logistic equation $\varphi(x) = 4x(1-x) = -4x^2 + 4x$ as a simple second-degree equation, we can extract the reciprocal function $\varphi^{-1}(x)$. Let us pose $y = 4x(1-x)$, then: $y = 4x - 4x^2 \Leftrightarrow x \pm (x - \frac{y}{4})^{\frac{1}{2}} = 0 \Leftrightarrow x \pm \frac{1}{2}(2x - y)^{\frac{1}{2}} = 0$. And

since the general solution of the logistic equation is $S(\frac{1}{2}, 1)$, the expression above becomes: $\frac{1}{2} \pm \frac{1}{2}(1-y)^{\frac{1}{2}} = 0.$[7] Thus, the domain of $\varphi^{-1}([0,x])$ is as follows:

$$\varphi^{-1}([0,x]) = \left[0, \frac{1}{2} - \frac{1}{2}(1-y)^{\frac{1}{2}}\right] \cup \left[\frac{1}{2} + \frac{1}{2}(1-y)^{\frac{1}{2}}, 1\right]. \tag{4.26}$$

Consequently, the generic equation:

$$Pf(x) = \frac{d}{dx} \int_{\varphi^{-1}([a,x])} f(u)du, \tag{4.27}$$

thus becomes on the interval:

$$Pf(x) = \frac{d}{dx} \int_0^{[\frac{1}{2} - \frac{1}{2}(1-x)^{\frac{1}{2}}]} f(u)du + \frac{d}{dx} \int_{[\frac{1}{2} + \frac{1}{2}(1-x)^{\frac{1}{2}}]}^1 f(u)du. \tag{4.28}$$

Consequently, for our map we obtain:

$$Pf(x) = \frac{1}{(4(1-x))^{\frac{1}{2}}} \left[f\left(\frac{1}{2} - \frac{1}{2}(1-x)^{\frac{1}{2}}\right) + f\left(\frac{1}{2} + \frac{1}{2}(1-x)^{\frac{1}{2}}\right)\right]. \tag{4.29}$$

The steps 2 and 3 lead to the following limit: $Lim_{n \to \infty} P^n f = f^l$ and with $Pf^l = f^l$ towards the invariant density associated with the system:

$$f^l = \frac{1}{\pi(x(1-x))^{\frac{1}{2}}}. \tag{4.30}$$

The function above gives the invariant density associated with the dynamical system, also known as the *theoretical density*. For the dynamical system resulting from our initial logistic function, the Frobénius–Perron equation is written in a generic way on our interval:

$$Pf(x) = [1/(4(1-x))^{\frac{1}{2}}] \left[f\left(\frac{1}{2} - \frac{1}{2}(1-x)^{\frac{1}{2}}\right) + f\left(\frac{1}{2} + \frac{1}{2}(1-x)\right)^{\frac{1}{2}}\right]. \tag{4.31}$$

---

[7] If we are interested in the logistic equation as with a simple equation of the second-degree, we can carry out a general study of the function. Remember that for an equation of the second-degree $ax^2 + bx + c$, $a \neq 0$, the equation behaves in the following way for $a < 0$:

x:	$-\infty$	$-\frac{b}{2a}$	$+\infty$	x:	$-\infty$	$1/2$	$+\infty$
$\varphi'(x)$:	$+$	0	$-$	$\varphi'(x)$:	$+$	0	$-$
$\varphi(x)$:	$-\infty \nearrow$	$\frac{4ac-b^2}{4a}$	$\searrow -\infty$	$\varphi(x)$:	$-\infty \nearrow$	1	$\searrow -\infty$

The solution of a second degree equation with $a < 0$ is $S(-\frac{b}{2a}, \frac{4ac-b^2}{4a})$, and for $\varphi(x) = 4x(1-x)$ the solution is $S(\frac{1}{2}, 1)$.

For an initial elementary density that can be selected equal to:

$$f(x) = \mathbf{1}_{[0,1]}(x) = \begin{cases} 1 & \text{if } x \subset [0,1], \\ 0 & \text{otherwise,} \end{cases} \tag{4.32}$$

the Frobénius–Perron equation becomes:

$$Pf(x) = \frac{1}{2(1-x)^{\frac{1}{2}}}. \tag{4.33}$$

The following iteration:

$$P(Pf(x)) = P^2 f(x) = \frac{1}{(4(1-x))^{\frac{1}{2}}} \left[\frac{1}{2(1 - \frac{1}{2} + \frac{1}{2}(1-x)^{\frac{1}{2}})^{\frac{1}{2}}} \right. \tag{4.34}$$

$$\left. + \frac{1}{2((1 - \frac{1}{2} - \frac{1}{2}(1-x)^{\frac{1}{2}})^{\frac{1}{2}})} \right],$$

it gives:

$$P^2 f(x) = \frac{2^{\frac{1}{2}}}{8(1-x)^{\frac{1}{2}}} \left[\frac{1}{(1 + (1-x)^{\frac{1}{2}})^{\frac{1}{2}}} + \frac{1}{(1 - (1-x)^{\frac{1}{2}})^{\frac{1}{2}}} \right]. \tag{4.35}$$

And so on until the convergence of the density. In the following first graph (Fig. 4.1a), before the convergence of the density, we provide a representation of $Pf(x)$ and $P(Pf(x)) = P^2 f(x)$, which is stopped before the final step of the convergence to show the process in the course of the convergence of the iterations:
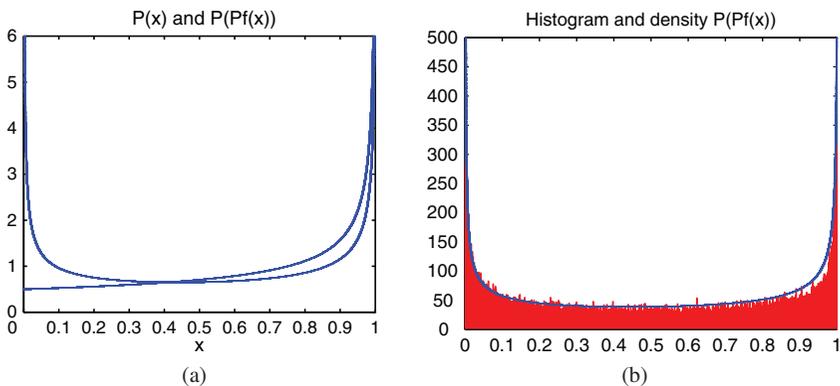
Fig. 4.1 (**a**) First and second iterations. (**b**) Histogram and second iteration
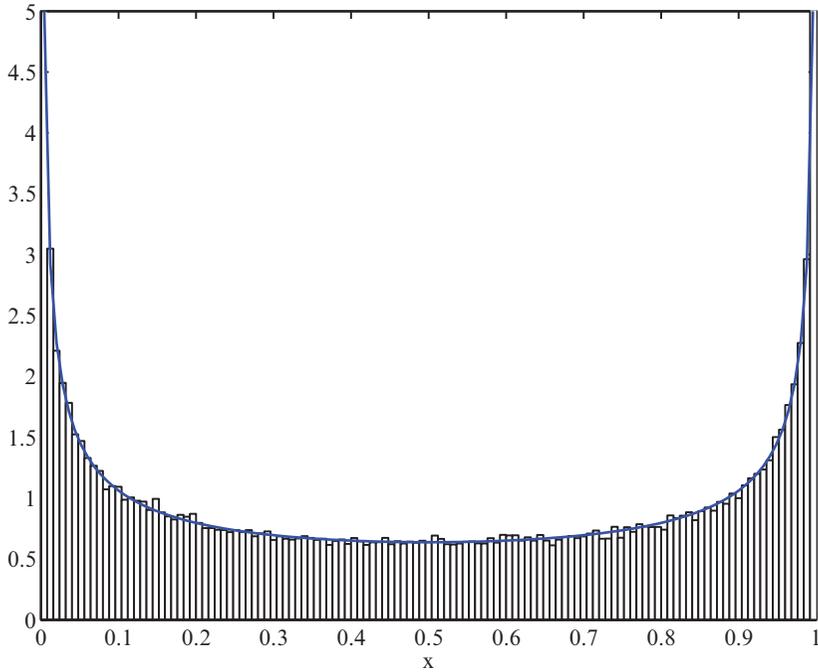
Fig. 4.2 Convergence of the invariant density associated with the dynamical system

In the Fig. 4.1b we observe the histogram and second-iteration $P(Pf(x)) = P^2f(x)$ towards the calculation of the limit invariant density associated with the dynamical system. We observe the incomplete adequacy of $P(Pf(x))$ to the empirical distribution. Indeed, in Fig. 4.1b the iterated density $P(Pf(x))$ is not completely adapted to the histogram and the convergence towards the limit density is not completed. However the convergence is fast, and the continuous iterative process, until the most adapted (limit) invariant density to the empirical distribution, leads to a convergence shown in Fig. 4.2.

It is possible to choose a starting density different from $f(x) = \mathbf{1}_{[0,1]}(x)$, however the convergence towards the (limit) invariant density will be completed in a similar way.

Chapter 5
Spectral and Time–Frequency Analyses and Signal Processing

5.1 Fourier Theory and Wavelets

5.1.1 Contribution of the Fourier Analysis to Regular and Stationary Series: An Approach of Linearities

Even if the Fourier analysis dates back before the Fourier's work and even if the different Fourier analysis developments have been done after him, Fourier is an icon whose influence is fundamental still today. In 1822, Fourier[1] in his work entitled "Analytical Theory of Heat", explained the way in which the *linear equations of partial derivatives could describe the propagation of Heat in a simple form*. In brief, it stated that any periodic function can be expressed as a sum of sinusoids, i.e. sines and cosine of different frequencies: This is the Fourier series. Then, by extension it is said that any periodic curve, even if it is discontinuous, can be *decomposed into a sum of smooth curves*. Consequently, an irregular or jagged curve and the sum of sinusoids are representations of the same thing, but one of them has an empirical nature and the other is the result of an algebraic decomposition. The decomposition method uses the amplitude of sinusoids by assigning to them coefficients and uses the phases. It is important to underline that we can reconstruct the function from the Fourier series (or Fourier transform) without loss of information.

The Fourier transform is an operation that consists in the decomposition of a function according to its frequencies, i.e. we transform a function that depends on time into a function that depends on the frequency. This new function, which depends on frequencies, shows how much sine and cosine of each frequency are contained in the original function. The new function obtained is called the Fourier series.[2] For stock market fluctuations for example, or signals which vary in the

---

[1] Although the first writings date back to 1807.

[2] *Fourier series for a periodic function:* The Fourier series of a periodic function contain only the sinusoids of frequencies equal to multiple integers of the fundamental frequency. When function is not periodic, but when it decreases rather quickly ad infinitum so that the area located under its graph is finite, it is possible to describe it by a superimposition of sine and cosine; but it is however

course of time, the frequency is generally expressed in cycles per second or in hertz. A function and its Fourier transform are thus the two faces of the same information. The first one favors the information about time and masks the information about frequencies, while the second highlights the information about frequencies and hides the information about time. *One of the fundamental stakes of this method is to be able to consider a curve of empirical origin as a function, even if it is irregular.*

The Fourier method was applied in many technical and scientific fields such as geophysics, astronomy, acoustics, medical research and more generally signal analysis.[3] The analyzed signals are often of irregular nature, and the Fourier analysis of these signals translates these curves into an analytical form. This analytical form transforms a chronicle which varies in the course of the time into a function which is the Fourier transform, and more exactly the Fourier series giving the quantity of sine and cosine of each frequency contained in the signal.

At this stage, it is important to underline that the different sinusoids composing a signal, highlighted by the Fourier analysis, can represent the *veritable physical waves* which compose the signal. Indeed, they can be for example an acoustic signal, radio or electromagnetic waves, which are individually identifiable by their respective frequencies. The Fourier method has demonstrated its capacity to analyze, decompose and decode the natural phenomena. In Physics the *space of Fourier transforms* (i.e. "Fourier space") makes it possible to *study the characteristics of an elementary particle* and more exactly *its quantity of motion* (i.e. momentum[4]), which makes it possible to *assimilate a particle to a wave*. Whereas "spontaneously", we tend to characterize a particle by its position in spatial and temporal coordinates. However, one of the main difficulties is that it is not possible to have simultaneously for the particle its precise position and its precise quantity of motion (momentum). This principle, called the Heisenberg uncertainty principle, is a consequence of the Fourier analysis. Besides, it is possible to transpose this principle to any signal analysis, in Finance for example. Indeed, in the Fourier space the results of the Fourier transform on a time series give information about the frequencies which compose it (and not about the quantity of motion). But obviously, the frequencies are measured only on a time segment, i.e. a variable duration or period, which can correspond to all the length of the signal until one of its segments which cannot be reduced to a point. Thus it is not possible to have at the same time a precise frequency and the position of this frequency at a point of the studied time series.

This same observation often leads to use the Fourier transform on segments of the time series.[5] Segments that slide along the signal to identify the various frequencies contained in this segment. Then, usually, this segment (i.e. duration) finds its transposition in the Fourier space in the form of another segment which expresses

---

necessary to calculate the coefficients at all the possible frequencies; this decomposition is called then the Fourier series of the original function.

[3] Used also in the following scientific fields: Seismology, ground analysis, radio-telescope imaging, medical imaging.

[4] *Momentum:* In Physics it is a measure of the motion of a body equal to the product of its mass and velocity. Also called linear momentum. Impetus of a physical object in motion.

[5] This method is called the "Windowed Fourier analysis" proposed by Gabor (1945).

a range of frequencies from the lowest to the highest contained in the considered duration. These segments, respectively in time and frequency, depict rectangles that replace the points, because we cannot have a frequency in a point. The technique is simply called the *short term Fourier transform*. Obviously the risk incurred in this case is that (due to the limitation of the duration through the time segment) the long frequencies are not captured (i.e. depicted) by this type of Fourier decomposition. The graphic illustration of this type of representation in the time–frequency planes is often named: *Heisenberg boxes* (or *Heisenberg rectangles*).

At an intermediate stage during the calculation of the Fourier transform, the sinusoids are provided with Fourier coefficients. The transformation of an arbitrary signal into a sum of sinusoids can create difficulties because we do not know the number of terms contained in the sum. We could even imagine an infinite number of terms. The number of coefficients obtained would be then infinite also and the calculations would become unrealizable. However, more often, a limited number of coefficients is sufficient, and the more the frequencies increase, the more the coefficients tend towards zero. This observation leads to establish the relation with the *sampling theorem*[6] in the information theory, which is a consequence of the Fourier analysis. In substance, if we position in the Fourier space, the theorem is equivalent to saying that if a signal contains only a limited number of frequencies, it is possible to reproduce it with a finite number of points. Generally, this allows the calculation and digitalization.

The Fourier method does not adapt to all the signals or time series, nor to all the difficulties. Indeed, it is advisable to say that *the Fourier analysis is adapted to the resolution of linear problems, i.e. to phenomena where* "*the effect is proportional to the cause*". The writing of a Fourier series is linear, and in front of a problem of non-linear nature, recently yet, we could treat the difficulty as if it were of a linear nature, and thus omitting the true problems related to the nonlinearities, i.e. the existence of transitory or turbulent behavior and also the instability of solutions. Furthermore, it is a problem with which the economists had to face in the past, in particular due to the limited capacities of calculation, which led to reduce a phenomenon which is not linear to a linear model. *In the nonlinear system, the most negligible parametric variations can modify more than proportionally the results*. Thus, forecasts and anticipations become more difficult. As explained before, the *Fourier analysis uses the sines and the cosine which oscillate infinitely*, each one with a fixed frequency. However, the signals which have "changing frequencies are not well adapted to this context of infinite time". Thus, the Fourier space, privileging the frequencies and hiding information about time, can use the same sinusoids to represent very different moments of a time series. That means that any moment of the signal in the Fourier space is confused with any other moment. Even if information about time is not lost, because we can rebuild it by an inverse transformation, there are however

---

[6] **Theorem (Sampling Theorem).** *In order for a band-limited (i.e. one with a zero power spectrum for frequencies $\eta > A$) baseband ($\eta > 0$) signal to be reconstructed fully, it must be sampled at a rate $\eta \geq 2A$. A signal sampled at $\eta = 2A$ is said to be Nyquist sampled, and $\eta = 2A$ is called the Nyquist frequency. No information is lost if a signal is sampled at the Nyquist frequency, and no additional information is gained by sampling faster than this rate.*

indeed a confusion or a dissembling of information about time in the Fourier space. For example, a completely flat signal except for its extremities where there are very strong oscillations, will be represented only by the sum of sinusoids oscillating individually in an infinite manner, which will have completely to be compensated and canceled in phase for the flat part of the signal, but however which should be able to represent the strong variations of the signal localized at the extremities. *This type of difficulty is symptomatic of the maladaptation of the Fourier analysis to the signals which have abrupt (or sharp) variations.* However, these abrupt variations contain crucial information, which if they are deformed or not well restored, damage the comprehension of subjacent phenomena.

We can express the problem in a different way, it is possible to say that *the constraint of the infinite oscillation of each sinusoid* at each frequency tends to *diffuse* a *singularity* of the signal *among all the frequencies of the transform of the signal* so that this singularity is represented (with more or less fidelity). That means that each sinusoid will be able to contain partial information on the aforementioned "singularity", which will be restored "the least inaccurately" possible only by the cumulative set of all the sinusoids. A discontinuity for example will be represented only by the cumulative set of all possible frequencies. However, the short term Fourier transform, mentioned previously, is a technique which makes it possible to avoid partially the pitfall of the diffusion of abrupt variations. But the principle of the signal segmentation in "sliding windows" creates a problem, indeed, this method cannot depict the long frequencies which exceed the size of the segment or the window chosen, and thus also penalizes the restitution of the signal.

5.1.2 Contribution of the Wavelet Analysis to Irregular and Non-Stationary Time Series: An Approach of Nonlinearities

At this stage, the wavelets are intervened.[7] They have the characteristic to decompose a signal, at the same time, into time and frequency, which in the Fourier analysis was not realizable in a simultaneous way. The origin and the history of wavelets are difficult to reconstitute, so much the sources are different, however we positions it in 1930. But it is at the same time in mathematics and physics that the subject found an elaborated structure.[8] The creation of wavelets is attributed to a geophysicist Jean Morlet, within the framework of petroleum prospecting. The former techniques (of the 1960s) used the Fourier analysis of echoes to analyze the soil layers. However, the Fourier analysis tends to diffuse the components of the different layers, the ones in the others, "mixing up" the decomposition of echoes as by interference of layers. These phenomena of *diffusion* and *interference* have been

---

[7] We will note the fundamental contributions of J. Morlet, A. Grossmann and Y. Meyer in the analysis and the comprehension of the wavelets.

[8] The first description in 1971 by K. Wilson, Nobel Prize of physics. Other very different work made it possible to build wavelets called the "self-similar functions of Gabor".

resolved in 1975 by the (previously mentioned) use of a segmented analysis of the signal by temporal sliding-windows which are also sometimes overlapped with the aim of having a finer local definition. Thus, in this technique the window slid along the signal but always preserved the same length. And inside the window, it is the frequency or the number of oscillations which changed.

A new step was initiated by Morlet, which consisted in preserving constant the number of oscillations while varying the size of the (centered) window. In order to increase the size of the window, as by dilation, by keeping the constant number of oscillations, has the consequence to decrease the frequency inside the window. Indeed, the stretching of the window stretches the oscillations. We can say that this is one of the principles which allows to understand the wavelet technique. *The objective is to locate the high frequencies with small windows and the low frequencies with large windows.* One calls a "mother wavelet" the basic shape which is the subject of the dilation or compression. One of the characteristics of mother wavelets is that they are not necessarily composed of a continuous sinusoids at a unique frequency but can be (now) composed of a combination of various frequencies.

At this stage, it is advisable to distinguish the Gabor functions (sometimes called "gaborets" or "gaborettes") also called "Gabor wavelets", which are an intermediate stage between the *short term Fourier analysis* and the wavelets created by Morlet and presented above which have the name of *constant form wavelets.*

It is through the operation of convolution the between the signal and the wavelets, at successive stages of the dilation, that the analysis and decomposition work is carried out. Each *dilation level* (or stretching) is often called *resolution level*. A large window, i.e. a mother wavelet largely stretched (large wavelet), highlights the components of long duration, of low frequency; the term used is the *low resolution*. And a compressed "mother wavelet" (narrow wavelet) highlights the high frequencies, i.e. the components of short duration and transients; the term used is the *high resolution*. This technique is called the *multiresolution*. It was also proved that during the wavelet transform the energy of the signal remains unchanged. The Energy is measured by the average of the square of the amplitude. This means that convolution and the deconvolution, i.e. the return to the initial signal, preserves the energy of the studied time series.

Unlike the *traditional Fourier analysis, which transforms a signal into a function with one variable, which is the frequency, the wavelet transformation produces a transformation through two variables; the time and the frequency.* Thus, the convolution must use "double integrals". However, the approximation methods allowed to avoid the double integration facilitating thus largely the calculations. Furthermore, it was proved that these approximation methods do not generate miscalculation.[9] Like what occurs in the Fourier analysis where the objective of the transformation is to calculate the coefficients assigned to each sinusoid, the wavelet analysis produces also coefficients. These coefficients can be used besides through different manners, such as filters for example. A coefficient corresponds to the calculation of the *integral* of the product of the signal with a wavelet at a given resolution level. As

[9] *Proof: see* J. Morlet et A. Grossmann.

explained above, in order to facilitate the calculations, approximation methods are preferred in order to avoid the double integration. The set of coefficients obtained at each resolution level synthesizes the set of *information* of the studied signal. By convention, each resolution level is twice finer than the previous (and usually five or six different resolution levels are used). When we say that the resolution level is twice finer, that means that the doubling of the resolution reveals the elements of the signal which have frequencies twice higher. On the other hand, it is possible to decompose the time series in intermediate levels of resolution. The term of "resolution" can also be understood as the number of wavelets used to decompose a signal, indeed the more the resolution increases, the more the wavelet number increases.

In the literature about the wavelets, the term "resolution" corresponds to the terms of "scale" and "octave". The term "scale" corresponds to the "correlation" between the size of the wavelet and the size of the elements of the signal that we can identify and highlight.

At this stage, it is essential to redefine that *a wavelet does not have necessarily a precise frequency* unlike a sine or a cosine. Indeed, the phenomenon of dilatation or compression of the mother wavelet, related to the size of the window which is used, modifies its frequency. A second element essential for the comprehension of the structure of a wavelet is that the integral of a wavelet is equal to zero (zero integral). This means that during the convolution between the signal and the wavelet, for example, if the signal is perfectly stationary, the coefficient emanating from the calculation will be also equal to zero. The direct consequence of *this observation is that the wavelets highlight the changes in a signal*. However, it is logical to think that the changes in a signal contain important information interesting to analyze. Obviously, the stationary elements are also interesting, but they are generally apprehended by former techniques belonging to approaches of linear nature: i.e. linear systems or Fourier analysis.

Thus, in the decomposition of a signal, we will be able to understand the need for using the "hybrid" methods (or "mixed methods") which utilize at the same time the properties of the wavelet analysis and those of the Fourier analysis: The first one to highlight the non-stationarities which are potentially related to nonlinearities, and the second one to highlight the stationarities which are potentially related to the linearities.

5.1.3 A Statistical Theory of the Time–Frequency Analysis Remains to Be Developed

The methods of Fourier transforms and the wavelets transforms are merged within the scope of the time–frequency analysis (or time-scale analysis). This analysis framework is largely widespread in many technical fields. Today, we have thus new tools which are particularly interesting for signal analysis specialists. These methods have a common purpose: Capture the best part of a signal, which by direct

reading offers a weak readability. The extraction of its characteristics and intrinsic structures is not necessarily solved and treated in an exhaustive way by the statistical work that we can apply to a signal. The transformation of a sound, or a sonorous echo (or others), medical images for example, exceed the framework of the current statistical analysis. Indeed the statistical analysis does not bring all the answers to the full knowledge of a signal. Thus, there is still a qualitative work to do to assemble the time–frequency analysis and the statistical analysis. This concerns moreover deterministic signals or random signals. Indeed, the signal analysis specialists also work on random or "noized" times series containing sometimes possible non-stationary components. And these subjects can also touch the community of statisticians, which generally neglects the spectral analysis of time series. The use of wavelet bases is increasingly widely used, and many applications, in engineering in particular, tried to utilize the continuous Gabor transforms or the wavelet transforms for statistical purposes. For example to detect, to denoise or to reconstruct a signal, and even more important, to carry out the spectral analysis on non-stationary signals.

The statisticians did not yet really share the benefit of this type of work. The traditional spectral analysis of random or deterministic stationary processes constituted the core of the time–frequency analysis. Today, a very particular attention is focused on the sampling of stationary continuous signals, but also non-stationary, within the framework of time–frequency representations. The main question is to apprehend the importance of time–frequency representations resulting from the Gabor transformations and from the wavelet transforms. Indeed, their potential is large, but still requires an academic validation of the statisticians' community. A remarkable attempt to bring closer both fields has been presented by R. Carmona, W.L. Hwang and B. Torrésani (1998). This work tries to establish the relation between the contribution of this type of analysis and their statistical interest. Moreover, the authors believe in *the capacity of these methods to provide a toolbox with various advantages and talents for the spectral analysis of non-stationary signals*. Unfortunately, *the corresponding statistical theory is not completely developed yet*. The authors revisit the traditional elements of spectral theory of the stationary random processes in the light of the new tools of the time–frequency analysis.

To revisit the traditional elements of the spectral theory implies the evolution of the *Wiener's deterministic spectral theory* towards a deterministic spectral theory of the time series in general, by highlighting for example *the duality and the relation between correlogram and periodogram* (*power spectrum*).[10] This also implies the necessity to develop methods of non-parametric spectral estimation that aim for example to carry out the analysis of non-stationary processes which can be *locally stationary*, or the analysis of *stationary random processes*.

---

[10] See section entitled "Wiener theory and time–frequency analysis" which presents some elements of R. Carmona, W. Hwang and B. Torrésani works.

5.2 A Brief Typology of Information Transformations in Signal Analysis

5.2.1 Fourier, Wavelet and Hybrid Analyses

There exist different types of transformation which are either the expression of technological developments or an adaptation to the needs of different application fields. We can list the various methods:[11]

1. The *Fourier transform*. The analyzing functions are sines and cosines, which oscillate indefinitely. This method is suitable for stationary signals corresponding to constant (or predictable) laws.
2. The *windowed Fourier transform*, i.e. Fourier transform with sliding windows (but of fixed size), also called short term Fourier transform or Gabor transform. The analyzing function is a wave limited in time, multiplied by trigonometric oscillations. That is suitable for quasi-stationary signals, i.e. stationary at the scale of the window.
3. The *wavelet transform* (with variable window). The analyzing functions is a wave limited in time, with a fixed number of oscillations.
4. The *adaptative windowed Fourier transform,* i.e. Fourier transform with adaptive windows, with an *analyzing function of particular forms*, also called *Malvar wavelets*.
5. The *wavelet packets transform*. This method corresponds to a wavelet multiplied by trigonometrical functions. The frequency, position and scale are independent parameters. That is suitable for signals that combine non-stationary and stationary elements (e.g. fingerprint analysis).
6. The Transformation by means of the *Matching Pursuit* algorithm.The analyzing function is a gaussian of variable window size, multiplied by trigonometric functions. The frequency, position of the window and size of the window can change independently. That is suitable for highly non-stationary signals composed of very different elements.

These types of transformation come in a great number of variants.

5.3 The Fourier Transform

5.3.1 Fourier Series and Fourier Transform

An arbitrary periodic function, even if it is discontinuous, can be represented by a sum of sinusoids of different frequencies, each one is provided with a coefficient. The set of these coefficients makes it possible to reconstitute the function or the

---

[11] Refer to Hubbard (1998) for an interesting report of the information transformations methods.

studied initial series. This is the principle of the Fourier series. The Fourier series allow to transform a time series or a function into a series of independent (differential) equations. Each one of these equations shows the chronological evolution of the coefficients of one of the sinusoids which will compose the initial series. Thus, any periodic curve can be decomposed into a sum of *smooth curves*, because they result from sinusoidal functions. Fourier series can be written for example (without cosine):

$$f(x) = c_1 \cdot \sin \varphi_1(x) + c_2 \cdot \sin \varphi_2(x) + c_3 \cdot \sin \varphi_3(x) + \cdots \tag{5.1}$$

or also $f(x) = \sin x + \frac{1}{a} \sin a \cdot x + \frac{1}{b} \sin b \cdot x + \cdots$. The Fourier coefficients of this series are $(1, 1/a, 1/b, \ldots)$. We can illustrate this remark by means of a numerical and graphic example, i.e.: $f(x) = (-0.4 \sin(5x)) + (0.3 \sin(3x)) + (\sin(x)) + (\sin(x/2))$. We observe that this series is written as the sum of sinusoidal functions. The series and the functions are represented in the graph below:
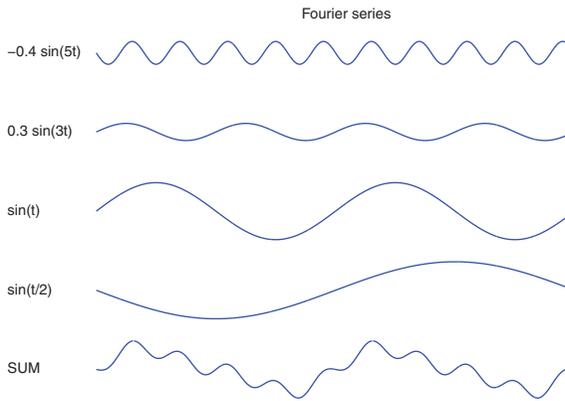
The measurement of Fourier coefficients at a given frequency, is done using the calculation of the integral of the function or series:

The continuous Fourier transform is written:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \tag{5.2}$$

or, $e^{i\omega t}$ are sinusoids.[12] And *the discrete Fourier transform $\hat{f}[k]$* is written:

$$\hat{f}[k] = \sum_{n=0}^{N-1} f[n] \exp\left(\frac{-i2\pi kn}{N}\right) \qquad \text{for} \quad 0 \leqslant k \leqslant N. \tag{5.3}$$

The traditional Fourier transform compares all the signal with infinite sinusoids of various frequencies. On the other hand, the Fourier transform with "sliding window" compares a *segment* of the signal with segments of sinusoids of different frequencies. We can also speak of *local frequencies.* when a segment of the signal was

---

[12] When we are in the *discrete mode*, the notation [.] is used to symbolize the *sequences*.

analyzed, we can slide the window along the signal, and analyze the following part. Gabor introduced the *Fourier atoms* with window, which had vocation to measure the variations of frequencies:

$$g_{u,\xi}(t) = e^{i\xi t}g(t-u). \tag{5.4}$$

The corresponding *continuous transform with window*[13] of f is written:

$$S\hat{f}(\omega) = \langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t)g(t-u)e^{-i\xi t}dt. \tag{5.5}$$

The *discrete transform* with window[14] is written:

$$Sf[m,l] = \langle f, g_{m,l} \rangle = \sum_{n=0}^{N-1} f[n]g[n-m]\exp\left(\frac{-i2\pi kn}{N}\right). \tag{5.6}$$

5.3.2 Interpretation of Fourier Coefficients

5.3.2.1 Frequencies

For the functions, or signals which vary with time (as it is the case with *the stock exchange fluctuations), the frequency is usually expressed in Hertz, or in cycles per second*. Below two sinusoids of different frequencies:

1. The frequency $\sin(2\pi t)$ corresponds to 1 cycle per second (1 Hz) (Fig. 5.1a).
2. The frequency $\sin(2\pi 2t)$ corresponds to 2 cycles per second (2 Hz) (Fig. 5.1b).
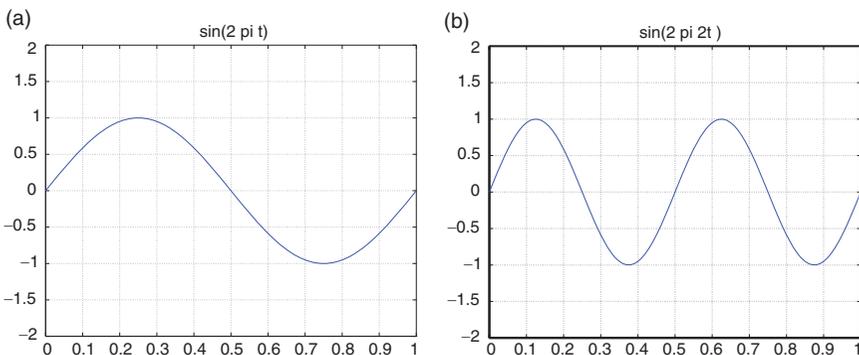
Fig. 5.1 (a) $\sin(2\pi t)$. (b) $\sin(2\pi 2t)$

---

[13] Also called "windowed continuous Fourier transform".

[14] Also called "windowed discrete Fourier transform".

We transform also functions which vary with space. The transform of a finger-print would present maxima in the neighborhood of the "spatial frequency" equal to 15 *striae per* centimeter. The temporal frequency is the inverse of time. The spatial frequency also called *number of waves*, is the inverse of the length. A function and its Fourier transform are two aspects of the same information. *The Fourier transform however reveals information on frequencies and hides information on the temporal evolution*.

5.3.2.2 Series and Coefficients of Fourier

Series

Any periodic function can be written in Fourier series. If the period is 1 (i.e. if $f(t) = f(t+1)$), this series takes the form:

$$f(t) = \frac{1}{2}a_0 + (a_1 \cos 2\pi t + b_1 \sin 2\pi t) + (a_2 \cos 2\pi 2t + b_2 \sin 2\pi 2t) + \cdots. \quad (5.7)$$

Coefficients

The coefficient a_k measures the "quantity" of cosine function $\cos 2\pi kt$ of frequency k contained in the signal; the coefficient b_k measures the "quantity" of sine function $\sin 2\pi kt$ of frequency k contained in the signal. The Fourier series includes only sinusoids of frequencies equal to multiple integers of the fundamental frequency (this fundamental frequency is the inverse of the period). We can write the formula above in a more usual way:

$$f(t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos 2\pi kt + b_k \sin 2\pi kt), \quad (5.8)$$

where k represents the frequency. The formula above makes it possible to recon-struct a function using its Fourier coefficients: we multiply each sinusoid by its coefficient (we modify its amplitude) and we add, point by point, the functions thus obtained; the first term is divided by 2. By convention, usually ξ corresponds to x for the transform of a signal which varies with space. All 2π make heavier the formulas but are inevitable to express the periodicity: The function $(sin\ 2\pi t)$ is of period 1, while $sin(t)$ is of period 2π (Fig. 5.2).

(a)

(b)
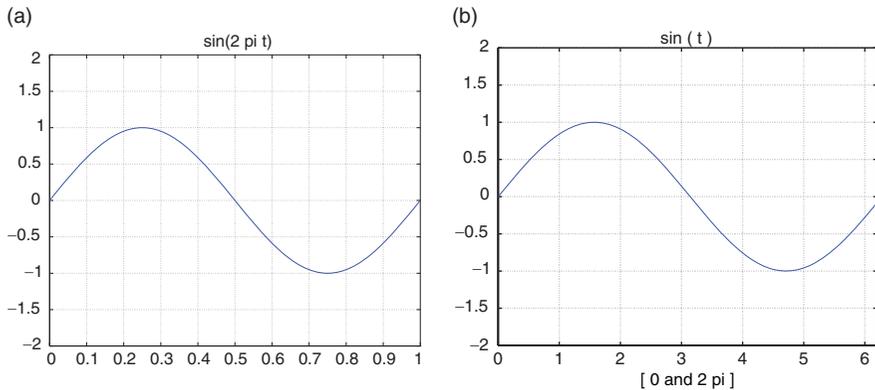
Fig. 5.2 (**a**) $\sin(2\pi t)$. (**b**) $\sin(t)$, between 0 and 2π

We obtain Fourier coefficients of a function $f(t)$ of period 1 using the integrals:

$$a_k = 2 \int_0^1 f(t) \cos 2\pi kt\, dt \quad \text{and} \quad b_k = 2 \int_0^1 f(t) \sin 2\pi kt\, dt \qquad (5.9)$$

that means that we calculate the product of the function f by the sine or cosine of frequency k, we integrate it, and we multiply the result by 2. Integrate a function comes down to measure the area located under a portion of its curve.

5.4 The Gabor Transform: A Stage Between the Short Term Fourier Transform and the Wavelet Transform

5.4.1 The Gabor Function

In the *windowed Fourier analysis* the *size of the window remains fixed*. It is the *number of oscillations inside the window which varies*, at the same time making vary the frequency. In the short term Fourier analysis the window is not fixed, but it moves along the signal to decompose it by successive segments. This technique has inspired the conception of a new technique. It is the *Gabor transform, which is the first step towards the wavelet analysis. It is an intermediate stage between the short term Fourier analysis and the wavelet analysis*.

The method uses a Gabor function (also known as *gaboret* or also *Gabor atom*).[15] The length of the window of the Gabor function remains constant, unlike what occurs with a wavelet. For a wavelet the stretching of the window modifies its length and in certain cases its amplitude also. In fact the length and amplitude of a Gabor function do not change. The only element which changes is the frequency inside the interval of the Gabor function. This principle induces a type of signal analysis by

---

[15] The Gabor function is also called the "gaboret" or "gaborette".
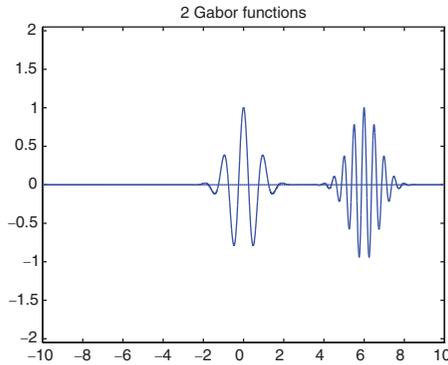
Fig. 5.3 Two Gabor functions: $g_{(0,2\pi)}(x)$ and $g_{(6,4\pi)}(x)$

segment. The Gabor function moves along the signal, it combines with the signal and produces its coefficients which are as many sampled decompositions of the signal by frequency. The displacement of the Gabor function along the signal is done by means of the window and *centering parameter b*. Thus, b indicates the *position in time* of the Gabor function and of the window

$$g_{(b,\omega)}(x) = g(x-b)e^{i\omega(x-b)}. \tag{5.10}$$

Using an example of window such as:

$$g(x-b) = exp^{-(t-b)^2}. \tag{5.11}$$

We can illustrate the subject with a quantified example and Fig. 5.3.

$$g_{(0,2\pi)}(x) = g(x)e^{i2\pi x} = e^{-t^2}e^{i2\pi x} \tag{5.12}$$

$$g_{(6,4\pi)}(x) = g(x-6)e^{i4\pi(x-6)} = e^{-(t-6)^2}e^{i4\pi\,(x-6)}. \tag{5.13}$$

5.4.2 The Gabor Transform with a Sliding Window: The "Gabor Wavelet"

The principle of the *Gabor function* such as it is quickly described above, shows that its use for the analysis of a signal will be done by sliding the Gabor function (Gaboret or gaborette) along the signal and each segment of the aforesaid signal will be analyzed by a "range (or scale)" of Gabor functions of different internal frequencies, but of identical amplitude and length. The Gabor transform illustrates one of the types of windowed transform. The Gabor transform of an unspecified function $f(x)$ is written:

$$\int f(x)g(x-b)e^{-i\omega x}dx \tag{5.14}$$

or can be written in a more practical form:

$$G_f(b,\omega) = \int f(x)g(x-b)^* e^{-i\omega(x-b)} dx = \langle f, g_{(b,\omega)} \rangle \qquad (5.15)$$

with a *Gabor wavelet* $g_{(b,\omega)}$:

$$g_{(b,\omega)} = g(x-b)e^{i\omega(x-b)}. \qquad (5.16)$$

The *reconstruction* of the function or initial series is carried out by the inversion of the Gabor transform.

5.5 The Wavelet Transform

5.5.1 A Wavelet ψ Is a Function of Zero Average, i.e. Zero-Integral: $\int_{-\infty}^{+\infty} \psi(t)dt = 0$

The above condition is a characteristic of wavelets and of all the functions which are potentially used as such. Although we can find some exceptions (e.g. Gauss pseudo-wavelet that is not considered as a true wavelet), the fact that the integral of a wavelet is equal to zero allows during the operation of convolution between the signal and the wavelet to identify the signal components which are of a *non-stationary nature*. In this case the wavelet coefficient resulting from the convolution will be different from zero (Fig. 5.4).

5.5.1.1 Conditions for Obtaining a Good Wavelet

It is said commonly that any function with zero-integral can be regarded as a wavelet. However additional conditions make it possible to obtain good wavelets. Firstly, it is said that they must be "regular", i.e. to have a support limited in
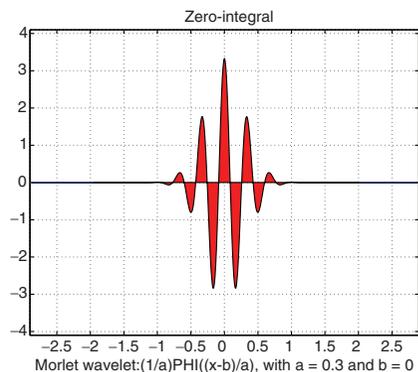
Fig. 5.4 Zero integral, zero area

Zero-integral

Morlet wavelet:(1/a)PHI((x-b)/a), with a = 0.3 and b = 0

frequency. It is necessary also that they have a support limited in time. Secondly, beyond the basic condition stated above: $\int_{-\infty}^{+\infty} \psi(t)dt = 0$, we add the following constraints relative to the *moments*:

$$\int_{-\infty}^{+\infty} t\psi(t)dt = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} t^n \psi(t)dt = 0. \tag{5.17}$$

Note that all the wavelets do not satisfy necessarily these conditions.

5.5.2 Wavelets and Variable-Window

The principle inherent to the wavelets is the *variable window* which is applied to the *mother wavelet*. The size of a variable-window is dilated or compressed. Thus, only the frequency and sometimes the amplitude of oscillations contained in the window vary. This technique is truly at the root of the construction of a wavelet transform.

5.5.3 The Wavelet Transform

5.5.3.1 Scale of Frequencies and Level of Resolution: Illustrations

The principle of the wavelet transform is different from the transformation by the *Gabor function* (also called "Gabor atom"). Indeed, contrary to the principle of the construction of the Gabor function, which varies its frequency inside the same segment which moves on the signal, the wavelet, in this case, has a variable amplitude, or more exactly it can be "dilated" (i.e. extended) or "contracted" (i.e. compressed). The purpose is to analyze the signal by a range of wavelets more adapted to the shape of the signal. Thus, by using a type of basic wavelet (of which there is an infinity) adapted to specificities of the signal, we can obtain a certain number of sub-wavelets which are only the variation through the amplitude, height and length of the initial wavelet. Thus, we contract or we dilate the wavelet to change the size of the "window" and therefore the size of the "scale" to which we observe the signal; the "frequency" of the wavelet changes at the same time as the size of the window. *The scale is the level of resolution of the observation*. Indeed, with large wavelets the examination of the signal is carried out at a coarse resolution (with a small number of coefficients). That is why the wavelets have sometimes been called the "mathematical microscope". The compression of the wavelet increases the enlargement (i.e. magnification) and the resolution which highlights the increasingly precise details of the signal. The level of resolution is also called the "number of octaves" (usually six or five). As we will see again in next sections, it is possible to analyze a signal with intermediate resolutions by the octaves which offer partially "redundant" information. *Scale*, *resolution* or *octave* express the same reality. The transformation gives a "number of samples" of the signal. The compute of each transform at a given

scale provides a "sample" (or observation) of the signal. The method can be applied to the entire signal (or possibly to segments of it by using sliding-windows). Here are two examples of wavelet transform.

Example 5.1. (Discrete wavelet transform). For an arbitrary signal, see Fig. 5.5.

    Figure 5.5 shows a discrete wavelet transform on five scales (with cubic spline wavelet). Note that the transforms, somehow, stationarize the signal (see supra).

Example 5.2. (Continuous wavelet transform). For French stock index, see Fig. 5.6.

    Figure 5.6 shows a continuous wavelet transform using a Gauss pseudo-wavelet which is a very particular case where the signal is not stationarized. The finest resolution of the wavelet transform at the octave 5 is in the higher part of the figure, it provides the most details. We observe the transform at given scale.
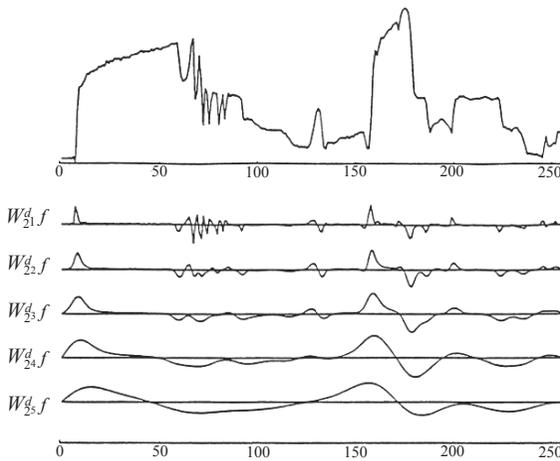
Fig. 5.5 Discrete (dyadic) wavelet transform of an arbitrary signal
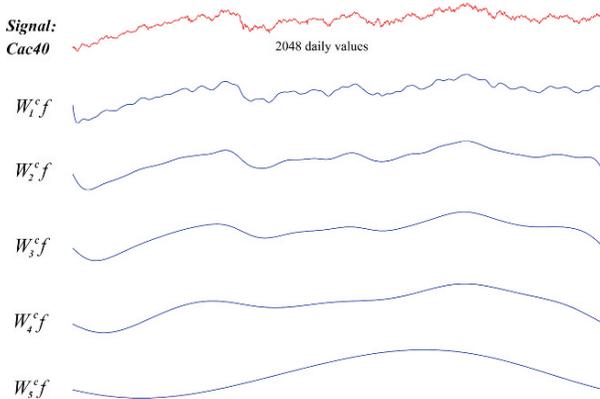
Fig. 5.6 Continuous wavelet transforms of a stock index by Gauss pseudo-wavelet

The term *scale* indicates the *correlation between the size of the wavelet and the size of the components that one can see*. Note that the scales are usually sampled in geometric progression. The term *octave* indicates that the doubling of the resolution causes to raise the frequency of wavelets in order to observe the components of double frequencies. The effective work of the window of the *wave* consists in extracting the "transitory" components of the signal at the high frequencies with a narrow window and the long duration components at the low frequencies with a large window. Thus, the size of the window is *variable*. Thus, the variable elements (after the window) are the scale and the frequency. We'll give later two examples of this variation from a Morlet basic wavelet (i.e. Morlet mother wavelet). If we write:

$$\text{Mother wavelet: } \psi(x), \tag{5.18}$$

$$\text{Wavelet family: } \psi_{(a,b)}(x) = \frac{1}{a}\psi\left(\frac{x-b}{a}\right), \tag{5.19}$$

obviously, we do not calculate a wavelet transform of a finite length signal for each possible value of the two parameters a and b, we restrict to finite sets and regular scales. The length of the signal divided by the length of the *wavelet support* noted I_ψ provides the maximum of the *scale parameter* "a", that we note here A_{max}. The floor of the scale parameter is provided by dividing the *central frequency* of the wavelet by the sampling rate of the signal (Torresani 1995). The extraction of the transform at each value of the scale provides a "sample" of the signal. The size of the wavelet is variable. It is possible to create by using for example a Morlet "mother wavelet" written as follows:

$$\psi(x) = e^{-x^2/2}e^{i\omega x}, \quad \text{with } \omega = 5.5 \tag{5.20}$$

a *wavelet family* by means of the variation of the parameter couple (b,a). Due to the fact that the lateral displacements of the wavelet according to the variation of b are simple to imagine, we keep the parameter of "centering" b equal to zero. However, it is interesting to observe the form of the wavelet, if we vary the scale parameter a and for example between values $1/5$ and 1.

5.5.3.2 Variation of the Dilation Parameter

Thus, the *wavelet family* for $a = (1/5)\ldots 1$, is written:

$$\psi_{(a,0)}(x) = \frac{1}{a}\psi\left(\frac{x}{a}\right) \tag{5.21}$$

and we can write for better represent them, by way of illustration, two sub-wavelets extracted from the family expressed above.

- For $a = 1/3$, the *sub-wavelet,* also called "daughter wavelet", is written:

$$\psi_{((1/3),0)}(x) = \frac{1}{(1/3)}\psi\left(\frac{x}{1/3}\right) = 3 \times e^{-(3x)^2/2}e^{i\,(5.5)3x}. \tag{5.22}$$

Morlet wavelet
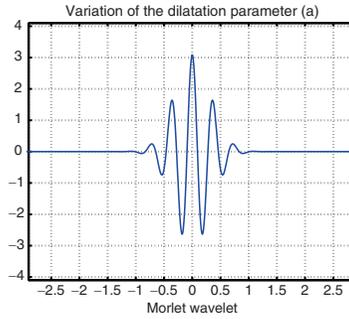
- For $a = 0.83$ the wavelet is written:

$$\Psi_{((0.83),0)}(x) = \tfrac{1}{0.83}\, \Psi\left(\tfrac{x}{0.83}\right) = \tfrac{1}{0.83} \times e^{(-x/0.83)^2/2} e^{i\,(5.5)\,(x/0.83)}. \qquad (5.23)$$

Morlet wavelet
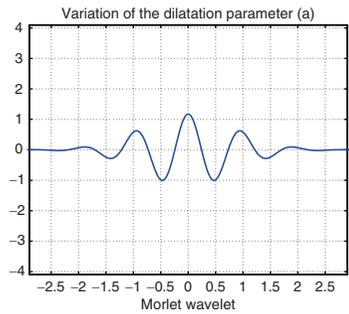
Generally, in this type of construction, when the scale parameter increases, the length of the "central frequency" of the wavelet tends to increase and, on the other hand, the height tends to decrease.

5.5.3.3 Concomitant Variation of the Parameters (b,a)

This provides a complete range, or a family of wavelets, which is written (as described previously): $\Psi_{(b,a)}(x) = e^{-\left(((x-b)/a)^2/2\right)} e^{i\omega x/a}$. We can represent two wavelets of this family, for example $\Psi_{(0,1)}$ and $\Psi_{(6,2)}$.

   In Fig. 5.7, we observe the lateral displacement, the change of amplitude, but also the shrinking of the segment inside of which the central frequency is observable. Another important aspect of the use of sliding windows is related to the good adequacy of the size of the window to the size of the studied object. (Recall: a is the dilation parameter or *scale* and b is the *position*.)
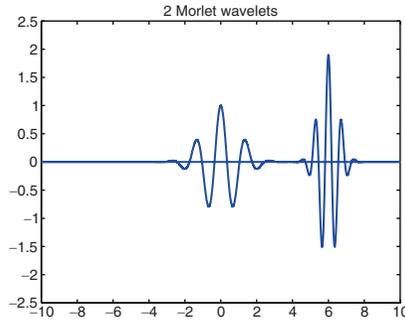
Fig. 5.7 $\left[\psi_{(0,1)}(x); \psi_{(6,2)}(x)\right]$: two Morlet wavelets $\psi_{(b,a)}(x)$, with $\omega = 5.5$

5.5.3.4 Transform and Inverse Transform

The wavelet transform can be written:

$$T_f(b,a) = \langle f, \psi_{(b,a)} \rangle = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{a}} \psi^* \left(\frac{x-b}{a} \right) dx. \qquad (5.24)$$

Like for the inverse Fourier transform, we can write the inverse wavelet transform.[16] The inverse transform is written:

$$f(x) = c_\psi \int_{-\infty}^{+\infty} \int_0^{+\infty} T_f(b,a) \frac{\psi_{(b,a)}}{a} da\,db, \qquad (5.25)$$

where the constant c_ψ is written: $c_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|}{\omega} d\omega < +\infty$.

5.5.4 Wavelet Transform and Reconstruction

5.5.4.1 Continuous Wavelet Transform and Reconstruction

The wavelet model has first been formalized by Grossmann and Morlet. For the sake of simplification, let us use the following notation: Let $\psi_s(x)$ be the dilation of the wavelet $\psi(x)$ by a factor s:

$$\psi_s(x) = \frac{1}{s} \psi \left(\frac{x}{s} \right). \qquad (5.26)$$

The wavelet transform of a function $f(x)$ at the scale s and position x is given by the convolution product:

$$W_s f(x) = f * \psi_s(x). \qquad (5.27)$$

---

[16] Theorem of Calderon, Grossmann, Morlet.

Given $\widehat{\psi}(\omega)$ the *Fourier transform* of $\psi(x)$. Morlet and Grossmann proved that if the wavelet $\psi(x)$ has a Fourier transform equal to zero at $\omega = 0$, then the wavelet transform satisfies an *energy conservation equation* and $f(x)$ *can be reconstructed* from its wavelet transform. The wavelet $\psi(x)$ can be understood as the *impulse response of a band-pass filter* and the *wavelet transform as a convolution with a band-pass filter which is dilated*. When the scale s is large, $W_s f(x)$ detects the lower frequency components of the signal $f(x)$. When the scale s decreases, the support of $\psi_s(x)$ decreases so the wavelet transform $W_s f(x)$ is sensitive to finer details. The scale s determines the size and the regularity of signal features extracted by the wavelet transform.

5.5.4.2 Dyadic Wavelet Transform and Reconstruction Formula

The wavelet transform depends on two parameters s and x which vary continuously over the set of real numbers. For practical applications, s and x must be discretized. For a specific class of wavelets, the scale parameter can be sampled along the dyadic sequence $[2^j]_{j\in\mathbb{Z}}$, without modifying the properties of the transform. The wavelet transform at the scale 2^j is given by:

$$W_{2^j} f(x) = f * \psi_{2^j}(x). \tag{5.28}$$

At each scale 2^j, the function $W_{2^j} f(x)$ is continuous since it is equal to the convolution of two functions in $\mathbf{L}^2(\mathbb{R})$. The *Fourier transform* of $W_{2^j} f(x)$ is

$$\widehat{W}_{2^j} f(\omega) = \widehat{f}(\omega)\widehat{\psi}(2^j\omega). \tag{5.29}$$

When the following *constraint* is imposed:

$$\sum_{j=-\infty}^{+\infty} \left|\widehat{\psi}(2^j\omega)\right|^2 = 1 \tag{5.30}$$

then the whole frequency axis is covered by a dilation of $\widehat{\psi}(\omega)$ by the scales factors $[2^j]_{j\in\mathbb{Z}}$. Any wavelet satisfying the constraint is called a *dyadic wavelet*. The dyadic wavelet transform corresponds also to the sequence of functions:

$$\left[W_{2^j} f(x)\right]_{j\in\mathbb{Z}}. \tag{5.31}$$

Let "$\mathscr{W}$" be the *dyadic wavelet operator* defined by $\mathscr{W} f = \left[W_{2^j} f(x)\right]_{j\in\mathbb{Z}}$. From the equation of the Fourier transform of $W_{2^j} f(x)$ with its constraint (supra), and by using the Parseval theorem, we can write an *energy conservation equation*:[17]

$$\|f\|^2 = \sum_{j=-\infty}^{+\infty} \|W_{2^j} f(x)\|^2. \tag{5.32}$$

---

[17] *Remark:* The norm (energy) of $f(x) \in \mathbf{L}^2(\mathbb{R})$ is given by $\|f\|^2 = \int_{-\infty}^{+\infty} |f(x)|^2\, dx$.

We pose $\overline{\psi}_{2^j}(x) = \psi_{2^j}(-x)$. The function $f(x)$ is *reconstructed* from its dyadic wavelet transform by the *reconstruction formula*:

$$f(x) = \sum_{j=-\infty}^{+\infty} W_{2^j} f * \overline{\psi}_{2^j}(x). \tag{5.33}$$

Given **V**, the space of the dyadic wavelet transforms $[W_{2^j} f(x)]_{j \in \mathbb{Z}}$, for all the functions $f(x) \in \mathbf{L}^2(\mathbb{R})$.[18] Then, $\mathbf{l}^2(\mathbf{L}^2)$ the Hilbert space of all sequences of functions $[h_j(x)]_{j \in \mathbb{Z}}$, such that $h_j(x) \in \mathbf{L}^2(\mathbb{R})$ and $\sum_{j=-\infty}^{+\infty} \|h_j(x)\|^2 < +\infty$. The energy conservation equation (see supra) proves that **V** is a subspace of $\mathbf{l}^2(\mathbf{L}^2)$. Then, $\mathscr{W}^{-1}$ denotes the operator from $\mathbf{l}^2(\mathbf{L}^2)$ to $\mathbf{L}^2(\mathbb{R})$ defined by: $\mathscr{W}^{-1}[h_j(x)]_{j \in \mathbb{Z}} = \sum_{j=-\infty}^{+\infty} h_j * \overline{\psi}_{2^j}(x)$. The reconstruction formula (see supra) shows that the restriction of $\mathscr{W}^{-1}$ to the wavelet space **V** is the inverse of the dyadic wavelet transform operator $\mathscr{W}$. Every sequence of functions $[h_j(x)]_{j \in \mathbb{Z}} \in \mathbf{l}^2(\mathbf{L}^2)$ is not – a priori – the dyadic wavelet transform of some function $f(x) \in \mathbf{L}^2(\mathbb{R})$. In fact, if there is a function $f(x) \in \mathbf{L}^2(\mathbb{R})$ such that $[h_j(x)]_{j \in \mathbb{Z}} = \mathscr{W} f$, then we should have $\mathscr{W}\left[\mathscr{W}^{-1}[h_j(x)]_{j \in \mathbb{Z}}\right] = [h_j(x)]_{j \in \mathbb{Z}}$. When we replace the operators $\mathscr{W}$ and $\mathscr{W}^{-1}$ by their expression given in the equations $W_{2^j} f(x) = f * \psi_{2^j}(x)$ (supra: wavelet transform) and $\mathscr{W}^{-1}[h_j(x)]_{j \in \mathbb{Z}} = \sum_{j=-\infty}^{+\infty} h_j * \overline{\psi}_{2^j}(x)$ (see supra), then it comes:

$$\sum_{l=-\infty}^{+\infty} h_l * K_{l,j}(x) = h_j(x) \text{ with } j \in \mathbb{Z} \text{ and } K_{l,j}(x) = \overline{\psi}_{2^j} * \psi_{2^l}(x). \tag{5.34}$$

The sequence $[h_j(x)]_{j \in \mathbb{Z}}$ is a dyadic wavelet transform if and only if the aforementioned equations $\sum_{l=-\infty}^{+\infty} h_l * K_{l,j}(x) = h_j(x)$ and $K_{l,j}(x) = \overline{\psi}_{2^j} * \psi_{2^l}(x)$ hold. These equations are known as *reproducing kernel equations*, and show the correlation between the functions $W_{2^j} f(x)$ of a dyadic wavelet transform. The correlation between the functions $W_{2^j} f(x)$ and $W_{2^l} f(x)$ at two different scales 2^j and 2^l can be understood by observing their Fourier transform: $\widehat{W}_{2^j} f(\omega) = \widehat{f}(\omega) \widehat{\psi}(2^j \omega)$ and $\widehat{W}_{2^l} f(\omega) = \widehat{f}(\omega) \widehat{\psi}(2^l \omega)$. The redundancy of $\widehat{W}_{2^j} f(\omega)$ and $\widehat{W}_{2^l} f(\omega)$ depend on the overlap of the functions $\widehat{\psi}(2^j \omega)$ and $\widehat{\psi}(2^l \omega)$. The energy of this overlap is equal to the energy of the kernel $K_{l,j}(x)$. (It is maximum for $l = j - 1$, $l = j + 1$.) Let $\mathbf{P_V}$ be an operator defined by

$$\mathbf{P_V} = \mathscr{W} \circ \mathscr{W}^{-1} \tag{5.35}$$

this operator is a projector from $\mathbf{l}^2(\mathbf{L}^2)$ on the **V**space. It is possible to prove that any sequence of functions $[h_j(x)]_{j \in \mathbb{Z}} \in \mathbf{l}^2(\mathbf{L}^2)$ satisfies $\mathbf{P_V}[h_j(x)]_{j \in \mathbb{Z}} \in \mathbf{V}$, and any element of **V** is invariant under the action of this operator.

---

[18] *Recall:* The following spaces correspond to the respective functions or associated signals: (1) $\mathbf{L}^2(\mathbb{R})$: Finite energy functions: $\int |f(t)|^2 dt < +\infty$; and consequently is the space of integrable square functions. (2) $\mathbf{L}^P(\mathbb{R})$: Functions such that $\int |f(t)|^p dt < +\infty$. (3) $\mathbf{l}^2(\mathbb{Z})$: Discrete finite energy signals: $\sum_{n=-\infty}^{+\infty} |f(t)|^2 < +\infty$. (4) $\mathbf{l}^P(\mathbb{Z})$: Discrete signals such that: $\sum_{n=-\infty}^{+\infty} |f(t)|^p < +\infty$.

5.5.4.3 Dyadic Wavelet Transform and Maxima

A signal is usually measured with a finite resolution which imposes a finer scale when computing the wavelet transform. For practical purposes, the scale parameter must also vary on a finite range. We are going to explain how to interpret mathematically a dyadic wavelet transform on a finite range. In both previous sections (for the sake of simplification) the model was based on functions of a continuous parameter x, but we have to discretize the abscissa x and explain efficient algorithms for computing a discrete wavelet transform and its inverse.

Interpretation of a Dyadic Wavelet Transform

For practical purposes, it is not possible to compute the wavelet transform at all scales 2^j for j varying between $-\infty$ and $+\infty$. In fact, we are limited by a finite larger scale and a non-zero finer scale. In order to normalize, we suppose that the finer scale is equal to 1 and 2^j is the largest scale. (With $f(x) \in \mathbf{L}^2$). Between the scales 1 and 2^j, the wavelet transform $[W_{2^j} f(x)]_{1 \leq j \leq J}$ can be interpreted as the details available when smoothing $f(x)$ at the scale 1 but which have disappeared when $f(x)$ at the larger scale 2^j. At this stage, we introduce a function $\phi(x)$ whose Fourier transform is:

$$\left| \widehat{\phi}(\omega) \right|^2 = \sum_{j=1}^{+\infty} \left| \widehat{\psi}(2^j \omega) \right|^2. \tag{5.36}$$

Because we know that the wavelet $\psi(x)$ verifies $\sum_{j=1}^{+\infty} |\widehat{\psi}(2^j \omega)|^2 = 1$, so we have $\lim_{\omega \to 0} |\widehat{\phi}(\omega)|^2 = 1$. In addition, the energy distribution of the Fourier transform $\widehat{\phi}(\omega)$ is localized in the low frequencies, thus $\phi(x)$ is a smoothing function. Given S_{2^j} the smoothing operator defined as follows:

$$S_{2^j} f(x) = f * \phi_{2^j}(x) \quad \text{where} \quad \phi_{2^j}(x) = \frac{1}{2^j} \phi_{2^j} \left(\frac{x}{2^J} \right). \tag{5.37}$$

The larger the scale 2^j, the more details of $f(x)$ are removed by the smoothing operator S_{2^j}. The dyadic wavelet transform $[W_{2^j} f(x)]_{1 \leq j \leq J}$ between the scale 1 and 2^j give the details available in $S_1 f(x)$ but not for $S_{2^j} f(x)$. The Fourier transforms $\widehat{S_1 f}(\omega)$, $\widehat{S_{2^J} f}(\omega)$, $\widehat{W_{2^j} f}(\omega)$ of $S_1 f(x)$, $S_{2^j} f(x)$, $W_{2^j} f(x)$ are respectively given by:

$$\widehat{S_1 f}(\omega) = \widehat{\phi}(\omega) \widehat{f}(\omega), \quad \widehat{S_{2^J} f}(\omega) = \widehat{\phi}(2^J \omega) \widehat{f}(\omega), \quad \widehat{W_{2^j} f}(\omega) = \psi(2^j \omega) \widehat{f}(\omega). \tag{5.38}$$

The first equation $|\widehat{\phi}(\omega)|^2 = \sum_{j=1}^{+\infty} |\widehat{\psi}(2^j \omega)|^2$ (i.e. Fourier transform of $\phi(x)$) gives:

$$\left| \widehat{\phi}(\omega) \right|^2 = \sum_{j=1}^{J} \left| \widehat{\psi}(2^j \omega) \right|^2 + \left| \widehat{\phi}(2^J \omega) \right|^2. \tag{5.39}$$

From the equation above and from the three equations of Fourier transforms (see supra) $\widehat{S_1}f(\omega)$, $\widehat{S_{2^J}}f(\omega)$, $\widehat{W_{2^j}}f(\omega)$ and by using Parseval's theorem, we obtain the following energy conservation equation:

$$\|S_1 f(x)\|^2 = \sum_{j=1}^{J} \|W_{2^j} f(x)\|^2 + \|S_{2^J} f(x)\|^2. \tag{5.40}$$

Such an equation shows that the higher frequencies of $S_1 f(x)$ which disappeared in $S_{2^J} f(x)$ can be recovered from the dyadic wavelet transform $[W_{2^j} f(x)]_{1 \leq j \leq J}$ between the scales 1 and 2^J. The functions $\{S_{2^J} f(x), [W_{2^j} f(x)]_{1 \leq j \leq J}\}$ are known as the *finite scale wavelet transform* of $S_1 f(x)$. For practical purposes, the signals are given by discrete sequences of values, and we know that any *discrete signal* D of finite energy can be interpreted as the uniform sampling of some function smoothed at the scale 1. (*Remark: Given* $D = [d_n]_{n \in \mathbb{Z}}$ *a discrete signal of finite energy:* $\sum_{n=-\infty}^{+\infty} |d_n|^2 < +\infty$. *Suppose that the Fourier transform* $\hat{\phi}(\omega)$ *verifies:* $\exists C_1 > 0, \exists C_2 > 0, C_1 \leq \sum_{n=-\infty}^{+\infty} |\hat{\phi}(\omega + 2n\pi)|^2 \leq C_2$ *where* $\omega \in \mathbb{R}$. *There exists a function* $f(x) \in L^2(\mathbb{R})$ *which is not unique, such that for any* $n \in \mathbb{Z}$, $S_1 f(n) = d_n$.) Thus the discrete signal D can be rewritten as follows $D = [S_1 f(n)]_{n \in \mathbb{Z}}$. For a specific class of wavelets $\psi(x)$, the samples $[S_1 f(n)]_{n \in \mathbb{Z}}$ allow to compute a uniform sampling of the finite scale wavelet transform of $S_1 f(x)$: $\{[S_{2^J} f(n)]_{n \in \mathbb{Z}}, [[W_{2^j} f(n+\rho)]_{n \in \mathbb{Z}}]_{1 \leq j \leq J}\}$ where ρ is the sampling shift which depends on the wavelet $\psi(x)$ (Meyer 1989, Appendix 2, pp. 267–268). Then, we pose: $W_{2^j}^d f = [W_{2^j} f(n+\rho)]_{n \in \mathbb{Z}}$ and $S_{2^j}^d f = [S_{2^j} f(n)]_{n \in \mathbb{Z}}$. The sequence of discrete signal $\{S_{2^J}^d f, [W_{2^j} f]_{1 \leq j \leq J}\}$ is known as a *discrete dyadic wavelet transform* of the signals $D = [S_1 f(n)]_{n \in \mathbb{Z}}$. Let $\mathscr{W}^d$ be the discrete wavelet transform operator which associates to a signal D the discrete wavelet transform previously defined. (*Remark: This operator uses a fast algorithm; note that if the signal possesses N non-zero samples, the level of complexity of such an algorithm is $O(N \log(N))$; the algorithm uses a cascade of convolutions with two discrete filters. It is also possible to compute the discrete inverse wavelet transform W^{-1d} which allows the reconstruction of the signal D from its discrete dyadic wavelet transform*). A model of a cubic spline wavelet $\psi(x)$ is shown in Fig. 5.8 (left). The shape of the function $\phi(x)$ corresponding to the wavelet $\psi(x)$ is shown in Fig. 5.8 (right) ($\phi(x)$ is also a cubic spline but with a compact support of size 4).
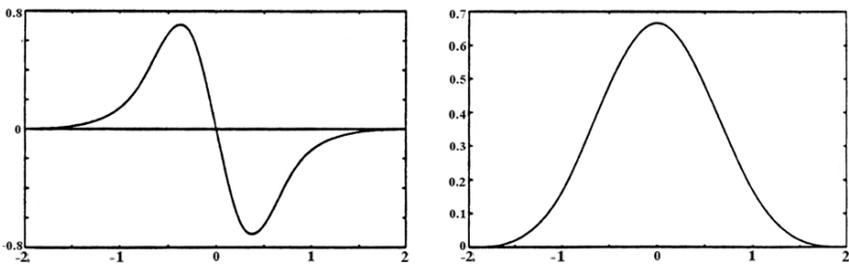
Fig. 5.8 Left: $\psi(x)$. Right: $\phi(x)$
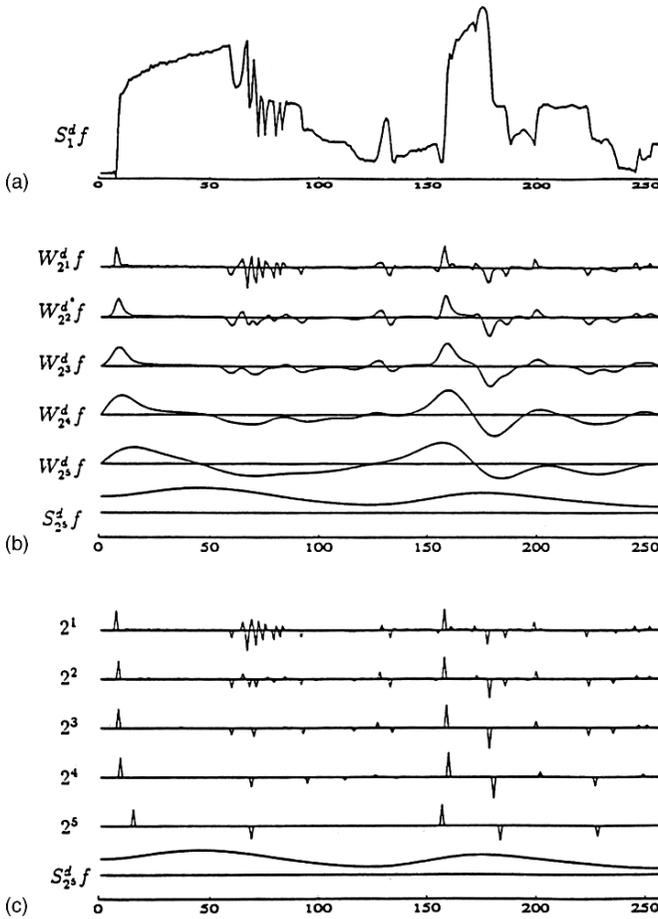
Fig. 5.9 (**a**) An arbitrary signal, (**b**) wavelet transforms (**c**) maxima (ref. to Mallat, Meyer 1989)

Figure 5.9a shows a *signal* (256 samples), Fig. 5.9b is the *discrete wavelet transform* of the signal decomposed on five scales with a cubic spline wavelet. Figure 5.9c corresponds to the *maxima representation* of the signal. At each scale, the *Diracs* show the position and amplitude of a local maximum of the wavelet transform given in Fig. 5.9b.

The curve denoted $S_{2^5}^d f$ is the coarse signal. Since the wavelet used is the first derivative of a smoothing function, the maxima of the wavelet transform show the points of sharper variation at each scale. The wavelet maxima representation of a signal D is defined by the discrete maxima of its discrete wavelet transform $W_{2^j}^d f$ for each scale $1 \leq 2^j \leq 2^J$, plus the coarse signal $S_{2^J}^d f$. The maxima show at different scales the position of the signal sharper variations. (Note that for the functions $\phi(x)$ whose support is larger than 2, the discrete maxima detection produces errors; note also that the function $\phi(x)$ used previously has a compact support of size 4.)
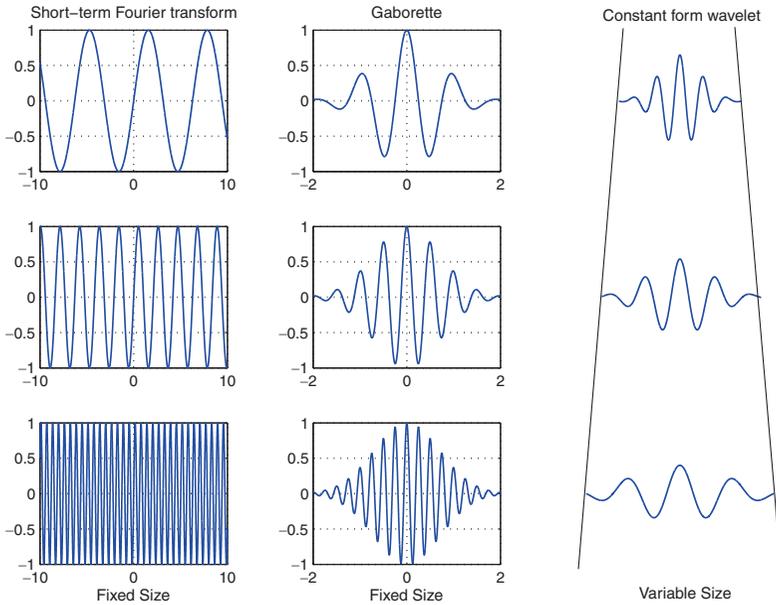
Fig. 5.10 Distinction between the different window mechanisms

5.6 Distinction of Different Window Mechanisms by Type of Transformation

In previous sections, three types of basic transformation have been distinguished, i.e. the Fourier transform (of short term), the Gabor transform and the wavelet transform. In order to distinguish the different window mechanisms, it seems useful to represent them in a graph. We observe clearly in Fig. 5.10, that for *the short term Fourier transform* and *the Gabor transform*, the window has a fixed size and consequently, it slides along the signal. On the other hand, *the size of the window is variable* for *the wavelet transform*. Indeed, we clearly distinguish the dilation of the wavelet by increase in the size of the window; the mechanism is easily visible.

5.7 Wavelet Transform of Function or Time Series

5.7.1 The Wavelets Identify the Variations of a Signal

The wavelets are a natural extension of the Fourier analysis. The purpose is always to transform a signal into coefficients, from which we can reconstruct the initial signal. The transformation of a signal is carried out with the sum of different wavelets.

If a variable size window of a wavelet is chosen narrow, we capture the *sharp* and short *variations* of the signal, i.e. events of the peak type, or the discontinuities. If the window is large, we capture *long oscillations* and long variations, i.e. low frequencies. The choice of the window is conditioned by what we want to extract. The wavelets highlight with precision the variations of the signal. An absence of variation or a very slow variation will provide a very small coefficient or equal to 0. A *wavelet* has *shapes more technical,* more complex and more adapted than a *sinusoid*. A wavelet coefficient measures the correlation between the wavelet and the signal (or portion of signal to which we apply it). Recall: a wavelet has a zero integral, and the positive areas neutralize negative areas. The wavelets are also a way of highlighting *changes of state* of a signal (i.e. transitions) with a reduced number of non-zero coefficients. The size of a window allows to capture the different components of a signal.

Figure 5.11 is shown to understand the mechanism of signal analysis and decomposition by a wavelet. A signal is artificially built, then a wavelet. Then we displace along the signal, a wavelet whose window size remains constant, and in so doing we carry out a convolution between the wavelet (whose size remains constant) and the signal. The convolution (*product*) of each segment of the signal with the wavelet provides a *new curve*. The *area* (i.e. the *integral* of this *curve*) gives the *wavelet coefficient*. The *segment* of the signal, which has "*similarities*" *with the wavelet*, will have a product (segment ∗ wavelet) providing a curve with a *high coefficient*.
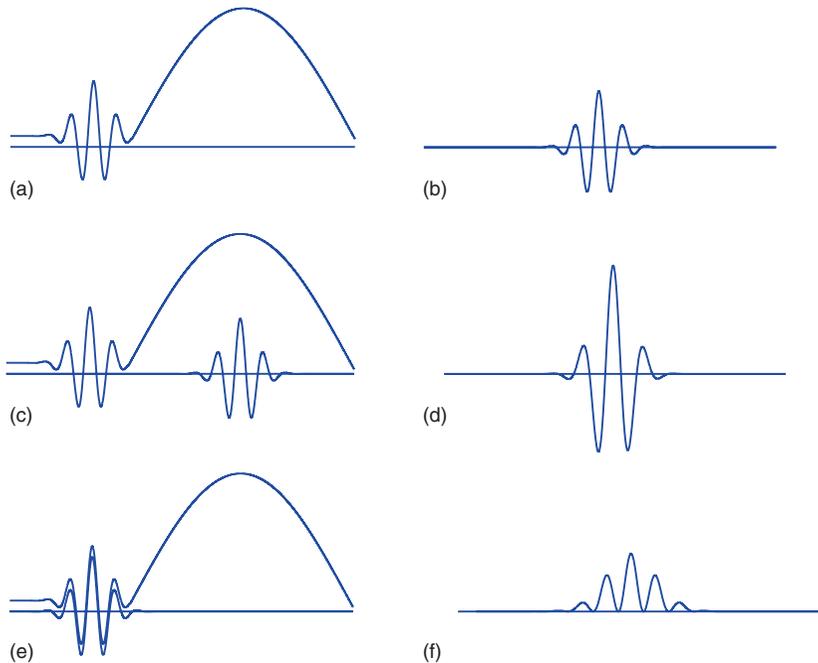
Fig. 5.11 (**a**) An arbitrary signal, (**b**) a wavelet, (**c, e**) signal analyzed by a wavelet, (**d, f**) product = signal ∗ wavelet
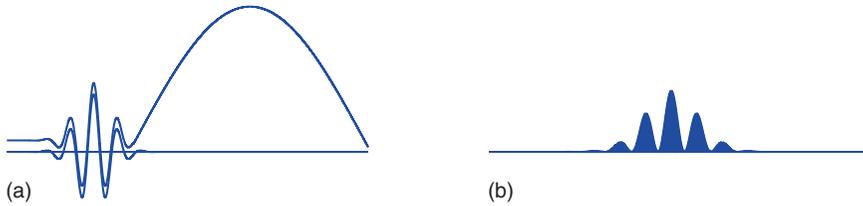
Fig. 5.12 (**a**) Signal analyzed by a wavelet, (**b**) *Area = coefficient*

Conversely, the integral of the product will be small or zero, if there is *no similarity*, because the negative values of the integral will be compensated by the positive values of areas. The coefficients of the wavelet applied to the signal, correspond to the surface located between the curve and the axis centered at zero. Note that when the wavelet analyzes the signal which has the same form as the wavelet itself, their product is entirely positive (curves of real parts) and thus the area is also positive and has a high enough value (Fig. 5.12). On the other hand, if the wavelet analyzes a *flat signal*, then the *integral* of the product (signal by wavelet) will be *zero*, because by definition the integral of the wavelet is zero. If the signal offers small curves without strong variation, the integral of the product will provide low coefficients. The displacement of the wavelet along the signal is carried out by means of the parameter (u):

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{5.41}$$

Usually, five or six groups of wavelet transforms are used to represent the initial signal, we can also speak of octaves. Each group is decomposed into intermediate transforms, usually six also. Thus, a resolution 6 provides 36 wavelet transforms of the initial signal.

5.7.2 Continuous Wavelet Transform

The continuous wavelet transform of a function or a signal f at the scale s and at the position u is calculated by "correlating" f with a wavelet also called "wavelet atom":

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right)dt, \tag{a}$$

where ψ^* is the *complex conjugate* of ψ in $\mathbb{C}$. It is also possible to write this equation as the following convolution product:

$$Wf(u,s) = f \star \bar{\psi}_s(u), \tag{b}$$

where $\bar{\psi}_s(t) = \frac{1}{\sqrt{s}}\psi^*\left(-\frac{t}{s}\right)$. The Fourier transform of $\bar{\psi}_s(t)$ is given by:

$$\hat{\bar{\psi}}_s(\omega) = \sqrt{s}\hat{\psi}^*(s\omega) \tag{5.42}$$

with

$$\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(t)dt = 0. \tag{5.43}$$

It appears that $\hat{\psi}$ is the *transfer function* of a *band-pass filter* of frequencies. This convolution computes the wavelet transform with *band-pass filters*. The graphic expression of a continuous wavelet transform is in a plane whose abscissa (u) is the unit of time and ordinate $(Log2(s))$ is the frequency scale.

5.7.3 Discrete Wavelet Transform

$$Wf[n,a^j] = \sum_{n=0}^{N-1} f[m]\frac{1}{\sqrt{a^j}}\psi_j^*[m-n] \tag{c}$$

with a wavelet:

$$\psi_j[n] = \frac{1}{\sqrt{a^j}}\psi\left(\frac{n}{a^j}\right). \tag{5.44}$$

The discrete transformation can also be written as a *circular convolution* like in the continuous case:

$$Wf[n,a^j] = f \circledast \bar{\psi}_j[n] \tag{d}$$

with

$$\bar{\psi}_j[n] = \psi_j^*[-n] \tag{5.45}$$

this circular convolution is calculated with the Fast Fourier Transform which requires $O(N\log 2N)$ operations.[19] If

$$a = 2^{\frac{1}{v}} \tag{5.46}$$

there are $v\log 2(N/2K)$ scales $a^j \in [2N^{-1}, K^{-1}]$. The number of operations necessary to compute the transformation is $O(vN(\log 2N)^2)$.

5.7.4 Wavelet Models: "Gauss Pseudo-Wavelet", Gauss-Derivative, Morlet and Sombrero

5.7.4.1 The Case of "Gauss Pseudo-Wavelet" or "Gauss Window"

Construction of a "Gauss pseudo-wavelet" (constructed by Gabor):

$$g(t) = e^{-\omega_o^2 t/2}. \tag{5.47}$$

---

[19] $f[n] = O(g[n])$, there exists K such that $(f[n] \le Kg[n])$. $(f[n]) = o(g[n])$, $\lim_{n \to +\infty} \frac{f[n]}{g[n]} = 0$.

Fig. 5.13 Gauss pseudo-wavelet, also called Gauss window
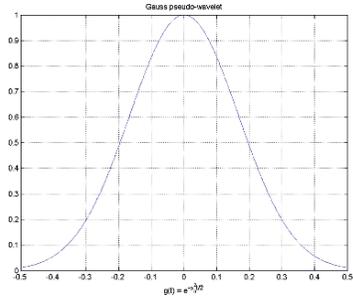
Fig. 5.14 Gauss window and its derivative x10^2
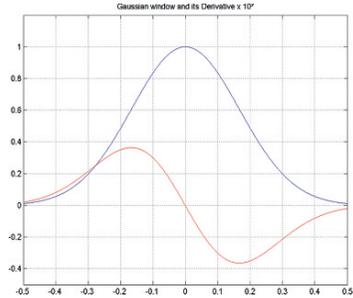
A *Gauss pseudo-wavelet* is an *isolated structure* in the set of wavelet forms which were constructed subsequently (Fig. 5.13). Indeed, its bell shape does not offer many oscillations around the *x*-axis (abscissa). It is an almost rudimentary form which offers a weak adaptation to signals with complex structures. Certain works still evoke this structure as a wavelet, however we will prefer to consider it as a *Gauss window* or as a *filter*. Indeed, it does not satisfy all the conditions necessary to the wavelet construction.

A Particularity of the Gauss Window and Gauss-Derivative Wavelet

It will be noted, first, that the derivative of a Gauss window can be regarded as a wavelet (the integral of this derivative is zero). The Gauss window and its derivative are represented in Fig. 5.14.

Let us consider the equivalent of the differential of an arbitrary signal denoted $f'(x)$ and a Gauss window $g(x)$ (i.e. "Gauss pseudo-wavelet"), then we can write by carrying out an *integration by parts*:[20]

$$\int f'(x)g(x)dx = -\int f(x)g'(x)dx + [\cdot],$$ (5.48)

where $[\cdot] = 0$. Thus:

$$\int f'(x)g(x)dx = -\int f(x)g'(x)dx.$$ (5.49)

---

[20] *Integration by parts:* $\int f'g = [fg] - \int fg'$.

Applied to a financial signal (French stock index: Cac40), thus:

$$\int cac'(x)g(x)dx = -\int cac(x)g'(x)dx. \tag{5.50}$$

This particularity is interesting because this shows that the Gauss pseudo-wavelet transform of the first-differences of the Cac40 is equivalent to the Gauss-derivative wavelet transform of the Cac40 itself (up to a sign). And their representations in the time–frequency plane are similar.

5.7.4.2 Construction of a Morlet Wavelet

A Morlet wavelet (Fig. 5.15) is written:

$$\psi(t) = e^{-i\omega_0 t}e^{-t^2/2} . \tag{5.51}$$

5.7.4.3 Construction of a Sombrero Wavelet

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}}(\frac{t^2}{\sigma^2} - 1)e^{-t^2/2\sigma^2} \tag{5.52}$$

(Fig. 5.16).
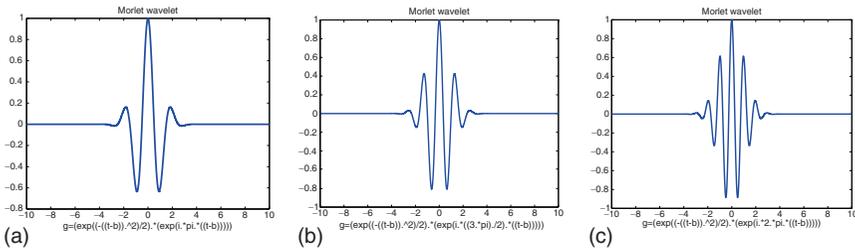
Fig. 5.15 (a) $\omega = \pi$, (b) $\omega = \frac{3\pi}{2}$, (c) $\omega = 2\pi$
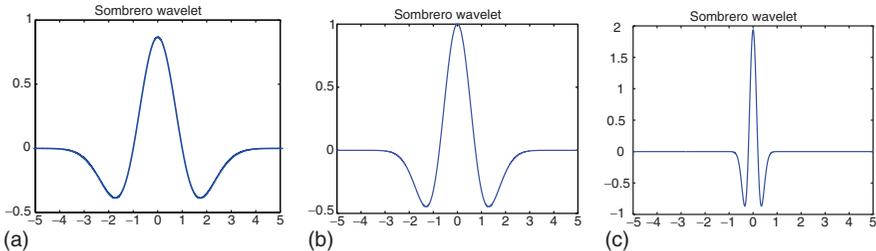
Fig. 5.16 (a) $\sigma = 1$, (b) $\sigma = 3/4$, (c) $\sigma = 1/10$

5.8 Aliasing and Sampling

In the figure below,[21] we show at the top a few "wave functions" $f(t)$ in the space of positions and at the bottom the corresponding Fourier transforms in the space of impulse responses. The graphs (2),(3),(6),(7) shows wave packets.
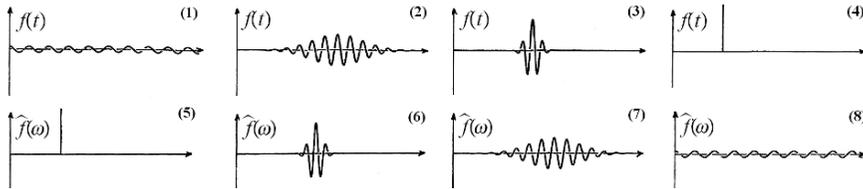
A signal f is discretized by recording its sample values $\{f(nT)\}_{n \in \mathbb{Z}}$ at interval T. A discrete signal can be represented as a sum of Diracs. Any sample $f(nT)$ is associated with a Dirac $f(nT)\delta(t - nT)$ located at $t = nT$. So, a uniform sampling of f corresponds to the weighted Dirac sum $f_d(t) = \sum_{n=-\infty}^{+\infty} f(nT)\delta(t - nT)$. The Fourier transform of $\delta(t - nT)$ is $e^{-inT\omega}$. Thus the Fourier transform of f_d is a Fourier series $\widehat{f}_d(\omega) = \sum_{n=-\infty}^{+\infty} f(nT)e^{-inT\omega}$. $\widehat{f}$ and $\widehat{f}_d$ are related by $\widehat{f}_d(\omega) = (1/T)\sum_{k=-\infty}^{+\infty} \widehat{f}(\omega - (2k\pi/T))$. The sampling theorem is given by:

Theorem 5.1 (Shannon, Whittaker). *If the support of the Fourier transform $\widehat{f}$ is included in $[-\pi/T, \pi/T]$ then $f(t) = \sum_{n=-\infty}^{+\infty} f(nT)\vartheta_T(t - nT)$ with $\vartheta_T(t) = \sin(\pi t/T)/(\pi t/T)$.*

This theorem imposes that the support of $\widehat{f}$ is included in $[-\pi/T, \pi/T]$ which prevents abrupt variations of f between consecutive samples. However the sampling interval is generally imposed and the support of $\widehat{f}$ is not included in $[-\pi/T, \pi/T]$. So the formula of the theorem does not recover f. A filtering technique to reduce error is then used, i.e. *aliasing*.[22] Consider a support of $\widehat{f}$ going beyond $[-\pi/T, \pi/T]$, usually the support of $\widehat{f}(\omega - (2k\pi/T))$ intersects $[-\pi/T, \pi/T]$ for several k, see infra Fig (ii). When there is aliasing, the interpolated signal $\vartheta_T * f_d(t) = \sum_{n=-\infty}^{+\infty} f(nT)\vartheta_T(t - nT)$ possesses a Fourier transform $\widehat{f}_d(\omega)\widehat{\vartheta}_T(\omega) = T\widehat{f}_d(\omega)\mathbf{1}_{[-\pi/T, \pi/T]}(\omega) = \mathbf{1}_{[-\pi/T, \pi/T]}(\omega)\sum_{k=-\infty}^{+\infty} \widehat{f}(\omega - (2k\pi/T))$ which can be very different from $\widehat{f}(\omega)$ over $[-\pi/T, \pi/T]$. The signal $\vartheta_T * f_d(t)$ can be an inappropriate approximation of f, see infra (†) in Fig. (ii). In Fig. (i): (a) shows a signal with its Fourier transform; (b) a uniform sampling of the signal generates a *periodic* Fourier transform; (c) low-pass filter; (d) the filtering of (b) with (c) recovers f. In Fig. (ii): (a)′ shows a signal with its Fourier transform; (b)′ aliasing generated by an overlapping of $\widehat{f}(-2k\pi/T)$ for different k, see dashed lines; (c)′ low-pass filter; (d)′ the filtering of (b)′ with (c)′ generates a low-frequency signal different from f, see (†).

---

[21] *Remark*: The Heisenberg uncertainty relations are illustrated by the fact that a large uncertainty in position is associated with a low uncertainty in impulse response and vice versa.

[22] **Definition (Aliasing).** Introduction of error into the computed amplitudes of the lower frequencies in a Fourier analysis of a function carried out using discrete time samplings whose interval does not allow the proper analysis of the higher frequencies present in the analyzed function.

Fig.(i)

Fig.(ii)
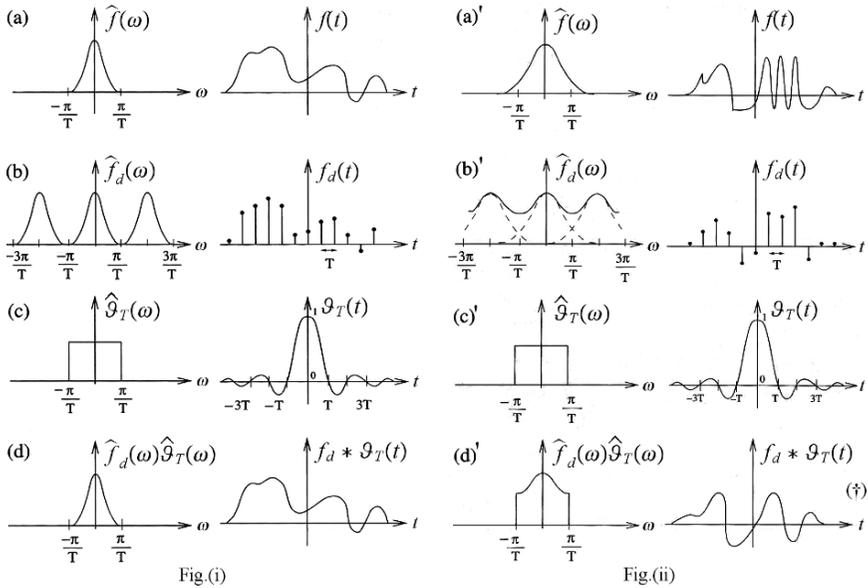
5.9 Time-Scale Plane (b,a), Cone of Influence

5.9.1 Cone of Influence and Time-Scale Plane

Let $f(x)$ be a function and $\psi(x)$ a wavelet which occupies the space: $I_\psi = [-X, X]$. The behavior of the function $f(x)$ in a precise point, x, is mainly observed on the wavelet transform in the cone $b \in a\, I_\psi + x = [-aX + x, aX + x]$ (Fig. 5.17).

$\psi(x)$ is a mother wavelet and $\psi_{(a,b)}(x)$ is a wavelet family created by the changes of the parameters a, b. As described in previous sections, "b" represents the position in time of the wavelet, i.e. its position along the analyzed signal. "a" represents the
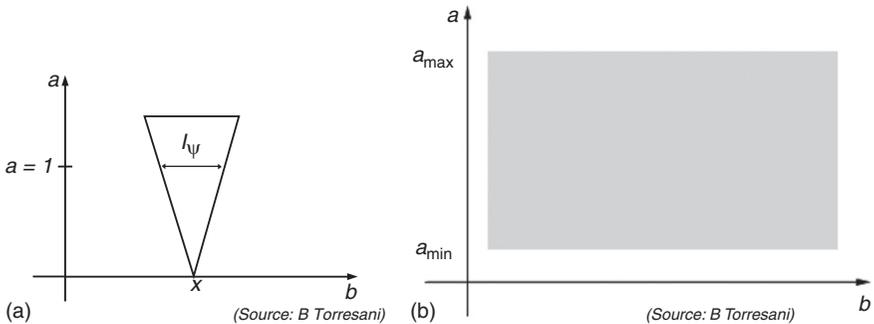
Fig. 5.17 (**a**) Cone of influence of a point x of b. (**b**) (b,a)-plane
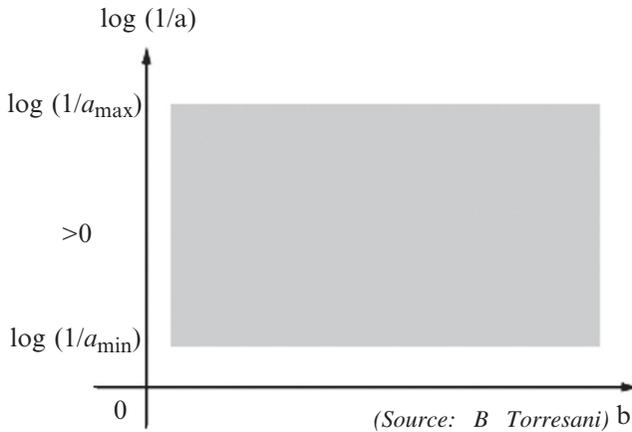
Fig. 5.18 Phase-space (hyperbolic plane)

dilation and contraction of the wavelet. The wavelet transform $T_f(b,a)$ is repre-
sented in the *hyperbolic space* of transforms (Fig. 5.18) (*Time–frequency plane =
Phase space*). We mentioned it in the preceding section, in certain types of construc-
tion, when the scale parameter "a" increases, the length of the "central frequency"
of the wave tends to increase. This is an analysis which is valid for example within
the framework of the continuous wavelet transform. Thus, in the plane which pre-
cedes, the coarsest or less fine "long waves" calculated by the transformation, will
be in the upper part of the plane. On the contrary, the finest wavelets, which analyze
with more precision the signal, will be represented in the lower part of the plane.
Hyperbolic planes such as these allow a form of symmetry around the abscissa (x-
axis). We often uses the scale "log 2", which offers the property to represent the
transforms of a signal in the positive plane above the abscissa, if the scale parameter
is taken as a divider, i.e. log 2(1/a). In the opposite case, the image of transforms will
be represented in the plane below the abscissa. The plane itself will obviously have
a minimum and a maximum that determine thus the limits of the image (Fig. 5.19).

Note that if we select a scale parameter as divider of transforms, the ordinates of
the plane will be positive, otherwise, the ordinates will be negative. The representa-
tion in negative ordinates is rather frequent (see Torrésani or Mallat). If the signal
which is the subject of a transformation has a dyadic length, we will use *log 2* rather
than *log*. We cannot numerically calculate a wavelet transform for all the values
of b and a. We sample the wavelet transform. The extreme values of the scale are
computed starting from the following extrema, as follows:

$$a_{max} = \frac{\text{Size of the signal}}{\text{Size of the wavelet}},$$

$$a_{min} = \frac{\text{Central frequency of the wavelet}}{\text{Sampling frequency of the signal}}.$$
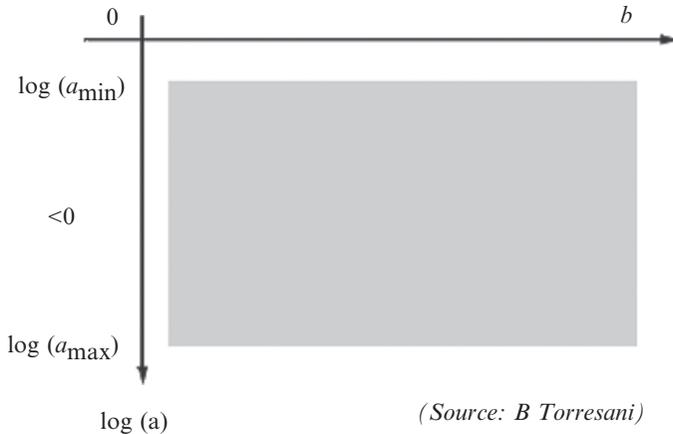
Fig. 5.19 Hyperbolic plane, logarithmic scale

5.9.2 Time–Frequency Plane

The time–frequency plane is used in particular in the transformations with sliding window of the Gabor type, or in the wavelet multiresolution programs. The Mallat and Zhang "Pursuit" algorithm uses also the time–frequency plane. The general principle: Since the wavelet and the window are localized along the signal, it is possible – by using segments which are sampled and localized in time – to make correspond to these segments a frequency which will occupy by projection a segment on the ordinate axis. Thus, we construct "time–frequency" rectangles or "time–frequency" squares localized such as "tiles" in the plane. This process is close to the principle of Heisenberg boxes described in the section which follows. The construction of such time–frequency planes requires to constitute a grid intended to receive these time–frequency rectangles: $(\Delta t, \Delta \omega)$ or $(\sigma_t, \sigma_\omega)$. The choice of the type of grid can be large enough, we can speak of the Fourier grid or Gabor grid. Then, we establish a scale corresponding to frequency levels, for example: gray nuances. Finally, we assign to each "time–frequency tile" a nuance of gray corresponding to its eigen frequency. There are other types of construction of time–frequency plane, in particular the plane of the Wigner–Ville distribution, the plane of the energy distribution, or the spectrograms, for which representations are more diffuse and segmented, just as for the representation of the energy distribution resulting from the "Pursuit" algorithm.

5.10 Heisenberg Boxes and Time–Frequency Plane

Briefly, we introduce the Heisenberg theorem which leads to the construction of Heisenberg boxes, which then allows to underline the two different approaches between the wavelets and "Fourier", and this by using *the bivalent concept of*

time–frequency atom which makes it possible to unify, but also to have in hand equivalences from a field to another of the analysis.

Theorem 5.2 (Heisenberg uncertainty). *The temporal variance σ_t^2 and the frequency variance σ_ω^2 of the function f (which belongs to the L^2 space), verify:*

$$\sigma_t^2 \sigma_\omega^2 \geq \frac{1}{4} \tag{5.53}$$

or, in a similar way by means of the standard deviations $\sigma_t \sigma_\omega \geq \frac{1}{2}$. The inequality becomes an equality if and only if there exist $(a, b, u, \xi) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that $f(t) = a \exp(i\xi t - b(t - u)^2)$.

This observation leads to a kind of compromise between the *temporal* resolution and the *frequential* resolution. *The localization in time–frequency can be reached only in standard deviation*. This localization is representable by means of a Heisenberg box. The cover of the time–frequency plane is carried out through different manners in accordance with the Fourier atoms or wavelet atoms.

5.10.1 Concept of Time–Frequency Atom: Concept of Waveform Family

We know that the transformation of an unspecified signal in the time–frequency plane can be carried out either by sines and cosines or by wavelets, thus, we can indicate in a general way these forms which enter into the mechanics of transformation as *waveforms*; a *family of waveforms* is also called *time–frequency atoms* and is written: $\{\phi_\gamma\}$ where γ is a vector of parameters. Generally, it is agreed that ϕ_γ such that $\|\phi_\gamma\| = 1$.[23] The transformation of a signal $f(t)$ by this family of atoms ϕ_γ, can be written:[24]

$$Tf = \int_{-\infty}^{+\infty} f(t)\phi_\gamma^*(t)dt = \langle f, \phi_\gamma \rangle. \tag{5.54}$$

For a *windowed Fourier transform*, *a Fourier atom* by translation and modulation is written:

$$\phi_\gamma(t) = g_{u,\xi}(t) = g(t - u)e^{i\xi(t-u)}. \tag{5.55}$$

For a *wavelet transform*, a *wavelet atom* by translation and modulation are written:

$$\phi_\gamma(t) = \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{5.56}$$

---

[23] ϕ_γ: belongs to the integrable functions $L^2(\mathbb{R})$.

[24] By the well-known Parseval formula we know that:
$Tf = \int_{-\infty}^{+\infty} f(t)\phi_\gamma^*(t)dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\widehat{f}(\omega)\widehat{\phi}_\gamma^*(\omega)d\omega.$

5.10.2 Energy Density, Probability Distribution and Heisenberg Boxes

After having defined the concept of time–frequency atom, we present the concept of density of energy (Mallat 1998). These notions are increasingly diffused out of the Physics where they come from.

5.10.2.1 Energy Density and Probability Distribution of a Time–Frequency Atom

For any parameter of translation and modulation (u, ξ), we suppose that there is a single atom $\phi_{\gamma:(u,\xi)}$ which is centered at (u, ξ) in the time–frequency plane. The time–frequency boxes of this atom determines a neighborhood of (u, ξ) where the *energy* of f is measured by

$$P_T f(u, \xi) = \left| \langle f, \phi_{\gamma:(u,\xi)} \rangle \right|^2 = \left| \int_{-\infty}^{+\infty} f(t) \phi^*_{\gamma:f(u,\xi)}(t) \, dt \right|^2. \tag{5.57}$$

The generic transformation $Tf = \langle f, \phi_\gamma \rangle$ in the time–frequency plane (t, ω) of a signal depends on the nature of the atom ϕ_γ and depends on its *time–frequency "spread"*, i.e. as well in time as in frequency. Thus, one attempts to approach this notion of *spread* of a time–frequency atom. We know that:

$$\|\phi_\gamma\|^2 = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 \, dt = 1, \tag{5.58}$$

the element which is the subject of the integration, i.e. the square of the absolute value of the atom, is understood as *a probability distribution* centered at u_γ, whose *center* (or average) is:

$$u_\gamma = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 \cdot t \, dt. \tag{5.59}$$

And its *variance* measures its *spread around u:*

$$\sigma_t^2(\gamma) = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 (t - u_\gamma)^2 dt. \tag{5.60}$$

So that we can symbolically write that the atom follows a *probability law:*

$$\phi_\gamma \sim Law(u_\gamma, \sigma_{t(\gamma)}^2). \tag{5.61}$$

If we are located in the space of frequencies (of the trigonometrical functions or wavelets), we can wonder what becomes the center, the variance and the spread of this atom which (it) is expressed simultaneously in time and frequency. It is by using the Plancherel transfer formula, which *provides an equivalence in both fields frequency and time, that we can determine the center frequency of the atom and*

its spread in the frequency (frequency support). The Plancherel formula, which is written $\int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 \, dt = 1/2\pi \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 \, d\omega$, provides this equivalence between both fields, and applied to ϕ_γ, this gives:

$$\int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 \, d\omega. \tag{5.62}$$

In addition, we know that for an unspecified[25] function f we have $\|f\|^2 = \langle f, f \rangle = \int_{-\infty}^{+\infty} |f(t)|^2 \, dt$, consequently for the atom ϕ_γ, there is a similar result: $\|\phi_\gamma\|^2 = \langle \phi_\gamma, \phi_\gamma \rangle = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 \, dt$, thus the Plancherel formula for the atom ϕ_γ is written:

$$\int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 \, d\omega, \tag{5.63}$$

this can also be written:

$$2\pi \|\phi_\gamma\|^2 = \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 \, d\omega. \tag{5.64}$$

Thus, endowed with *these equivalences*, we can determine the *center frequency* of the atom and *its "frequency spread" (measured by the variance)*. The *center frequency* ξ_γ of $\hat{\phi}_\gamma$ is written:

$$\xi_\gamma = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega \left|\hat{\phi}_\gamma(\omega)\right|^2 \, d\omega, \tag{5.65}$$

then the *spread* around the center frequency ξ_γ is written:

$$\sigma_\omega^2(\gamma) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\omega - \xi_\gamma)^2 \left|\hat{\phi}_\gamma(\omega)\right|^2 \, d\omega. \tag{5.66}$$

We are thus endowed with the necessary tools for a representation in the time–frequency plane of a basic atom ϕ_γ by means of the two following pairs (u_γ, σ_t^2) and $(\xi_\gamma, \sigma_\omega^2)$, i.e. respectively the temporal center and spread, then the center frequency and spread in the frequency. The time–frequency resolution is depicted in the time–frequency plane (t, ω) by means of the Heisenberg boxes centered at (u_γ, ξ_γ) of which the width along time is $\sigma_t(\gamma)$ and the width along frequency is $\sigma_\omega(\gamma)$. Moreover, we note that $\sigma_t \sigma_\omega$ is the area of a Heisenberg box. *By referring to the Heisenberg uncertainty theorem* which explains that the temporal variance σ_t^2 and the frequency variance σ_ω^2 of the function f, verify:

$$\sigma_t^2 \sigma_\omega^2 \geq \frac{1}{4}, \tag{5.67}$$

---

[25] Even if it is a *discontinuous* function f.
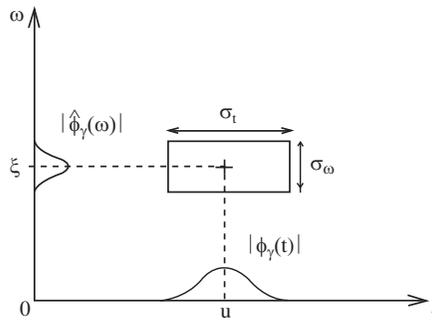
Fig. 5.20 Heisenberg Boxes representing an atom ϕ_γ

or, in an equivalent way: $\sigma_t \sigma_\omega \geq 1/2$, this *constraint* on the variances says that the *surface of the rectangle representing the atom is higher or equal to 1/2*. Thus, we do not represent the atom by a point but by rectangles, we will be able to refer to the figure which follows in the next section.

5.10.2.2 Heisenberg Boxes of Time–Frequency Atoms: Fourier Atoms and Wavelet Atoms

Generic presentation of the Heisenberg box: One represents the time–frequency localization of a basic atom (time–frequency atom) by means of a Heisenberg box, located in the time frequency plane, which is a rectangle of dimensions σ_t and σ_ω, centered at the point of coordinates: (temporal center, center frequency), as described in the previous section (Fig. 5.20).

Presentation of the Heisenberg box for a windowed Fourier atom: Remember that the principle of the windowed Fourier transform can be defined as follows:

$$Sf(u,\xi) = \langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t)g(t-u)e^{i\xi(t-u)}dt. \tag{5.68}$$

The *atom* used is a *sinusoid multiplied* by a *window g*. The analysis vector family is obtained by translation and modulation of the window as follows: $g_{u,\xi}(t) = g(t - u)e^{i\xi(t-u)}$. This function is centered for the frequencies at ξ and is symmetric in relation to u. The standard-deviation in frequency is constant. The family is thus obtained by translation in time and frequency of a single window. Here is an example of Heisenberg boxes of windowed Fourier atoms (Fig. 5.21).

Presentation of the Heisenberg box for the "wavelet atoms": Let us recall the principle of the wavelet transform which can be defined as follows:

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right)dt, \tag{5.69}$$
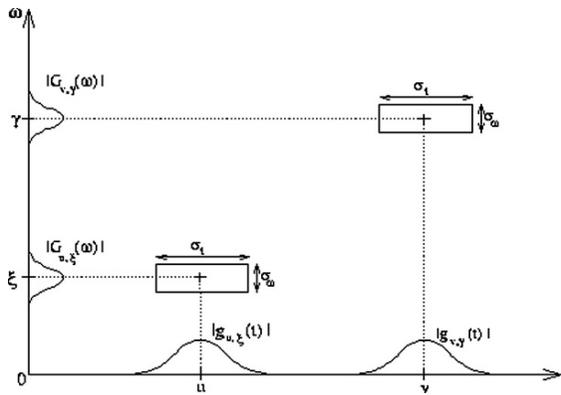
Fig. 5.21 Heisenberg rectangles, or time–frequency boxes, symbolizing the *energy spread* of two Gabor atoms

where the basic atom, or the basic wavelet, ψ is a function of zero average, centered at the neighborhood of 0 and of a finite energy. The family of vectors is obtained by translation and dilation of the atom $\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi \left(\frac{t-u}{s} \right)$. This function is centered at the neighborhood u, as the windowed Fourier atom. If the center frequency of ψ is indicated by η, the center frequency of the dilated function is at η/s. The temporal standard-deviation is proportional to s. The standard-deviation in frequency is inversely proportional to s. Thus, one can present an example of Heisenberg boxes of wavelet atoms (Fig. 5.22).

5.10.3 Spectrogram, Scalogram and Energy Conservation

The square of the modulus of the windowed Fourier transform described in the preceding section corresponds to the spectrogram:

$$P_S f(u, \xi) = |S f(u, \xi)|^2 = \left| \int_{-\infty}^{+\infty} f(t) g(t-u) e^{i\xi(t-u)} dt \right|^2. \qquad (5.70)$$

In order to calculate the *scalogram*, which is the equivalent for the *wavelets* of the *spectrogram* for the *Fourier transform,* one takes for example η as the center frequency of the wavelet, then the center frequency of a dilated wavelet is $\xi = \eta/s$. The scalogram is written then:

$$P_W f(u, \xi) = |W f(u, s)|^2 = \left| W f \left(u, \frac{\eta}{\xi} \right) \right|^2. \qquad (5.71)$$
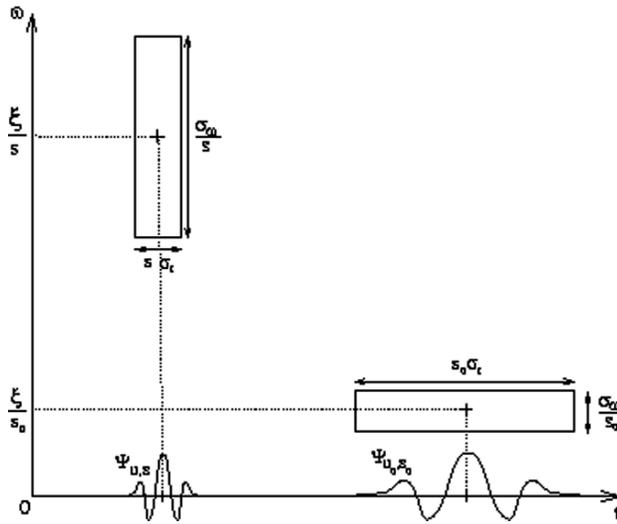
Fig. 5.22 Example for two wavelets $\psi_{u,s}(t)$ and $\psi_{u_0,s_0}(t)$. When the scale s decreases, the time-segment is reduced but the "spread" of the frequency increases

The *scalogram* is also called *energy density*. This interpretation of the scalogram as energy density is shown by the following theorem, which shows that the energy is preserved during the transformation of a real unspecified signal f:

Theorem 5.3 (The scalogram is an energy density). [26] *For any $f \in \mathbf{L}^2(\mathbb{R})$*

$$W f(u,s) = \frac{1}{2} W f_a(u,s) \tag{5.72}$$

if the admissibility condition of the wavelet written:

$$C_\psi = \int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|}{\omega} d\omega < +\infty \tag{5.73}$$

is verified and if f is real, then:

$$f(t) = \frac{2}{C_\psi} \operatorname{real} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} W f(u,s) \psi_s(t-u) du \frac{ds}{s^2} \right] \tag{5.74}$$

and

$$\|f(t)\|^2 = \frac{2}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |W f(u,s)|^2 du \frac{ds}{s^2}. \tag{5.75}$$

---

[26] Proof in Mallat (1998).

5.10.4 Reconstruction Formulas of Signal: Stable and Complete Representations

The objective of this section is to point out the reconstruction formulas of a signal, which result from the Fourier transform with window and from the wavelet transform.

5.10.4.1 Reconstruction by the Windowed Fourier Transform

For the Fourier atoms $g_{u,\xi}(t)$ when the couple of indexes (u,ξ) varies on $\mathbb{R}^2$, the Heisenberg boxes of the atoms cover the whole of the time–frequency plane. This assertion has a corollary, indeed it is possible to rebuild the signal by means of its windowed Fourier transform $Sf(u,\xi)$. The *reconstruction formula* of the signal is written:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u,\xi) g(t-u) e^{it\xi} d\xi du, \qquad (5.76)$$

with $\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |Sf(u,\xi)|^2 g(t-u) e^{it\xi} d\xi du$. Moreover, the *reconstruction formula preserves identical the energy of the signal*. The reconstruction formula can also be rewritten as follows:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle f, g_{u,\xi} \rangle g_{u,\xi}(t) d\xi du. \qquad (5.77)$$

5.10.4.2 Reconstruction by the Wavelet Transform

Similarly, it is possible to rebuild the signal from the wavelet transform. The (real) wavelet transform is complete and preserves the quantity of energy (as long as the wavelet verifies the weak admissibility condition written: $C_\psi = \int_0^{+\infty} (|\widehat{\psi}(\omega)|/\omega) d\omega < +\infty$). The *reconstruction formula is given by:*

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) du \frac{ds}{s^2}, \qquad (5.78)$$

with

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(u,s)|^2 du \frac{ds}{s^2}. \qquad (5.79)$$

5.11 Wiener Theory and Time–Frequency Analysis

5.11.1 Introduction to the Correlogram–Periodogram Duality: Similarities and Resemblances Researches

The correlogram and periodogram are the base of spectral analysis. Spectral analysis can be interpreted as the decomposition of the variance of a time series into the field of frequencies. It is commonly said that the function of the temporal correlation corresponds to the Fourier transform of the power spectrum of a time series. The purpose of spectral analysis is to identify similarities and resemblances of a signal (on itself). One will choose here in this preamble a simplified presentation of these periodogram and correlogram concepts, which are also described within the framework of the Wiener–Khintchine theorem. Consider a process $\{x_t\}$ with a temporal index $t = 1,\dots,T$. If one chooses a stochastic process weakly stationary, it will satisfy the conditions:

$$E\,|x_t|^2 < \infty, \quad \forall t, \tag{5.80}$$

$$Ex_t = \mu, \quad \forall t, \tag{5.81}$$

$$E(x_t - \mu)(x_s - \mu) = E(x_{t+\tau} - \mu)(x_{s+\tau} - \mu), \quad \forall s,t,\tau. \tag{5.82}$$

5.11.1.1 Autocovariance and Autocorrelation Functions

Definition 5.1 (Autocovariance). The *autocovariance function* at the given delay τ of a process x_t is defined for $\tau \in \{0,\dots,T-1\}$ such that:

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=\tau+1}^{T} (x_t - \bar{x})(x_{t-\tau} - \bar{x}), \tag{5.83}$$

where $\bar{x} = \frac{1}{T}\sum_{T=1}^{T} x_t$ is an arithmetic mean of x_t. The autocorrelations $\hat{\rho}(\tau)$ are defined by standardizing the autocovariance function $\hat{\gamma}(\tau)$ by $\hat{\gamma}(0)$ (i.e. the variance of the process):

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\gamma(0)}. \tag{5.84}$$

(The confidence interval on the correlograms is written $\pm 2/\sqrt{T}$.) Another measure often used is *the partial autocorrelation*. A partial autocorrelation with order $\tau \geq 2$ is calculated as a correlation of two residuals obtained after regression of $x_{\tau+1}$ and x_1 on the intermediate observations $x_2,\dots,x_\tau$. The partial autocorrelation at the given delay equal to 1 is defined as the correlation between x_1 and x_2.

5.11.1.2 Estimator of the Spectral Density

The periodogram is the equivalent, in the field of frequencies, to the autocorrelation function in the field of time. The analysis of the frequency field is carried out by the decomposition of the observed series into periodic components. The principal tool for the spectral analysis of a series is the *spectrum* which is defined by:

$$f(\lambda) = \frac{1}{2\pi} \left[\gamma(0) + 2 \sum_{\tau=1}^{\infty} \gamma(\tau) \cos(\lambda \tau) \right],$$
(5.85)

with λ the *angular frequencies* $[-\pi, \pi]$ and $\gamma(\tau)$ the theoretical autocovariances $(\tau = 1, \ldots, \infty)$. The *spectrum is symmetric around zero*, the analysis is restricted at the frequencies $[0, \pi]$. For a sample of T observations, one considers the harmonic frequencies, or Fourier frequencies $\lambda_j = 2\pi j/T$, $j = 1, \ldots, [T/2]$. The *periodogram*, by using the spectrum, is defined by:

$$I(\lambda) = \frac{1}{2\pi} \left[\hat{\gamma}(0) + 2 \sum_{\tau=1}^{T-1} \hat{\gamma}(\tau) \cos(\lambda \tau) \right].$$
(5.86)

An *estimator* is said to be "*consistent*" *if its variance tends towards zero* when the number of observations tends towards the infinite. The consistency implies that the estimator becomes gradually more precise when the quantity of acquired information increases. If one chooses *a stationary random process of order two*, it can be shown that the *estimator obtained is centered but does not converge*. It is appropriate only for the time series which have strict periodicities and its graph (periodogram is equivalent to the autocorrelation function in the field of time) has an "abrupt" shape which makes it difficult to interpret it. In order to solve this problem of readability of the graph, spectral windows are used. Indeed, in order to solve this readability problem of graphs, we do not estimate for each frequency the value $I(\lambda)$, *but an average value on equal frequency bands*, whose juxtaposition covers the interval given by λ. In fact, we carry out a *smoothing* of the *spectrum*, i.e. a kind of filtering. We *take, around a number of equidistant points of the frequency axis*, a weighted average on the "neighboring frequencies", i.e. one builds a window in the space of the frequency. The choice of the window is important. According to its form, i.e. according to the chosen weighting factors, *spectral "leaks" at the adjacent frequencies for a given frequency are possible*. For example, if at a frequency λ_j of the spectrum of a time series there exists a *"peak" with a high spectral power*, in relation to the adjacent frequencies, *the secondary lobes inherent to the spectral window, can generate, at their frequencies, significant spectral powers. It is thus difficult to choose a good spectral window,* because *it is necessary that the power of secondary lobes is low* and that *the power at the adjacent frequencies are not correlated*. There are two types of window often used in spectral analysis, they are the *Tukey–Hanning* and *Parzen* windows. For the *first*, it is said that *the secondary peaks are lower or*
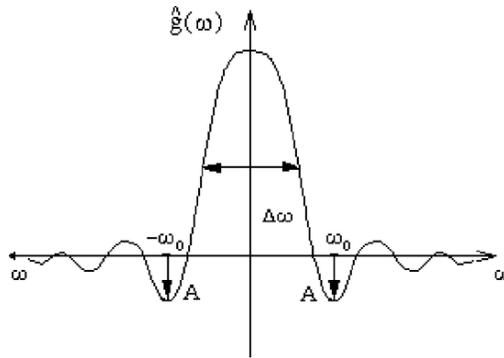
Fig. 5.23 *Plane (x-axis: frequency; y-axis: window function ĝ):* The energy *spread of* $\hat{g}$ measured by its *bandwidth* $\Delta\omega$ and the *maximum amplitude A* of the *first side-lobes* (localized at $\omega = \pm\omega_0$)

equal to 2% of the main peak and the spectral estimators at the frequencies λ_j *and* λ_{j+2} *are not correlated.*[27]

Construction of a Fourier Windows

Let g be the window in the (windowed) Fourier transform, whose energy is concentrated around 0 (Fig. 5.23). There is *three parameters which determine the energy spread*:

(1) The *root mean square bandwidth* $\Delta\omega$ written as follows:

$$\frac{|\hat{g}(\Delta\omega/2)|^2}{|\hat{g}(0)|^2} = \frac{1}{2}. \tag{5.87}$$

If $\Delta\omega$ is small, t*he energy of the window is well concentrated* around 0.
(2) The *maximum amplitude A* of its *first side-lobes*, measured in decibels:

$$A = 10\log_{10}\frac{|\hat{g}(\omega_0)|^2}{|\hat{g}(0)|^2}. \tag{5.88}$$

It is possible that these side-lobes create *shadows* on each side of the center frequency.
(3) The *polynomial exponent* which describes *the decay of the window of the Fourier transform for broad frequencies*:

$$|\hat{g}(\omega_0)| = O(\omega^{-p-1}) \tag{5.89}$$

It represents the behavior of the Fourier transform beyond the first side-lobes.

---

[27] For a discuss about the subject one will be able to refer to Bourbonnais and Terraza (1998).

Traditional Spectral Windows: Rectangular, Gaussian, Hamming, Hanning, Blackman

As presented previously the Heisenberg uncertainty theorem imposes that the standard-deviation in time σ_t and in frequency σ_ω of a function verify: $\sigma_t \sigma_\omega \geq 1/2$. This observation leads to a compromise between the temporal resolution and the frequential resolution. The localization in time–frequency can be reached only in standard deviation (or variance). This localization is representable by the Heisenberg box. However, it will be noted that *the limit $\sigma_t \sigma_\omega = 1/2$ is reached only if the window is Gaussian*. The Gaussian, Hamming, Hanning, Blackman windows obviously have very close structures, but nevertheless different. The following table gives the values of parameters which were presented in the previous section for traditional windows, normalized so that $g(0) = 1$.

Window	$g(t)$	$\Delta\omega$	A_{max} (dB)
Rectangular	$g_{rec}(t) = 1,$	0.89	-13
Gaussian	$g_g(x) = \exp(-18t^2),$	1.55	-55
Hamming	$g_{hm}(t) = 0.54 + 0.46.\cos(2\pi t),$	1.36	-43
Hanning	$g_{hn}(x) = \cos^2(\pi t),$	1.44	-32
Blackman	$g_b(x) = 0.42 + 0.5.\cos(2\pi t) + 0.08.\cos(4\pi t).$	1.68	-58

A_{max} is the maximum amplitude measured in decibels (dB) and $\Delta\omega$ the bandwidth. Hereafter their graphs (the rectangular window is not represented) (Fig. 5.24).
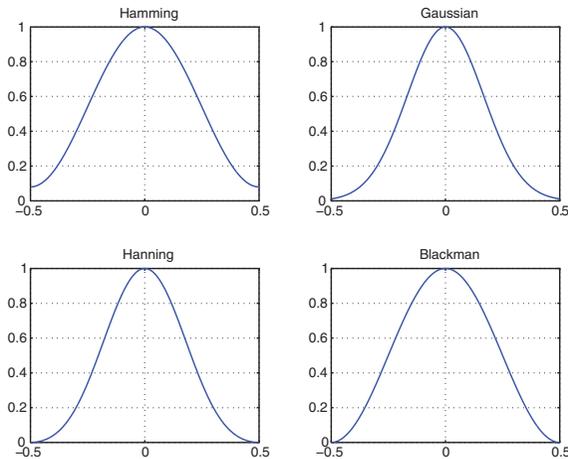
Fig. 5.24 Four windows on the support: $[-0.5, 0.5]$

5.11.2 Elements of Wiener Spectral Theory and Extensions

In this section we consider the time–frequency analysis for signals of dimension one and we present basic tools. These tools are particularly interesting for the specialists who work on random or noized (disturbed) signals with possible non-stationary characteristics. These subjects can also concern the statisticians who usually are rarely involved in the spectral time series analysis. The utilization of *wavelet bases* is increasingly widespread and many applications in particular in engineering have tried to exploit the *continuous Gabor transforms* or the *wavelet transforms* with *statistical goals*. It could be a question of *detecting*, *denoizing* or *reconstruction* of signals and more important *to carry out the spectral analysis of non-stationary signals*. It seems that the statisticians did not share yet the benefit of this type of work.

The *traditional spectral analysis of stationary processes (random or deterministic)* constitutes the hard core of the analysis. But, a very particular interest can be granted to t*he sampling of stationary or non-stationary continuous signals* within the framework of time–frequency representations. One of the important questions is to apprehend the time–frequency representations resulting from the *Gabor transformations* and wavelet transform. Their potential is vast but still requires an academic validation.

A synthetic work has been presented in 1998 by R. Carmona, W.L. Hwang and B. Torrésani (Carmona et al. 1998). This work tries to establish the link between the contribution of this type of analysis and their statistical interest. Moreover, the authors believe in the capacity of these methods to provide tools with advantages and varied talents for the spectral analysis of non-stationary signals. Unfortunately, the corresponding statistical theory is not completely developed yet. The authors *revisit* in their works the traditional elements of the spectral theory of stationary random processes in the light of these new tools.

5.11.2.1 Wiener's Deterministic (Classical) Spectral Theory

This theory of deterministic signals is a set of mathematical concepts which are presented in the sub-sections which follow. First, one considers continuous signals. Let us consider a function f, one can write the *autocovariance function $C_f(\tau)$ of f* for all τ, as follows:

$$C_f(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f(x+\tau)f(x)dx. \tag{5.90}$$

Its value at the origin is written:

$$C_f(0) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f(x)^2 dx. \tag{5.91}$$

This value at the origin is called the *power* of the signal f, which is finite. The function of autocovariance of a signal with *real values* makes it possible to write

also $C_f(\tau) = C_f(-\tau)$. If one confines oneself with a class of function f *with finite power* and possibly complex valued, it comes:

$$C_f(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} |f(x)^2| dx. \tag{5.92}$$

And if this value exists, then one can write the inner product of two functions:

$$\langle f, g\rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f(x)\overline{g(x)}dx. \tag{5.93}$$

And the *Schwartz inequality* is written:

$$|C_f(\tau)| \leq C_f(0). \tag{5.94}$$

The autocovariance is defined as non-negative:

$$\sum_{j,k=1}^{n} z_j\overline{z_k}C_f(x_j - x_k) \geq 0 \tag{5.95}$$

whatsoever the choice of the complex numbers z_j and elements x_j. This equation implies the existence of non-negative finite measurements v_f satisfying:[28]

$$C_f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\tau\omega} v_f d\omega. \tag{5.96}$$

The *measure* v_f is called the *spectral measure* of the signal f. *The spectral analysis of a signal f consists* in finding the *properties* of this measure. In general, when we speak of measurements, we are interested in the significance of non-negative finite measurement. The following stage relates to *the Lebesgue decomposition into three elements of this measurement,* which is written:

$$v_f = v_{f,pp} + v_{f,sc} + v_{f,ac}, \tag{5.97}$$

- $v_{f,pp}$ is a *pure point* measure (i.e. a weighted sum of Dirac point masses).
- $v_{f,sc}$ is a *singular*, i.e. concentrated on a set of Lebesgue measure zero, and *continuous* because $v_{f,sc}(\{\omega\}) = 0$ for any singleton ω measure. Thus $v_{f,sc}$ is a *continuous singular*.
- $v_{f,ac}$ is *absolutely continuous*, i.e. is given by a density with respect to the Lebesgue measure, i.e. $dv_{f,ac}(\omega) = v_{f,ac}(\omega)d\omega$, for some non-negative integrable functions v_f.

In a similar way, *the Lebesgue decomposition* of the spectral measure provides a decomposition of the autocovariance function, such that:

$$C_f = C_{f,pp} + C_{f,sc} + C_{f,ac} \tag{5.98}$$

---

[28] **Theorem (Bochner's theorem).** *Among the* continuous functions *on* $\mathbb{R}^n$, *the positive definite functions are those functions which are the* Fourier transforms *of* finite measures.

with

$$C_{f,..}(\tau) = \int_{-\infty}^{+\infty} e^{i\tau\omega} v_{f,...}(d\tau).$$

(5.99)

This *decomposition means that the signal can be understood as the sum of three orthogonal signals,* i.e. *uncorrelated*, with *pure spectra* given by the components of original spectral components in the Lebesgue decomposition (Carmona et al. 1998). The following function $\rho_f(\omega) = v_f(\{\omega\})$ is called the *spectral mass* function, and the density $v_{f,ac}(\omega)$ is called the *spectral density function*. The interpretation of the component $C_{f,sc}$ of the covariance is difficult. However, in practice this component relative to a measure $v_{f,sc}$ is rare and usually is equal to zero: $v_{f,sc} = 0$. The measure $v_{f,pp}$ is written:

$$v_{f,pp} = \sum_k \rho_k \delta_{\omega k},$$

(5.100)

considering the non-negative weights $\rho_k = \rho_f(\omega_k)$ and the possibly *unit point masses* $\delta_{\omega k}$, then the *pure point* part of the autocovariance function has the following form of a potentially complex trigonometrical polynomial when the number of frequencies ω_k is finite,

$$C_{f,pp}(\tau) = \frac{1}{2\pi} \sum_k \rho_k e^{i\omega_k \tau}$$

(5.101)

and it has generally the form of an almost periodic function. It will be noted that: $\frac{1}{2\pi}\sum_k \rho_k = C_{f,pp}(0) < +\infty$.

Case of the Spectrum of an Almost Periodic Function

If one considers an *almost periodic function* of the following form:

$$f(x) = \sum_{j=-\infty}^{+\infty} c_j e^{i\lambda_j x}$$

(5.102)

λ_j are distinct *real numbers* and the coefficients c_j are *potentially complex* numbers, satisfying:

$$\sum_{j=-\infty}^{+\infty} |c_j|^2 < \infty.$$

(5.103)

One can write:

$$C_f(\tau) = \sum_{j=-\infty}^{+\infty} |c_j|^2 e^{i\lambda_j \tau},$$

(5.104)

this formula expresses that the autocovariance function is a pure point, which means that $C_{f,sc}(\tau) = C_{f,ac}(\tau) = 0$, and the spectrum is concentrated on a set of λ_j. Thus, the *spectral mass function* $\rho_f(\omega)$ can be written:

$$\rho_f(\omega) = \begin{cases} |c_j|^2 & \text{if } \omega = \lambda_j, \\ 0 & \text{otherwise.} \end{cases}$$

(5.105)

It is highlighted that there is an information loss of phases. That means that the autocovariance function $C_f(\tau)$ is only a function of the modulus of the coefficient c_j (i.e. $|c_j|^2$) and the arguments of these *complex numbers* cannot be obtained from the knowledge of the spectrum. (Because "these phases values measure the displacements of *harmonic components* relative to a fixed time origin, and they cannot be retained by the spectral analysis because the latter is based on a definition of the autocovariance function which wipes out any natural notion of time origin" [Carmona et al. 1998, p. 44].)

The Wiener Theory is Extended to Random Signals

The Wiener theory is extended to random signals if the limit defining the auto-covariance function $C_f(\tau) = \lim\limits_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f(x+\tau)f(x)dx$ *exists and if the limit is non-random.* This is indeed the case for *ergodic* and stationary *processes*. But spectral theory of these processes has advantages which exceed the spectral Wiener theory of deterministic signals considered previously.

Indeed, it admits a representation of a signal *as a random superposition of complex exponentials* and this representation is very important in spectral analysis of these random signals (Carmona et al. 1998, pp. 74–79). But there is no decomposition adapted in the Wiener theory of deterministic signals.

5.11.2.2 Extensions: Deterministic Spectral Theory for Time Series

Note that this type of analysis could be illustrated by taking for example time series resulting from *stock exchange* or *financial markets*.[29]

Correlogram and Memory of a Series

In this section, let us consider the *finite time series*

$$f = \{f_0, f_1, \ldots, f_{N-1}\} \tag{5.106}$$

of real numbers and *it is discussed, in different ways, of quantifying the statistical correlations between the successive values of* the f_j. The most used *dependence measure* is the *sample autocovariance function*, defined as follows:

$$c_f(j) = \frac{1}{N} \sum_{K=0}^{N-1-|j|} \left(f_k - \bar{f}\right)\left(f_{k+|j|} - \bar{f}\right), \tag{a}$$

---

[29] It is in particular about the evolution of the logarithms of the prices of a *"futures"* contracts of *Japanese yen* that R. Carmona, W.L. Hwang and B. Torrésani develop their argument.

where $\bar{f}$ is the average of the sample which is simply written:

$$\bar{f} = \frac{1}{N} \sum_{j=0}^{N-1} f_j. \tag{5.107}$$

The writing (a) of *the autocovariance function above* is usually preferred to the following writing (b):

$$\tilde{c}_f(j) = \frac{1}{N - |j|} \sum_{K=0}^{N-1-|j|} (f_k - \bar{f})(f_{k+|j|} - \bar{f}) = \frac{N}{N - |j|} c_f(j). \tag{b}$$

j in the expression above is considered as a *lag* and often corresponds to *non-negative values*. From this autocovariance function we are led to the notion of *sample autocorrelation function* of the signal f, which is written:

$$\rho_f(j) = \frac{c_f(j)}{c_f(0)} = \frac{\sum_{j'=0}^{N-1-j} (f_{j'} - \bar{f})(f_{j+j'} - \bar{f})}{\sum_{j'=0}^{N-1-j} (f_{j'} - \bar{f})^2}, \quad j = 0,1,\dots,N-1. \tag{c}$$

The *graph* of the *autocorrelation function* $\rho_f(j)$ of j is called a *correlogram*. (The *maximum lag corresponds to* $N-1$, even if we use, in practice, values which usually are much smaller.) *Commonly, it is said that a series has a "long memory" when the correlogram decreases slowly, and it said that a series has a short memory when the correlogram decrease quickly.* However when the series is "detrended" by differentiation for example, the phenomenon of long memory disappears and different new correlograms take place. The *correlograms of stock exchange indexes* on the long term generally highlight these *phenomena of long memories* where a *slow decrease of correlograms* is observed.

5.11.2.3 Periodogram: Interpretations

The objective of spectral theory is to represent a signal as *a sum of trigonometric functions with specific phases and amplitudes*. All the phases involved in the representation are called the *spectrum* of the series and the *size of amplitudes* is summarized in what is usually called the *periodogram*. If one quickly points out the expression of the *discrete Fourier transform* (DFT) one writes

$$\hat{f}_j = \sum_{k=0}^{N-1} f_j e^{-2i\pi j \omega_k}, \quad k = 0,1,\dots,N-1, \tag{5.108}$$

ω_k is the *natural frequency* $\omega_k = k/N$. Usually in the Fourier transform it is ξ which is equivalent to the relation: $\xi = 2\pi\omega$. It is also pointed out that the original signal

can be *reconstituted* from *an inverse Fourier transform:*

$$f_j = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}_j e^{2i\pi j \omega_k}, \quad k = 0, 1, \dots, N-1,$$ (5.109)

the inversion formula above expresses that any finite series f *can be written as a linear combination of complex exponentials.* From this decomposition, in particular when the original series is real, one can write that: $a_k = \frac{1}{N} \mathscr{R} \hat{f}_k$ and $b_k = \frac{1}{N} \mathscr{I} \hat{f}_k$. Moreover, we have:[30]

$$f_j - \bar{f} = \sum_{k=2}^{N-1} e_j(k),$$ (5.110)

the *trigonometric functions* e_j are defined by

$$e_j(k) = r_k \cos(2\pi j \omega_k - \varphi_k),$$ (5.111)

and $r_k = \sqrt{a_k^2 + b_k^2}$, $\varphi_k = \tan^{-1}(b_k/a_k)$. Thus, $k = 1, \dots, N/2$, and we have

$$a_k = a_{N-k} \quad \text{and} \quad b_k = -b_{N-k},$$ (5.112)

and then,

$$r_k^2 = r_{N-k}^2 \quad \text{and} \quad e_j(k) = e_j(N-k).$$ (5.113)

What precedes implies the following trigonometrical representation:

$$f_j - \bar{f} = \begin{cases} 2\sum_{k=1}^{N/2} r_k \cos(2\pi j \omega_k) & \text{if } N \text{ odd,} \\ 2\sum_{k=1}^{(N/2)-1} r_k \cos(2\pi j \omega_k) + r_{N/2} \cos(2\pi j \omega_{N/2}) & \text{if } N \text{ even.} \end{cases}$$

The purpose of the equation above is to "*untangle the dependence* among the complex exponentials appearing in the *inverse Fourier transform*". Such a *dependence is not present any more in the equation above because the trigonometric functions* $e_j(k) = r_k \cos(2\pi j \omega_k)$ *appearing in the function above are mutually orthogonal.* As expressed previously, any series f can be written as a sum of cosine functions of frequencies $\omega_1 = 1/N, \omega_2 = 1/N, \dots, \omega_{N/2} = (N/2)/N$. (Remember that $r_k^2 = \frac{1}{N^2} \left| \sum_{j=0}^{N-1} f_j e^{-2i\pi j \omega_k} \right|^2$, with $k = 0, 1, \dots, N/2$ and $\omega_k = k/N$.) The graph of $N_{r_k^2}$ for ω_k is called the *periodogram* of the series f, and the function $v_f(\omega)$ defined for $\omega \in [0, 1]$, by:

$$v_f(\omega) = \frac{1}{N} \left| \sum_{j=0}^{N-1} f_j e^{-2i\pi j \omega} \right|^2 \quad : \textit{Sample spectral density}$$ (d)

considering $\omega \in [0, 1/2]$ and $v_f(\omega) = v_f(1 - \omega)$, considering $\omega \in [1/2, 1]$ this expression is called the "sample *spectral density*" of f. Often some values of r_k are

---

[30] Note that $a_1 = \bar{f}$ and $b_1 = 0$.

very large compared with the others, it is one of the reasons *which creates difficulties to analyze the information contained in the periodogram*. However to represent the *logarithm* of the values r_k^2, instead of the values themselves, helps to see more elements relating to the frequencies. Furthermore, we have:

$$\frac{1}{N} \sum_{j=1}^{N-1} \frac{N r_k^2}{\sigma^2} = 1. \tag{5.114}$$

If σ^2 is expressed as a variance

$$\sigma^2 = \frac{1}{N} \sum_{j=0}^{N-1} (f_j - \bar{f})^2, \tag{5.115}$$

then in practice we have more graphic information about the contents of the set of frequencies, it is represented graphically as follows:

$$\left[Log \left(\frac{N r_k^2}{\sigma^2} \right) \ against \ \omega_k \right]. \tag{5.116}$$

The remarks which follow are useful for the interpretation of periodograms, which are:

- *"Smooth" when the amplitudes of cosine functions of low frequencies are large compared with the other frequencies.*
- *"Erratic" when the amplitudes of cosine functions of high frequencies are large compared with the other frequencies.*
- *With a "peak" at the frequency ω when the signal is a cosine or a sine function of period $1/\omega$.*
- *With a "peak" at the frequency ω and "peaks" at multiple frequencies of ω (harmonic) when the signal is a periodic function of period $1/\omega$ without being a sine or cosine function.*

Comments on the Comparison Between Correlogram and Periodogram

Usually a debate is open between the two possible definitions $[(a)$ or $(b)]$ seen previously of the autocovariance function of finite time series. The two definitions differ only when the size N of the data is large, however one can choose. The second definition $\tilde{c}_f$ (corresponding to the equation (b)) is adapted to the point of view of the statistical analysis in particular when samples f_j of random variables form *a stationary stochastic process*. This second definition provides *an unbiased estimation of the true autocovariance function*. But it is generally (a), the first definition c_f which is selected as the autocovariance function. Usually one chooses this definition because (except for the particular cases for which all the f_j are equal to each other): the sequence $c_f = \{c_f(j)\}_j$ defined by $c_f(j) = \frac{1}{N} \sum_{K=0}^{N-1-|j|} (f_k - \bar{f}) (f_{k+|j|} - \bar{f})$ when

$|j| < N$ and 0 otherwise the sequence is *defined non-negative*. Which implies due to the *Bochner's theorem*[31] the existence of a finite non-negative measure v_f on [0,1] whose Fourier transform is the sequence c_f (such a measure is absolutely continuous). One writes this density $v_f(\omega)$

$$v_f(\omega) = \sum_{j=-(N-1)}^{N-1} c_f(j)e^{2i\pi j\omega}, \quad \omega \in [0,1]. \tag{5.117}$$

Such a $v_f(\omega)$ is called the *spectral density* due to the fact that it is equal to the quantity defined previously having the same name. One can use simple trigonometrical handling to show that the function $v_f(\omega)$ defined above satisfied:

$$v_f(\omega) = \frac{1}{N}\left|\sum_{j=0}^{N-1} f_j e^{-2i\pi j\omega}\right|^2. \tag{5.118}$$

That means that the *autocovariance function c_f and the spectral density $v_f(\omega)$ form a "Fourier pair", because they are Fourier transforms the one each other. This correspondence is of a quite essential interest.*

5.11.2.4 Remark About the Non-Parametric Spectral Estimation: Case of the Non-Stationary Process Locally Stationary

If a *stochastic process is stationary*, then its *autocovariance* defines what is called a *convolution operator* which is *diagonalized* by *Fourier expansions*. It is one of the starting points of all the non-parametric spectral estimation methods. In a number of practical situations, the function of autocovariance does not allow to see a convolution operator and the problem of the spectral estimation cannot be solved by the standard Fourier analysis. The stationarity may be broken in various manners. One can give two examples of time–frequency or time-scale representations to illustrate the subject:

- The first example is *the case of "locally stationary process"*, i.e. processes for which there is *a local spectrum varying slowly*. In such case the time–frequency representations provide efficient tools for these local spectral estimation subjects.
- The second example is the case of *self-similar process* such as *fractional Brownian motions* which can have *stationary increases*, and after the *wavelets transform the non-stationarities disappear.*

---

[31] **Theorem (Bochner theorem).** *Among the continuous functions on $\mathbb{R}^n$, the* positive definite functions *are those functions which are the Fourier transform of finite measures.*

5.11.2.5 Spectral Theory of Stationary Random Processes

Stationary Processes

Consider a signal made up of real numbers such as: $f = \{f_0, f_1, \ldots, f_{N-1}\}$, this signal is *random* if the numbers f_j can be regarded as results of a finite sequence of random variables. An example is provided through observations of *a deterministic signal in the presence of additional noise:* signal + noise. Consequently, the observations can be written in the following way:

$$f_j = f_j^* + \varepsilon_j \qquad (5.119)$$

where $f_j^* = \{f_0^*, f_1^*, \ldots, f_{N-1}^*\}$ is of "deterministic origin" and $\varepsilon = \{\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{N-1}\}$ is the "noise perturbation". The field of the random series searches to develop tools to carry out forecasts from observed time series. The usual methods use models made by series generators and the majority of these models is based on the notion of *temporal homogeneity* of signals. This notion is regarded as relative to the *stationnarity*. The following lines provide some elements of the theory of stationary processes, but it is important to say *that generally the most interesting signals are non-stationary* and for these signals the theory is inoperative. (It will be also noticed that beyond the *non-stationarity*, the majority of *complex signals* are rather *non-Gaussian.*)

The most adapted way to analyze the finite signals is to regard the set of N samples of f_j as a part of a *doubly infinite sequence* $\ldots f_{-1}, f_0, f_1, \ldots$ which can be seen as a set of regular samples which comes from a continuous signal but discretized. In fact, one supposes that there is a function f of a continuous variable $x \in \mathbb{R}$, such that:

$$f_j = f(x_j), \quad \text{with } j = 0, 1, \ldots, N-1 \qquad (5.120)$$

the x_j corresponds to the time at which the measurements are made. One considers the regular samples provided by the observations at times $x_j = x_0 + j\Delta x$ for an fixed sampling interval Δx which allows to sample the signal. One can define the concept of stationnarity in the following way:

Stationarity in the Strict Sense

A random signal $f = \{f(x); x \in \mathbb{R}\}$ is *stationary in the strict sense* if for any choice x and $x_1 < \cdots < x_k$ of real numbers, the random vectors:

$$[f(x_1), \ldots, f(x_k)] \quad \text{and} \quad [f(x_1 + x), \ldots, f(x_k + x)] \qquad (5.121)$$

have the same distribution in $\mathbb{R}^k$ (or in $\mathbb{C}^k$ if it is a complex signal). Any statistic obtained by translation or shift of such random processes is *invariant*. One can say that *the first moment $m(x) = E\{f(x)\}$ is constant and independent of the variable x.*

Considering the case of a finite signal, the *stationnarity* can be defined without the notion of continuous variables functions. Thus, it is written that *a random signal*

$f = \{f_0, f_1, \ldots, f_{N-1}\}$ is known as *stationary* if for any choice of j and $j_1 < \cdots < j_k$ of integers in $\{0, 1, \ldots, N-1\}$, the following random vectors

$$\{f_{j_1}, \ldots, f_{j_k}\} \quad \text{and} \quad \{f_{j_1+j}, \ldots, f_{j_k+j}\} \tag{5.122}$$

take advantage of the same distribution in $\mathbb{R}^k$ or in $\mathbb{C}^k$ for the complex signals.[32] If one considers the case of a *signal* to which one *adds a noise* as previously: $f_j = f_j^* + \varepsilon_j$, the stationarity hypothesis is satisfied for the *noise component*, i.e. for $\varepsilon = \{\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{N-1}\}$, but if $f^* = \{f_0^*, f_1^*, \ldots, f_{N-1}^*\}$ is not constant, the stationarity cannot be proved for the *other component* $f = \{f_0, f_1, \ldots, f_{N-1}\}$.

Stationarity in the Wide Sense

There exists *a weaker definition of the stationarity*. Consider a *random process* $f = \{f(x); x \in \mathbb{R}\}$ *of order 2* (for a continuous variable x), that means that all the random variables $f(x)$ are *square integrables*. This random process *is known as stationary in the wide sense* if

$$m_f(x) = E\{f(x)\} \tag{5.123}$$

is independent of x, that means $m_f(x) = m$ for constant m, and if its autocovariance

$$C_f(y, x) = E\left\{(f(y) - m)\overline{(f(x) - m)}\right\} \tag{5.124}$$

is a function of the difference $(y - x)$, i.e. for a one-variable non-negative function noted C_f, if:

$$C_f(y, x) = C_f(y - x). \tag{5.125}$$

By means of the *Bochner theorem* it is possible to say that there exists a non-negative finite measure ν_f on $\mathbb{R}$ which satisfies the definition $C_f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\tau\omega} \nu_f d\omega$. This measure ν_f is called the *spectral measure* of the stationary process $f = \{f(x); x \in \mathbb{R}\}$. The expression $C_f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\tau\omega} \nu_f d\omega$ allows to say that *the covariance function C_f is the Fourier transform of the spectral measure ν_f*.

Usually, the covariance function C_f is integrable on $\mathbb{R}$, then it is possible to introduce the spectral measure without the Bochner theorem. Thus, the following *Fourier transformation* (under the integrability condition) is defined as a function on $\mathbb{R}$:

$$\vartheta_f(\xi) = \int C_f(x) e^{-ix\xi} d\xi. \tag{5.126}$$

Because of the *autocovariance function C_f* is defined non-negative, the function ν_f is non-negative. This can be called the *spectral function* or *spectral density*.[33] The

---

[32] The addition of integers is understood as *mod N*.

[33] It is the density of the spectral measure which is possible by means of the Bochner theorem because of $\nu_f(d\xi) = \nu_f(\xi)d\xi$, and which allows the utilization of the same notation to express the *spectral measure* and the *spectral function*.

autocorrelation function ρ_f defined as follows

$$\rho_f(y,x) = \frac{C_f}{(C_f(y,y))^{1/2}(C_f(x,x))^{1/2}} \tag{5.127}$$

is also written as a function of the difference $(y-x)$ when the process is stationary, and in such a case there is $\rho_f(y,x) = \rho_f(y-x)$ with:

$$\rho_f(y-x) = \frac{C_f(y-x)}{C_f(0)}. \tag{5.128}$$

In a similar way, it is possible to provide a definition of *the stationarity in the wide sense for the finite random signals*. In all these approaches it is always assumed that the average of a stationary signal in the wide sense is zero.

Comparison with the Wiener Spectral Theory

It is interesting to compare the elements of the spectral theory of this type of process with the Wiener spectral theory, presented previously. *Ergodic theorem*[34] explains that the limit:

$$C_f(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f(x+\tau)f(x)dx \tag{5.129}$$

exists, if f is a *stationary process*. Furthermore, if the process f is *Ergodic, i.e. if the shift-invariant functions of signal are constant,* the limit above results from an almost sure convergence, and is deterministic, and is equivalent to the autocovariance $C_f(\tau)$ seen previously defined when $y = x + \tau$, as follows:

$$C_f(y,x) = E\left\{(f(y) - m)\,\overline{(f(x) - m)}\right\} \tag{5.130}$$

and

$$C_f(y,x) = C_f(y-x). \tag{5.131}$$

That means that the "Spectral theory of stationary processes contains the Wiener theory when the analyzed processes are Ergodic". ("It is known that the estimates calculated from $\bar{f} = \frac{1}{N}\sum_{j=0}^{N-1} f_j$ or from the sample autocovariance function $c_f(j) = \frac{1}{N}\sum_{K=0}^{N-1-|j|} (f_k - \bar{f})(f_{k+|j|} - \bar{f})$ or from the sample autocorrelation function $\rho_f(j) = \frac{c_f(j)}{c_f(0)} = \frac{\sum_{j'=0}^{N-1-j}(f_{j'}-\bar{f})(f_{j+j'}-\bar{f})}{\sum_{j'=0}^{N-1-j}(f_{j'}-\bar{f})^2}$ are estimations of their theoretical counterparts: m, C_f, and ρ_f".)

---

[34] Regarded as a generalization of the law of large numbers.

5.12 The Construction of Orthonormal Bases and Riesz Bases

5.12.1 Signal of Finite Energy

The time series or signals are often regarded as vectors, and can also be regarded as measurable functions. Then, one poses the integration Lebesgue theorem, which explains that *a function f is integrable*, if *f* is of *finite energy:*

$$f_{-\infty}^{+\infty} |f(t)| \, dt < +\infty. \tag{5.132}$$

The space of integrable functions is noted $\mathbf{L}^1(\mathbb{R})$ if $f_{-\infty}^{+\infty} |f_1(t) - f_2(t)| \, dt = 0$. This means that f_1 and f_2 can differ only on a set of points of measure zero. Roughly, it is said that they are equal almost everywhere. Moreover, the time series or signals which are the subject of transformations in the time–frequency planes are offered to the handling of scientists who practise the signal processing and *define a metric* and exploit the properties of *vector spaces*. The concepts of distance, norm, convergence, integration, orthogonality, projection, basis, are largely used and enriched.

5.12.2 Reminders: Norms and Banach Spaces

Banach space: A *complete normed vector space*, i.e. a normed space in which any Cauchy sequence converges, is called *Banach space*. In order to define a distance we work inside a vector space $\mathbf{H}$ which admits a norm. The properties of a norm are the following:

$$\forall f \in \mathbf{H}, \|f\| \geq 0 \ \text{ and } \ \|f\| = 0 \Leftrightarrow f = 0, \tag{5.133}$$
$$\forall \lambda \in \mathbf{C} \, \|\lambda f\| = |\lambda| \, \|f\|, \tag{5.134}$$
$$\forall f, g \in \mathbf{H}, \|f + g\| \leq \|f\| + \|g\|. \tag{5.135}$$

With such a norm, the convergence of a family (or sequence) of positive functions $\{f_n\}_{n \in N}$ towards f in $\mathbf{H}$ means that:

$$\lim_{n \to +\infty} f_n = f \Leftrightarrow \lim_{n \to +\infty} \|f_n - f\| = 0. \tag{5.136}$$

One can impose a complete state of properties using the Cauchy sequence. A family or sequence $\{f_n\}_{n \in N}$ is a Cauchy sequence if for all $\varepsilon > 0$, if n and p are large enough, then $\|f_n - f_p\| < \varepsilon$. *The vector space* $\mathbf{H}$ *is said complete if any Cauchy sequence in* $\mathbf{H}$ *converges towards an element of* $\mathbf{H}$. One can evoke the particular case of a space $\mathbf{L}^P(\mathbb{R})$. The space $\mathbf{L}^P(\mathbb{R})$ is composed of measurable functions on $\mathbb{R}$ for which:

$$\|f\|_p = \left(\int_{-\infty}^{+\infty} |f(t)|^p \, dt \right)^{1/p} < +\infty. \tag{5.137}$$

$\mathbf{L}^P$ is a Banach space and the integral above determines a norm.

5.12.3 Reminders: Inner Products and Hilbert Spaces

A Hilbert space is a space on $\mathbb{R}$ or $\mathbb{C}$, provided with a *scalar product* whose *associated normed space is complete*. The elements of these spaces historically were functions coming from the *formalization of oscillatory phenomena* and from the *calculus of variations* where the searched solutions (generally integrals) appear as sums of a series of functions, often trigonometric, which one approaches by orthogonal polynomials for a scalar product. *A complete prehilbert space* (for the norm associated with the scalar product) is called *Hilbert space*.

The *Hilbert* space is a *Banach* space provided with an inner product. The inner product of two vectors $\langle f, g \rangle$ is linear and respects:

$$\forall \lambda_1, \lambda_2 \in \mathbf{C}, \langle \lambda_1 f_1 + \lambda_2 f_2, g \rangle = \lambda_1 \langle f_1, g \rangle + \lambda_2 \langle f_2, g \rangle. \tag{5.138}$$

And there is an *Hermitian symmetry:*[35]

$$\langle f, g \rangle = \langle g, f \rangle^* \tag{5.139}$$

Moreover, $\langle f, f \rangle \geq 0$ and $\langle f, f \rangle = 0 \Leftrightarrow f = 0$. It can be also shown that $\|f\| = \langle f, f \rangle^{1/2}$ is a norm. The inequality of Cauchy–Schwarz can be also shown:

$$|\langle f, g \rangle| \leq \|f\| \, \|g\|, \tag{5.140}$$

and if f and g are *linearly independent*, there is the following equality:

$$|\langle f, g \rangle| = \|f\| \, \|g\|. \tag{5.141}$$

5.12.4 Orthonormal Basis

Orthogonal vectors: A family of vector noted $\{e_n\}_{n \in N}$ of a Hilbert space is *orthogonal* if for $n \neq p$:

$$\langle e_n, e_p \rangle = 0. \tag{5.142}$$

(It will be noted that from a statistical point of view, *two orthogonal vectors are uncorrelated*.)

Orthogonal basis: If for $f \in \mathbf{H}$, there exists a sequence $\lambda[n]$ such that:[36]

$$\lim_{N \to +\infty} \left\| f - \sum_{n=0}^{N} \lambda[n] e_n \right\| = 0, \tag{5.143}$$

---

[35] *Property of Hermitian symmetry:* For $f(t) \in R$, on the Fourier transform we obtain: $\widehat{f}(-\omega) = \widehat{f}^*(\omega)$.

Property of complex conjugates: For $f^*(t)$, on the Fourier transform we obtain: $\widehat{f}^*(-\omega)$.

[36] $\lambda[n]$: Corresponds to a discrete sequence.

then $\{e_n\}_{n \in N}$ is an *orthogonal basis* of $\mathbf{H}$. The orthogonality implies that $(\lambda[n]$ is a sequence):

$$\lambda[n] = \frac{\langle f, e_n \rangle}{\|e_n\|^2},\qquad(5.144)$$

and one can write:

$$f = \sum_{n=0}^{+\infty} \frac{\langle f, e_n \rangle}{\|e_n\|^2} e_n.\qquad(5.145)$$

The basis is orthonormal if $\|e_n\| = 1$ for any $n \in N$. If one calculates the inner product of $g \in \mathbf{H}$ with the *Parseval equation* for orthonormal bases:

$$\langle f, g \rangle = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \langle f, e_n \rangle^*.\qquad(5.146)$$

When $g = f$, one obtains an *Energy Conservation* derived from the *Plancherel formula*:

$$\|f\|^2 = \sum_{n=0}^{+\infty} |\langle f, e_n \rangle|^2.\qquad(5.147)$$

It is possible to construct *orthonormal bases* by means of *local cosine (or sine) functions and wavelets* or *wavelet packets*.

5.12.5 Riesz Basis, Dual Family and Biorthogonality

A Riesz basis: A family of vectors denoted $\{e_n\}_{n \in N}$ is a Riesz basis of the Hilbert space $\mathbf{H}$, if it is *linearly independent* and if there exists $A > 0$ and $B > 0$ such that for any $f \in \mathbf{H}$. It is possible to find a sequence $\lambda[n]$ with $f = \sum_{n=0}^{+\infty} \lambda[n]e_n$, which satisfies

$$\frac{1}{B}\|f\|^2 \leq \sum_n |\lambda[n]|^2 \leq \frac{1}{A}\|f\|^2.\qquad(5.148)$$

The Riesz theorem proves that there exists $\tilde{e}_n$ such that $\lambda[n] = \langle f, \tilde{e}_n \rangle$, and furthermore the preceding inequations $\frac{1}{B}\|f\|^2 \leq \sum_n |\lambda[n]|^2 \leq \frac{1}{A}\|f\|^2$ implies:

$$\frac{1}{B}\|f\|^2 \leq \sum_n |\langle f, \tilde{e}_n \rangle|^2 \leq \frac{1}{A}\|f\|^2.\qquad(5.149)$$

Moreover, it is also possible to write that for any $f \in \mathbf{H}$,

$$A\|f\|^2 \leq \sum_n |\langle f, e_n \rangle|^2 \leq B\|f\|^2,\qquad(5.150)$$

And

$$f = \sum_{n=0}^{+\infty} \langle f, \tilde{e}_n \rangle e_n = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \tilde{e}_n,\qquad(5.151)$$

$\{\widetilde{e}_n\}_{n \in N}$ is a linearly independent *dual family* and is also a *Riesz basis*. If for example $f = e_s$, we have $e_s = \sum_{n=0}^{+\infty} \langle e_s, \widetilde{e}_n \rangle e_n$. The linear independence of $\{e_n\}_{n \in N}$ implies a biorthogonality of dual bases, which are called biorthogonal bases:

$$\langle e_n, \widetilde{e}_s \rangle = \delta[n-s]. \tag{5.152}$$

5.12.6 Orthogonal Projection

Let **V** be a subspace of the space **H**. A projector P_V on **V** is a linear operator which satisfies:

$$\forall f \in \mathbf{H}, P_V f \in \mathbf{V} \quad \text{and} \quad \forall f \in \mathbf{V}, P_V f = f. \tag{5.153}$$

The projector P_V is orthogonal if:

$$\forall f \in \mathbf{H}, \forall g \in \mathbf{V}, \quad \langle f - P_V f, g \rangle = 0. \tag{5.154}$$

The following properties of projectors can be proposed:

- If $\{e_n\}_{n \in N}$ is an orthogonal basis of **V**, then we have:

$$P_V f = \sum_{n=0}^{+\infty} \frac{\langle f, e_n \rangle}{\|e_n\|^2} e_n. \tag{5.155}$$

- If $\{e_n\}_{n \in N}$ is a Riesz basis of **V** and $\{\widetilde{e}_n\}_{n \in N}$ is the orthogonal basis then we have:

$$P_V f = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \widetilde{e}_n = \sum_{n=0}^{+\infty} \langle f, \widetilde{e}_n \rangle e_n. \tag{5.156}$$

5.12.7 The Construction of Orthonormal Basis and Calculation of the "Detail" Coefficient on Dyadic Scale

Subsequently, it will be developed the notions of orthonormal basis, Fourier basis and wavelet basis, which are fundamental in the time–frequency analysis, but before to familiarize with this notion, one shows quickly how to build a basis from a dyadic scale. Let $\psi(t)$ be a wavelet whose dilations and translations generate an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$:

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j}\right) \right\}_{(j,n) \in \mathbb{Z}^2}. \tag{5.157}$$

All the *signals of finite energy* can be *decomposed* on this type of *wavelet basis* $\{\psi_{j,n}(t)\}_{(j,n)\in\mathbb{Z}^2}$, which is written:

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \, \psi_{j,n}. \tag{5.158}$$

We know that $\psi(t)$ has a zero-integral or a zero-average and we can write:

$$d_j(t) = \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \, \psi_{j,n}(t), \tag{5.159}$$

which corresponds to the variations of "detail" at a dyadic scale 2^j. The approximation of f is constructed from the addition of these variations of detail on each scale, whose sum allows to reconstruct the signal f. Note that if the signal has structures rather *smooth*, one can then approach the signal with an approximation from which the "details" at fine scale have been removed, thus the writing of $f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \, \psi_{j,n}$ becomes:

$$f_J(t) = \sum_{j=J}^{+\infty} d_j(t). \tag{5.160}$$

5.13 Concept of Frames

Before quickly presenting the concept of frames, it seemed fundamental to re-introduce two elements of the Fourier theory, without which (even reduced to their rudiments) the time–frequency analysis and the wavelet analysis seem difficult or even impossible to understand. These elements are the Fourier transform in $\mathbf{L}^2$ and the *Parseval and Plancherel formulas* that one has for the latter already mentioned.

5.13.1 The Fourier Transform in $L^2(\mathbb{R})$

Briefly, remember that the following spaces correspond to the respective functions or associated signals:

- $\mathbf{L}^2(\mathbb{R})$: Finite energy functions: $\int |f(t)|^2 \, dt < +\infty$. And consequently is the space of integrable square functions.
- $\mathbf{L}^P(\mathbb{R})$: Functions such that $\int |f(t)|^p \, dt < +\infty$.
- $\mathbf{l}^2(\mathbb{Z})$: Discrete finite energy signals: $\sum_{n=-\infty}^{+\infty} |f(t)|^2 < +\infty$.
- $\mathbf{l}^P(\mathbb{Z})$: Discrete signals such that: $\sum_{n=-\infty}^{+\infty} |f(t)|^p < +\infty$.

5.13.1.1 Convergence Limit of Fourier Transform of Function in $\mathbf{L}^1(\mathbb{R})$

As seen several times, the Fourier transform of a function $f(t)$ is carried out by the integration of the function f and trigonometric functions $e^{-i\omega t}$. The latter are used to decompose and rewrite the studied function into a sum of sine curves, whose transformation in the Fourier space provides Fourier coefficients, which characterize the studied function in the space of the frequencies (ω). In other words, *the number of sine or cosine functions and their respective frequencies define f*. The Fourier transform is written:

$$\widehat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}\,dt. \tag{5.161}$$

If the function f is integrable (i.e. if f is of finite energy $f_{-\infty}^{+\infty}|f(t)|\,dt < +\infty$), i.e. if $f \in \mathbf{L}^1(\mathbb{R})$, then there exists a bound for the Fourier transform, i.e. there is convergence:

$$\left|\widehat{f}(\omega)\right| \le \int_{-\infty}^{+\infty} |f(t)|\,dt < +\infty. \tag{5.162}$$

Thus, the Fourier transform through the absolute value $\left|\widehat{f}(\omega)\right|$ is bounded by the finite integral of the absolute value of the function f:

$$\int_{-\infty}^{+\infty} |f(t)|\,dt. \tag{5.163}$$

5.13.1.2 Fourier Transform in the Space $\mathbf{L}^2(\mathbb{R})$ of Square Integrable Functions

Certain functions are not integrable, because they are not continuous. This case is not rare and can also be transposed to signals themselves; *discontinuities* or *singularities* can create problems concerning the integrability of a function and by extension of a signal. *A non-integrable function f is not continuous, however its square is integrable*. One can for example evoke the indicator function which is non-continuous.[37] This is on the base of this observation that the Fourier transform of integrable functions (and thus continuous) is extended to the Fourier transform of square integrable functions, in order to compensate the difficulties related to problems of discontinuities. The extension of the Fourier transform is done in the space $\mathbf{L}^2(\mathbb{R})$ of finite energy functions $\int_{-\infty}^{+\infty}|f(t)|\,dt < +\infty$ which is also the space of square integrable functions. Consequently, we are located in the *Hilbert* space, where we can be endowed with an *inner product* and a *norm*. It is pointed out that the inner product of f and g belonging to $\mathbf{L}^2(\mathbb{R})$ is written

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(t)g^*(t)\,dt, \tag{5.164}$$

---

[37] *Example*. The Fourier transform of indicator function: $f = \mathbf{1}_{[-1,1]}$ is written as a multiple of cardinal sine: $\widehat{f}(\omega) = \int_{-1}^{+1} e^{-i\omega t}\,dt = (2\sin(\omega))/\omega$.

and that the norm in the Hilbert space $\mathbf{L}^2(\mathbb{R})$ is:

$$\|f\|^2 = \langle f, f \rangle = \int_{-\infty}^{+\infty} |f(t)|^2 \, dt. \qquad (5.165)$$

5.13.1.3 Parseval and Plancherel Formulas: Norm and Inner Product are Conserved by the Fourier Transform up to a Factor of 2π

Theorem 5.4 (Parseval and Plancherel transfer formulas). *If f and g belong to* $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$, *then*

$$\int_{-\infty}^{+\infty} f(t)g(t)dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \widehat{f}(\omega)\widehat{g}^*(\omega)d\omega, \qquad (5.166)$$

which is the Parseval formula. When $f = g$, *it follows the well-known Plancherel formula:*

$$\int_{-\infty}^{+\infty} |f(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \widehat{f}(\omega) \right|^2 d\omega. \qquad (5.167)$$

5.13.2 Frames

The frames provide a stable and possibly redundant representation of a signal. *The frames are a generalization of the basic notion of a vector space.* A frame is a family of vectors which allows to represent any signal ($\mathbf{L}^2$) by its scalar products with these vectors. The frames provide a discrete and redundant representation of the signal.

Definition 5.2 (Frame). A sequence[38] $\{\phi_n\}_{n \in \Gamma}$ (with the index n belonging to a finite or infinite set noted: Γ) of vectors of a Hilbert space $\mathbf{H}$ is a frame of $\mathbf{H}$ if there exist two constants $A > 0$ and $B > 0$ such that, for any f of $\mathbf{H}$,

$$A\|f\|^2 \leq \sum_{n \in \Gamma} |\langle f, \phi_n \rangle|^2 \leq B\|f\|^2, \qquad (5.168)$$

If $A = B$, it is said that the frame is tight.[39] (A Riesz basis is a frame of independent vectors.)

If the vectors of frame are independent, then $A \leq 1 \leq B$.[40] The frame is an orthonormal basis if and only if $A = B = 1$. If $A > 1$, the frame is *redundant*.

---

[38] Or family of vectors.

[39] *Example.* We are located in the plane and we consider a family of three unit vectors resulting the ones from the others by rotation of the third of a turn. It forms a tight (adjusted) frame of the plane, with $A = B = 3/2$.

[40] It is taken vectors of frame normalized to 1.

5.13.2.1 Windowed Fourier Frames and Fourier Atoms

It is pointed out that the windowed Fourier transform is written

$$\langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t)g(t-u)e^{i\xi(t-u)}dt. \tag{5.169}$$

The atom used is a sinusoid multiplied by a window g. The family of vectors of the analysis is obtained by translation and modulation of the window:

$$g_{u,\xi}(t) = g(t-u)e^{i\xi(t-u)}. \tag{5.170}$$

This function is *centered* for the frequencies at ξ and symmetrical in relation to u. *The temporal standard deviation is constant. The standard deviation in frequency is constant.* The family is thus obtained by translation in time and frequency of a single window. The Heisenberg boxes of windowed Fourier atoms have dimensions independent of the center in time and frequency. A frame is obtained by covering the time–frequency plane by discrete boxes. The grid used is uniform rectangular
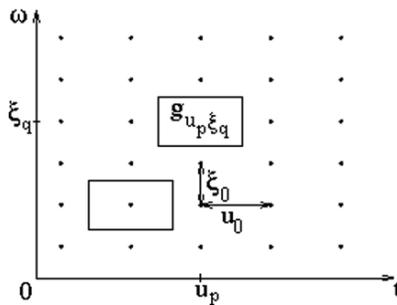
To obtain a frame, the following condition on the paving (tiling) of the zone proves to be necessary:

$$\frac{2\pi}{u_0\xi_0} \geq 1. \tag{5.171}$$

5.13.2.2 Frames of Wavelets and Atoms of Wavelets

One points out the writing of the wavelet transform

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right)dt, \tag{5.172}$$

where the atom or the wavelet ψ is a function of zero average, centered at the neighborhood of 0 and of finite energy. The family of vectors is obtained by translation and dilation of the atom:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{5.173}$$

This function is centered at the neighborhood of u, like the windowed Fourier atom. If the center frequency of ψ is indicated by η, the center frequency of the dilated function is in η/s. The standard deviation in time is proportional to s. The standard deviation in frequency is inversely proportional to s. In order to obtain a complete cover, the scale is sampled in an exponential way and denoted a^j and the parameter of translation is uniformly distributed $u = nu_0$. The transform is written:

$$\psi_{j,n}(t) = \frac{1}{\sqrt{a^j}} \psi\left(\frac{t - nu_0\, a^j}{a^j}\right).$$
(5.174)

In order to pave the time–frequency plane with *wavelet Heisenberg boxes*, one does not use a grid with fixed segmentation, but with time-segments inversely proportional to the frequential segments, which are themselves proportional to the scale.
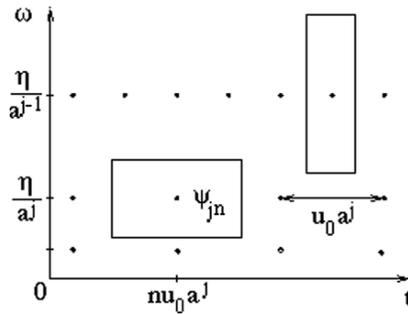
The wavelet must satisfy the condition which follows:

$$\int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|}{\omega} d\omega < +\infty$$
(5.175)

which guarantees the inversibility of the wavelet transform. One will refer to the work of Daubechies or Mallat to know the necessary conditions to the construction of a wavelet frame.

5.13.2.3 Frames and Bases of Time–Frequency Atoms

The windowed Fourier transform and the wavelet transform are written as inner products in $\mathbf{L}^2(\mathbb{R})$ with each one their time–frequency atom, that means respectively:

$$Sf(u,\xi) = \langle f, g_{u,\xi}\rangle = \int_{-\infty}^{+\infty} f(t) g_{u,\xi}^*(t)dt,$$
(5.176)

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\psi_{u,s}^*(t)dt.$$
(5.177)
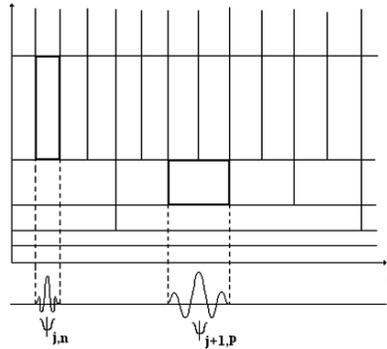
Fig. 5.25 Tiling of the plane by time–frequency boxes of a wavelets basis

A signal can be reconstructed from linear combinations of windowed Fourier atoms or wavelet atoms, which provide a *complete representation* of the signal. The *frame theory* treats conditions of signal reconstructions, stability, completeness and elimination of redundancies in the construction of bases, many notions which are the subject of important developments of which we provided here only preliminary elements.

5.13.3 Tiling of the Time–Frequency Plane by Fourier and Wavelets Bases

A complete orthogonal wavelet basis allows the tiling by time–frequency boxes of wavelets of the whole surface of the plane. Figure 5.25 illustrates this tiling of the plane and highlights two wavelets boxes constructed from $\psi_{j,n}(t)$ and its translation by $2^j n$, i.e. $\psi_{j+1,p}(t)$, which contributes both to the cover of the plane. It is a simple example and a multitude of possibilities is offered to us to cover the plane.

   The cover of the time–frequency plane can be carried out from a *Fourier basis of local sinusoids* (i.e. *local cosine bases*) (Fig. 5.26). At this stage, one utilizes *Malvar bases of local sinusoids*, which are constructed with windows $g_p(t)$ on intervals $[a_p, a_{p+1}]$, that one multiplies by cosine functions of the type $cos(\lambda t + \phi)$ whose frequencies vary. The set is thus written in the following way: $[g_p(t) \cdot cos(\lambda t + \phi)]$, and allows to carry out translations through the axis of frequencies. The time axis itself is divided by the intervals $[a_p, a_{p+1}]$. This method makes it possible to build a basis of local cosines which divide the time axis by the windows $g_p(t)$.
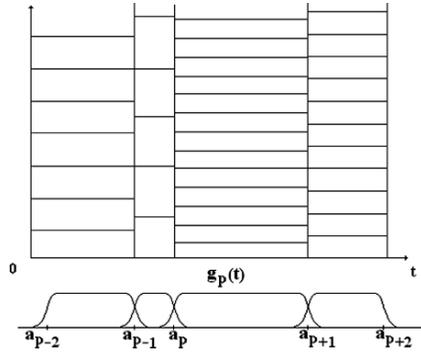
Fig. 5.26 Basis of local sinusoids which cover the time–frequency plane

5.14 Linear and Nonlinear Approximations of a Signal by Projection on an Orthonormal Basis

5.14.1 General Framework of the Linear Approximation and Karhunen–Loève Optimal Basis

Within the framework of the time–frequency analysis, a linear approximation consists in projecting the studied signal f on M vectors (the M first) a priori selected in an orthonormal basis $\{g_m\}_{m\in\mathbb{N}}$. The projection, called *the first M*, is written:

$$f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle\, g_m, \tag{5.178}$$

and the approximation error is written as *the square of the norm of the difference between the signal and its approximation* $(f - f_M)$, or as *the sum of the remaining squared inner products*:

$$Error = \|f - f_M\|^2 = \sum_{m=M}^{+\infty} |\langle f, g_m \rangle|^2. \tag{5.179}$$

In a Fourier basis the reduction of the error is related to the *general regularity* of the signal. A *Fourier basis* provides a *good linear approximation* of *uniformly smooth* signals. They are projected on the M first low frequency sinusoids. The linear approximations of these signals have the same properties in the Fourier or wavelet bases.

Choice of an Optimal Basis: This choice consists in minimizing the approximation error from a basis called "Karhunen–Loève basis", which diagonalizes the covariance matrix of the signal. (This method has to be connected with the projection method of the Singular Spectrum Analysis described in the first part of this book

which consists in operating a projection of the signal on the covariance matrix of this signal.)

5.14.2 Nonlinear Approximation and Adaptive Basis Dependent on the Signal: Regularity and Singularity

For nonlinear structures, the Karhunen–Loève basis is not suitable, it is indeed rather simple to find bases which have errors of approximation smaller than that of Karhunen–Loève. However, there is no in the nonlinear field standard procedure to find an optimal basis as in the linear field. The approximation process is improved by the (no longer "a priori" but) "a posteriori" choice of the M vectors of g_m. These M vectors are identified by their index which belongs to a set $I_M : m \in I_M$. The choice of I_M is crucial, it is done by taking the largest modulus of the inner product: $|\langle f, g_m \rangle|$, i.e. to maximize the amplitude of the inner product. The approximation of f is written:

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle g_m. \tag{5.180}$$

And the approximation error is expressed as the sum of the absolute value of squared inner products, but on the vectors $n \notin I_M$:

$$Error = \|f - f_M\|^2 = \sum_{n \notin I_M} |\langle f, g_m \rangle|^2. \tag{5.181}$$

During the construction of the basis, the choice of I_M maximizes the amplitude of the inner product and also aims to minimize the error. The notion of the nonlinear approximation comes from the fact that the choice of approximation vectors is modified according to the nature of the signal. (In the previous section in the linear framework, we have seen that when M increases and the approximation error decreases, this can be connected with the global regularity of the signal.) The *global* or *local regularity* and singularity notions of signals are central within the nonlinear approximations. Indeed (see section about the *regularity* concept), the *amplitudes* (i.e. *modulus*) of inner products in a *wavelet* basis are related to the notion of *local regularity* of a signal.

The *irregularities* and *singularities* are symptomatic of nonlinear signals. Because of irregularities of a signal and in particular at the places where this signal is irregular, it is necessary to have an *adaptive approximation grid*, by which the resolution increases where the signal is irregular, which corresponds to a nonlinear approximation having the largest possible inner products of wavelets (maximizing the amplitude of inner products locally). At the places *where the signal has singularities,* the *nonlinear approximation is much more precise* and *adapted* than the *linear method* which does not change resolution on the entire signal. Thus, we repeat again, the *amplitude of inner products* in a *wavelet basis* must be linked with the *local regularity* of the signal.

The nonlinear signals can be also approximated by families of vectors which are not built from a unique basis. It will be also noted that the technique of the *best basis* is constructed from *bases* of *wavelets packets* or from *local sinusoidal functions*. These bases of wavelets packets or sinusoidal local functions are families of orthogonal bases which thus contain various types of time–frequency atoms. The choice of the "best basis" is carried out from a *dictionary of bases*, by the *minimization of a "concave cost function"*. A wavelet packet "best basis" or a best basis of local sinusoidal functions decompose the signal with atoms which are adapted to the time–frequency structures of the signal.

It is also possible to slacken the orthogonality constraint for the choice of bases but that can lead to non-effective or explosive constructions. An approximation of the signal can be produced from M non-orthogonal vectors $\{g_{\gamma_m}\}_{0 \leq m \leq M}$, selected in a redundant dictionary $D = \{g_\gamma\}$.

$$f_M = \sum_{m=0}^{M-1} a_m g_{\gamma_m}. \tag{5.182}$$

One can summarize the idea by saying that to optimize the approximation of non-linear signals, *it is interesting to choose, in an adaptive way, a basis dependent on the signal itself.*

5.14.2.1 Concave Cost Function, Best Basis, Ideal Basis and Entropy

A basis B^* is a *better basis* than the basis B to approximate a signal f if and only if for any concave function $\Phi(u)$ there is:[41]

$$\sum_{m=1}^{N} \Phi \left(\frac{|\langle f, g_m^* \rangle|^2}{\|f\|^2} \right) \leq \sum_{m=1}^{N} \Phi \left(\frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \right). \tag{5.183}$$

In fact, the comparison of two bases is done by using a single concave function $\Phi(u)$. The *cost of the approximation of a signal f* in a basis B^λ is determined by the sum as follows:[42]

$$C(f, B^\lambda) = \sum_{m=1}^{N} \Phi \left(\frac{|\langle f, g_m^\lambda \rangle|^2}{\|f\|^2} \right). \tag{5.184}$$

Coifmann and Wickerhauser construct a *best basis B^*, by minimizing the cost of a signal f*:

$$C(f, B^*) = \min_{\lambda \in \Lambda} C(f, B^\lambda). \tag{5.185}$$

Even if there is not better basis than B^*, there can exist several bases, which minimize in an equivalent way the cost of f; then the choice is done according to the

---

[41] For a demonstration see Mallat (1998, p. 407, Lemma of Hardy, Littlewood, Polya).

[42] *Schur-concave sum (Ostrowski):* To refer for example to Tin-Yau Tam (2001). (This concept is also linked to the concept of invariance, S_n-invariance.)

concave function Φ. After having presented how the choice of the "best basis" was carried out, the approximation theory introduces another concept which is the "ideal basis". An ideal basis is a basis which has one of its vectors proportional to the signal itself, i.e. one can write: $g_m = \alpha\, f$ with $\alpha \in \mathbb{C}$. Thus the signal f can be approximated and rebuilt from a single vector which is a basis. Any basis B is worse than the ideal basis and better than a "diffusing" basis to approximate a signal f. If $\Phi(0) = 0$, then:

$$\Phi(1) \leq C(f,B) \leq N\Phi\left(\frac{1}{N}\right). \tag{5.186}$$

The concept opposed to the "ideal basis" is the "diffusing basis" for which the approximation quality is lower. All the information contained in the signal is not restored in its integrity. There is an information loss by diffusion, which is linked to the *Entropy concept*. It is written that the entropy $\Phi(x) = -x\log_e x$ is concave for $x \geq 0$. The resulting *cost is regarded as the entropy of the energy distribution*:

$$C(f,B) = -\sum_{m=1}^{N} \frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \log_e \left(\frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \right). \tag{5.187}$$

Whence it is possible to deduce the bounds of the cost function:

$$0 \leq C(f,B) \leq \log_e N. \tag{5.188}$$

5.14.2.2 Tree Structure, Time–Frequency Plane and Best Basis

A local Fourier "best basis" divides the time–frequency plane in atoms according to the signal itself. The constitution of a dictionary of local sinusoids or wavelet packets requires more than $2^{N/2}$ bases for a signal whose length is N, these are considerable sizes. A best basis minimizes the cost function:

$$C(f,B^*) = \min_{\lambda \in \Lambda} C(f,B^\lambda), \tag{5.189}$$

however to calculate this minimum by a systematic comparison of the cost of each base requires almost $N2^{N/2}$ calculations, which is even more considerable. A "dynamic programming algorithm" developed by *Coifman* and *Wickerhauser* constructs the *best basis* with $O(N\log_2 N)$ calculations. This *algorithm uses the tree structures* constructed from *dictionaries of local Fourier or wavelets bases*. The *trees divide the space into subspaces*, which admit themselves an orthonormal basis of wavelet packets or local sinusoids. In these binary trees of wavelet packets or local sinusoids, each node (at the place of the junction or bifurcation of the fork) corresponds to a space $\mathbf{W}_j^p$ with which an orthonormal basis is associated B_j^p either with wavelet packets, or local sinusoids. The space $\mathbf{W}_j^p$ is divided into two subspaces localized at the nodes:

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \bigoplus \mathbf{W}_{j+1}^{2p+1}. \tag{5.190}$$
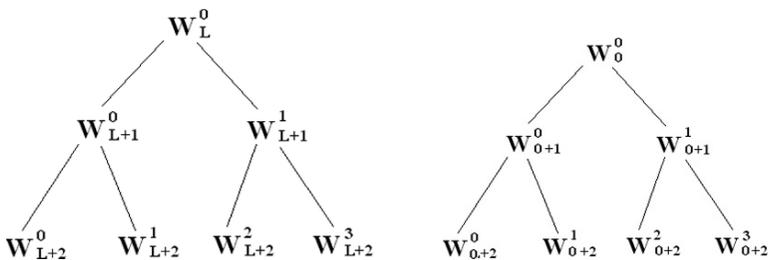
Fig. 5.27 Tree of wavelet packet space (*left*). Tree of local sinusoids space (*right*)

The space $\mathbf{W}_j^p$ is thus divided into two subspaces $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$, which admit themselves an orthogonal basis each one. The union of the orthogonal bases of $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$ forms thus the basis associated with $\mathbf{W}_j^p$. And so on... during the divisions. According to the considered basis (i.e. local sinusoids or the wavelet packets) the initial space is denoted $\mathbf{W}_0^0$ or $\mathbf{W}_L^0$, this space covers[43] all the length N of signal, it is the starting point of bifurcations which follow (Fig. 5.27).

The "best basis" of the space $\mathbf{W}_j^p$ is obtained by minimization of the cost function. We have already presented how to write the cost function of a signal f within a basis: $B = \{g_m\}_{0 \le m < M}$, which is constituted of M vectors. The cost must be calculated according to the division in subspace. Whatever the bases B^0 and B^1, the idea of Coifman and Wickerhauser is to say that if C is an additive cost function, i.e. if $C(f, B^0 \cup B^1) = C(f, B^0) + C(f, B^1)$, then the *best basis* noted Θ_j^p, corresponds to:

$$\Theta_j^p = \begin{cases} \Theta_{j+1}^{2p} \cup \Theta_{j+1}^{2p+1} & \text{if} \quad C(f, \Theta_{j+1}^{2p}) + C(f, \Theta_{j+1}^{2p+1}) < C(f, B_j^p), \\ B_j^p & \text{if} \quad C(f, \Theta_{j+1}^{2p}) + C(f, \Theta_{j+1}^{2p+1}) \ge C(f, B_j^p). \end{cases} \tag{5.191}$$

This method makes it possible to build the best basis of the space at the root of the tree structure, by calculating in a recursive way the best bases of all spaces $\mathbf{W}_j^p$ inside the tree, from the root downwards. The best bases of the spaces $\{\mathbf{W}_j^p\}_p$ are calculated from the best bases of spaces $\{\mathbf{W}_{j+1}^p\}_p$ by means of the cost calculation method presented previously. The operation is thus repeated until obtaining the best basis on $\mathbf{W}_0^0$ for the bases of sinusoids and $\mathbf{W}_L^0$ for the bases of wavelets. The depth of the tree is lower than $\log_2 N$, and the quantity of calculations necessary to obtain the best basis is $O(N \log_2 N)$.

Example of Tree and Wavelet Packet Basis

The multiresolution analysis is a good support to illustrate the wavelet packet trees. Indeed, in the multiresolution analysis (MRA), the space $\mathbf{V}_j$ is reduced to a lower resolution space $\mathbf{V}_{j+1}$ and to a space called *the "detail" space* $\mathbf{W}_{j+1}$. What means

---

[43] Where $2^L = N^{-1}$.

to *divide an orthogonal basis* $\{\phi_j(t-2^j n)\}_{n\in\mathbb{Z}}$ *belonging to* $\mathbf{V}_j$ *into two orthogonal bases* (of scaling functions and wavelets):

$$\{\phi_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}} \in \mathbf{V}_{j+1}, \quad \text{and} \quad \{\psi_{j+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}} \in \mathbf{W}_{j+1}. \quad (5.192)$$

In a generic way, a theorem stated by Coifmann, Meyer and Wickerhauser shows that the "conjugate mirror filters" transform an orthogonal basis $\{\theta_j(t-2^j n)\}_{n\in\mathbb{Z}}$ into two families of orthogonal bases: $\{\theta_{j+1}^0(t-2^{j+1}n)\}_{n\in\mathbb{Z}}$ and $\{\theta_{j+1}^1(t-2^{j+1}n)\}_{n\in\mathbb{Z}}$. If we denote $\mathbf{U}_{j+1}^0$ and $\mathbf{U}_{j+1}^1$ the spaces associated with these two families, it is said that these spaces are orthogonal and:

$$\mathbf{U}_{j+1}^0 \bigoplus \mathbf{U}_{j+1}^1 = \mathbf{U}_j. \quad (5.193)$$

Consequently, it is possible to establish that $\mathbf{U}_j = \mathbf{W}_j$ and thus to divide these "detail" spaces $\mathbf{W}_j$ into subspaces which have new associated bases. These successive divisions produce (see illustration supra) the binary trees (because division is a division by two), and if the scale for the approximation is dyadic of the type 2^L, it is possible to associate with it the approximation space $\mathbf{V}_L$ and an orthogonal basis of *scaling functions:*

$$\{\phi_L(t-2^L n)\}_{n\in\mathbb{Z}}, \quad (5.194)$$

(where $\phi_L = 2^{-L/2}\phi(2^{-L}t)$, with $j \geq L$). $j-L \geq 0$ corresponds to the depth of the binary tree. Each junction point (or node) of the binary tree is referred by a couple (j,p) where j is the resolution level and p the *"number" of the node* starting from the left at the same depth $j-L$. To each one of the node of the tree (indexed by the couples (j,p)) corresponds one space $\mathbf{W}_j^p$ and one basis $\{\psi_j^p(t-2^j n)\}_{n\in\mathbb{Z}}$. To the starting point of the tree, the space $\mathbf{V}_L$ corresponds to the space $\mathbf{W}_j^0$, $(\mathbf{V}_L = \mathbf{W}_j^0)$ and the scaling function ϕ_L corresponds to ψ_L^0. And if one chooses an unspecified division of the tree, one says that the bases $B_{j+1}^{2p} = \{\psi_{j+1}^{2p}(t-2^{j+1}n)\}_{n\in\mathbb{Z}}$ and $B_{j+1}^{2p+1} = \{\psi_{j+1}^{2p+1}(t-2^{j+1}n)\}_{n\in\mathbb{Z}}$ are orthonormal bases of two orthogonal spaces $\mathbf{W}_{j+1}^{2p}$ and $\mathbf{W}_{j+1}^{2p+1}$, which respect the equality mentioned previously

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \bigoplus \mathbf{W}_{j+1}^{2p+1}. \quad (5.195)$$

Illustration of Tree Structure and Best Basis for a Transient Signal

Tree and Best basis of sinusoids packets space of the signal. Observe in Fig. 5.28 the signal and image of the Heisenberg boxes of Fourier atoms (Fig. 5.28).

Hereafter, in the upper part of the figure on the left: the *sinusoids packets decomposition* within the framework of the *best basis* (level of division: D = 9 = $\log_2(512)$). In the upper part of the figure on the right: the tree of the *best basis of sinusoids packets*. In the lower part on the left: Heisenberg boxes of the Fourier atoms in the time–frequency plane. In the lower right part: *Image* of the *Heisenberg boxes* of *Fourier atoms:*
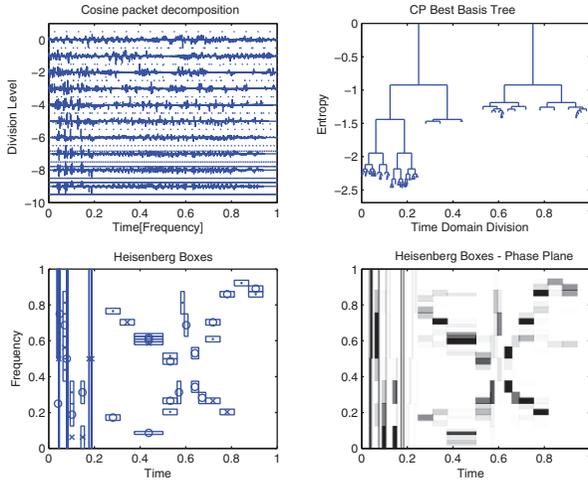
Fig. 5.28 Heisenberg boxes of Fourier atoms
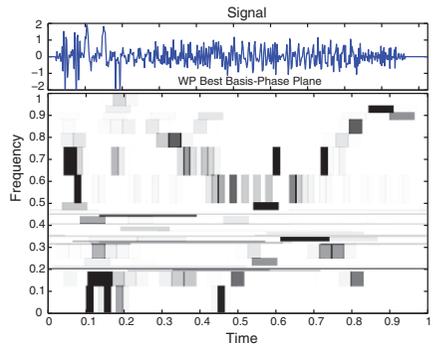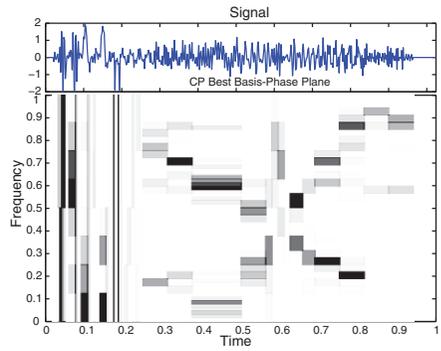
Fig. 5.29 Heisenberg boxes of wavelet atoms

Tree and Best Basis of wavelet packets space of the signal. Observe in Fig. 5.29 the signal and image of the Heisenberg boxes of wavelet atoms (Fig. 5.29).

Hereafter in the upper part on the left: the wavelet packets decomposition of the signal with the best basis (calculated with a Daubechies-12 filter). In the upper right part: Tree of the best basis of wavelet packets. In the lower left part: Heisenberg

boxes of wavelet atoms in the time frequency plane. In the lower right part: Image of the Heisenberg boxes of wavelet atoms:
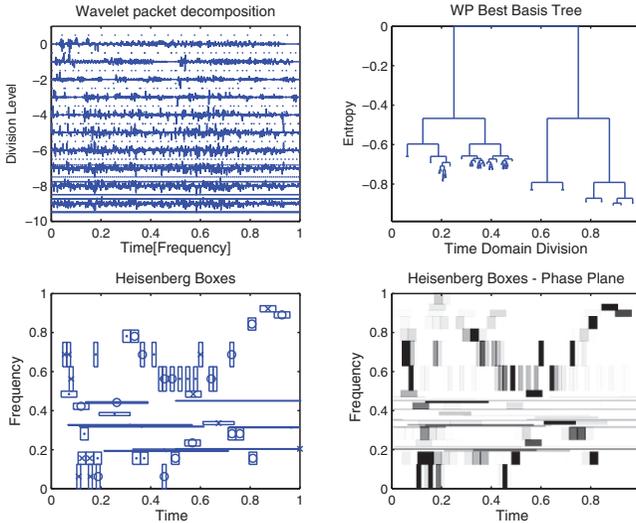
Illustration of Tree Structure and Best Basis of the Stock Exchange Index: Cac40

Tree of the wavelet packet space: The decomposed signal is the French stock exchange index (Cac40) for 2,048 observations of the daily growth rate.
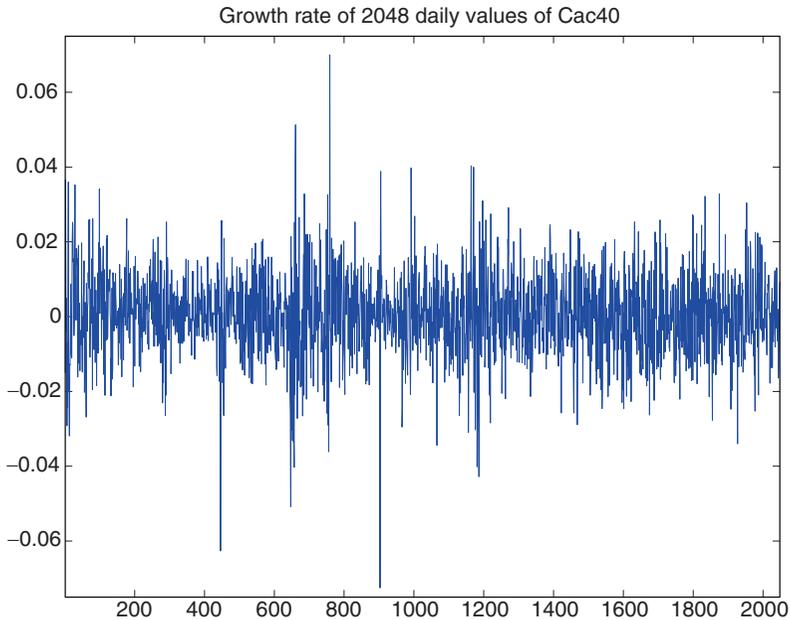
Fig. 5.30 Mirror filter: Coiflet
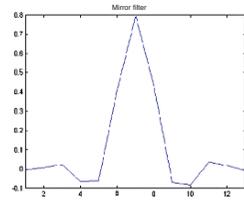
Fig. 5.31 Wavelet packet tree
(frequency domain division)
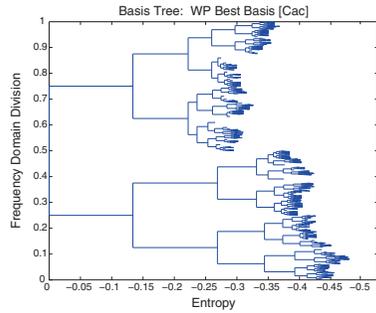
For the wavelet packet decomposition, a mirror filter (of the type Coiflet 3) is used, one gives its representation (Fig. 5.30). Then a convolution is operated between the signal and the filter. The calculation of the best basis (6.5285) gives the following tree with a depth level of $D = 11 = log_2(2,048)$ (Fig. 5.31).

Tree of the sinusoid packet space: The representation of Heisenberg boxes in the time–frequency plane has, in this case, a weak visual interest and thus will not be presented (Fig. 5.32).

5.14.3 Donoho and Johnstone Nonlinear Estimation: Algorithm with Threshold

The problem of the estimation is to find a structure in a signal which apparently does not let it show through. Either a signal is really random without any structure, or the structure is hidden due to the presence of a noise which covers it. (It is obvious that the structures of nonlinear origin are more difficult to isolate.) A good part of the debate regarding the *estimation* is related to *the notion of threshold* between *noise* and *signal*. Thus, it is possible to conceive a signal as the addition of a structure and a noise:

$$X = f + noise. \tag{5.196}$$

The signal f is estimated by the transformation of the disturbed (i.e. noized) signal X by an operator D, one poses $\widetilde{F}$ *the estimator of* f:

$$\widetilde{F} = DX. \tag{5.197}$$

Fig. 5.32 Fourier basis tree
(time domain division)
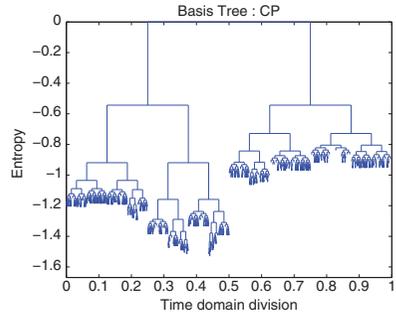
The average error provides *the risk of the estimator*:

$$r(D, f) = E \left\{ \|f - DX\|^2 \right\}. \tag{5.198}$$

The choice of a *nonlinear operator D* is *more effective* to decrease the risk *than a linear operator*. Donoho and Johnstone show that a nonlinear estimator close to the optimality is obtained by means of a *thresholding:*

$$\widetilde{F} = DX = \sum_{m=0}^{N-1} \rho_T \left(\langle X, g_m \rangle \right) g_m. \tag{5.199}$$

$\rho_T(x)$ is a function with *threshold T* which is worth x for $|x| \geq T$ and 0 for $|x| < T$. The method proceeds to a *smoothing* which is a function of the threshold but also of the selected basis.

5.14.4 Nonlinear Estimators are More Efficient to Minimize the Bayesian Risk: Optimization by Minimax

The estimate of the operator D in the expression described previously $\widetilde{F} = DX$ is a vast subject and obviously the objective here is not to be exhaustive. We will describe the current practices in the field, which use in particular the bayesian analysis. See also the (statistics) part II of the present book.

5.14.4.1 Bayesian Approach: A Posteriori Law Specification of the Parameter and Choice of the A Priori Law

A statistical model is a couple $(\mathscr{Y}, \mathbf{P})$, where $\mathscr{Y}$ is the set of possible observations and $\mathbf{P}$ is a family of probability laws on $\mathscr{Y}$. Such a model is known as dominated, if

all the laws of the family admit a density with respect to a same measure μ, which is called the dominating measure. Given a parametric (dominated) model:

$$[\mathscr{Y}, \mathbf{P} = \{P_\theta = l(y;\theta) \cdot \mu; \theta \in \Theta\}], \qquad (5.200)$$

the Bayes approach provides the set Θ of the possible values of the parameter θ of a law Π, this law is known as the *a priori law (or prior law)*, and *it is made as if the parameter were random of law Π*. Moreover, when we possess the observations y, the "a priori" *idea* about the parameter is then modified. Indeed, we consider these observations, and have to replace the (*a priori*) *marginal law* of θ by the *conditional law* of θ *knowing y*, which is the "*a posteriori*" *law*. The transition of the *a priori law* to the *a posteriori law* is expressed by *the Bayes formula*:

$$\pi(\theta/y) = \frac{l(y;\theta)\pi(\theta)}{\int_\Theta l(y;\theta)\pi(\theta)v(d\theta)} = \pi(\theta)\frac{l(y;\theta)}{l(y)}, \qquad (5.201)$$

this formula supposes that the a priori law admits a density $\pi(\theta)$ with respect to a measure v. Consequently, the a posteriori law admits a density $\pi(\theta/y)$ with respect to this measure. The a *posteriori law is obtained by multiplying the a priori density by the ratio of the conditional density* of y knowing θ (*likelihood of y knowing θ*) *to the marginal density* of y (*predictive density*). In this approach, the *most difficult* is to *choose the most adapted a priori law*, even if obviously the calculation of the transition formula of the a priori law towards the a posteriori law is often also difficult.

In this type of Bayesian approach, there is thus an *a priori law Π on Θ*, and for a statistical test φ, it is possible to define the *bayesian risk* and to compare the tests. The bayesian risk is written:

$$r_\Pi(\varphi) = \int_\Theta R(\varphi, \theta)d\Pi(\theta). \qquad (5.202)$$

The Bayes analysis benefits from the information contained in the signal to calculate an a posteriori law. (There is a kind of recursivity in the Bayes approach.) The estimate of the operator D, in the expression described previously $\widetilde{F} = DX$, will be optimized if one can benefit from the information that the signal contains. Within the framework of the terminology used in time–frequency analysis, the *risk* is written:

$$r(D, \pi) = E_\pi\{r(D, F)\}, \qquad (5.203)$$

this risk takes into account the empirical law, i.e. the *probability distribution π* of the signal. An optimal operator which is taken among all the possible estimators D, must provide the *minimum bayesian risk*, which is written:

$$r(\pi) = \inf_D r(D, F). \qquad (5.204)$$

These techniques are difficult to use and the *choice* of the *a priori law predetermines the results*.

5.14.4.2 Minimax and Nonlinear Operators

The framework of the *Game* and *Decision Theory* allows to introduce simpler techniques, where *we do not specify the probability distribution*. Here are signals modelled as "particular elements of a set Θ". In order to control the risk, we evaluate its maximum for any signal belonging to the set Θ. The maximum risk is written:

$$r(D,\Theta) = \sup_{f \in \Theta} r(D,f), \tag{5.205}$$

And the *Minimax risk* is written:

$$r_n(\Theta) = \inf_D r(D,\Theta). \tag{5.206}$$

In a general way, in the field of the possible estimators for D, the nonlinear estimators prove to have a "minimax risk" lower than the linear estimators (except for particular cases of convexity).[44] Thus, a good operator D will be provided by a nonlinear estimate using the minimax to reach the weakest possible risk.

5.14.5 Approximation by the "Matching Pursuit": A General Presentation

For many signals made up of complex structures, the *best bases* of local sinusoids or wavelet packets *are not the most powerful approximations*. As underlined previously, by the slackening of the orthogonality constraint, an approximation of the signal can be produced from M non-orthogonal vectors $\{g_{\gamma_m}\}_{0 \leq m \leq M}$, selected in a redundant dictionary $D = \{g_\gamma\}$. And the approximation of the signal was written:

$$f_M = \sum_{m=0}^{M-1} a_m g_{\gamma_m}. \tag{5.207}$$

For a signal length N, the number of vectors of a redundant dictionary is calculated by $P = N \log_2 N$. Thus, the dictionary is constituted of P vectors at the most, which contains at least N linearly independent vectors. These N linearly independent vectors are selected to build *a set of bases*. In addition, note that if a *set of non-orthogonal bases* is constructed, this set usually will be *larger* than *the set made up from orthogonal bases*. For a dictionary $D = \{g_p\}_{0 \leq p < P}$, whatsoever M (with $N < P$, $0 \leq m \leq M$, and $M \leq N$), the approximation is the result of a linear combination of the M vectors of the dictionary.[45] And the approximation is written

---

[44] For a discussion about the optimality of the minimax within the framework of the time–frequency analysis, we can refer to Mallat (1999, pp. 469–500).

[45] Ref. to *Theory* concerning the *linear programming*, the *minimax*, and the algorithm of the *simplex* is used in the time–frequency analysis and is also used by *economists*.

in a discrete way:

$$f_M = \sum_{m=0}^{M-1} a[p_m] g_{p_m}. \tag{5.208}$$

The calculations of f_M and the minimization of $\|f - f_M\|$ are heavy and difficult. Mallat and Zhang have developed the version of the Matching Pursuit algorithm which has the property to reduce the calculations by sacrificing the "optimality" to the "efficiency". Starting from the initial redundant dictionary $D = \{g_\gamma\}$, the algorithm "pursuit"[46] allows to select N vectors to construct a basis $\{g_{\gamma_m}\}_{0 \le m < N}$. The matching pursuit is connected with the *pursuit algorithms* of *projection* developed in *statistics*.

5.14.5.1 The "Matching Pursuit" Structure

Given the dictionary $D = \{g_\gamma(t)\}_{\gamma \in \Gamma}$ of P vectors (endowed with a unit norm such that $\|g_\gamma\| = 1$), which contains N linearly independent vectors defining a basis. Thus, a matching pursuit calculates a linear expansion of f, by successions of approximations of f, by means of orthogonal projections of the signal on elements of D. Given $g_{\gamma_0} \in D$, the vector f can be decomposed as follows:

$$f = \langle f, g_{\gamma_0} \rangle g_{\gamma_0} + Rf. \tag{5.209}$$

$R f$ is the *residual vector after approximation of f* in the direction of g_{γ_0}. And g_{γ_0} is orthogonal to $R f$ and is normalized at 1. Since Rf is orthogonal to g_{γ_0}:

$$\|f\|^2 = |\langle f, g_{\gamma_0} \rangle|^2 + \|Rf\|^2. \tag{5.210}$$

Minimize the residual vector $\|Rf\|$ comes down to choose g_{γ_0} in order to maximize $|\langle f, g_{\gamma_0} \rangle|$. From the point of view of calculations, the most effective is to find a sub-optimal vector g_{γ_0}:

$$|\langle f, g_{\gamma_0} \rangle| \ge \alpha \sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|, \tag{5.211}$$

[with $0 \le \alpha \le 1$]. The Matching pursuit is an iterative algorithm which with successive stages decomposes the residue Rf of a previous projection.[47] If we are located at an unspecified moment of the iteration process and if the mth residue $R^m f$ has been calculated, then the following stage is to choose an element g_{γ_m} of the dictionary which approximates the $R^m f$ residue:

$$|\langle R^m f, g_{\gamma_m} \rangle| \ge \alpha \sup_{\gamma \in \Gamma} |\langle R^m f, g_\gamma \rangle| \tag{5.212}$$

---

[46] Chen and Donoho's algorithm (Chen and Donoho 1995).
[47] $R^0 f = f$.

the residue $R^m f$ is decomposed:

$$R^m f = \langle f, g_{\gamma_m} \rangle g_{\gamma_m} + R^{m+1} f, \tag{5.213}$$

Since $R^{m+1} f$ is orthogonal to g_{γ_m}, we can write:

$$\|R^m f\|^2 = \left| \langle R^m f, g_{\gamma_m} \rangle \right|^2 + \|R^{m+1} f\|^2. \tag{5.214}$$

We obtain by summation:

$$f = \sum_{m=0}^{M-1} \langle R^m f, g_{\gamma_m} \rangle g_{\gamma_m} + R^M f. \tag{5.215}$$

and then:

$$\|f\|^2 = \sum_{m=0}^{M-1} \left| \langle R^n f, g_{\gamma_m} \rangle \right|^2 + \|R^M f\|^2. \tag{5.216}$$

$\|R^m f\|$ converges (Mallat 1998, p. 422) exponentially towards 0, when m tends towards the infinite:

$$\lim_{m \to \infty} \|R^m f\| = 0. \tag{5.217}$$

The approximation carried out by the matching pursuit is largely improved by the orthogonalization of the directions of projection. This is done by means of the Gram–Schmidt procedure (Pati et al. 1993). The *convergence* of the *orthogonal matching pursuit* is carried out with *a finite number of iterations*, whereas it is not the case with the *non-orthogonal matching pursuit*.

Orthogonal Matching Pursuit

In the previous procedure the vectors g_{γ_m} selected by the matching pursuit are a priori non-orthogonal to vectors earlier selected, which are written: $\{g_{\gamma_p}\}$ with $0 \leq p < m$. The previous matching pursuit selected the g_{γ_m} which verified:

$$\left| \langle R^m f, g_{\gamma_m} \rangle \right| \geq \alpha \sup_{\gamma \in \Gamma} \left| \langle R^m f, g_\gamma \rangle \right|. \tag{5.218}$$

In order to distinguish with the former procedure, we pose $\xi_0 = g_{\gamma_0}$. The Gram–Schmidt algorithm orthogonalizes the g_{γ_m} with the $\{g_{\gamma_p}\}_{0 \leq p < m}$ and we pose:

$$\xi_m = g_{\gamma_m} - \sum_{p=0}^{m-1} \frac{\langle g_{\gamma_m}, \xi_p \rangle \, \xi_p}{\|\xi_p\|^2}, \tag{5.219}$$

Previously $R^m f$ was projected on g_{γ_m}, but now, $R^m f$ is projected on ξ_m:

$$R^m f = \frac{\langle R^m f, \xi_m \rangle \, \xi_m}{\|\xi_m\|^2} + R^{m+1} f. \tag{5.220}$$

By summation on m, we obtain:

$$f = \sum_{m=0}^{k-1} \frac{\langle R^m f, \xi_m \rangle\, \xi_m}{\|\xi_m\|^2} + R^k f \tag{5.221}$$

$$= P_{V_k} f + R^k f. \tag{5.222}$$

The $\{\xi_m\}_{0 \leq m < k}$ generates the space V_k, and P_{V_k} is an orthogonal projector on V_k. For $m = k$, we have:

$$\langle R^m f, \xi_m \rangle = \langle R^m f, g_{\gamma_m} \rangle, \tag{5.223}$$

and it exists M such that $f \in V_M$, thus $R^M f = 0$. And for $k = M$ by combining the preceding expression with the equation of f, we have:

$$f = \sum_{m=0}^{k-1} \frac{\langle R^m f, g_{\gamma_m} \rangle\, \xi_m}{\|\xi_m\|^2}. \tag{5.224}$$

In the preceding section, we have introduced the convergence of the orthogonal matching pursuit, this convergence is done at the end of M iterations. The square of the norm of f is written as the following sum:

$$\|f\|^2 = \sum_{m=0}^{M-1} \frac{\left|\langle R^m f, g_{\gamma_m} \rangle\right|^2}{\|\xi_m\|^2}. \tag{5.225}$$

Finally, in order to develop f on the initial dictionary $\{g_{\gamma_m}\}_{0 \leq m < M}$, we write:

$$\xi_m = \sum_{p=0}^{m} b[p, m] g_{\gamma_p}$$

which gives:

$$f = \sum_{m=0}^{M-1} a[\gamma_p] g_{\gamma_p}, \tag{5.226}$$

with:

$$a[\gamma_p] = \sum_{m=p}^{M-1} b[p, m] \frac{\langle R^m f, g_{\gamma_m} \rangle}{\|\xi_m\|^2}. \tag{5.227}$$

Hereafter, we provide the representations in the time–frequency plane of both decompositions of a transitory signal (transient).

Figure 5.33 corresponds to the decomposition of the signal by the matching pursuit with dictionaries of windowed Fourier atoms. One will notice the "spread" of each atom in the time–frequency plane.

Figure 5.34 corresponds to the decomposition of the same signal by the matching pursuit with dictionaries of wavelet atoms. Observe the "spread" and the localization of atoms compared with Fig. 5.33.

Fig. 5.33 Heisenberg boxes
of Fourier atoms
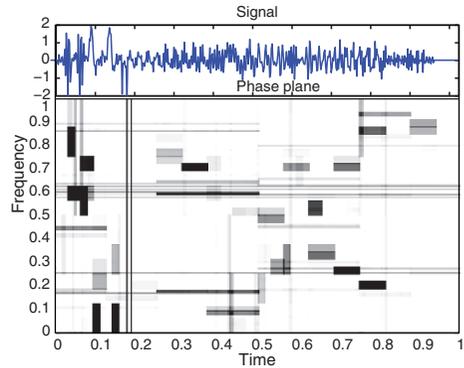
Fig. 5.34 Heisenberg boxes
of wavelet atoms
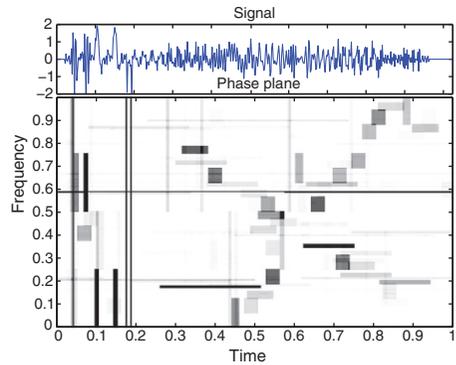
Gabor Atom, Function and Dictionary

An improvement of the Matching Pursuit was made by Mallat and Zhang, as well
as by Qian and Chen by means of the Gabor dictionary. This is by modulating, relo-
cating and spreading (i.e. scaling) a Gaussian window which is used for its *qualities
of energy (and frequency) distribution.*[48] The window is written:

$$g(t) = 2^{1/4} e^{-\pi t^2}, \tag{5.228}$$

in a discrete mode for a period denoted N and a parameter denoted α_j, this window
is sampled and modulated on a scale 2^j, and becomes:

$$g_j[n] = \alpha_j \sum_{p=-\infty}^{+\infty} g\left(\frac{n - pN}{2^j}\right), \tag{5.229}$$

$\|g_j\| = 1$ is obtained by means of α_j. Then, the window is translated in time and
frequency via an index denoted:

$$\gamma = (p, k, 2^j) \tag{5.230}$$

---

[48] See proof by the Heisenberg theorem.

with j belonging to $[0, \log_2 N]$ and (p, k) belonging to $[0, N-1]^2$. We obtain a Gabor atom:

$$g_\gamma[n] = g_j[n-p]e^{\frac{2\pi ikn}{N}}, \tag{5.231}$$

We denote $D = \{g_\gamma\}$ the Gabor dictionary of these Gabor atoms. More precisely, the matching pursuit uses the following set of indexes:

$$\gamma_\pm = (p, \pm k, 2^j) \tag{5.232}$$

and *provides groups of complex Gabor atoms*: $g_{\gamma-}$ and $g_{\gamma+}$ (also called *polarized* atoms). In this case and for a real signal, the matching pursuit is a projection of $R^m f$ on:

$$g_\gamma^\Omega[n] = \alpha_{j,\Omega} g_j[n-p] \cos\left(\frac{2\pi ikn}{N} + \Omega\right). \tag{5.233}$$

The approximation of the signal is thus written as the following decomposition:

$$f = \sum_{m=0}^{+\infty} \left\langle R^m f, g_{\gamma m}^{\Omega m} \right\rangle g_{\gamma m}^{\Omega m}, \tag{5.234}$$

which is represented by an *energy distribution* resulting from the addition of *the Wigner–Ville density* of *complex atoms*:

$$P_M f[n,k] = \sum_{m=0}^{+\infty} \left| \left\langle R^m f, g_{\gamma m}^{\Omega m} \right\rangle \right|^2 \mathbf{D} g_{\gamma m}[n,k]. \tag{5.235}$$

(This matching pursuit version will be presented in detail later on in another section.) Hereafter, one provides the representation of the decomposition in the time–frequency plane of the (same) transient signal by Matching Pursuit with Gabor atoms. It will be noticed the shape, the spread and the localization of atoms in the time–frequency plane (Fig. 5.35).
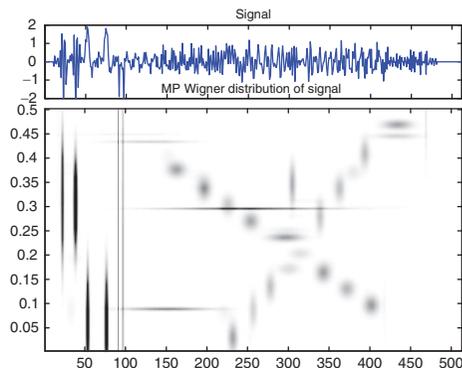
Fig. 5.35 Wigner–Ville distribution of Gabor atoms

5.14.6 Comparison of Best Bases and Matching Pursuits

The presentations of approximation methods of the Best Basis and Matching Pursuits previously made have been illustrated by means of the same signal that is of transient type (i.e. intermittent with some Diracs), this one now makes it possible to compare the results in Fig. 5.36.
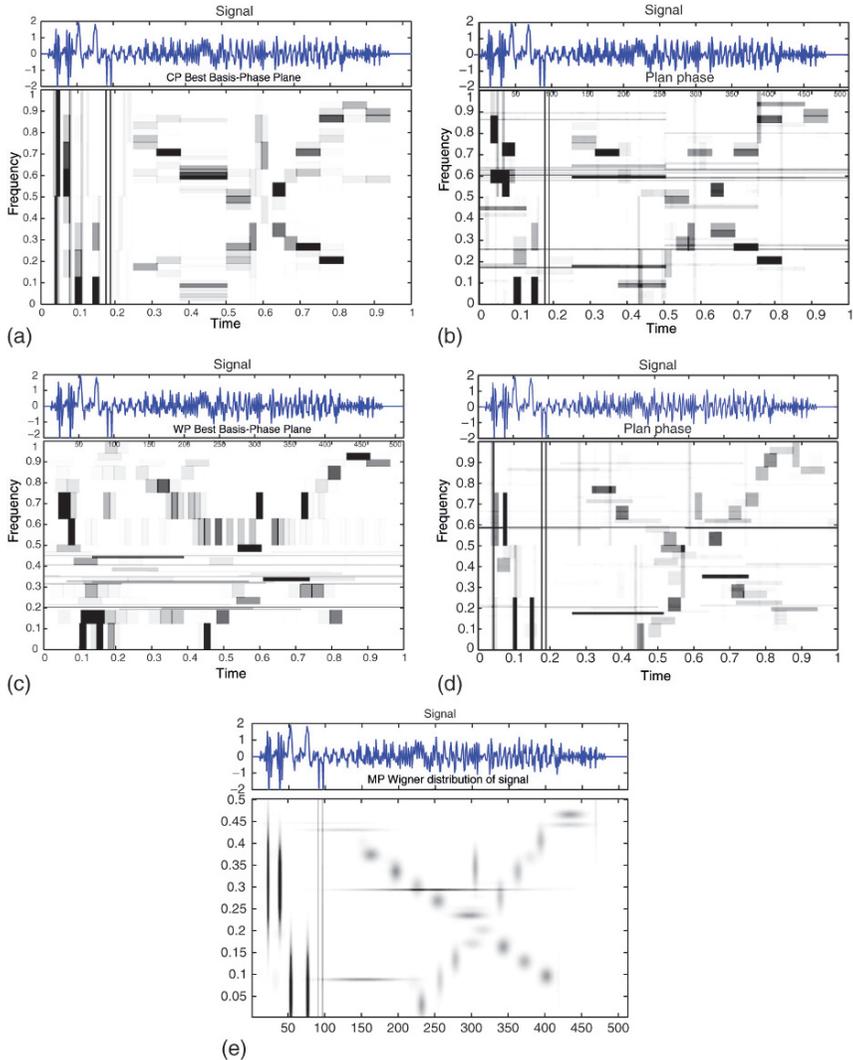
Fig. 5.36 (a) Best basis and Fourier, (b) matching pursuit and Fourier, (c) best basis and wavelets, (d) matching pursuit and wavelets, (e) matching pursuit and Gabor atoms

Although the parameter setting (choice of filter mirror for the wavelets, split level, etc.) in the construction of approximations, plays an important role and can change the quality of representations, it is easy to observe the evolution of approximation methods towards the latest method presented in Fig. 5.36e.

5.15 The Multiresolution Analysis Notion

It is obviously necessary to present the multiresolution analysis (MRA: Multiresolution analysis) whose first corollary is the discrete wavelet transform (DWT). Beforehand, we quickly introduce the notion of (quadratic) *conjugate mirror filter* which is used in the construction of fast algorithms of the wavelet transform.

5.15.1 (Quadratic) Conjugate Mirror Filter

This type of filter is used in various applications and gave rise to the development of a true theory. In particular for the construction of fast algorithms of calculations of the (orthogonal) wavelet transforms and the improvement of methods of signal reconstruction. It contributes also to the construction of the algorithms of the multiresolution analysis (MRA). Its role is important, it is perfectly presented and developed in the works of S. Mallat. The presentation of the history and the construction of these filters allows to approach the multiresolution analysis in an interesting way. It will be simply noted here that the fast algorithms of wavelet transforms calculations decompose the signals into *low-pass* and *high-pass components* which are sub-sampled in two subsets. The signal is filtered by a low-pass filter denoted $\bar{h}[p] = h[-p]$ and a high-pass filter $\bar{g}[p] = g[-p]$, these filters make it possible by convolution to sub-sample a signal f into two outputs:

$$a_1[p] = f \star \bar{h}[2p] \quad \text{and} \quad d_1[p] = f \star \bar{g}[2p]. \tag{5.236}$$

In fact, such a transformation decomposes each approximation $P_{V_j}f$ into a coarse approximation $P_{V_{j+1}}f$ and into wavelet coefficients $P_{W_{j+1}}f$. During the reconstruction the $P_{V_j}f$ are "re-found" from $P_{V_{j+1}}f$ and $P_{W_{j+1}}f$. It is known that:

$$\{\phi_{j,n}\}_{n\in\mathbb{Z}} \quad \text{and} \quad \{\psi_{j,n}\}_{n\in\mathbb{Z}} \tag{5.237}$$

are orthonormal bases of $\mathbf{V}_j$ and $\mathbf{W}_j$. Furthermore, it is known that the projections on these bases are defined by:

$$a_j[n] = \langle f, \phi_{j,n} \rangle \quad \text{and} \quad d_j[n] = \langle f, \psi_{j,n} \rangle. \tag{5.238}$$

These coefficients are calculable by means of a cascade of discrete convolutions and sub-samplings. If we pose that $\bar{x}[n] = x[-n]$ and:

$$\hat{x}[n] = \begin{cases} x[p] & \text{if } n = 2p, \\ 0 & \text{if } n = 2p+1, \end{cases} \tag{5.239}$$

then, it is known that *a signal is decomposable* into:

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] = a_j \star \bar{h}[2p], \tag{5.240}$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g[n-2p]a_j[n] = a_j \star \bar{g}[2p], \tag{5.241}$$

and the *signal reconstruction* is written:

$$a_j[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_{j+1}[n] + \sum_{n=-\infty}^{+\infty} g[n-2p]d_{j+1}[n] \tag{5.242}$$

$$= \hat{a}_{j+1} \star \bar{h}[p] + \hat{d}_{j+1} \star g[p]. \tag{5.243}$$

The *mirror filters* satisfy the following (Mallat and Meyer) quadratic condition:

$$\left|\hat{h}(\omega)\right|^2 + \left|\hat{h}(\omega+\pi)\right|^2 = 2 \quad \text{and} \quad \hat{h}(0) = \sqrt{2}, \tag{5.244}$$

where ϕ is an integrable scaling function belonging to $\mathbf{L}^2(\mathbb{R})$, where $h[n] = \langle 2^{-1/2}\phi(t/2), \phi(t-n) \rangle$ and $\hat{h}$ is its Fourier transform $\forall \omega \in \mathbb{R}$. The mirror filters $h[n]$ are of different types, we will mention without presenting them the following types: "Haar", "Beylkin", "Coiflet", "Daubechies", "Symmlet", "Vaidyanathan", "Battle".

5.15.2 Multiresolution Analysis

Definition 5.3 (Multiresolution). A sequence[49] $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $\mathbf{L}^2(\mathbb{R})$ is a multiresolution approximation, if the six following properties are verified:

$$\forall (j,k) \in \mathbb{Z}^2, f(t) \in \mathbf{V}_j \Leftrightarrow f(t-2^jk) \in \mathbf{V}_j, \tag{5.245}$$

$$\forall j \in \mathbb{Z}, \mathbf{V}_{j+1} \subset \mathbf{V}_j, \tag{5.246}$$

$$\forall j \in \mathbb{Z}, f(t) \in \mathbf{V}_j \Leftrightarrow f\left(\frac{t}{2}\right) \in \mathbf{V}_{j+1}, \tag{5.247}$$

$$\underset{j \to +\infty}{Lim}\ \mathbf{V}_j = \bigcap_{j=-\infty}^{+\infty} \mathbf{V}_j = \{0\}, \tag{5.248}$$

---

[49] $\mathbb{Z}$: Space of the (positive and negative) integers, $Z = \{\ldots, -3, -2, -1, 0, +1, +2, +3, \ldots\}$ which is an extension of the space of the integers $\mathbb{N}$.

$$\underset{j \to -\infty}{Lim} \, \mathbf{V}_j = Closure \left(\bigcup_{j=-\infty}^{+\infty} \mathbf{V}_j \right) = \mathbf{L}^2(\mathbb{R}), \tag{5.249}$$

There exists θ such that $\{\theta(t - n)\}_{n \in \mathbb{Z}}$ is Riesz basis of $\mathbf{V}_0$. (5.250)

These properties are interpreted in the following way:
- $\mathbf{V}_{j+1}$ is the image of $\mathbf{V}_j$ by a dilation of a factor 2: There exists a subjacent frequential grid in geometric progression.
- For any j, $\mathbf{V}_{j+1}$ is a sub-space of $\mathbf{V}_j$.
- $\mathbf{V}_j$ is invariant by translation of 2^j: There exists a subjacent temporal grid by step of 2^j.
- The intersection of $\mathbf{V}_j$ is reduced to 0 in $\mathbf{L}^2$: At a minimal resolution, one loses all the image. This hypothesis is formulated by convention, because it is always verified.
- The union of the $\mathbf{V}_j$ is dense in $\mathbf{L}^2$: At an infinite resolution, any signal is perfectly reproduced.
- There exists a function θ such that the whole translations of θ form a Riesz basis of $\mathbf{V}_0$: Each resolution is generated by a basis of atoms translated of 2^j. A Riesz basis is a frame of independent vectors.

The dilated and translated functions $\phi_{j,k}(t)$:

$$\{\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k)\}_{k \in \mathbb{Z}} \tag{5.251}$$

constitute also a Riesz basis for V_0. It is fundamental to understand that to carry out a multiresolution analysis of a time series $x(t)$, this is successively to project the latter on each approximation subspaces V_j:

$$Approx_j(t) = \text{Proj}_{V_j} x(t) = \sum_k a_{j,k}\phi_{j,k}(t). \tag{5.252}$$

As presented previously $V_j \subset V_{j-1}$, consequently we know that the approximation: $approx_j(t)$ of the series $x(t)$ is an approximation coarser than $approx_{j-1}(t)$. Thus, another fundamental idea of the multiresolution analysis (MRA), consists in the analysis of the information loss which is indicated by the term: "$detail_j$", it expresses the information loss related to an increasingly coarse approximation, as follows:

$$Detail_j(t) = approx_j(t) - approx_{j-1}(t). \tag{5.253}$$

The multiresolution analysis (MRA) computes the "$detail_j$" by projections of $x(t)$ on a collection of subspaces noted W_j, called *the wavelet subspaces*. Furthermore, the multiresolution analysis proves that there is *a function ψ_0*, called *mother wavelet* (built as the scaling function ϕ_0) which is used to construct the functions:

$$\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\} \tag{5.254}$$

which constitute also a Riesz basis for W_j:

$$Detail_j(t) = \text{Proj}_{W_j} x(t) = \sum_k d_{j,k} \psi_{j,k}(t). \tag{5.255}$$

The principle of the multiresolution analysis is to transform the information contained in the time series $x(t)$ in a collection, which gathers on the one hand, an approximation with low resolution J, and on the other hand, the set of "$detail_j$" at different resolution $j = 1, \ldots, J$ (noted $\sum_{j=1}^{J} detail_j$). Then it is possible to write:

$$x(t) = approx_J + \sum_{j=1}^{J} detail_j(t) = \sum_k a_{J,k} \phi_{J,k}(t) + \sum_{j=1}^{J} \sum_k d_{j,k} \psi_{j,k}(t). \tag{5.256}$$

In order to construct $approx_J$ which is a coarse representation at a low resolution of the series, it is necessary that the *scaling function* ϕ_0 (from which we calculate $\phi_{J,k}$ and $\sum_k a_{J,k} \phi_{J,k}(t)$) *is a low-pass filter.* As regards the terms "$detail_j$" which correspond to the differences $approx_j(t) - approx_{j-1}(t)$, the function ψ_0, which enters in their construction, is *a band-pass filter. The band-pass filter* (which in spite of its source because it historically belongs to the Fourier analysis) *corresponds also to a wavelet through its form.*

Scaling function: Within the general framework of the signal analysis, it is known that a *scaling function* is interpreted as the *impulse response* of a *low-pass filter. From the gauge of a generic filter band-pass*, we construct filters of the low-pass and high-pass type, which filter by means of their specific forms the Fourier coefficients, either from the high frequencies towards the low frequencies or from the low frequencies towards the high frequencies. If we note $\phi_s(t) = \frac{1}{\sqrt{s}} \phi\left(\frac{t}{s}\right)$ (and with $\overline{\phi}_s(t) = \phi_s^*(-t)$), then the *low frequency approximation* of a time series $x(t)$ *at the scale s* is written: $Lx_{u,s} = \left\langle x(t), \frac{1}{\sqrt{s}} \phi\left(\frac{t-u}{s}\right)\right\rangle = x \star \phi_s^*(u).$

Definition of a scaling function: As presented previously, the approximation of a time series $x(t)$ at the resolution 2^j is defined as an orthogonal projection of the time series $x(t)$ on $\mathbf{V}_j$: $\text{Proj}_{V_j} x(t)$. To calculate this projection, we have to be endowed with a basis of $\mathbf{V}_j$. The Multiresolution theory says that it is possible to construct an orthonormal basis from each space $\mathbf{V}_j$ by *dilation* and *translation* of a simple function ϕ_0 called the *scaling function. A scaling function can be defined coarsely as an aggregation of wavelets at scales higher than 1.*

Approximation of coefficients: The projection of $x(t)$ on $\mathbf{V}_j$ is obtained by a progressive dilation of the scale:

$$\text{Proj}_{V_j} x = \sum_{k=-\infty}^{+\infty} \left\langle x, \phi_{j,k} \right\rangle \phi_{j,k}. \tag{5.257}$$

The inner products $a_j[k] = \left\langle x, \phi_{j,k} \right\rangle$ provide a discrete approximation at the scale 2^j, from which we can rewrite them as the result of a convolution: $a_j[k] = \int_{-\infty}^{+\infty} x(t)$

$\frac{1}{\sqrt{2^j}}\phi\left(\frac{t-k}{2^j}\right)dt$, or $a_j[k] = \int_{-\infty}^{+\infty} x(t)\, 2^{-j/2}\phi_0(2^{-j}t - k)dt$. The discrete approximation $a_j[k]$ is *low-pass filtering* of $x(t)$ sampled on the interval 2^j. We express the $a_{J,k}$ as follows:

$$a_{J,k} = \int x(t) 2^{-j/2}\phi_0(2^{-j}t - k)dt. \tag{5.258}$$

By analogy, we obtain the coefficients $d_{j,k}$ *no more by means of a scaling function of the low-pass filter type, but by a band-pass filter*, which *also corresponds to a wavelet built from a mother wavelet* ψ_0 (note *this correspondence between filter band-pass resulting from the Fourier analysis and the wavelet analysis*). The projection of $x(t)$ on $\mathbf{W}_j$ is obtained by the filter of the *wavelet function:* $2^{-j/2}\psi_0(2^{-j}t - k)$. The inner products $\langle x, \psi_{j,k}\rangle$ provide an approximation at the scale 2^j. The coefficient at the scale 2^{-j} is written:

$$d_{j,k} = \int x(t) 2^{-j/2}\psi_0(2^{-j}t - k)dt. \tag{5.259}$$

The index j corresponds at the resolution level of the wavelet analysis, j is also called octave. The octave j is the logarithm of the scale 2^j, and k plays the role of time.[50] n is the length of the time series. The number of available coefficients at the octave j is denoted $n_j = 2^{-j}n$.

Discrete Wavelet Transform: The *discrete non redundant transformation* allows the transition from a Hilbert space $\mathbf{L}^2(\mathbb{R})$ *constructed on the real numbers to* the Hilbert space $\mathbf{l}^2(\mathbb{Z})$ *constructed on the integer numbers* $\mathbb{Z}$.[51] Provided with a *scaling function* ϕ_0 and a *mother wavelet* ψ_0, the discrete transform products from the time series $x(t)$ the set $\{\{a_{J,k}\}_{k\in\mathbb{Z}}, \{d_{j,k}, j = 1,\dots,J\}_{k\in\mathbb{Z}}\}$. We observed previously that these coefficients are the result of the inner products $\langle x, \phi_{j,k}\rangle$ and $\langle x, \psi_{j,k}\rangle$ and where $\phi_{j,k}$ and $\psi_{j,k}$ are obtained by *dilation and translation respectively* of the *gauge* of a *scaling function* ϕ_0 and of *a mother wavelet* ψ_0. The algorithm used to produce the discrete wavelet transform is the *recursive pyramidal algorithm* (APR).

General remark about the notations: It is noted that $\mathbf{W}_j$ is a closed subspace generated by an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$ (i.e. here a family of wavelets at the scale 2^{-j}) :$\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\}_{j\in\mathbb{Z}}$. We can write:

$$L^2(R) = \bigoplus_{j\in\mathbb{Z}} W_j. \tag{5.260}$$

The analysis of a time series by scale level is carried out through the introduction of the approximation spaces noted $\mathbf{V}_j$, which are defined by:

$$V_{l+1} = \bigoplus_{j\le l} W_j,\; l \in Z. \tag{5.261}$$

---

[50] The *dyadic wavelet transforms* are of large interest (they are discrete in scale but continuous in time) because they make it possible not to sub-sample the signals when we pass on to coarse scales and thus not to deteriorate a signal with low resolutions.

[51] Generic writing of the DWT: $Wf[n, a^j] = \sum_{n=0}^{N-1} f[m]\frac{1}{\sqrt{a^j}}\psi_j^*[m - n]$ with a wavelet $\psi_j[n] = \frac{1}{\sqrt{a^j}}\psi_j(\frac{n}{a^j})$.

5.16 Singularity and Regularity of a Time Series: Self-Similarities, Multifractals and Wavelets

We had the opportunity to underline several times the property of wavelets to provide information simultaneously in time and frequency. This ability to locate in frequency the events is interesting. The *singularities* and the *irregularities* of a signal provide information as fundamental as the regularities or periodicities for example. We know that a (positive or negative) *peak very localized* in a *financial series* or in an *electrocardiogram contains fundamental information*. Mallat (1998, p. 163) textually highlights that "the local signal regularity is characterized by the decay of the wavelet transform amplitude across scales". Moreover, he underlines that "non-isolated singularities appear in complex signals such as multifractals". Multifractals are subjacent in numerous signals of natural origin. "The wavelet transform takes advantage of multifractals self-similarities" in order to calculate the distribution of their singularities. *Thus, the time-scale analysis (with its singularity spectrum) and the wavelets contribute to the knowledge of the properties of multifractals.*

In order to present the *singularity* and the *regularity* of a signal it is necessary before to point out *the Lipschitz conditions*, which were already presented in the part I. With this intention, as a preliminary, we present the Taylor approximation formula and the associated approximation error. Given the function f which is n times differentiable on a interval centered at x_0: $[x_0 - h, x_0 + h]$, the Taylor formula is written:

$$p_{x_0}(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(t - x_0)^k. \tag{5.262}$$

And the approximation error $\varepsilon(t) = f(t) - p_{x_0}(t)$ verifies with $u \in [x_0 - h, x_0 + h]$:

$$\forall t \in [x_0 - h, x_0 + h], |\varepsilon(t)| \leq \frac{|t - x_0|}{n!} \sup |f^n(u)|, \tag{5.263}$$

The nth order differentiability of f around x_0 provides the upper bound of the error $\varepsilon(t)$ when $t \to x_0$. At this stage, *the Lipschitz exponent allows to specify this upper bound*.

5.16.1 Lipschitz Exponent (or Hölder Exponent): Measurement of Regularity and Singularity by Means of the Hölder Functions $\alpha(t)$

The *Fourier analysis* makes it possible to characterize the *global regularity* of a function and the *wavelet transform* makes it possible to analyze the *pointwise regularity*. This distinction exists also in the *definition* of *the regularity in the Lipschitz sense* (also called Hölder regularity), we speak of *pointwise or uniform regularity*. This *Lipschitz or Hölder regularity concept* was extended recently to more robust approaches such as the *regularity 2-microlocal* (not presented here). Moreover, there

are multiple manners of carrying out a *fractal analysis* of a signal, *the calculation of the pointwise regularity is one of them,* and the *multifractal analysis* is an another one also. In the first case, we associate with a signal $f(t)$ another signal $\alpha(t)$, which is called the Hölder function of f, which measures the regularity of f *at each point* t. The *regularity* of f can be *evaluated* by means of different methods, the *pointwise Lipschitz (or Hölder) exponent* α of f at x_0, is one of those methods, another one is the *local exponent* which is often written $\alpha_L(x_0)$. But let us present the initial definition of the regularity in the Lipschitz sense.

Definition 5.4 (Regularity in the Lipschitz sense).

- *A function f is pointwise Lipschitz $\alpha \geq 0$ at x_0, if there exist $K > 0$ and a polynomial p_{x_0} of degree n, where n is equal to the largest integer $n \leqslant \alpha$ such that:*

$$\forall t \in \mathbb{R}, \ \left| f(t) - p_{x_0}(t) \right| \leq K \left| t - x_0 \right|^{\alpha}. \tag{5.264}$$

- *A function f is uniformly Lipschitz-α over $[a,b]$, if it satisfies the preceding condition for all $x_0 \in [a,b]$, with a constant K which is independent of x_0.*
- *The Lipschitz regularity of the function f at x_0 or over $[a,b]$ is the sup of the α such that f is Lipschitz-α.*

If f is n time continuously differentiable (with n equal to the largest integer) in a neighborhood of x_0, then the polynomial p_{x_0} is the Taylor approximation of f at x_0.[52]

It is remarkable to note that "it is possible to construct multifractal functions with non-isolated singularities", for which the signal has a different Lipschitz regularity at each point. However, the uniform Lipschitz exponents provide a more global measure of regularity on the entire interval. With a signal f *uniformly Lipschitz $\alpha > n$* in the neighborhood of x_0, it is necessary to verify that f is n times continuously differentiable. Otherwise, i.e. for a f non-differentiable, α will characterize the type of singularity. We will note simply that if the condition: f is n times continuously differentiable is verified for $0 \leq \alpha < 1$, then $p_{x_0}(t) = f(x_0)$ and the Lipschitz condition $\forall t \in \mathbb{R}, \left| f(t) - p_{x_0}(t) \right| \leq K \left| t - x_0 \right|^{\alpha}$ becomes:

$$\forall t \in \mathbb{R}, \ \left| f(t) - f(x_0) \right| \leq K \left| t - x_0 \right|^{\alpha}. \tag{5.265}$$

The main idea could be summarized as follows: *For the pointwise regularity we associate a signal $f(t)$ with another signal $\alpha(t)$ which is the Hölder function of f and which measures the regularity of f at each point t.*

This *pointwise regularity* is evaluated by means of:

1. The *pointwise Lipschitz exponent* α of f at x_0, which can be defined as follows:

$$\alpha(x_0) = \lim_{\rho \to 0} \sup \{ \alpha : \exists K > 0, |f(t) - f(x_0)| \leq K |t - x_0|^{\alpha}, |t - x_0| < \rho \}$$

   if α is non-integer and if f is non-differentiable. Otherwise, it is necessary to replace in the expression above the $f(x_0)$ term by a polynomial $p_{x_0}(t)$.

---

[52] For each x_0 the polynomial is unique.

2. The *local exponent* $\alpha_L(x_0)$ such that:

$$\alpha_L(x_0) = \lim_{\rho \to 0} \sup\{\alpha \ : \ \exists K > 0, |f(t) - f(y)| \leq K|t - y|^{\alpha}, |t - x_0| < \rho,$$
$$|y - x_0| < \rho\}.$$

It will be noted that α and α_L have *different properties and usually do not correspond,*[53] we give the following example, if we consider the signal f as a function such that: $f(t) = |t|^{\alpha} \sin\left(1/|t|^{\gamma}\right)$, then $\alpha(0) = \alpha$ and $\alpha_L(0) = \alpha/(1+\gamma)$.

Generally the value of $\alpha(t)$ give us the following indications about the *regularity* and the *continuity* of f:

- If $\alpha(t)$ is small, then the function (or signal) f is *irregular* at t.
- If $\alpha(t) > 1$: f is *at least differentiable once* at t.
- If $\alpha(t) < 0$: there is a *discontinuity* on f.

The Lipschitz (or Hölder) exponent allows to *intuitively* represent the *regularity concept* and to *characterize a time-series* by means of their Hölder regularities. It is *used* in particular *in the signal processing to study the "turbulence phenomena"*. These methods are of a strong *interest* when the *irregularities* of a signal *contain important information*, as it is frequent besides.

5.16.2 n Wavelet Vanishing Moments and Multiscale Differential Operator of Order n

In order to study and measure the local regularity of a function or a signal, it is fundamental to use the wavelet vanishing moments. (Recall: A wavelet ψ has N vanishing moments if $\int t^k \psi(t)dt = 0$, $0 \leq k < N$, and $\int t^N \psi(t)dt \neq 0$.) Indeed, the wavelet transform of a signal which has n vanishing moments is interpreted as a multiscale differential operator of order n.[54]

The Lipschitz condition presented previously ($\forall t \in \mathbb{R}$, $|f(t) - p_{x_0}(t)| \leq K|t - x_0|^{\alpha}$) makes it possible to approximate f by a polynomial p_{x_0} in the neighborhood of x_0: $f(t) = p_{x_0}(t) + \varepsilon(t)$ with $|\varepsilon(t)| \leq K|t - x_0|^{\alpha}$. However, *a wavelet transform estimates the exponent* α ignoring the polynomial p_{x_0}. In order to do this, a wavelet is selected which has $n > \alpha$ vanishing moments: $\int t^k \psi(t)dt = 0$, $0 \leq k < n$. A wavelet with n vanishing moments is orthogonal to the polynomials of degree $n - 1$. And because $n > \alpha$, the polynomial p_{x_0} has a degree at most equal to $n - 1$. If one carries out a change of variable you $t' = (t - u)/s$, we verify the following transform:

$$W p_{x_0}(u, s) = \int p_{x_0}(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t - u}{s}\right) dt = 0. \tag{5.266}$$

because $f = p_{x_0} + \varepsilon : W p_{x_0}(u, s) = W\varepsilon(u, s)$.

---

[53] **Stable by differentiation:** One of the properties which is respected by α_L (and which is not respected by α) is the property *to be stable by differentiation* $\alpha_L(f', x_0) = \alpha_L(f, x_0) - 1$.

[54] See in the appendix the *definition of the differentiable operators in Banach spaces*.

The transformation of a signal by a wavelet which has n vanishing moments is interpreted as a (multiscale) differential operator of order n. There is a relation between the differentiability of f and the decay of its wavelet transform at the fine scales. And it is possible to write that *a wavelet* with *n vanishing moments* can be written *as the derivative of order n of a function θ.*

Theorem 5.5 (Vanishing moments of wavelet with a fast decay). *A wavelet* ψ *with a fast decay has n vanishing moments if and only if there exists* θ *with a fast decay such that:*

$$\psi(t) = (-1)^n (d^n \theta(t)/dt^n). \tag{5.267}$$

Consequently:

$$Wf(u,s) = s^n \frac{d^n}{dt^n}(f \star \overline{\theta}_s)(u), \quad \text{with } \overline{\theta}_s(t) = (\theta(-t/s)/\sqrt{s}). \tag{5.268}$$

And ψ *have no more than n vanishing moments if and only if* $\int_{-\infty}^{+\infty} \theta(t)dt \neq 0$.

5.16.3 Regularity Measures by Wavelets

In the prolongation of what was presented above, the decay of the *amplitude* of the wavelet transform along the scales is associated with the uniform and pointwise Lipschitz regularities of a signal. *To measure this decay comes down to observe the structure of the signal while varying the scale.* If we choose a n times differentiable wavelet with n vanishing moments, it comes: $\exists C_\gamma$, with $\gamma \in \mathbb{N}$, $\forall t \in \mathbb{R}$, $\left| \psi^{(k)}(t) \right| \leq C_\gamma/(1+|t|^\gamma)$.

Theorem 5.6 (Lipschitz uniform). *If* $f \in \mathbf{L}^2(\mathbb{R})$ *is uniformly Lipschitz* $\alpha \leq n$ *over* $[a,b]$, *then there exists* $A > 0$ *such that:*

$$\forall (u,s) \in [a,b] \times \mathbb{R}^+, |Wf(u,s)| \leq As^{\alpha+(1/2)}. \tag{5.269}$$

Reciprocally, let us suppose that f is bounded and that $Wf(u,s)$ *satisfy the condition above for an* $\alpha < n$ *which is not an integer, then f is uniformly Lipschitz* α *over* $[a+\varepsilon, b-\varepsilon]$ *for any* $\varepsilon > 0$.

Theorem 5.7 (Jaffard). *If* $f \in \mathbf{L}^2(\mathbb{R})$ *is Lipschitz* $\alpha \leq n$ *at* x_0, *then there exists A such that:* $\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+$,

$$|Wf(u,s)| \leq As^{\alpha+(1/2)}(1+|(u-x_0)/s|^\alpha). \tag{5.270}$$

Reciprocally, if $\alpha < n$ *is not an integer and there exist A and* $\alpha' < \alpha$ *such that* $\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+$,

$$|Wf(u,s)| \leq As^{\alpha+(1/2)}(1+|(u-x_0)/s|^{\alpha'}), \tag{5.271}$$

then f is Lipschitz α *at* x_0.

This condition establishes a relationship between the pointwise regularity of a signal and the decay of the modulus of its wavelet transform. To give an illustration of the condition above we can represent the cone of influence of a point x_0. If we consider a mother wavelet with compact support $[-C, C]$, we obtain by modulations of scales $\psi_{(u,s)} = \psi((t-u)/s)/\sqrt{s}$, and the modulations on the compact support provide $\psi((t-u)/s) : [u - Cs, u + Cs]$.

The cone of influence of x_0 is written $|u - x_0| \leq Cs$. If u is in the cone of influence of x_0, there is the wavelet transform $Wf(u,s) = \langle f, \psi_{u,s} \rangle$ which depends on the value of f in the neighborhood of x_0. We observe the cone of influence of x_0 in the *scale-plane* (u,s).
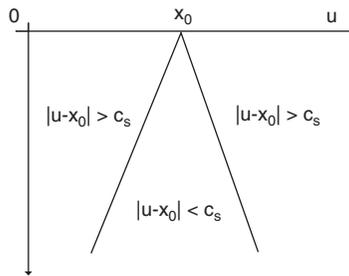
Since $|u - x_0|/s \leq C$, the conditions of the theorem above is written thus:

$$| Wf(u,s) | \leq A' s^{\alpha+(1/2)} \tag{5.272}$$

which corresponds to the *first theorem* of a *uniformly Lipschitz function f*.

The two preceding theorems show that the Lipschitz regularity of a function f at x_0 depends on the decay of the modulus $|Wf(u,s)|$ in the neighborhood of x_0.

5.16.4 Detection of Singularities: The Maxima of the Modulus of Wavelet Transform are Associated with the Singularities

The theorem of Hwang and Mallat shows that there is a maximum at fine scales when a signal contains singularities. Indeed, it is shown that *there cannot be singularity without local maximum of the modulus of the wavelet transform at fine scales.* The theorem is thus interested in the modulus of the wavelet transform and obviously in the abscissa to which the singularity is identified on the signal. In general, we detect *a succession of modulus maxima converging towards the singularity.* The notion of *modulus maximum $|Wf(u,s)|$* is described by means of the derivative at u of the transform:

$$\frac{\partial Wf(u,s)}{\partial u} = 0. \tag{5.273}$$

with s given, we search a local maximum at u.[55] (Remember that ψ is a wavelet, f is the signal-function and θ is a function.)

Theorem 5.8 (Hwang, Mallat). *Suppose that ψ is of the class C^n with a compact support and that $\psi = (-1)^n \theta^{(n)}$ with $\int_{-\infty}^{+\infty} \theta(t)dt \neq 0$. Let $f \in \mathbf{L}^1[a,b]$. If there exists $s_0 > 0$ such that the modulus $|Wf(u,s)|$ has no local maximum for $u \in [a,b]$ and $s < s_0$, then f is uniformly Lipschitz n on $[a+\varepsilon, b-\varepsilon]$ for any $\varepsilon > 0$.*

Thus, the function-signal f can be singular at a point x_0, which means not Lipschitz-1, if there exists *a sequence of wavelet maxima points* (u_α, s_α) that converges towards x_0 at the fine scales,[56] i.e. if $\alpha \in \mathbb{N}$:

$$lim_{\alpha \to +\infty}(u_\alpha, s_\alpha) = (x_0, 0). \tag{5.274}$$

The decay rate of maxima on the curves indicates the order of the isolated singularities.[57] In general, we make appear the logarithm of the modulus of the wavelet transform $\log_2 |Wf(u,s)|$ in a graph in the plane $[log_2(s), \log_2 |Wf(u,s)|]$. That means that it is possible to express the modulus maxima according to the scale in a log-log plane and *the slope obtained provides the estimated order of singularity.*

5.16.4.1 Examples of Local Maxima for the Continuous Wavelet Transform of the Stock Exchange Index: *Cac40*

The calculation of the local modulus maxima of the continuous wavelet transform for 2,048 daily values of the *Cac40 index*, then for its daily growth rate (Fig. 5.37).

The modulus maxima correspond to the "*ridges* curves" of images. Figure 5.38 represents in a log–log plane, the amplitude along the ridges of the continuous wavelet transform for a selection of six ridges.

5.16.5 *Self-Similarities, Wavelets and Fractals*

The most familiar way to approach the concept of *self-similarity* or more exactly the concept of *self-affinity*, that it is possible to observe for example *in the high frequencies of the financial series* or in *the Internet traffic*, can be presented from

---

[55] In detail, *the wavelet transform is rewritten as a multiscale operator of order n ($\psi(t) = (-1)^n (d^n \theta(t)/dt^n)$; ψ with n vanishing moments) as presented previously: $Wf(u,s) = s^n \frac{d^m}{du^m}(f \star \overline{\theta}_s)(u)$. The multiscale modulus maxima are used to analyze the discontinuities* of a signal (see S. Mallat).

[56] Stephane Mallat underlines that they are modulus maxima of the transform. Whereas the instantaneous frequencies are detected while following the maxima of the normalized scalogram $(\xi/n)P_W f(u,\xi)$.

[57] $\log_2 |Wf(u,s)| \leq \log_2 A + (\alpha+1)\log_2 s$.
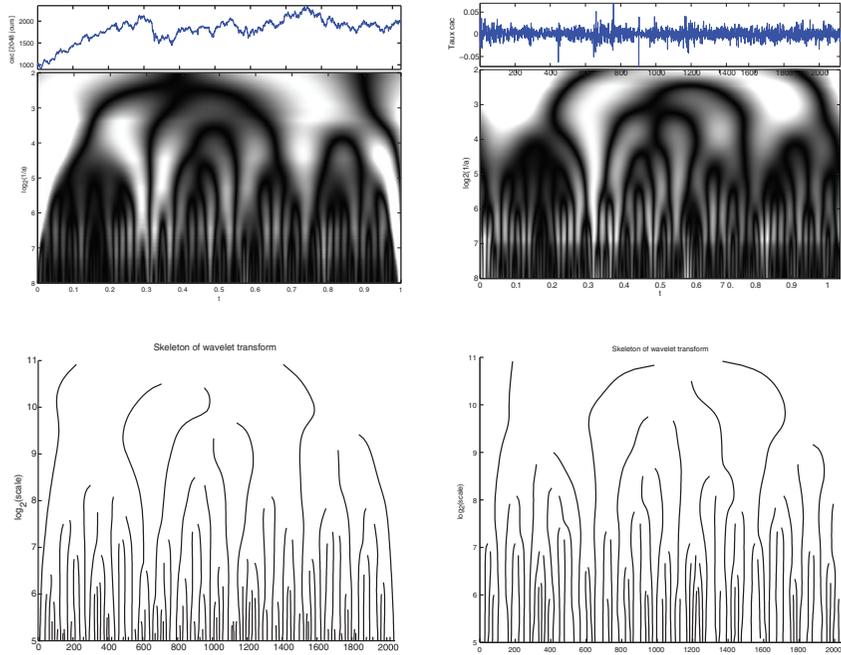
Fig. 5.37 Local maxima of the cwt
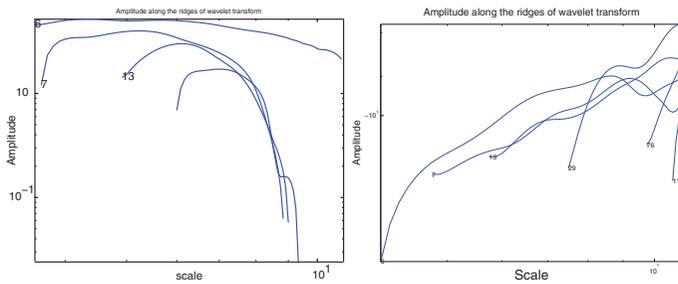
Fig. 5.38 Amplitude along the ridges

the *Hurst exponent* in the following way,[58] for a stochastic process $x(t), t \in \mathbb{R}^+$, $\exists$ $H > 0$:

$$x(t) = a^{-H} x(at) \quad \text{for any } a > 0. \tag{5.275}$$

The topological approach of the *definition of self-similarities* states that a set $S \in \mathbb{R}^n$ is *self-similar* (or self-affine), *if the union of the disjoint subsets $S_1, \ldots, S_m$ can be*

---

[58] Where the equality is understood as a equality by distribution $\overset{d}{=}$.

obtained from S by scales of translation and rotation. The examples of sets of this type are numerous, such the *Cantor set* presented in another section. This concept must be associated with the concepts of the *fractal dimension, Hausdorff dimension or capacity dimension,* which is a simplification of the Haussdorff dimension. It is possible to describe the generic concept of capacity dimension in the following way. Let us suppose a set of points P, in a *space of dimension d,* then if we imagine for example a *first return map* which *intersects the trajectory* of an unspecified dynamics; consequently the dynamics is lying on this plane. We imagine then a cube (or a "hypercube" of unspecified dimension) denoted ε, and we measure the number $N(\varepsilon)$ of ε which is necessary to *cover the set of points* P (explaining thus its other name of *box counting*). Then, D the capacity dimension of the point set P is equal to:

$$D = \lim_{\varepsilon \to 0} [\log N(\varepsilon)/\log(1/\varepsilon)]. \tag{5.276}$$

We can illustrate the subject while choosing for the set P a *single point,* consequently the number of ε necessary to cover P, is $N(\varepsilon) = 1$, and the dimension D is equal to 0. In the case where P is a *segment* of which the length by convention is equal to 1, then $N(\varepsilon) = 1/\varepsilon$ and $D = 1$. And in the case where P is a simple *plane,* by convention of side equal to 1, then $N(\varepsilon) = 1/\varepsilon^2$ and $D = 2$, and so on.

If we uses again the topological approach of the self-similarity, and if we replace the cubes (or hypercubes) of the "box-counting" approach by "radial structures" (that we could presage above) by the action of the rotation scale. Then we can consider our set S (bounded on $\mathbb{R}^n$) and we can *count the minimum number $N(s)$ of radial structures of radius s* which it is necessary to cover the set S. If S is a set of dimension D of finite size (such that $D = 1, D = 2, D = 3$) then,

$$N(s) \sim s^{-D}, \tag{5.277}$$

consequently, we have: $D = -\lim_{s \to 0}(\log N(s)/\log(1/s))$. Thus, the *capacity dimension D,* or fractal dimension, of our set is defined in a generic way by:

$$D = -\liminf_{s \to 0} \frac{\log N(s)}{\log s}. \tag{5.278}$$

And the measure of S is written: $\limsup_{s \to 0} N(s)\, s^D$.

Self-similarity: Beyond the definition of selfsimilar processes $x(t) = a^{-H}x(at)$ given above, the wavelet analysis of signals and self-affine functions, offers a very interesting perspective to introduce the subject. Let us pose a continuous function f with a compact support S. This function is selfsimilar, if there exist disjoint subsets $S_1, \ldots, S_m$ such as the representation of f on each subset S_i is an affine transformation of f. Thus, we have for any $t \in S_i$ an affine transformation by a translation a_i, a weight b_i, a scale d_i, with a constant c_i:

$$f(t) = b_i f(d_i(t - a_i)) + c_i. \tag{5.279}$$

It is said that the affine invariance of f on S_i produces an affine invariance for any wavelet whose support belongs to S_i. Moreover, It is said that if a function is self-similar, or self-affine, its wavelet transform is also self-similar. Moreover, *the self-similarity of the wavelet transform* means that the *positions* and *the modulus maxima are also self-similar*.

If an affine transformation ℓ of f is selected, it is written:

$$\ell(t) = bf(d(t-a)) + c, \tag{5.280}$$

the wavelet transform of ℓ is written:

$$W\ell(u,s) = \int_{-\infty}^{+\infty} \ell(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt. \tag{5.281}$$

Since a wavelet had an integral equal to zero $\int_{-\infty}^{+\infty} \psi(t) dt = 0$, by a change of variable $t' = d(t-a)$, it comes:

$$W\ell(u,s) = \frac{b}{\sqrt{d}} Wf(d(u-a), ds). \tag{5.282}$$

If we take a wavelet with non-infinite support, i.e. compact, for example on $[-C, C]$, the affine invariance of f on $S_i = [m_i, n_i]$ provides an affine invariance for any wavelet whose support is included in S_i. For any $s < (n_i - m_i)/C$ and for any $u \in [m_i + Cs, n_i - Cs]$:

$$Wf(u,s) = b_i(d_i^{-1/2}) Wf(d_i(u-a_i), d_i s). \tag{5.283}$$

5.16.6 Spectrum of Singularity: Multifractals, Fractional Brownian Motions and Wavelets

Definition 5.5 (Singularity spectrum). Let S_α be a set of points $t \in \mathbb{R}$ where the Lipschitz regularity of f is equal to α. The spectrum of singularity $D(\alpha)$ of f is the fractal dimension of S_α.

The singularity spectrum measures the global distribution of singularities having different Lipschitz regularities. And the distribution of singularities in a multifractal signal is used for the study of its properties. It is in the *complex phenomena of turbulences*, in particular in fluid dynamics and in geophysics that this type of work was born. The spectrum of singularity makes it possible to obtain the *proportion of Lipschitz-α singularities* which are present at any scale s. Indeed, the capacity dimension or fractal: $D = -\lim_{s\to 0} \inf(\ln N(s)/\ln(s))$, is based on a principle which consists in proceeding to a disjoint cover of the support of f by "objects" (i.e. radial structures, intervals, or hypercubes) of length s, of which the number on the set S_α is:

$$N_\alpha(s) \sim s^{-D(\alpha)}. \tag{5.284}$$

A multifractal function or signal is known as "homogeneous", if all the singularities have the same Lipschitz exponent α_0. That means that the support of $D(\alpha)$ is reduced to α_0. In order to illustrate the subject we will present the case of fractional Brownian motions which are homogeneous multifractals. It was shown previously that the *multifractals have non-isolated singularities*, and thus it is impossible to obtain their pointwise Lipchitz regularity. However, *it is possible to analyze the singularity spectrum of multifractals by means of wavelet transforms of the local maxima*. It is known that for a signal f, if the pointwise Lipschitz regularity $\alpha_0(<n)$ at x_0 (n is the vanishing moments of the wavelet), then the wavelet transform $Wf(u,s)$ has a sequence of modulus maxima $|Wf(u,s)|$ at the fine scale that converges towards x_0.

The maxima at the scale s can be taken as a "cover of the singular support" of f by means of the wavelet transform at the scale s. At these maxima locations we have:

$$|Wf(u,s)| \sim s^{\alpha_0+(1/2)}. \tag{5.285}$$

The locations of all these modulus local maxima of transforms at the scale s are represented by the sequence $u_\theta(s)$ (with $\theta \in \mathbb{Z}$: integers). By means of a partition[59] we can proceed to the sum of these modulus maximum at a power $q \in \mathbb{R}$, this sum is written: $\sum_\theta |Wf(u,s)|^q$. And the length between two consecutive maxima has a lower limit for a $\Delta > 0$, such that $|u_{\theta+1} - u_\theta| > (s\Delta)$. If the length between two maxima was shorter, i.e. equal or lower than the limit length $s\Delta$, the sum $\sum_\theta |Wf(u,s)|^q$ will include only the maxima of larger amplitude to "avoid the redundancies". If we pose the exponent $\tau(q)$ which evaluates the decay of the sum, we have:

$$\tau(q) = \liminf_{s\to 0} \frac{\log \sum_\theta |Wf(u,s)|^q}{\log s}, \tag{5.286}$$

which provides:

$$\sum_\theta |Wf(u,s)|^q \sim s^{\tau(q)}. \tag{5.287}$$

Under several conditions concerning the wavelet and its vanishing moments, $\tau(q)$ is the Legendre transformation[60] of the singularity spectrum $D(\alpha)$.

Theorem 5.9 (Self-similar signal, Arneodo, Bacry, Jaffard, Muzy). *Let* $\Lambda = [\alpha_{min}, \alpha_{max}]$ *be the support of $D(\alpha)$ and ψ is a wavelet with $n > \alpha_{max}$ vanishing moments. If f is a selfsimilar function or signal, then we obtain (the Legendre transformation of the spectrum of singularity $D(\alpha)$):*

$$\tau(q) = \min_{\alpha\in\Lambda}(q(\alpha+1/2) - D(\alpha)). \tag{5.288}$$

---

[59] **Definition (Partition).** If $\bigcup_{i\in I} A_i = G$ and if all the A_i are supposed to be different from $\emptyset$ and are pairwise disjoint, then the set $\{A_i\}$ is called a partition of G.

[60] *Legendre transformation:* A mathematical procedure in which one replaces a function of several variables with a new function which depends on partial derivatives of the original function with respect to some of the original independent variables. Also known as Legendre contact transformation.
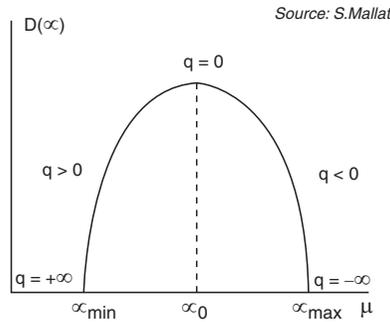
Fig. 5.39 Convex spectrum with $D(\alpha_0) = \max_{\alpha \in \Lambda} D(\alpha) = -\tau(0)$

It is noted that the scaling exponent $\tau(q)$ is the Legendre transformation $\tau(q) = \min_{\alpha \in \Lambda}(q(\alpha + 1/2) - D(\alpha))$ and is a convex function of q. This Legendre transformation is invertible (to recover the singularity spectrum $D(\alpha)$) if and only if $D(\alpha)$ is convex, thus:

$$D(\alpha) = \min_{q \in \mathbb{R}}(q(\alpha + 1/2) - \tau(\alpha)). \qquad (5.289)$$

Then it possible to write that "the $D(\alpha)$ spectrum of a selfsimilar signal is convex". Moreover, this observation can be applied to the majority of multifractals and fractional Brownian motions, under the condition of the convexity of the singularity spectrum. Refer to the works of A. Arneodo, E. Bacry, J.F. Muzy, S. Jaffard, Y. Meyer and S. Mallat about the singularity spectrum and its convexity. The Brownian motions are regarded as (Gaussian) selfsimilar processes and it is interesting to analyze their (wavelet and Fourier) singularity spectra (Fig. 5.39).

5.16.6.1 Power Spectrum and Wavelet Transform: Brownian Motion and Fractal Noise

The power spectrum is introduced to show that the differences or increments in a Brownian motion are stationary. But, let us note about the fractional Brownian motions that they:

- Are *non-stationary* and with *Gaussian increase* (i.e. *Gaussian increment*)
- Have *power spectra with fast decay*, (in spite of the difficulty in producing the spectrum because of non-stationarity)
- Are singular almost everywhere with the same Lipschitz regularity at all points

Definition 5.6 (Fractional Brownian motions). A fractional Brownian motion $x(t)$ of Hurst exponent $0 < H < 1$, is a process with zero-average Gaussian increase (increments), such that with $x(0) = 0$:

$$E[|x(t) - x(t - \tau)|^2] = \sigma^2 |\tau|^{2H}. \qquad (5.290)$$

It follows that a fractional Brownian motion "fBm" is singular almost everywhere with a pointwise Lipschitz regularity $\alpha = H$. There exists an additional property, which explains that the singularity of fBm decreases when H increases, and conversely when H decreases, the singularity increases. Moreover if $\tau = t$ is posed, it follows:

$$E[|x(t)|^2] = \sigma^2 |t|^{2H}. \tag{5.291}$$

Furthermore, it $\tau = t - u$, is posed, we obtain then:

$$E[x(t) \cdot x(u)] = \frac{\sigma^2}{2}(|t|^{2H} + |u|^{2H} - |t - u|^{2H}). \tag{5.292}$$

Remember *the autocovariance definition* for a *random process* and a *stationary process:*

Definition 5.7 (Autocovariance function of a random process). An autocovariance function of a random process x_t with finite variance, is written: $\gamma_k = Cov[x(t), x(t+k)] = E([x(t) - E(x(t))][x(t+k) - E(x(t+k))])$.

Definition 5.8 (Autocovariance function of a stationary process). An autocovariance function of a stationary process x_t with finite variance, is written:

$$\gamma_0 = Cov[x(t), x(t)] = E([x(t) - E(x(t))]^2) = Var(x(t)) = \sigma_x^2 \geqslant 0,$$
$$|\gamma_k| \leq \gamma_0 \text{ and } \gamma_k = \gamma_{-k}.$$

If we consider again the property $E[x(t) \cdot x(u)] = (\sigma^2/2) \cdot (|t|^{2H} + |u|^{2H} - |t - u|^{2H})$, it is said that the covariance of fBm does not depend solely on the argument $\tau = t - u$, which shows that a fBm is quite non-stationary.

Let us recall the definition of self-similar processes: $x(t) \overset{d}{=} a^{-H}x(at)$ or $x(at) \overset{d}{=} a^H x(t)$ (where $\overset{d}{=}$ means: Equal in distribution). It is by introducing a scale parameter on t which is noted a, that we verify the self-similarity characteristic. This way of making which consists in introducing a scale parameter on t, is familiar to the practitioners of the wavelet transform, although in general there are translations on t by u and modulations of the scale by $1/s$. In the case of the self-similarity property we note that the modulation of the scale over time is done with $1/s = a$. Beyond this remark, from the preceding expression: $E[x(t) \cdot x(u)] = (\sigma^2/2) \cdot (|t|^{2H} + |u|^{2H} - |t - u|^{2H})$, it follows:

$$E[x(at) \cdot x(au)] = E[a^H x(t) \cdot a^H x(u)], \tag{5.293}$$

and $x(at)$, $a^H x(t)$ are Gaussian with identical covariances and averages, and it is possible to write:

$$x(at) \overset{d}{=} a^H x(t). \tag{5.294}$$

As explained previously, one of the characteristics of the fBm is that they have power spectra with fast decay, in spite of the difficulty to produce the spectrum because of

non-stationariness. We circumvent the difficulty by *incrementing*,[61] thus the power spectrum is calculated on the increments of fBm which are stationary.

Power Spectrum of a Fractional Brownian Motion and of a Fractal Noise

If we pose, on the one hand, a fractional Brownian motion $fBm(t)$ and on the other hand $\Delta_\tau(t) = \delta(t) - \delta(t - \tau)$, the process resulting from the increments:

$$I\left[fBm(t)\right] = fBm(t) \star \Delta_\tau(t) = fBm(t) - fBm(t - \tau), \qquad (5.295)$$

is a stationary process whose spectrum is written:

$$\widehat{S}_{I\,[fBm]}(\lambda) = \frac{\sigma_H^2}{|\lambda|^{2H+1}} \, |\widehat{\Delta_\tau}(\lambda)|^2 . \qquad (5.296)$$

where λ is the frequency, $\widehat{\Delta_\tau}(\lambda)$ is the Fourier transform of $\Delta_\tau(t)$. And $(fBm(t) \star \Delta_\tau(t))$ is a continuous convolution. $\delta(t)$ a distribution of dirac.[62] From the lines which precede, it arises that *the increments of a (non-stationary) fractional Brownian motion are stationary*. And we can extend the subject by writing:

$$\widehat{S}_{fBm}(\lambda) = \frac{\widehat{S}_{I\,[fBm]}(\lambda)}{|\widehat{\Delta_\tau}(\lambda)|^2} = \frac{\sigma_H^2}{|\lambda|^{2H+1}} . \qquad (5.297)$$

The increments $I\left[fBm(t)\right]$ are stationary because in the expression of the power spectrum $\widehat{S}_{I\,[fBm]} = (\sigma_H^2 / |\lambda|^{2H+1}) \left|\widehat{\Delta_\tau}(\lambda)\right|^2$ *the multiplication by* $\left|\widehat{\Delta_\tau}(\lambda)\right|^2$ *allows to "remove" the energy explosion of low frequency*. And it is important to highlight that *the non-stationarity of* fBm *occurs in the energy burst at low frequency*. Note that $\left|\widehat{\Delta_\tau}(\lambda)\right|^2 = O(\lambda^2)$. (Recall: For two different discrete signals $f_1[n]$ and $f_2[n]$, to write $f_1[n] \sim f_2[n]$ is equivalent to write that $f_1[n] = O(f_2[n])$ and $f_2[n] = O(f_1[n])$. The function $O(.)$ means "of order #", indeed one can write that $f_1[n] = O(f_2[n])$, and there exists K such that $f_1[n] \le K \cdot f_2[n]$).[63] *The power spectrum is used to prove that the increments of a* fBm *are stationary*.

---

[61] In statistics we evaluate the degree of differentiation τth to measure the quality of the stationarization.

[62] *Dirac distribution:* These distributions are important in particular when one passes *from the continuous to the discrete*, and *during the transition between functions and real discrete series*. (They make it possible to relieve convergence problems.) A Dirac delta function $\delta(t)$ has a support reduced to $t = 0$, i.e. $\delta(0)$. And if one associates with a Dirac any function $f(t)$ its value in $t = 0$: $\int_{-\infty}^{+\infty} \delta(t)f(t)dt = f(0)$.

[63] In a similar way one defines the order $o(\cdot)$ of two discrete signals or function with $f_1[n] = o(f_2[n])$ by $\lim_{n\to\infty}(f_1[n]/f_2[n]) = 0$.

In a similar way to the Brownian motions, the power spectrum of a fractal noise $x(t)$ is decreasing and is written:

$$\widehat{S}_{x(t)}(\lambda) = \frac{\sigma_H^2}{|\lambda|^{2H+1}}. \tag{5.298}$$

This type of processes although generally non-Gaussian which can contain singularities of various types, have decreasing power spectra.

The Wavelet Transform of a Fractional Brownian Motion Is a Gaussian Stationary Process

The transformation is written:

$$W_{\text{fBm}}(u,s) = fBm \star \overline{\Psi}_s(u). \tag{5.299}$$

Around $\lambda = 0$, the modulus of the Fourier transform of a wavelet, which has at least one vanishing moment, is of order λ: $|\widehat{\psi}_s(\lambda)| = O(\lambda)$. *The purpose is to highlight the fact that the wavelet transform of a fBm on each scale is a Gaussian stationary process.*

$$\widehat{S}_{W_{\text{fBm}}}(\lambda) = s|\widehat{\psi}_s(s\lambda)|^2 \frac{\sigma_H^2}{|\lambda|^{2H+1}} = s^{2H+2}\widehat{S}_{W_{\text{fBm}}}(s\lambda). \tag{5.300}$$

The gaussianity and the self-similarity of the fBm show that the wavelet transform is selfsimilar on the scale: $W_{\text{fBm}}(u,s) = s^{H+1/2}W_{\text{fBm}}(u/s,1)$. The figure hereafter illustrates the wavelet transform of a fractional Brownian motion for $H = 0.7$.
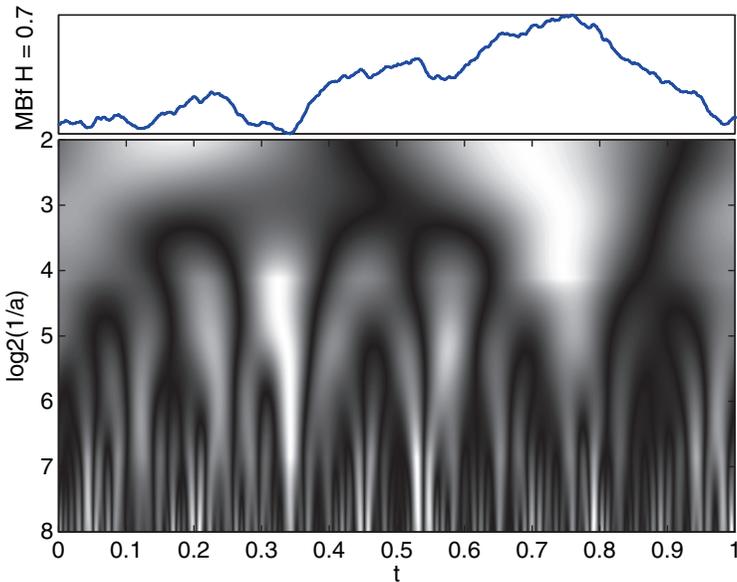
Hereafter two different white noises generated in an identical way and their wavelet transforms.
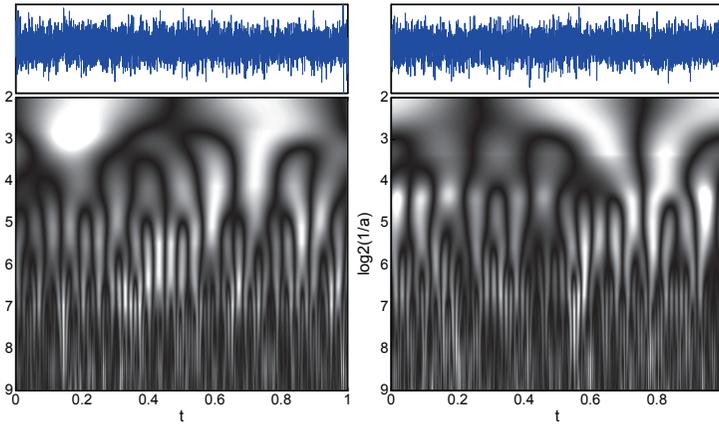
5.16.6.2 The Wavelet Transform of an Artificial Signal

The signal which is used as support here is successively made up of a Gauss curve, its derivative, then, its second derivative, a triangular, a sinusoid, a Dirac, a Morlet wavelet, a staircase, a random series and an increasing then decreasing oscillating structure ($n = 512$). Here is a continuous transforms by the Morlet wavelets and by the derivative[Gauss] wavelet in the time-scale plane (Figs. 5.40 and 5.41).
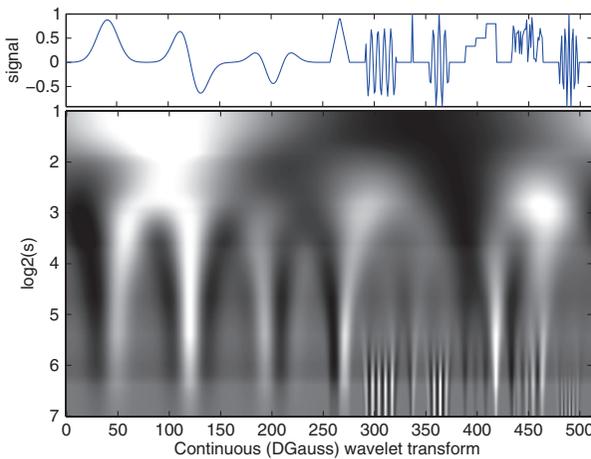
Fig. 5.40 Derivative[Gauss]-wavelet transform
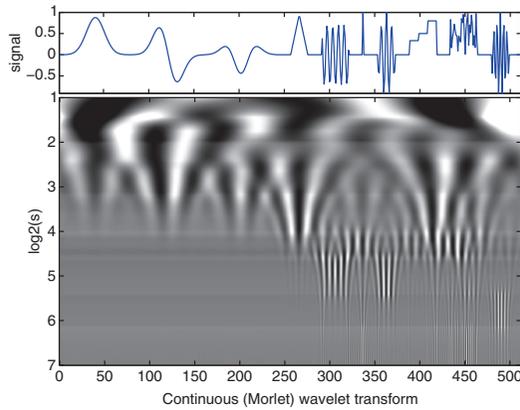
Fig. 5.41 Morlet-wavelet transform
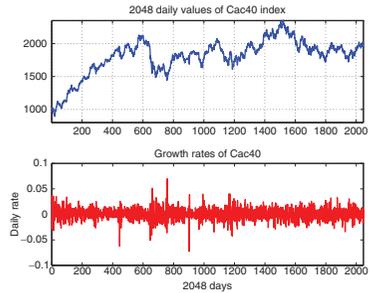
Fig. 5.42 2,048 daily values
and their growth rates

5.17 The Continuous Wavelet Transform

5.17.1 Application to a Stock Exchange Index: Cac40

The decomposed series is a sample of the *Cac40* index corresponding to 2,048 daily
values from January 1988 during more than 8 years (Fig. 5.42).

Let us recall that *the continuous wavelet transform* of a function (or a *signal*) f
at the scale s and at the position u *is calculated by correlating* f *with a wavelet* (or
with a wavelet atom):

$$W f(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt, \tag{a}$$

where ψ^* is the *complex conjugate* of ψ in $\mathbb{C}$. There is *equivalence* between the
expression above and the following equation which is written as a product of
convolution:

$$W f(u,s) = f \star \bar{\psi}_s(u) \tag{b}$$

with $\bar{\psi}_s(t) = \frac{1}{\sqrt{s}}\psi^*\left(-\frac{t}{s}\right)$. The Fourier transform of $\bar{\psi}_s(t)$ is:

$$\overset{\wedge}{\bar{\psi}}_s(\omega) = \sqrt{s}\hat{\psi}^*(s\omega) \tag{5.301}$$

with $\hat{\psi}(0) = \int_{-\infty}^{+\infty}\psi(t)dt = 0$, where $\hat{\psi}$ is the *transfer function*[64] of a *band-pass filter* of frequencies. Thus, *in practice* the *algorithms calculate the continuous wavelet transform by means of band-pass filters.*

The filtering is carried out in the space of the Fourier coefficients of the signal and in the space of the filter. The coefficients are successively filtered by the filter modulations. Then, we proceed to an inverse transform to return to the phase space of the signal at the chosen frequency scales, from the coarser to the finest. In this case, the length of the series which is the subject of the transformation is equal to 2,048, that means that it is a dyadic dimension ($n = 2^j$), i.e. $n = 2^j = 2^{11} = 2,048$. The finest scale of the decomposition of the signal is written: $\log 2(n) - 5 = \log 2(2,048) - 5 = \log 2(2,048) - \log 2(32) = 11 - 5 = 6$, and the coarsest scale is 2. We illustrate hereafter the method with a Gauss pseudo-wavelet at a fine scale, *we observe in the lower left part of Fig. 5.43 the coefficients filtered in the Fourier space.*

The result of the signal transformation at different scales is a matrix of vectors, it is represented in Fig. 5.44.
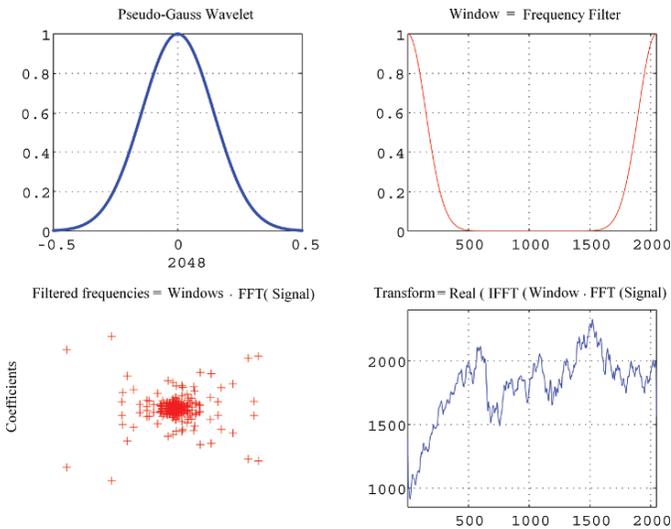
Fig. 5.43 Stages of the transformation by filtering

---

[64] *Transfer function:* The engineering terminology for a use of Fourier transforms. *By breaking up a wave pulse into its frequency spectrum:* $f_v = F(v)e^{2\pi i v t}$, the entire signal can be written as a sum of contributions from each frequency, $f(t) = \int_{-\infty}^{+\infty} f_v dv = \int_{-\infty}^{+\infty} F(v)e^{2\pi i v t}dv$.
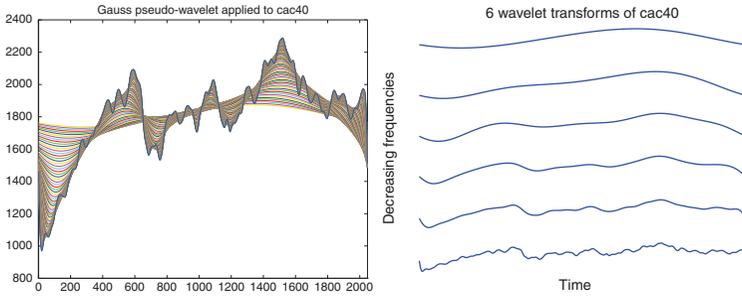
Fig. 5.44 Matrix of transforms (*left*). Range (scale) of transforms (*right*)
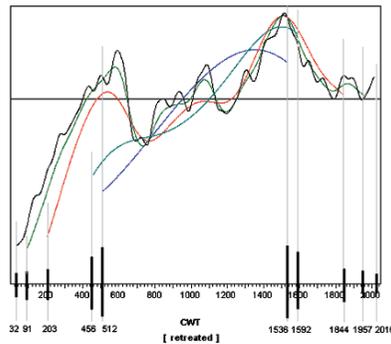
Fig. 5.45 Range of transforms used for the time-scale representation

The transform which considers at the same time the low frequencies and the high frequencies is the closest to the initial signal.

5.17.1.1 Images of Transforms in the Time-Scale Plane

The graphical representation of the continuous wavelet transform can be done in a time-scale plane with in abscissa u the time unit and in ordinate $Log2(s)$ the frequency scale. The axis of ordinates exploits the form of wavelet transforms while going from the lower part for the transforms at the high frequencies towards the upper part for the transforms at the low frequencies. (The colors or the scale of gray represents the amplitude.) Thus, the time, the frequency and the amplitude are depicted by means of this graphical representation. The scale denoted $\log 2(1/a)$ corresponds to a division of the frequency, with a minimum and a maximum respectively equal to 2 and $\log 2(n) - 5$. Without providing the detail of the construction of the wavelet transform matrix resulting from the algorithm used to build this image in the time-scale plane, we show simply through Fig. 5.45 that the divergent elements of the matrix at the extremities have been removed.

The amplitude of the transform is expressed by a scale of colors or a gray scale (if the color is not used). Here are the results of a transformation by a Gauss-window (i.e. a Gauss pseudo-wavelet) of the *cac40* (picture on the left) and of its growth rate (picture on the right):
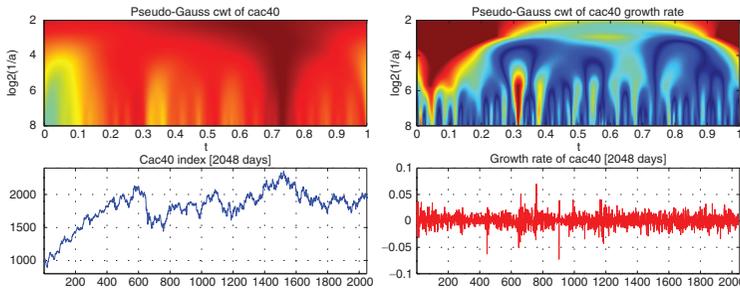
It is known that filtering by a Gauss-window (i.e. a Gauss pseudo-wavelet) does not stationarize the signal that contains a trend. Thus, the resulting image does not offer much information because the amplitudes are not centered and normalized. However, this Gauss filtering applied to the growth rate of the Cac40, offers more information. Indeed, the use of the growth rate (i.e. "differentiation") stationarizes the signal. Thus, the changes of amplitudes and the frequencies become more visible. In the time-scale planes, the amplitude differences at the limits of involved frequencies are depicted by means of a gray scale (or colors). In the left image the green and blue colors represent the Cac40 depressions, which are visible in particular at the beginning of the series. In the right image, although the *analyzing function* is the same one, the transformation is applied to the growth rate and provides more information. We observe the dark color (or red) in the upper left part and in the upper right part of the image, they depict the depressions (described previously) at the beginning and also at the end of the signal, which are not visible any more besides by means of a direct reading on the growth rate series. We now clearly observe the high frequencies which are in the lower part of the plane (and which are depicted by a blue scale) which express at the same time different amplitudes. Let us remark for example, at the middle of the frequency scale and around $t = 0.3$ and $t = 0.75$, a dark color (blue) with ramifications in the lower part of the image in the highest frequencies. This is the representation of the two signal rises, visible on the initial signal, but that we do not distinguish any more on the growth rate; they involve many frequencies but not all frequencies.

Image of the Mexican Hat-wavelet transform (or Sombrero): Hereafter, we show the rough matrix of wavelet transforms and different three-dimensional perspectives (3D) of this same matrix, with the contours projected on the bottom of the reliefs of the image. Moreover, a two-dimensional image of the continuous Sombrero-wavelet transform (scale: "color jet") is shown. The transformations by true wavelets center at zero the amplitudes and stationarize the signal, which is not the case with the Gauss pseudo-window (Figs. 5.46 and 5.47).

In the 2D image, the darkest colors (reds) express depressions of the signal. It is easy to identify on the left *a depression which involves almost all the frequencies*.
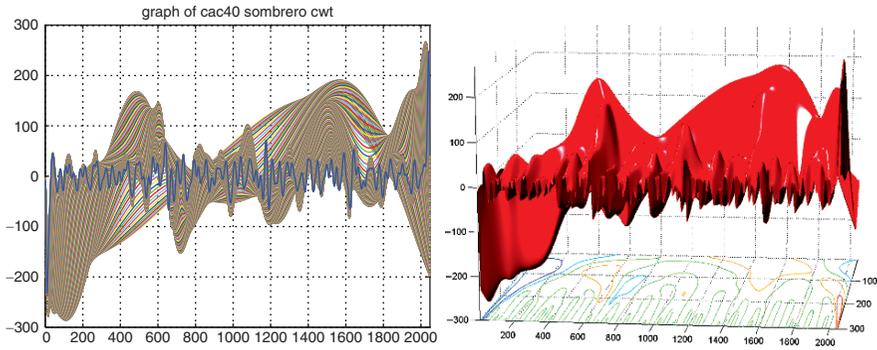
Fig. 5.46 Matrix of transforms (*left*). Perspective and contours (*right*)
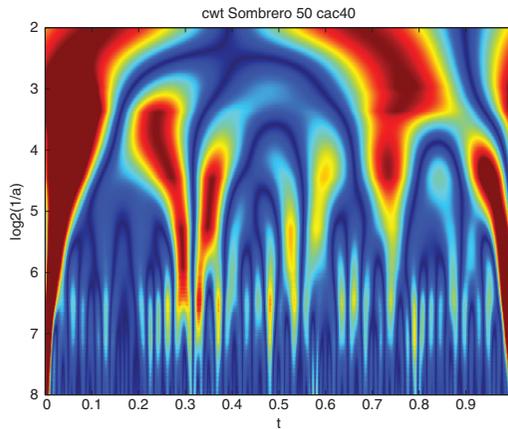
Fig. 5.47 Image of a CWT of the index Cac40 by sombrero wavelet (Jet version)

We identify smaller depressions that imply less frequencies. This is the case around $t = 0.75$ and around $t = 0.3$, for which the involved frequencies are different. We clearly observe at $t = 0.3$ that the covered frequencies are located on the scale between 3 and 7. Whereas for $t = 0.75$ the covered frequencies are located between 2 and 6. The high frequencies in the lower part of the image are numerous. Certain low frequencies in the upper part of the image are spread in the form of tree structures. It is the case for example of a (blue) dark tree structure at $t = 0.35$ which is divided into three very distinct parts, then is subdivided again until the highest frequencies in the lower part of the plane. That means that the observed "object", if it is indeed made of all the frequencies, does not cover however all the time axis but only sub-segments of this one.
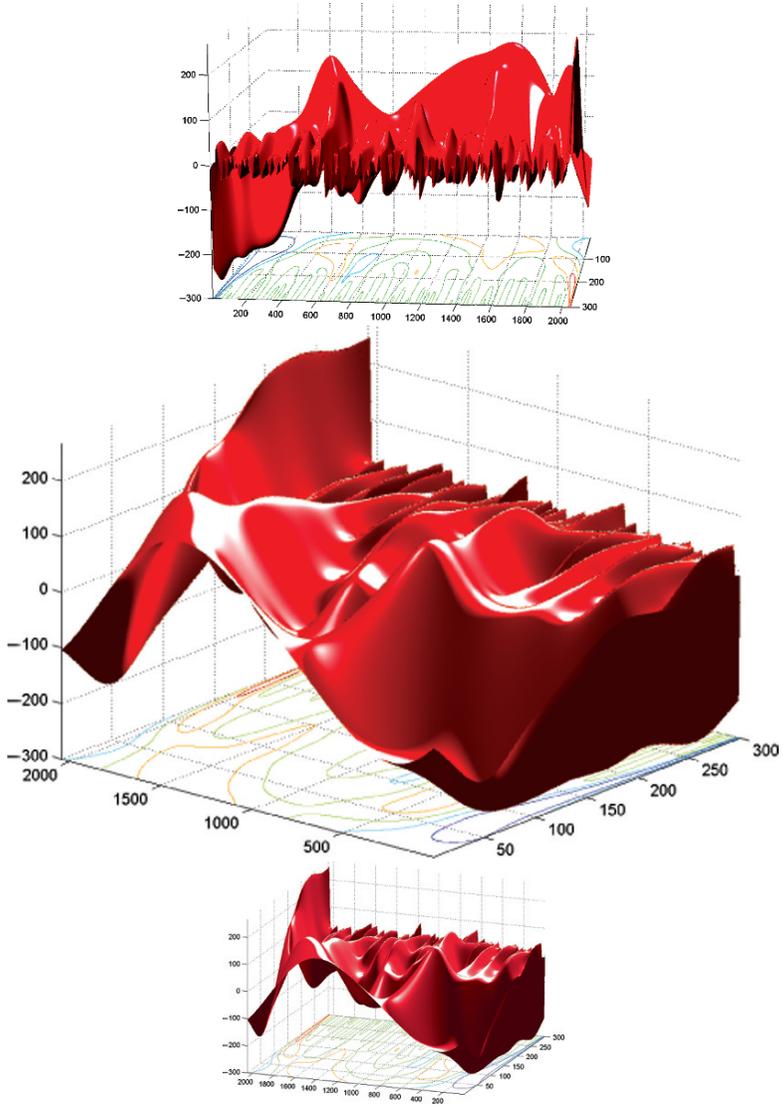
The option for the two images which follow is to use a gray scale to represent the amplitudes. The wavelet is a sombrero and the signal is the same French stock index as previously:
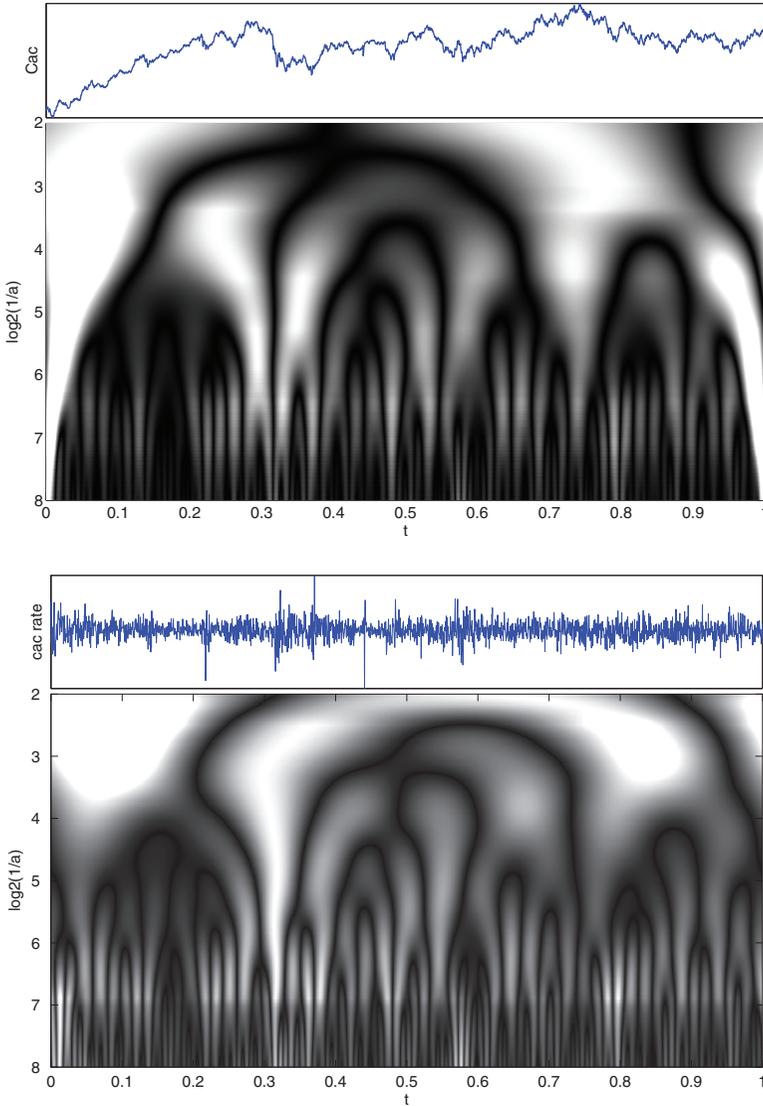
Compared images of transforms: The decomposed signal is the growth rate of the Cac40. The four types of decompositions were gathered in a same figure. We show a decomposition by means of a Gauss pseudo-wavelet, then by a true wavelet that

is the derivative of the Gauss pseudo-wavelet, then by a Morlet wavelet and finally by a sombrero wavelet. Note that the representations are rather different, but the fundamental subjacent structures can be still distinguished. The differences in the representation result from the shape of the analyzing wavelet.
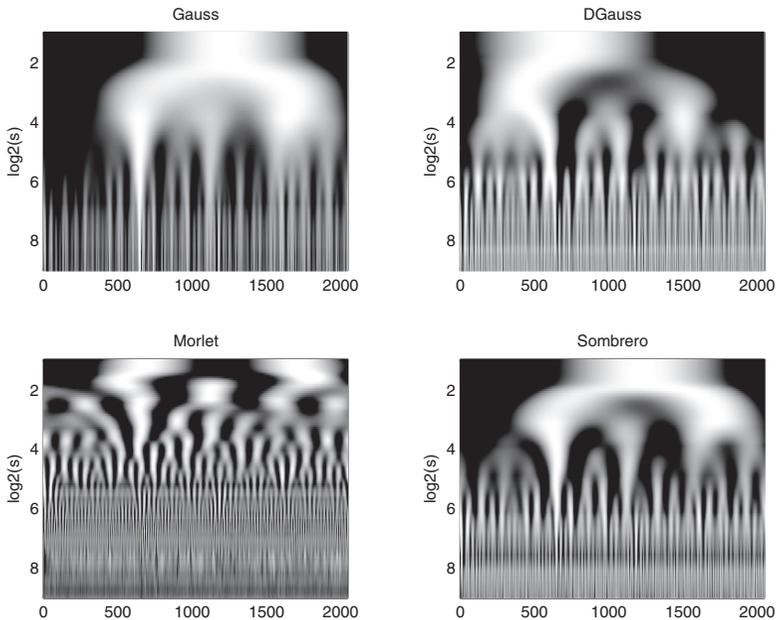
5.17.1.2 Details of the Transform Matrix

We show the elements of the transform matrix by means of two types of wavelets: the Gauss pseudo-wavelet and the Morlet wavelet. Note that the results highlight different structures of the signal. This results from the nature of the *analyzing waveform* but also from the fact that we initially analyze the rough signal (i.e. non-stationary) and then, its growth rate (i.e. stationary).

Each group of graphs represents the stock index (Cac40) or its *growth rate* and *its transform* at arbitrary octaves; The simple objective is to illustrate the evolution of the shape of transforms at different scales, from the finest towards the coarsest. We will notice how much for the highest octave (i.e. for the finest transform) the adaptation to the shape of the signal is strong, as an "interpolation", an "approximation", or an "estimation". Moreover, it is remarkable to note how much for the lowest octave, *the transform exhibits the shape of a wave whose cycle is spread over the entire length of the signal*.
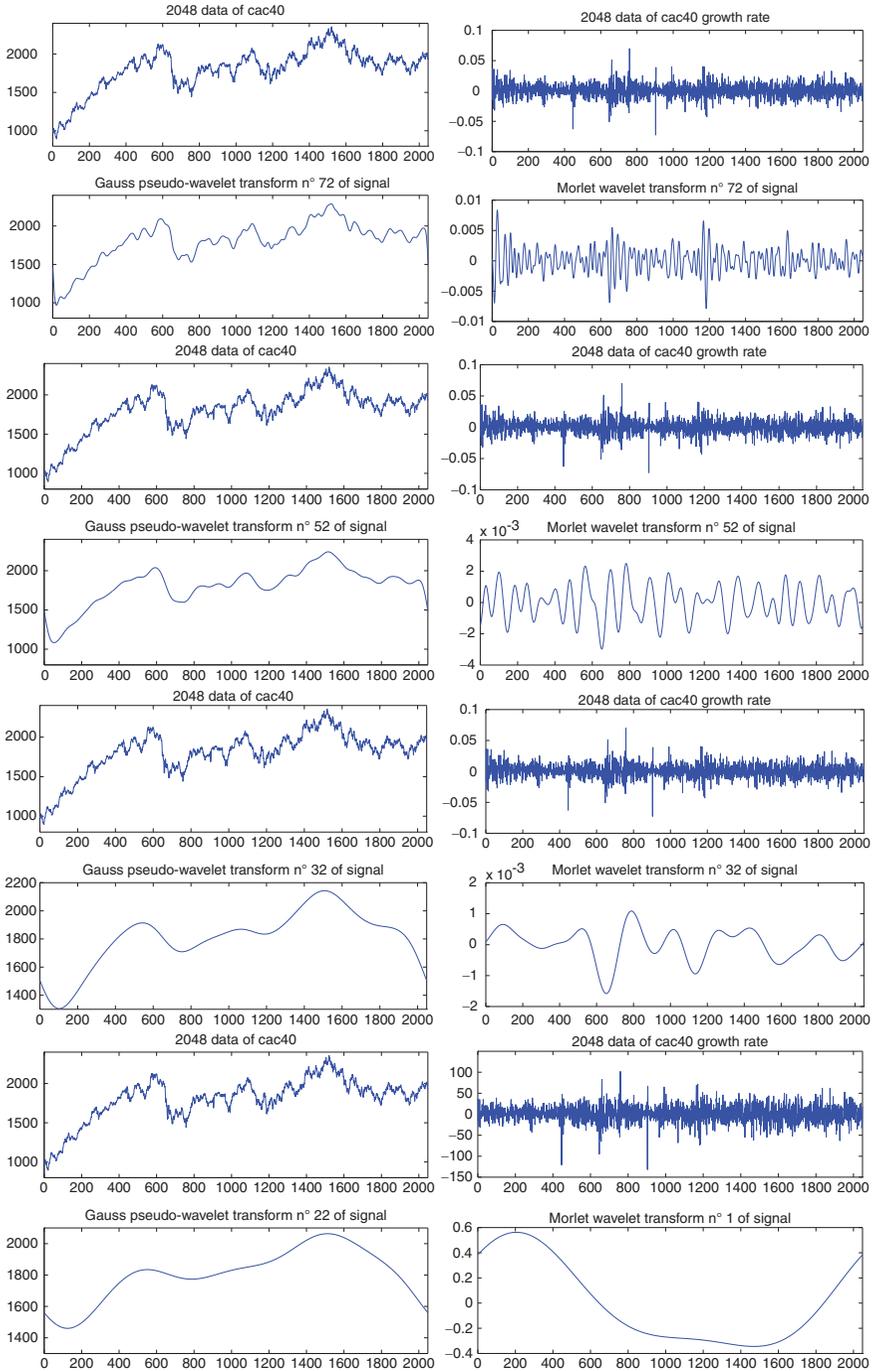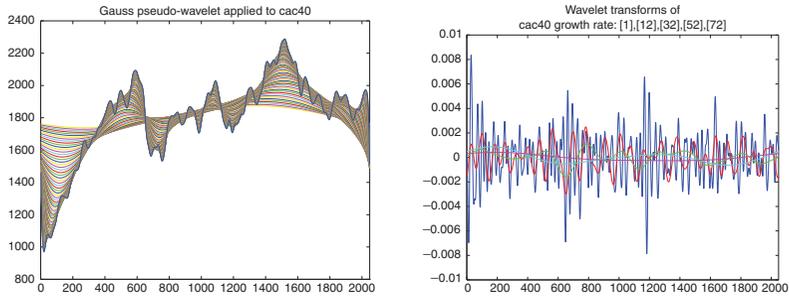
Fig. 5.48 Gauss pseudo-wavelet transforms (*left*). Morlet wavelet transforms (*right*)
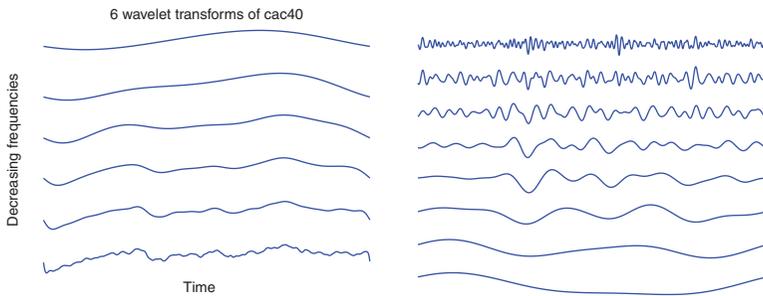
Fig. 5.49 Gauss pseudo-wavelet transforms (*left*). Morlet wavelet transforms (*right*)

Fig. 5.48 shows superimposed graphs of different transforms at different octaves.

Fig. 5.49 shows some transforms without superimposition. Note the decreasing scale of transforms in left-picture, and the increasing scale in the second.

5.18 Wigner–Ville Density: Representation of the Fourier and Wavelet Atoms in the Time–Frequency Plane

We will not describe the history of the Wigner–Ville density or distribution in spite of its great interest, which concerns physics and thermodynamics. As explained previously the concept of time–frequency atom allows to have a common notion to Fourier and wavelet analyses. The Wigner–Ville density is calculated by correlating the signal with a translation in time and frequency of itself. It is possible to do the same observation about time–frequency atoms. The Fourier transforms or wavelet transforms are also a correlation between the signal and its translation in time and frequency. Even if it is not the reason of its creation, since at the beginning the interest was the *representation* of the *instantaneous frequencies*, *nevertheless the Wigner–Ville density allows a type of representation in a time–frequency plane of time–frequency atoms, without resolution loss and without energy loss* (ref. to

Moyal's theorem[65]). The limit of resolution corresponds to the limit of time–frequency atoms themselves. The Wigner–Ville density or distribution $\mathbf{D}_{W.V}(u,\xi)$ (which is also a projector P_V on V) is written as follows:

$$\mathbf{D}_{W.V}f(u,\xi) = P_V f(u,\xi) = \int_{-\infty}^{+\infty} f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) e^{-i\pi\xi}d\tau. \qquad (5.302)$$

We can also write it in the frequency field by means of the Parseval formula:

$$\mathbf{D}_{W.V}f(u,\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \widehat{f}\left(\xi+\frac{\gamma}{2}\right)\widehat{f}^*\left(\xi-\frac{\gamma}{2}\right) e^{-i\gamma u}d\gamma. \qquad (5.303)$$

Remark 5.1. While leaving the framework of this section temporarily, it seems interesting to provide a simple illustration of the difference between the Wigner distribution and the wavelet transform in the time–frequency planes. An elementary Dirac function makes it possible to observe in Fig. 5.50 that the Wigner–Ville distribution does not spread the localization of the Dirac function, whereas the wavelet transform simultaneously spreads the Dirac function in time and frequency in the time-scale plane.

The Wigner–Ville distribution is recognized to be an important instrument of the time–frequency analysis. However, its main *critique* is to produce *interferences* because of *quadratic terms* (or *cross terms*) in its construction. These terms can be highlighted, by creating for example a signal built by means of two "sub-signals",
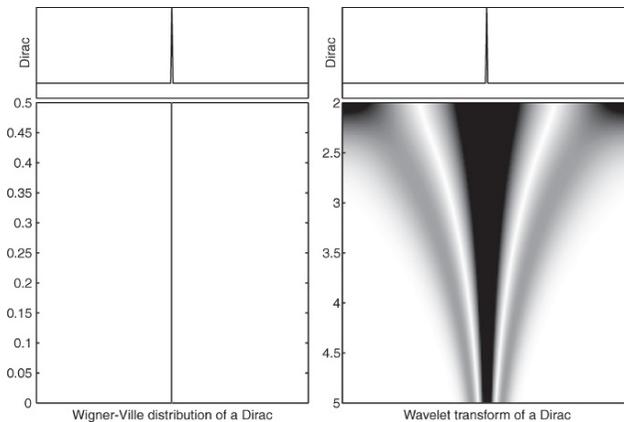
Fig. 5.50 Comparison (in *time–frequency* and *time-scale* planes) between the Wigner–Ville distribution and wavelet transformation of a *Dirac function*

---

[65] **Theorem (Moyal).** *For any f and g in* $L^2(\mathbb{R})$:

$$\left|\int_{-\infty}^{+\infty} f(t)g^*(t)\right|^2 = \frac{1}{2\pi}\int\int P_V f(u,\xi)\, P_V g(u,\xi)\, du\, d\xi.$$

For a demonstration see Moyal or Mallat.

the quadratic terms of the combination of the two sub-signals *reveal non-zero values which correspond to interferences:* $f = f_a + f_b$:

$$P_V f(u, \xi) = P_V f_a(u, \xi) + P_V f_b(u, \xi) + P_V[f_a, f_b](u, \xi) + P_V[f_b; f_a](u, \xi) \quad (5.304)$$

the *interference terms* are:

$$I(u, \xi) = P_V[f_a, f_b](u, \xi) + P_V[f_b; f_a](u, \xi). \quad (5.305)$$

The figure which follows represents a signal composed of two Gabor atoms of different frequencies. We will observe the localization in time and frequency of the energy of each atom, but also the interference at the center (either by means of a gray scale, or a color scale according to the representation mode).
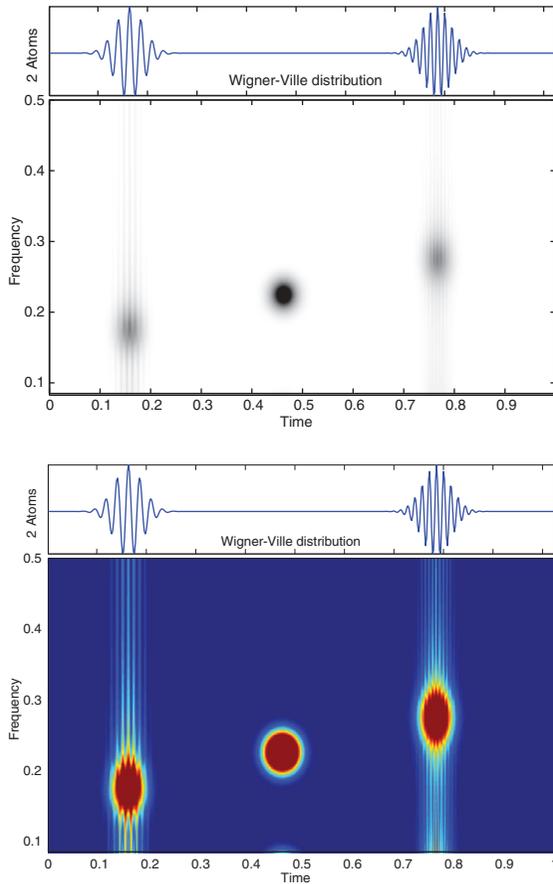
5.18.1 Cohen's Class Distributions and Kernels of Convolution

The *interference terms*, which are of *oscillatory nature* and with more or less complex structure, *can be "removed" or "attenuated" by the weighting action of kernels* θ (also called windows) in the Wigner–Ville distribution, which give in particular positive densities but nevertheless induced a resolution loss. Indeed, the interference terms contain positive and negative oscillations. They can be "removed" by means of the kernels θ. Then, it comes:

$$P_\theta f(u,\xi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P_V f(\overline{u},\overline{\xi})\theta(u,\overline{u},\xi,\overline{\xi})d\overline{u}d\overline{\xi}. \tag{5.306}$$

The invariance[66] by linear translation is a fundamental property of the time–frequency analysis.[67] And the theory of the time–frequency analysis shows that "the invariant operators of linear translations are convolution products" (Mallat 1998, p. 116). The energy distributions are translated of a quantity equivalent to the translation, because the *energy conservation*[68] *properties* are respected (ref. to Moyal's theorem). The kernel above can thus be written as a structure of translation in time and modulation in frequency, provided with the invariance properties:

$$\theta(u,\overline{u},\xi,\overline{\xi}) = \theta(u-\overline{u},\xi-\overline{\xi}), \tag{5.307}$$

which allow by convolution with the Wigner–Ville distribution to obtain a weighted or smoothed distribution, such as:

$$P_\theta f(u,\xi) = P_V f \star \theta(u,\xi) = \int\int P_V f(\overline{u},\overline{\xi})\theta(u-\overline{u},\xi-\overline{\xi})d\overline{u}d\overline{\xi}. \tag{5.308}$$

This type of distribution is known as of the "Cohen's class". The general expression above of the distribution is provided with a double integral on time and on frequency, and the kernels impact obviously the initial distribution on spreads corresponding to their own support in time and frequency. The types of kernels are rather numerous and condition obviously the result of the convolution. (In spite of the interest of the subject, we will not report it here.)

---

[66] Invariance by linear translation:
  $g(t) = Lf(t) \Rightarrow g(t-\tau) = Lf_\tau(t)$.

[67] Invariance property of the Wigner–Ville distribution:
  $f(t) = g(t-u_0) \Rightarrow P_V f(u,\xi) = P_V g(u-u_0,\xi)$,
  $f(t) = \exp(i\xi_0 t)g(t) \Rightarrow P_V f(u,\xi) = P_V g(u,\xi-\xi_0)$.

[68] *Energy conservation equation*: $\|f\|^2 = \int_{-\infty}^{+\infty}|f(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|\widehat{f}(\lambda)\right|^2 d\lambda$.

  Property of the Wigner–Ville distribution $P_V f(u,\xi)$*:* For any f, one has the following properties:

  $\int_{-\infty}^{+\infty} P_V f(u,\xi)du = |\widehat{f}(\xi)|^2$ and $\frac{1}{2\pi}\int_{-\infty}^{+\infty} P_V f(u,\xi)d\xi = |f(u)|^2$.

  These properties make it possible to regard the Wigner–Ville distribution as *a "time–frequency energy density"*.

When the distribution is computed again and the *interference terms removed* or more exactly *attenuated*, we can observe the image in the time–frequency plane of atoms (see Figs. 5.51 and 5.52).
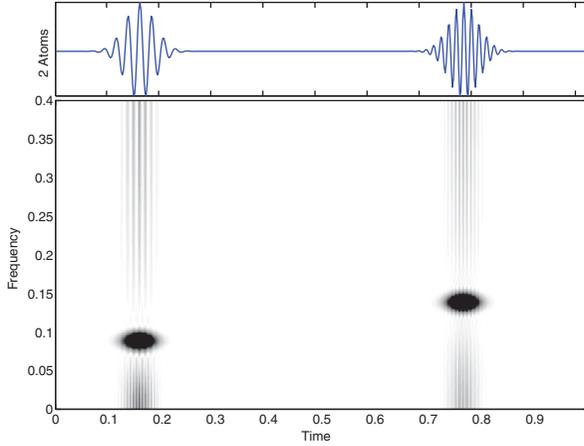
Fig. 5.51 Cohen's class distribution of two Gabor atoms (B&W)
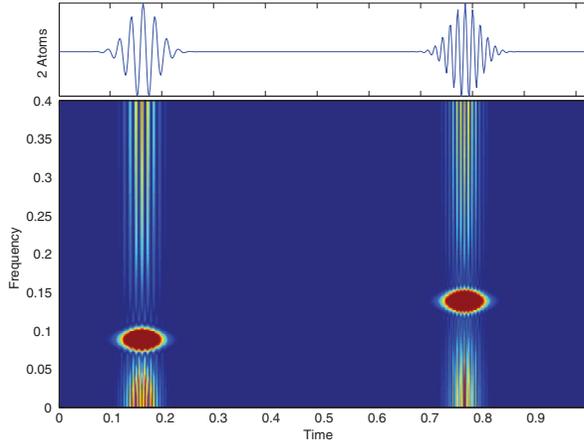
Fig. 5.52 Cohen's class distribution of two Gabor atoms (Jet)

5.18.1.1 Distribution of an Artificial Signal

We create an artificial signal, it is successively composed of a Gauss-curve, then, its derivative, its second derivative, a triangular function, a sinusoid, a Dirac function, a Morlet wavelet, a staircase, a random series and an increasing oscillating structure then decreasing ($n = 512$). Figures 5.53 and 5.54 show the Wigner–Ville distribution and the Cohen's class distribution in the time–frequency plane.

Hereafter, we present the Cohen's class distribution for the same signal. Note that the representation in the time frequency plane is refined and the interferences are attenuated.
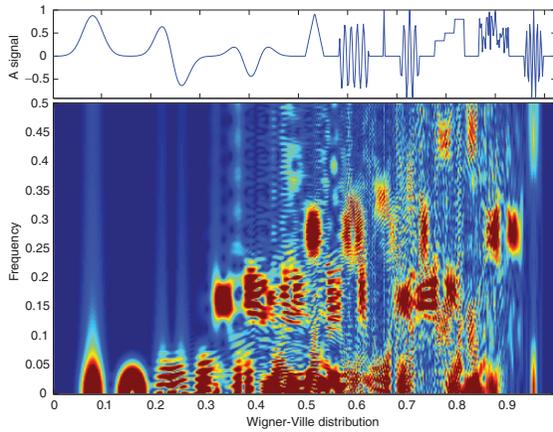
Fig. 5.53 Wigner–Ville distribution
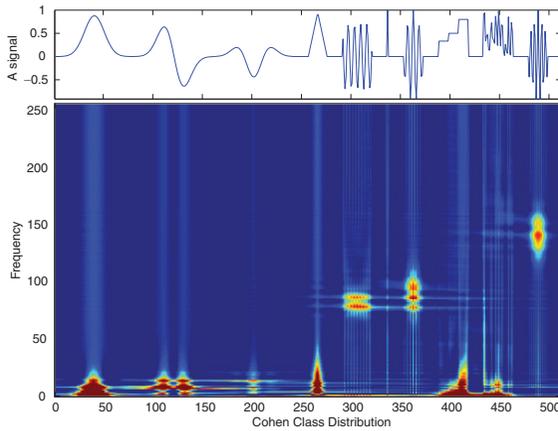
Fig. 5.54 Cohen's class distribution
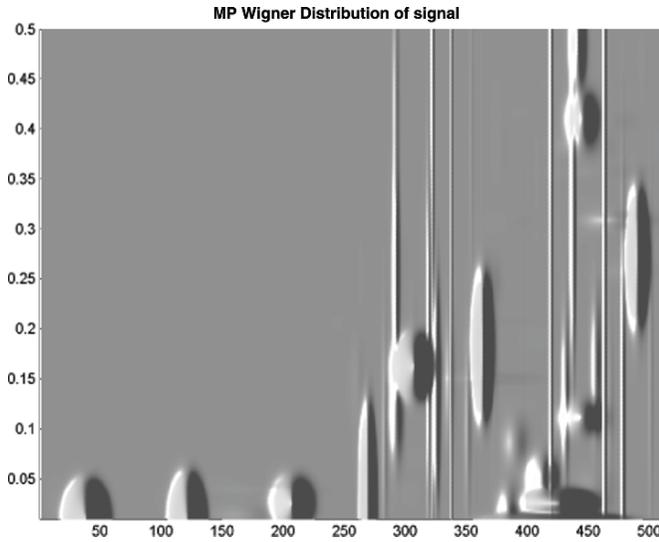
Fig. 5.55 Wigner distribution of the atomic decomposition by the MP

Wigner Distribution of the Decomposition by the "Matching Pursuit" of This
Artificial Signal

As explained previously, the Wigner distribution can be used to represent in the
time–frequency plane the decomposition by the "Matching Pursuit" with dictionar-
ies of time–frequency atoms of this artificial signal. The signal is almost the same, as
in the preceding section, except for the "random part" which is a new random sam-
ple. We will observe in Fig. 5.55 the result of the matching pursuit approximation
in the time–frequency plane that it will be possible to compare with the previous
Wigner–Ville distribution or with the Cohen's class distribution for the artificial
signal.

5.19 Introduction to the Polyspectral Analysis
     (for the Nonlinearities)

It is not possible to approach the domains of spectral analysis, time–frequency anal-
ysis or signal processing without providing a report about polyspectral analysis.
New estimators of the normalized *polyspectrum* of the *order 3* (*bicoherency, bis-
pectrum*) and of the normalized polyspectrum of the order 4 (*trispectrum*) have
been developed. Their *normalization terms take into account the fact that we search
to identify the "phase relations" between the involved frequencies*. The evaluation
of the performances of estimators are currently in process, in particular *about the
nonlinear dynamical systems* described by the *Zakharov equations*. This type of

study is to be supplemented by the *evaluations of the fractal dimension of turbulent processes*. Such approaches are carried out by the physics laboratories for the signal processing of the stellar plasma type for example, but also by some economic laboratories about the social data analysis. We refer to some studies supported by the OECD concerning economical and social data which exploit the polyspectral analysis but also wavelet analysis (Amin et al. 1997).

5.19.1 Polyspectral Analysis Definition for Random Processes with Zero-Average

In order to define the polyspectral analysis or the spectral analysis of a high order (i.e. order higher than two), we use an extension of the autocorrelation notion with multiple lags (delays), from the definition of the *cumulant* (or moments) of order n of a random process with zero-average (Mendel 1991; Pilgram et al. 1997),

$$C_n(\tau_1, \tau_2, \ldots, \tau_{n-1}), \tag{5.309}$$

the *cumulants of order 2* for the time series $x(t)$ is written:

$$C_2(\tau_1) = E[x(t)x(t+\tau_1)], \tag{5.310}$$

where $E[.]$ is the expectation. Thus, the second order cumulant of $x(t)$ is the autocovariance function of $x(t)$. The third order cumulant is:

$$C_3(\tau_1, \tau_2) = E[x(t)x(t+\tau_1)x(t+\tau_2)] \tag{5.311}$$

and the fourth order cumulant is written:

$$C_4(\tau_1, \tau_2, \tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)] - C_2(\tau_1)C_2(\tau_2 - \tau_3)$$
$$- C_2(\tau_2)C_2(\tau_3 - \tau_1) - C_2(\tau_3)C_2(\tau_1 - \tau_2). \tag{5.312}$$

If $x(t)$ is a *Gaussian random process, all the cumulants of an order higher than 2 are equal to 0*. This property is useful for example during a *cardiorespiratory analysis*, because it makes it possible to distinguish the *non-Gaussian components* (e.g. *a deterministic oscillation as a respiratory arrhythmia*) *in a Gaussian background noise, independently of the noise spectrum shape, by the calculation of cumulants of order higher than 2*. But, because the third order cumulants are equal to 0 for the Gaussian processes, but also for the processes which have a symmetrical distribution, then we use in such cases the fourth order cumulants. *The Fourier transform of cumulants are called spectra of high order, or polyspectra*. The Fourier transform of second order cumulants (corresponding to the autocovariance) is *obviously* the *power spectrum*. The Fourier transform of $C_3(\tau_1, \tau_2)$ is a *bispectrum*.

5.19.2 Polyspectra and Nonlinearities

The spectra of a high order are supposed to provide a tool to study the nonlinearities. *All the cumulants of order higher than two are equal to zero for a Gaussian process,* thus the polyspectra are particularly *useful to study the deviations (i.e. gaps) in comparison with the Gaussian behaviors*. But, this creates problems, because the intuitive interpretation of cumulants is not spontaneous. The *spectral distribution of the second order is linked with the autocorrelation function by the Fourier transform,*

$$S(\lambda) = \int_{-\infty}^{+\infty} C(\tau)e^{-i2\pi\,\lambda\tau}d\tau, \tag{5.313}$$

$$C(\tau) = \int_{-\infty}^{+\infty} S(\lambda)e^{+i2\pi\,\lambda\tau}d\lambda, \tag{5.314}$$

λ is the frequency and τ the lag. And we have in particular σ^2 the variance of the process written as follows:

$$\sigma^2 = C(0) = \int_{-\infty}^{+\infty} S(\lambda)d\lambda. \tag{5.315}$$

To calculate the bispectrum, we pose the "tri-variance" $T(\tau_1, \tau_2)$, written:

$$T(\tau_1, \tau_2) = E[(x(t) - \mu) \cdot (x(t + \tau_1) - \mu) \cdot (x(t + \tau_2) - \mu)], \tag{5.316}$$

to simplify the writings we poses μ as the average of the process $x(t)$, and we obtain the *bispectrum:*

$$B(\lambda_1, \lambda_2) = \int\int T(\tau_1, \tau_2)e^{-i2\pi\,(\lambda_1\tau_1 + \lambda_2\tau_2)}d\tau_1 d\tau_2. \tag{5.317}$$

Its inverse Fourier transform is written:

$$T(\tau_1, \tau_2) = \int\int B(\lambda_1, \lambda_2)e^{+i2\pi\,(\lambda_1\tau_1 + \lambda_2\tau_2)}d\lambda_1 d\lambda_2. \tag{5.318}$$

From the definition above, $T(0,0)$ is σ^3 times the *skewness* (asymmetry[69]), then the preceding equation becomes:

$$\sigma^3 \times skewness = \int\int B(\lambda_1, \lambda_2)d\lambda_1 d\lambda_2, \tag{5.319}$$

thus, $B(\lambda_1, \lambda_2)$ *is interpreted as a function which shows how the "skewness" is linked with the frequency pairs.*

It is known that the *power spectra,* i.e. *the correlations of order 2,* are *blind* as regards *the phases*. In addition, it is said that *the Gaussian random processes* can

---

[69] In statistics, the *Skewness* test (i.e. asymmetry) is written: $\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^3/\sigma^3$. And the *Kurtosis* test (i.e. flatness) is written: $\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^4/\sigma^4$.

be *completely specified by the knowledge of their statistics of the first and second order*. There exist many practical situations, where we have to look at correlations of a higher order, i.e. *cumulants of order higher than 2* of a signal, in order to extract *the information concerning the phase*, i.e. the *presence of nonlinearities or the deviation in relation to the "Gaussianity" of the signal, etc. The polyspectra (order >2) of Gaussian processes are equal to zero*. Thus, it is said in theory that *the polyspectra analysis are domains with high Noise–Signal Ratio*, where the identification of the system and the reconstruction of the signal can done and provide results. *The non-Gaussian processes are better apprehended or identified by the polyspectra*. However, the polyspectra have tardily received many critiques, in particular because the amount of data necessary to produce estimates of weak-variance and the number of involved calculations are very high.

Examples of Polyspectrum.

First example: Bispectrum and third order cumulants are calculated for an ARMA (2,1) process with $AR = [1, -1.5, 0.8]$ and $MA = [1, -2]$. Figures 5.56 and 5.57 depict respectively, Bispectrum, Cumulants in the plane, Cumulants in 3D and a sample of Trispectrum (contour plots).

Second example: The analyzed time series is segmented into several overlapping records. The signal corresponds to artificial data for the *Quadratic-Phase Coupling problem*. Signal is constructed from four unity amplitude harmonics with frequencies $0.1, 0.15, 0.25$ and 0.40 Hz. We add to the signal a white Gaussian noise with a variance of 1.5. The harmonics at $(0.1, 0.15, 0.25)$ are *frequency-coupled* and *phase-coupled*, however the harmonics at $(0.15, 0.25, 0.40)$ are frequency-coupled, but not phase-coupled. Hereafter the bispectrum amplitude is shown in a contour plot. The sharpness of peaks highlights the quadratic phase coupling (see Fig. 5.58a).

Third example: Figure 5.58b shows the Bispectrum of sunspots (Annual sunspots for years 1700–1987). The sunspot time series is positive and the bispectrum is calculated by means of differences. *Last examples: The first time series is the logistic model close to the chaotic regime. The second time series is a Bi-linear model (Signal* $= y_1 \cdot y_2 : y_1 = sin(2\pi \cdot f_1 \cdot t + \varphi_1)$ and $y_2 = (0.1) \cdot sin(2\pi \cdot f_2 \cdot t + \varphi_2)$, $f_1 = 60$ Hz, $f_2 = 4$ Hz, and φ_1, φ_2 are randomly chosen. In this case $\varphi_1 = 0.7$ and $\varphi_2 = 0.3$ *radians*). Each Fig. 5.59a and Fig. 5.59b depicts Averaged signal, Third order cumulant, Bispectrum magnitude and Bispectrum phase.
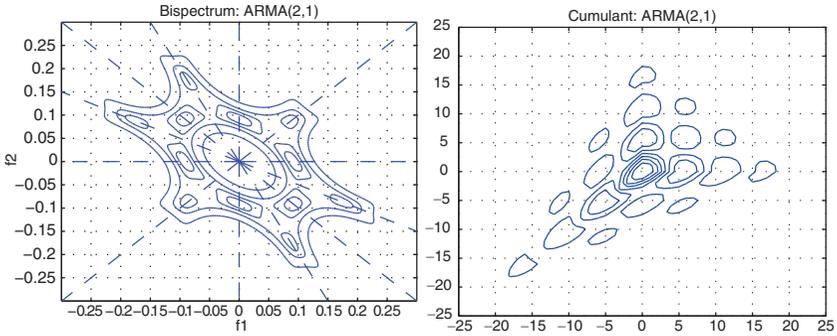
Fig. 5.56 ARMA(2,1) Bispectrum (*left*), ARMA(2,1) Cumulants (*right*)
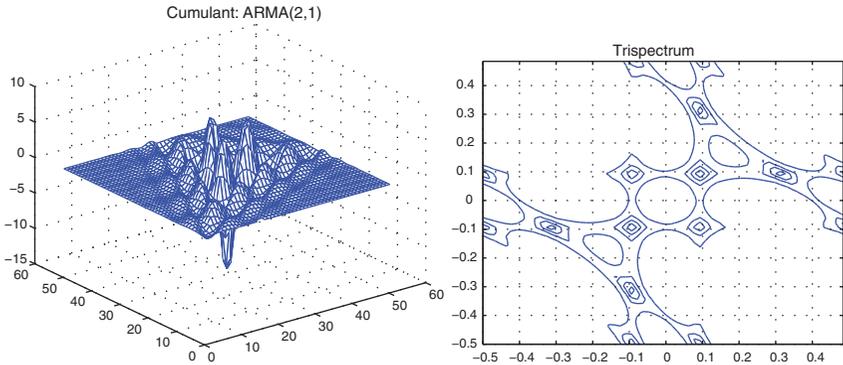
Fig. 5.57 3D: ARMA(2,1) Cumulants (*left*), ARMA(2,1) Trispectrum (*right*)
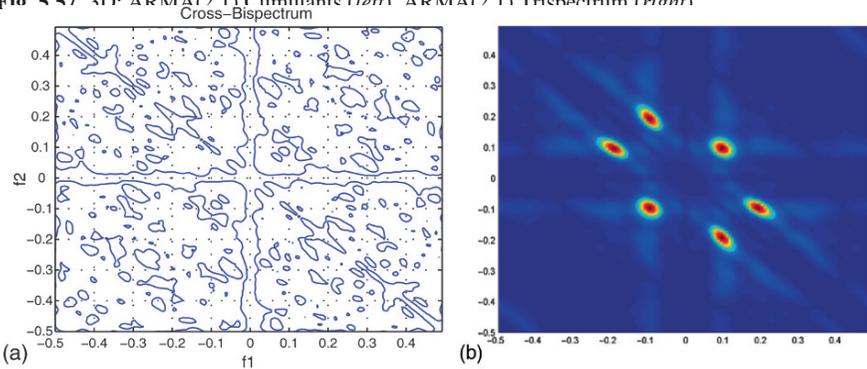
Fig. 5.58 (**a**) Cross-Bispectrum of signal, (**b**) Bispectrum of Sunspots

5.20 Polyspectral and Wavelet Bicoherences

5.20.1 A New Tool of Turbulence Analysis: Wavelet Bicoherence

Recently, a new tool for the analysis of turbulence has been introduced and investigated, it is the *wavelet bicoherence* (see van Milligen et al. 1995a). This tool

(a)   (b)
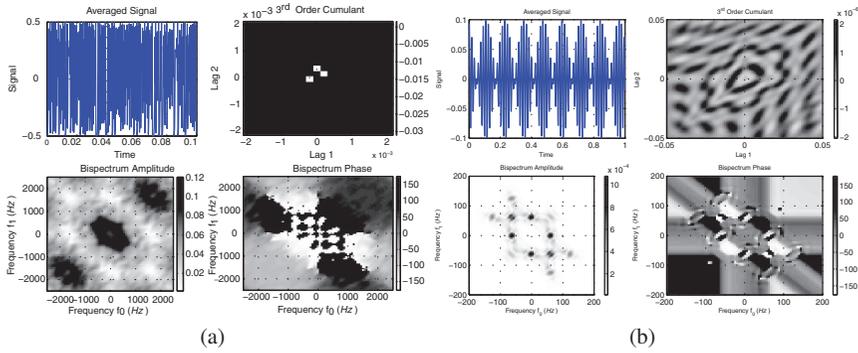
Fig. 5.59 (**a**) Logistic equation, (**b**) Bi-linear process

is able to detect the *phase coupling*, i.e. the *nonlinear interactions of the lowest "quadratic" order with time resolution*. Its potential is important and it was *applied in specific works (see footnote[70]) to numerical models of chaos and turbulence* and also applied to *real measures*. In the first case, the van der Pol model of chaos, this tool detected the *coupling interaction between two coupled van der Pol oscillators*. In the second case, the *drift wave turbulence model* concerning the plasma physics, it detected a *highly localized* "*coherent structure*". In the case of real measures, i.e. for the analysis of reflectometry measures concerning fusion plasmas, *it detected temporal intermittency* and *a strong increase in nonlinear phase coupling* which coincide with the Low-to-High[71] confinement mode transition. Three arguments plead for a new tool about the turbulent phenomenon analysis: (1) *First*, the tools of *chaos theory* (as fractal dimensions, Lyapunov exponent, etc.) are not always easily applicable to real phenomena, in particular if the noise level is high. Moreover, information recovered by these methods is not always adapted to physical understanding. Indeed, a low fractal dimension measured in a real phenomenon is helpful information, but a high dimension, in particular higher than 5 (for example often observed in fusion plasmas) does not offer interesting solution. (2) *Second*, the applications of the traditional analysis (i.e. which involve long-time averages of moments of data) of *standard spectral analysis* are limited concerning chaotic or turbulent real phenomena. Indeed, if we consider the transition from quasi-periodicity to chaos in theoretical dynamical systems, as we have explained in the first part of this book, there exist several possible routes to chaos which can be summarized as follows: *period-doubling*, *crises* and *intermittency*. In these three main routes (Ott 1993), *the transition to chaos are abrupt*. Then, an explanation of the chaotic regime by means of *a superposition of a large number of harmonic modes* (i.e. *oscillators which correspond to the Fourier analysis*[72]) does not seem suitable. Indeed,

---

[70] van Milligen et al. 1995b. Asociación EURATOM-CIEMAT, Madrid, Spain. B. Carreras. Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A. L. García.Universidad Carlos III, Madrid, Spain.

[71] Usually written by physicists: (L/H).

[72] See the Navier–Stokes equations sections in the present book.

Fig. 5.60 "Coherent structures" in turbulent phenomenon (ref: Haller)
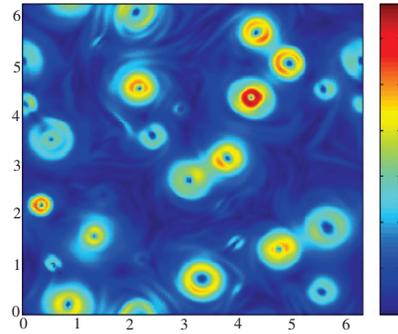
for some systems of which the equations do not have a complete set of solutions and for which therefore the observed behavior cannot be decomposed into eigenmodes. Therefore, *a decomposition into Fourier harmonics (i.e. modes), which is appropriate for the global description of turbulence or chaos by means of (for example) the decay of the spectrum in frequency, produces a confused picture at a finer scale, because it is not an optimal method in particular for the expansion* (van Milligen et al. 1995b, p. 2). (3) *Third*, numerous and important arguments (both numerical simulations and observation turbulence) lead to state that *turbulence usually is an intermittent phenomenon*, that means that *it is localized in time and often in space also. Such an observation is in opposition to the Fourier analysis*, indeed, *the Fourier analysis supposes the* "*homogeneity*" *in the phenomenon*, and thus does not seem suitable. Moreover, this concept of *non-homogeneity* in the turbulence phenomena is closely related to the concept of *Coherent structures*[73] (see for example, Lagrangian Coherent Structures in 2D Turbulence, Haller's research, Department of Mechanical Engineering, MIT).

Figure 5.60 depicts how *Coherent structures (and their possible interactions) could be distributed in a turbulence phenomenon picture.* These types of phenomena plead for the use of the wavelet analysis. As explained previously, the wavelet analysis can be taken as an extension of the Fourier analysis. Even if Fourier and wavelet analyses in many cases lead to similar conclusion, however *wavelet analysis has a fundamental additional property* which is the *time resolution.* (Even if we use the windowed Fourier transform or short time Fourier transform, because the defects and deficiencies of the Fourier methods persist.) It is important to highlight that many successful applications (Farge 1992; Hudgins et al. 1993; Meneveau 1991) of the wavelet analysis concern turbulent phenomena. At this stage, during the wavelet transformation the important purpose is to "extract the information relevant to non-linear interactions". In connection with what precedes, the higher-order spectra in the Fourier analysis were applied with interesting results (Meneveau 1991; Kim and Powers 1978). The paper (van Milligen et al. 1995b) that we evoke presently goes further, indeed, the concept of *the first higher-order spectrum*, i.e. the *bispectrum*

---

[73] "While the existence of these structures is clear from visual observations, their mathematical description is far more difficult" Haller and Iacono (2003).

was generalized to the wavelet analysis. The bicoherence, i.e. the normalized bispectrum, is a measure of the amount of "phase coupling" which occurs in a signal. The *phase coupling* occurs when two frequencies θ_1 and θ_2 are simultaneously present in a signal along with their sum (or difference) and the sum of the phases φ of these frequency components remains constant. The bicoherence measures this quantity and is a function of two frequencies θ_1 and θ_2 which is close to 1 when the signal contains three frequencies θ_1, θ_2 and θ that satisfy the relation $\theta_1 + \theta_2 = \theta$ and $\varphi_1 + \varphi_2 = \varphi + constant$ (otherwise it is close to 0). When the signal, which is analyzed, shows "*structures*", *it is highly probable that "some phase coupling occurs".* Therefore, the generalization of the bispectrum to wavelet analysis can be used to be able to *"detect temporal variations in phase coupling", i.e. "intermittent behavior"* or "structures localized in time".

In the paper evoked presently, this method is applied to the output of a nonlinear chaotic dynamical system (i.e. van der Pol), then to a model of drift wave turbulence relevant to plasma physics, and finally to real measurements, i.e. to the analysis of reflectometry measures made in (thermonuclear) fusion plasmas. The only one application which will be briefly presented in this section is the nonlinear van der Pol model, in the case of a system of two coupled van der Pol oscillators in the periodic and chaotic states.

5.20.1.1 Wavelet Bicoherence: Study of "Abrupt Frequency Change", "Phase Coupling" and Singularities

The objective of these works was to analyze the abrupt changes of frequencies, pulses, or singularities very localized in time (as in turbulent phenomena). *These types of problems cannot be analyzed and resolved by the Fourier analysis which does not have resolution in time,* because this type of problem involves an integral over time as well as on frequencies. With the Fourier analysis, abrupt variations and singularities localized in time are spread among all the signal decomposition, and the temporal information is eliminated during the reconstruction of a signal (such remark is valid also for the windowed Fourier transform to rebuild the singularities or abrupt variations). Wavelet analysis is based on (non-continuous) *"oscillating functions" (with compact supports)* which decay rapidly in the course of time, rather than sines and cosines in the Fourier analysis which do not have such a decay. Indeed, a wavelet transform is a function of the frequency and time. Let us present briefly these important works. If ψ indicates a wavelet, consider a *wavelet family* which can be written:

$$\psi_a(t) = \frac{1}{a^p} \psi \left(\frac{t}{a} \right),\qquad (5.320)$$

where a is a scale parameter and the factor p is a normalization choice, the authors selected $p = 0.5$ (the argument of their choice is to say that the $\mathbf{L}^2$-norm of the wavelet is independent of a). We know that the wavelet admissibility condition is written:

$$C_\psi = \int_{-\infty}^{+\infty} |\hat{\psi}(\omega)|^2 |\omega|^{-1} d\omega < \infty. \tag{5.321}$$

The wavelet transform of a function $f(t)$ is written:

$$W_f(a, \tau) = \int f(t)\psi_a(t - \tau) dt, \tag{5.322}$$

where τ is a translation parameter. And the inverse wavelet transform is written for $p = 0.5$:

$$f(t) = \frac{1}{C_\psi} \int \int W_f(a, \tau)\psi_a^*(\tau - t) d\tau \frac{da}{a^2}. \tag{5.323}$$

In the *wavelet analysis* we know that *time-resolved spectra (i.e. time resolution in spectra)* can be preserved. The *wavelet cross spectrum* is written:

$$C_{fg}^W = \int_T W_f^*(a, \tau)W_g(a, \tau) d\tau, \tag{5.324}$$

where T is a finite time interval and $f(t)$ and $g(t)$ are two time series. It is introduced a *delayed* wavelet cross spectrum:

$$C_{fg}^W(a, \Delta\tau) = \int_T W_f^*(a, \tau)W_g(a, \tau + \Delta\tau) d\tau \tag{5.325}$$

which is a *quantity* used to *detect "structures"* between two separated observations. A normalized delayed wavelet cross coherence is written:

$$\gamma_{fg}^W(a, \Delta\tau) = \frac{\left| \int_T W_f^*(a, \tau)W_g(a, \tau + \Delta\tau) d\tau \right|}{\left(P_f^W(a)P_g^W(a) \right)^{1/2}} \tag{5.326}$$

which is contained in the interval $[0, 1]$. The *wavelet power spectrum* can be written in term of the usual Fourier power spectra of the wavelet:

$$P_f^W(a) = \frac{1}{2\pi} \int P_{\psi_a}(\omega)P_f(\omega)d\omega \tag{5.327}$$

with $P_f(\omega)$ calculated on the interval T. The *wavelet*[74] *auto-power spectrum* can be written as follows: $P_f^W(a) = C_{ff}^W(a)$. The first higher-order spectrum along the same lines is introduced and the *wavelet cross-bispectrum* is defined by the authors of the paper in the following way:

$$B_{fg}^W(a_1, a_2) = \int_T W_f^*(a, \tau)W_g(a_1, \tau)W_g(a_2, \tau) d\tau \tag{5.328}$$

with $1/a = 1/a_1 + 1/a_2$ which is the frequency sum-rule. This *wavelet cross-bispectrum* measures the amount of *phase coupling* (inside T) which occurs between

---

[74] *According to the term used in the authors' article.*

wavelet components a_1 and a_1 of $g(t)$ and wavelet component a of $f(t)$ (in such a way that the sum-rule is verified). It is supposed that $\omega = 2\pi/a$, *wavelet cross-bispectrum* can be understood as the coupling between waveforms or more exactly between waves of frequencies, i.e. wavelets of frequencies such that $\omega = \omega_1 + \omega_2$.[75] In a similar way, the wavelet[76] auto-bispectrum is written:

$$B^W(a_1, a_2) = B^W_{ff}(a_1, a_2). \tag{5.329}$$

It is possible to write that the "squared wavelet cross bicoherence" is the "normalized squared cross-bispectrum":

$$\left(b^W_{fg}(a_1, a_2)\right)^2 = \frac{\left|B^W_{fg}(a_1, a_2)\right|^2}{\left(\int \left|W_g(a_1, \tau) W_g(a_2, \tau)\right|^2 d\tau\right) P^W_f(a)} \tag{5.330}$$

contained between the interval $[0, 1]$. Besides, the *squared wavelet auto bicoherence* is written:

$$\left(b^W(a_1, a_2)\right)^2 = \left(b^W_{ff}(a_1, a_2)\right)^2. \tag{5.331}$$

The square bicoherence $\left(b^W(a_1, a_2)\right)^2$ is usually plotted in a (ω_1, ω_2)-plane rather than (a_1, a_2)-plane.[77] At this stage, the *total bicoherence* is written:

$$\left(b^W\right)^2 = \frac{1}{S} \sum \sum \left(b^W(a_1, a_2)\right)^2, \tag{5.332}$$

where S the number of terms in the summation insures the inclusion in the interval $[0, 1]$. At this stage, the choice of one wavelet can be introduced. Usually the choice is often restricted to a wavelet which has Fourier transforms showing a single prominent peak. In this respect, a theoretical wavelet is constructed by the authors which is written as follows: $\psi_a(t) = \frac{1}{\sqrt{a}} \exp\left[i\frac{2\pi t}{a} - \frac{1}{2}(t/ad)^2\right]$ the choice of d provides the exponential decay of the wavelet, which is adapted to the time–frequency resolution.

5.20.1.2 Application to the Coupled *van der Pol Oscillators*

The coupled *van der Pol oscillators* can be written as follows:

$$\frac{dx_i}{dt} = y_i, \tag{5.333}$$

$$\frac{dy_i}{dt} = \left((a_i - (x_i + b_j x_j)^2)\right) y_i - (x_i + b_j x_j), \tag{5.334}$$

---

[75] See article about this postulate.

[76] *According to the term used in the authors' article.*

[77] *Remark:* $\omega_1 + \omega_2 = \omega$, this sum must be smaller than the Nyquist frequency (half sampling frequency). Moreover, it is postulated to plot that $\omega_1 \geqslant \omega_2$.

where $i = 1$, $j = 2$ correspond to the first oscillator and $i = 2$, $j = 1$ is the second. a_i correspond to the limit-cycles of oscillators and b_j correspond to the *nonlinear couplings* between oscillators. Two cases are studied, the first one is the *periodic state* and the second is the *chaotic state*.

Periodic state: The successive pictures below, show the *Fourier spectrum* (top-left), the cross phase probability (top-right), the *wavelet bicoherence* (bottom-left), the *wavelet cross bicoherence* (bottom-right). The Fourier spectrum shows some rare peaks of frequency and their associated harmonics. The bicoherence (bottom-left) shows *rectilinear horizontal and diagonal ridges* corresponding to a frequency at 0.34, which can be identified in the spectrum as the second peak. So, the two dominant peaks (i.e. the first and third) in the Fourier spectrum respectively at 0.17 and at 0.5 *couple (i.e. there is coupling)* with their difference frequency at 0.34. Similarly, the difference in frequency between the second and the fourth peak, between the fourth and the sixth peak, ... is always 0.34. Idem for the odd sequence of peaks. *The even peaks are the consequence of the coupling interaction between the two oscillators, whereas the odd peaks are the harmonics of the limit cycle of oscillators* (ref. to van Milligen et al. 1995b).
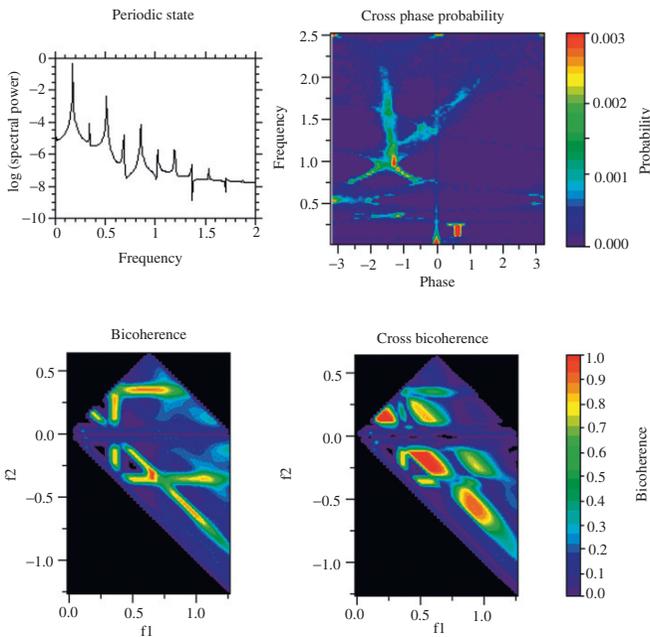
Chaotic state: Fourier spectrum (top-left) still shows some of former peaks and other of the periodic state which are shifted in frequency. *New peaks occur because of period-doublings* (on the route to chaos). Spectrum shows chaos on a broad band. The bicoherence (bottom-left) shows (weaker) *rectilinear horizontal and diagonal ridges* which appear at a frequency close to 0.25 corresponding to the 4th peak in the Fourier spectrum, which is related to the frequency at 0.34 highlighted in the

periodic-state spectrum. A modification of the parameter of the coupled model to obtain the chaotic regime induces a global frequency shift.
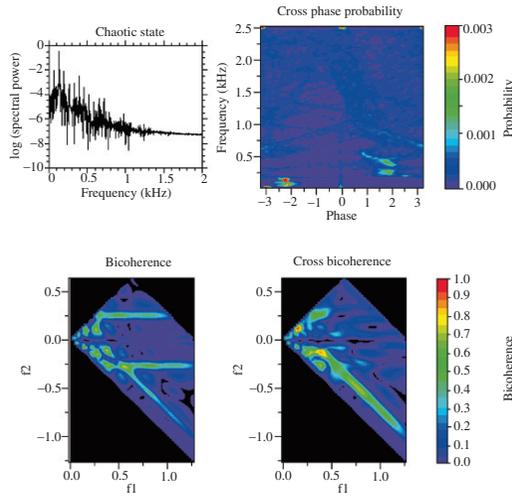
Extracting the *wavelet bicoherences* of the periodic and chaotic states from the preceding pictures, we *highlight* the comparison between two states:
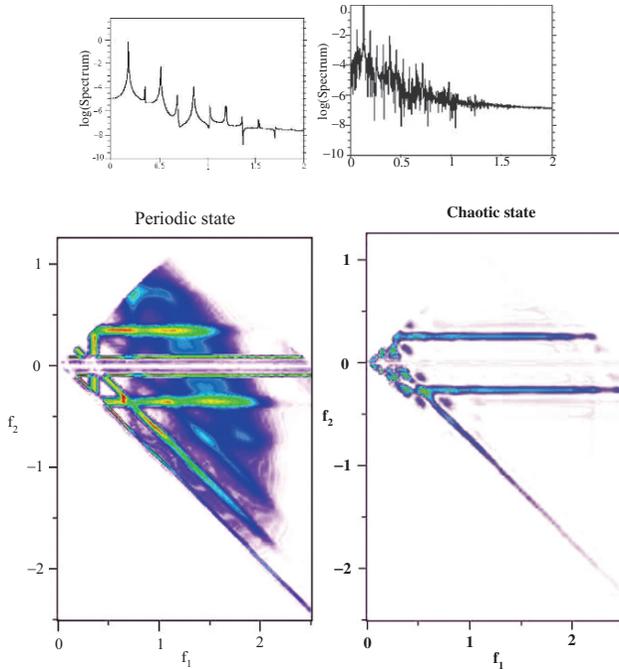
The comparison above shows that the images have similar strong main ridges, and except for these ridges the images are different. Indeed, the vertical line which

appeared around the frequency "0.34" (showing the *coupling* in the periodic-state bicoherence) is divided and partitioned into several small distinct parts. This remark is in connection with the period-doublings appeared in the chaotic-state. New couplings can be also observed in lower frequencies (around 0.2 and 0.13, see article for comments).

5.20.2 Compared Bicoherences: Fourier and Wavelet

In order to compare the bicoherences between wavelet and Fourier, the analyzed measures are taken at the "Advanced Toroidal facility" (ATF, Lyon et al. 1986) with "Langmuir" probes and with (known) strong Fourier bicoherence (Hidalgo et al. 1994).at 1 MHz. The *Fourier spectrum* of this signal (not shown) *is turbulent* and does not have preeminent peaks or modes (i.e. oscillations, periods or cycles). *Figure 5.61 show the Fourier and wavelet bicoherences.*[78] The wavelet picture shows less detail and the global aspects of the pictures are similar.

In order to demonstrate that the wavelet bicoherence shows more clearly elements which are less well detected by means of the Fourier spectrum, the FFT is computed on the raw data, the phase information is scrambled, and an inverse FFT is computed to obtain a new data series. We strongly recommend the reader to consult this interesting article to have comments, technical demonstrations and results.
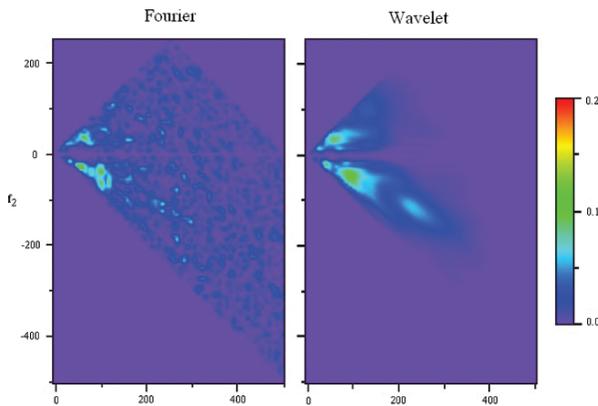
Fig. 5.61 Bicoherences (ref. to van Milligen et al, 1995a, 1995b)

---

[78] Computed on a frequency grid with 64 grid points from 0 to 500 kHz from a data section running from 1 to 16 ms.

5.21 Arguments in Favor of Wavelet Analysis Compared to Fourier Analysis

5.21.1 Signal Deformation by Diffusion of Peaks, Discontinuities and Errors in the Fourier Transform

In a Fourier transform the information about time is not lost because we can rebuild the signal from a transform, however it is hidden under the phases, i.e. the same sines and cosine can represent moments very different of the signal because they are shifted in phase to amplify or to cancel themselves. Consequently, the Fourier analysis is inadequate and unsuited to the signals which change abruptly and in an unpredictable way. However, such changes contain very interesting information. In theory, it is possible to extract information about time by computing the phases from the Fourier coefficients. In practice, to calculate them precisely is impossible. The fact that the *information* at *one moment* of the signal is *widespread among all the frequencies* of the *transform is a major problem*. A local characteristic of the signal as a peak or a discontinuity becomes a global characteristic of the transform. In connection with the peaks, we know for example that a narrow and high rectangular function has a very wide Fourier transform, as it is easy to observe in Fig. 5.62.

As mentioned before, a discontinuity for example, is represented by a superposition of all the possible frequencies. In the same way, as the white color is made of all the colors of the spectrum; It is not possible to deduce from such a superposition that the signal is discontinuous and moreover to locate this discontinuity. Consequently for all these reasons, in *the Fourier space* the *lack* of *information about time makes a Fourier transform very sensitive and vulnerable to the errors*. Indeed, *an error can corrupt all the Fourier transform. The (true or erroneous) information in a part of the signal spreads throughout the transform*; thus, *the errors of phases deform the signal* and can produce a signal very different from the original.
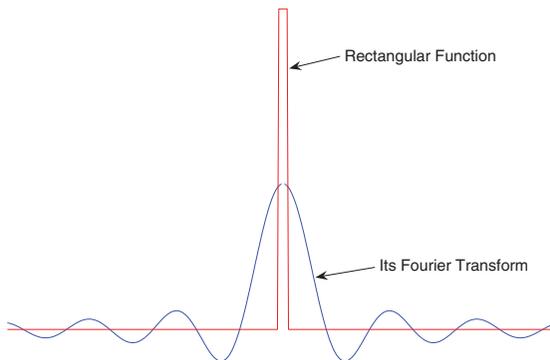
Fig. 5.62 The Fourier transform of a peak is extensive

5.21.2 Wavelets are Better Adapted to the Signal by Respecting Discontinuities and Peaks, Because they Identify the Variations

Because of the preceding observations, the wavelets have been conceived. *They avoid the diffusion and interference phenomena*, allowing also the decomposition of the signal and its reconstruction. In the *Fourier analysis, the window is fixed and the number of oscillations varies. A small window is blind for the low frequencies* which are too large to be captured. In a too large window, the information about *the fast or abrupt variations is drowned* in the information concerning the entire interval contained in the window. In the wavelet analysis, we analyze the signal at different scales, because we stretch and compress a mother wavelet. The broad wavelets provide an approximate image of the signal while the narrow wavelets allow to zoom in the details.[79] We reconstruct a signal from its wavelet transform, by adding wavelets of various sizes, just as for the Fourier transform we rebuild the signal by adding sines and cosines. In principle, the calculation of coefficients is done in the same way, we multiply the signal and the analyzing function, and we compute the integral of the product.[80] However, in practice we use different fast algorithms. (We have already explained that a wavelet[81] was built starting from the "Gaussian function", but this one does not satisfy the strict conditions for the construction of a wavelet, and such a wavelet was used by Gabor in the windowed Fourier analysis.) The stretching and compression of wavelets are the true inventions which made evolve the method. The wavelets are automatically adapted to the different components of the signal: they use a narrow window to look at the *transitory* components and a broad window for the *components* of *long duration* and *low frequency*.

It is essential to remember that, unlike the Fourier analysis, the information about frequencies is only approximate, indeed a sine and a cosine have a precise frequency, this is not the case for a wavelet. This is the characteristic of wavelets, due to their "constant shapes": The resolution, the scale and the frequency vary at the same time.

The set of coefficients at all the scales provide a good image of the signal. Unlike Fourier series, here, the wavelet coefficients translate, in simple, precise and exact ways, the properties of these functions, i.e. the properties which correspond to transitory periods, for example the ruptures, discontinuities. It is said that the *wavelets detect and encode only the variations*. A wavelet coefficient measures the correlation between the wavelet (its peaks and its hollows) and the corresponding interval of the signal and they allow to closely look at the details of a signal. A constant interval of the signal provides a coefficient equal to zero, because the wavelet by definition has a zero-integral. Thus, the integral of the product of the wavelet by the signal is equal to zero for a constant signal. *The wavelet analysis is a manner of expressing the sensitivity to the variations*.

---

[79] A technique which is sometimes called the "mathematic microscope".

[80] $\int (f * g) dt$.

[81] Created by Yves Morlet.

When we use the continuous wavelet transform, any function can be named "wavelet", on the condition that it has a zero integral. The Fourier transform decomposes a signal according to its frequencies; the wavelet transform decomposes a signal into its components at different scales. In both cases, we calculate the integrals: we multiply the signal by the analyzing function (sinusoids or wavelets) and we integrate. It is essential to explain that the transformation is robust, i.e. a small change in the wavelet representation induces a comparable change of size in the signal; *a small error or modification is not amplified in an exaggerated way*.

5.21.3 Wavelets are Adapted to Non-Stationary Signals

The wavelet analysis is thus appropriate to the non-stationary signals which present discontinuities and peaks. The Fourier analysis is appropriate to regular periodic signals and to stationary signals. If the signal is regular over a long duration and simply oscillates, it is not logical to analyze it with small wavelets, which capture only some oscillations. Moreover, the determination of the frequency by wavelets is imprecise in high frequencies.

5.21.4 Signal Energy is Constant in the Wavelet Transform

Grossmann and Morlet proved that when we represent a signal by means of the wavelet transform, the signal "energy" does not change. This *energy* corresponds to *the average value of the square of the amplitude*, which is different from the energy concept in physics.

5.21.5 Wavelets Facilitate the Signal "Denoizing"

The wavelets provide also a method to extract the signal from the white noise, which exists at all the frequencies. The procedure is simple, we transform the signal by means of wavelets, we eliminate for all the resolutions the coefficients lower than a "threshold value" and we rebuild the signal with the remaining values. This method requires few information about the signal. Previously we were supposed to guess the type of regularity of the signal.

5.21.6 Wavelets are Less Selective in Frequency than the Fourier Transform

In the Fourier analysis, the analyzing function is a sinusoid of precise frequency and when we multiply it by the signal, the obtained coefficient only refers to this

frequency. On the other hand, a wavelet is composed of a mixture of frequencies (which is indicated by its own Fourier transform). The wavelet coefficients refer to this mixture of frequencies.

5.21.7 The Hybrid Transformations Allow an Optimal Adaptation to Transitory Complex Signals

Starting from the analytic observation concerning wavelet and Fourier transforms, the creation of "hybrid" tools turned out a necessity for an optimal adaptation to the structure of transitory signals, i.e. signals composed of different sub-structures. *The wavelet packets*, for example, are the product of a *wavelet* with an *oscillating function*. The *wavelet detects the abrupt change*, whereas the *oscillations reveal the regular variations*. This technique is used for example to study the *turbulence phenomena*.

Moreover, the algorithm of the "best basis (Coifman)" identifies the signal and in so doing directs its analysis towards:

– *Fourier transforms* for the *periodic "patterns"*
– *Wavelets* for the *irregular signals* with *strong short variations*
– *Wavelet packets* mentioned above, or towards the *Malvar wavelets* which are an evolution of wavelet packets (i.e. an attack, a plate and a decreasing of the analyzing function)

The algorithm selects, for each signal, the basis which will encode it with an optimal concision.

Finally, the *Pursuit* algorithm is better adapted than the "best basis" to the decomposition of non-stationary signals. This algorithm finds for each part of the signal the wave which resembles the most to it. Instead of seeking the optimal representation for the entire signal, we seek the optimal representation for each characteristic of the signal. To this end, the algorithm browses a wave and wavelet dictionary and selects the most resembling to the part of the gradually analyzed signal. The window of the algorithm is a *Gaussian* of variable size which respects the non-stationarities, modified by sinusoids of different frequencies. Starting from this structure it is possible to build an infinity of bases. The dictionary can be very large.

Chapter 6
The Atomic Decompositions of Signals

6.1 A Hybrid Transformation: Evolution of the "Matching Pursuit" Towards the Mallat and Zhang Version

It is one of the applications of waveforms theory, which corresponds to the transformation by the "Pursuit" algorithm with adaptive window. This technique will be applied in this chapter to a stock index, i.e. the French stock index: Cac40. We know that this transformation decomposes the signal in a time–frequency plane, the analyzing function is usually Gaussian of a variable width. The variable elements in the decomposition are the frequency, the position of the window and the size of the window, and we know that these three elements are independent. *This transformation is particularly adapted to the strongly non-stationary signals which contain very different components.* The algorithm "Pursuit" seeks the best "accord" (i.e. concordance) for each component of the signal rather than for the entire signal. The encoding or the decomposition of a non-stationary signal with "Pursuit" is concise and invariant by translation.

The latest version of the Matching Pursuit (i.e. Mallat and Zhang), that we present at the end of this chapter, has allowed J.B. Ramsey and Z. Zhang to establish several statements about the nature of the analyzed signals. This also made it possible to have a new perspective concerning the time–frequency analysis due to the use of the time–frequency atom dictionaries (i.e. *wave form dictionaries*). A wave form dictionary is a class of transforms that generalizes both windowed Fourier transforms and wavelets. Each wave form is parameterized by location, frequency and scale. Such methods can analyze signals that have highly localized structures in either time or frequency spaces as well as broad band structures. The matching pursuit algorithm is used to implement the application of wave form dictionaries to decompose the signal in the stock market index. Over long period, a stock market index shows very localized bursts of high intensity energy (see Ramses and Zhang) and in the neighborhood of which the signal is on the contrary very stable. Later on we will see how these explosions are decomposed and analyzed by the algorithm.

The traditional statistics still recently concluded[1] that the long-period stock market indexes have a *random* nature, whether they are stationary or not. Generally, it is said that they follow a random walk and they are also compared to the behaviors of Brownian motions. In econometrics, it is said that the non-stationary series have heteroscedastic variances.[2]

The Ramsey and Zhang analysis is to consider according to their own terms that "the energy of a system is largely internally generated, rather than the result of external forcing". In fact, they consider that oscillations of quasi-periodic nature, inside of which exist all the frequencies, can appear in an explosion of a signal, in a very localized manner. They also consider that the first proof that the stock exchange signals do not follow a random walk is that the number of wave forms necessary to represent them is smaller than for an usual random series. The *wave form dictionaries*, which are also called time–frequency atoms dictionaries, are noted

$$D = \{g_\gamma(t)\}_{\gamma \in \Gamma} \tag{6.1}$$

and a sub-dictionary is noted:

$$D\alpha = \{g_\gamma\}_{\gamma \in \Gamma}. \tag{6.2}$$

Before this recent version, the algorithms of Matching Pursuit exploited separately the Fourier dictionaries and the wavelet dictionaries. We provide some representations of sinusoid packets (also called cosine packets) and wavelet packets in the following section. These collections of sinusoids and wavelets are used for the decomposition of signals.

6.1.1 Construction of Sinusoid and Wavelet Packets

Starting from the Fourier series, we produce sinusoid collections of different types, they are named sinusoid packets (sine or cosine packets) and will be used for the signal decomposition. The sinusoid characteristics of a collection are indexed.

A similar methods is used for the wavelets, we generate different wavelet collections, these are wavelet packets which will be used for the signal decomposition. The wavelet characteristics of a collection are indexed. We give an example of these sinusoid and wavelet packets in Figs. 6.1 and 6.2, first, in time, then their respective spreads in frequency.

The figures which follow are important (Figs. 6.3 and 6.4), because they make it possible to observe how each sinusoid and wavelet is represented in the time–frequency plane. We observe the spread in the time–frequency plane of the Fourier

---

[1] Until the work of Lo and MacKinlay.

[2] Ref: Fama and LeBaron.
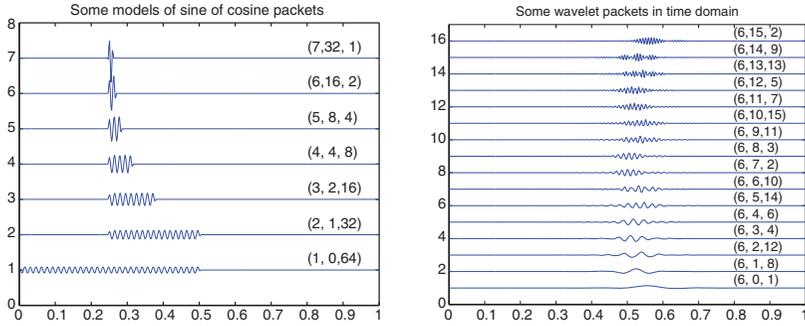
Fig. 6.1 *Left:* Sinusoid packets (time). *Right:* Wavelet packets (time)
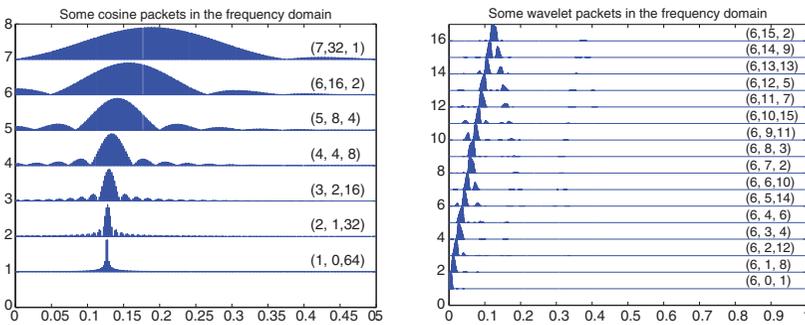
Fig. 6.2 *Left:* Sinusoid packets (frequency). *Right:* Wavelet packets (frequency)
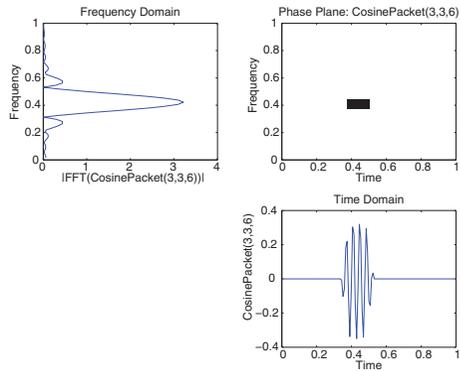
Fig. 6.3 A short sinusoid

atom or wavelet atom, i.e. their Heisenberg box, represented by a black rectangle in the upper right part of each group of figures. Hereafter, at the top-left of each group of figures, we provide the representation in the frequency-space and at the bottom-right we provide the representation in the phase-space.
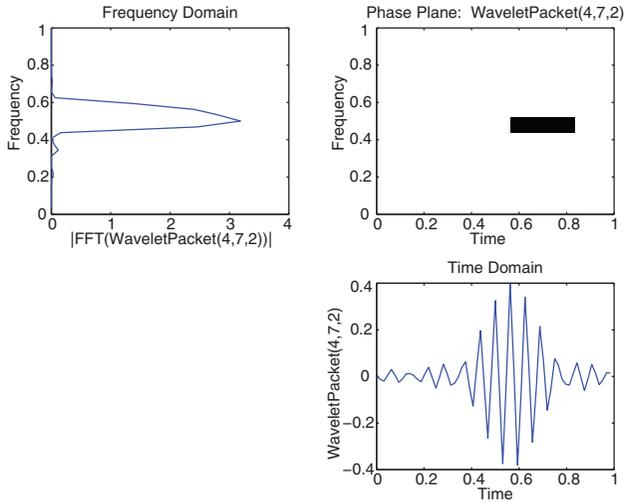
Fig. 6.4 A wavelet

Let us present hereafter, just for the record, some shapes of standard wavelets, Haar, Daubechies D4, coiflet C3 et symmlet S8:
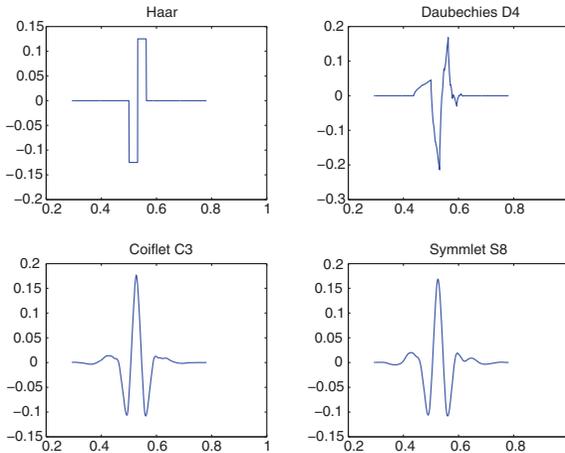
Now we are going to begin the presentation of the procedure of the (atomic) Matching Pursuit decomposition. Let us recall the definition of a time–frequency atom and the definition of the Matching Pursuit algorithm with redundant dictionary. Then, we present the version described by Mallat and Zhang of the Matching Pursuit algorithm with non-redundant time–frequency dictionaries.

6.1.2 Reminders About the Time–Frequency Atoms

We presented previously the decompositions of signals by families of well localized functions simultaneously in time and frequency; Such functions are called time–frequency atoms. The (short term) Fourier and wavelet transforms are examples of time–frequency decomposition of a signal. To extract information from complex signals, it is necessary to adapt the decomposition to the signal structure. S. Mallat and Z. Zhang describe an example of function with complex values belonging to a Hilbert space $L^2(\mathbb{R})$.[3] (See the section about the Frames, the sub-section about the Fourier transform in L^2 of square integrable functions: A discontinuous function or a function with singularities which is not integrable, has its square integrable.) Such a function can be noted: $\|g\|^2 = \int_{-\infty}^{+\infty} |g(t)|^2\, dt < +\infty$ with $g(t) \in L^2(\mathbb{R})$ $\hat{g}(\omega)$ is the Fourier transform of $g(t) \in L^2(\mathbb{R})$.

$$\hat{g}(\omega) = \int_{-\infty}^{+\infty} g(t) e^{-i\omega t}\, dt. \tag{6.3}$$

The inner product of $(g, f) \in L^2(\mathbb{R})^2$ is defined by $\langle g, f \rangle = \int_{-\infty}^{+\infty} g(t) \bar{f}(t) dt$, where $\bar{f}(t)$ is the complex conjugate of $f(t)$. A family of time–frequency atoms can be generated by scaling, translation and modulation of a windows function $g(t) \in L^2(\mathbb{R})$. It is supposed that the windows functions $g(t)$ are real and centered at zero. One poses $\|g\| = 1$, $\int g(t) \neq 0$, $g(0) = 0$. The atom is thus defined by:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t - \mu}{s}\right) e^{i\xi t}, \tag{6.4}$$

s = scaling, μ = translation, ξ = frequency modulation ω. $\gamma = (s, \mu, \xi)$ = index (The index γ is an element of the set $\Gamma = \mathbb{R}^+ \times \mathbb{R}^2$). The factor $\frac{1}{\sqrt{s}}$ normalizes the norm of $g_\gamma(t)$ towards 1. The function $g_\gamma(t)$ is centered at μ. Its Fourier transform $\hat{g}(\omega)$ is centered at the frequency $\omega = \xi$ and has an energy concentrated in a neighborhood of ξ (for which the size is proportional to $1/s$). The Fourier transform is:

$$\hat{g}(\omega) = \int_{-\infty}^{+\infty} g(t) e^{-i\omega t}\, dt \quad \text{and} \quad \hat{g}_\gamma(\omega) = \sqrt{s}\hat{g}(s(\omega - \xi)) e^{-i(\omega - \xi)u}. \tag{6.5}$$

The time–frequency dictionary of atoms is defined by $D = \{g_\gamma(t)\}_{\gamma \in \Gamma}$. A dictionary thus defined is very redundant, and contains collections of Fourier and wavelets windows. For any signal, and in particular when there exist complex structures or non-stationarities, we have to choose a (a priori) single general framework adapted to decompose it in a compact way by means of atom series of the dictionary. This adaptive framework of a redundant dictionary is described in the following section.

---

[3] See: sub-section about the "Concept of Frames" about the Fourier transform in L^2 of the square integrable functions. A discontinuous function or with singularities which is not necessary integrable, has its square integrable.

6.1.3 Reminders About the Matching Pursuit

Let H be an experimental signal and D a dictionary $g_\gamma(t)_{\gamma \in \Gamma}$ of vectors in H such that $\|g_\gamma\| = 1$. D is redundant and contains more vectors than necessary to build a basis. When the dictionary is redundant the representation of the signal is not single. An optimal approximation of $f \in H$, is written:

$$\tilde{f} = \sum_{n=1}^{N} a_n g_{\gamma n}, \qquad (6.6)$$

N is the number of terms; a_n and $g_{\gamma n} \in D$ are selected to minimize $\|f - \tilde{f}\|$. It is impossible to calculate an optimal solution numerically, thus, one calculates by means of a new algorithm a sub-optimal approximation of $f \in H$. The purpose is the calculation of successions of approximations of f, by means of orthogonal projections of the signal on elements of D. Given $g_{\gamma o} \in D$, the vector f can be decomposed into:

$$f = \langle f, g_{\gamma o} \rangle g_{\gamma o} + Rf. \qquad (6.7)$$

Rf is the residual vector after approximation of f in the direction of $g_{\gamma o}$. $g_{\gamma o}$ is orthogonal with Rf and is normalized to 1. Thus:

$$\|f\|^2 = |\langle f, g_{\gamma o} \rangle|^2 + \|Rf\|^2. \qquad (6.8)$$

to minimize $\|Rf\|$, we must take $g_{\gamma o}$ such that $|\langle f, g_{\gamma o} \rangle|$ is maximum. In some cases one finds $g_{\gamma o}$ "sub-optimal", i.e.

$$|\langle f, g_{\gamma o} \rangle| \geqslant \alpha \sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|, \qquad (6.9)$$

where α is an optimal factor which satisfies $0 \leq \alpha \leq 1$. The iterative algorithm decomposes the residue Rf of an antecedent (former) projection by a projection of the residue on a vector D, as carried out about f. This is repeated for each residue of a former projection. Given $R^0 f = f$, one supposes that the residue of the nth order $R^n f$ for any n (positive or zero) was calculated. Then, one chooses an element $g_{\gamma n} \in D$ which approximates the residue $R^n f$;

$$|\langle R^n f, g_{\gamma n} \rangle| \geqslant \alpha \sup_{\gamma \in \Gamma} |\langle R^n f, g_\gamma \rangle|, \qquad (6.10)$$

$R^n f$ is decomposed into:

$$R^n f = \langle f, g_{\gamma n} \rangle g_{\gamma n} + R^{n+1} f, \qquad (6.11)$$

which provides the residue of order $n + 1$. Thus $R^{n+1} f$ is orthogonal with $g_{\gamma n}$

$$\|R^n f\|^2 = |\langle R^n f, g_{\gamma n} \rangle|^2 + \|R^{n+1} f\|^2. \qquad (6.12)$$

if this decomposition is continuous at the order m, then f was decomposed into:

$$f = \sum_{n=0}^{m-1} \left(R^n f - R^{n+1} f \right) + R^m f \tag{6.13}$$

replacing the equation $\left| \langle R^n f, g_{\gamma n} \rangle \right| \geqslant \alpha \sup_{\gamma \in \Gamma} \left| \langle R^n f, g_\gamma \rangle \right|$ in the equation $\| R^n f \|^2 = \left| \langle R^n f, g_{\gamma n} \rangle \right|^2 + \| R^{n+1} f \|^2$, one obtains:

$$f = \sum_{n=0}^{m-1} \langle R^n f, g_{\gamma n} \rangle + R^m f. \tag{6.14}$$

$\| f \|^2$ is decomposed in:

$$\| f \|^2 = \sum_{n=0}^{m-1} \left(\| R^n f \|^2 - \| R^{n+1} f \|^2 \right) + \| R^m f \|^2. \tag{6.15}$$

The equation $|R^n f| = \langle f, g_{\gamma n} \rangle g_{\gamma n} + R^{n+1} f$, provides an *equation of energy conservation:*

$$\| f \|^2 = \sum_{n=0}^{m-1} \langle R^n f, \ g_{\gamma n} \rangle^2 + \| R^m f \|^2. \tag{6.16}$$

The decomposition of f is nonlinear, but the energy conservation is respected. We observe when m increases how evolves the residue $R^m f$. Mallat and Zhang (1993) checked that "Pursuit" converges in spaces of finite or infinite dimension.

6.1.3.1 Mallat and Zhang Theorem

Theorem 6.1 (Mallat and Zhang). *Given $f \in H$. The residue defined by the following induction equation* $\left| \langle R^n f, g_{\gamma n} \rangle \right| \geqslant \alpha \sup_{\gamma \in \Gamma} \left| \langle R^n f, g_\gamma \rangle \right|$ *satisfies* $\lim_{m \to +\infty} \| R^m f \| = 0$*, then it comes:*

$$f = \sum_{n=0}^{+\infty} \langle R^n f, g_{\gamma n} \rangle g_{\gamma n} \tag{6.17}$$

then

$$\| f \|^2 = \sum_{n=0}^{n-1} \langle R^n f, g_{\gamma n} \rangle^2 \tag{6.18}$$

when H has a finite dimension, $\| R^m f \|$ decays exponentially towards zero.

6.1.4 Improvement of the Algorithm

When the dictionary is very redundant, the search of the best residue-signal accord can be limited to a sub-dictionary:

$$D_\alpha = \{g_\gamma\}_{\gamma \in \Gamma} \subset D. \tag{6.19}$$

It is supposed that $\Gamma_\alpha \subset \Gamma : \forall f \in H$

$$\sup_{\gamma \in \Gamma_\alpha} \left| \langle f, g_\gamma \rangle \right| \alpha \geqslant \sup_{\gamma \in \Gamma} \left| \langle f, g_\gamma \rangle \right|. \tag{6.20}$$

We start by calculating $\left(\langle f, g_\gamma \rangle \right)_{\gamma \in \Gamma_\alpha}$, then after the calculation of $\left(\langle R^n f, g_\gamma \rangle \right)_{\gamma \in \Gamma_\alpha}$, we search in D_α an element $g_{\tilde\gamma n}$ such that:

$$\left| \langle R^n f, g_{\tilde\gamma n} \rangle \right| = \sup_{\gamma \in \Gamma_\alpha} \left| \langle R^n f, g_\gamma \rangle \right|, \tag{6.21}$$

where $g_{\tilde\gamma n}$ approximates f better than $g_{\tilde\gamma n}$. We seek via the Newton method for an index γ_n inside a neighborhood of $\tilde\gamma_n$ in Γ, where $\left| \langle f, g_{\gamma n} \rangle \right|$ *reached a local maximum. The optimum* provides the following inequalities:

$$\left| \langle R^n f, g_{\gamma n} \rangle \right| \geqslant \left| \langle R^n f, g_{\tilde\gamma n} \rangle \right| \geqslant \alpha \sup_{\gamma \in \Gamma} \left| \langle R^n f, g_\gamma \rangle \right| \tag{6.22}$$

when $g_{\gamma n}$ (better than $g_{\tilde\gamma n}$) is found, Mallat and Zhang calculate the inner product of the new residue $R^{n+1} f$ with any $g_\gamma \in D_\alpha$, by means of an update formula resulting from the equation $\left| \langle R^n f, g_{\gamma n} \rangle \right| \geqslant \alpha \sup_{\gamma \in \Gamma} \left| \langle R^n f, g_\gamma \rangle \right|$, i.e.:

$$\langle R^{n+1} f, g_{\tilde\gamma n} \rangle = \langle R^n f, g_\gamma \rangle - \langle R^n f, g_{\gamma n} \rangle \langle g_{\gamma n}, g_\gamma \rangle \tag{6.23}$$

during the updates only the calculation of $\langle g_{\gamma n}, g_\gamma \rangle$ is necessary. The dictionaries are generally built with a small number of operations. It is the desired precision ε which decides of the level and number of sub-decomposition. We write:

$$\|R^p f\| = \left\| f - \sum_{n=0}^{p-1} \langle R^n f, g_{\gamma n} \rangle \right\| \leqslant \varepsilon \|f\|, \tag{6.24}$$

where p is the number of iterations. The energy conservation equation $\|f\|^2 = \sum_{n=0}^{n-1} \langle R^n f, g_{\gamma n} \rangle^2 + \|R^m f\|^2$ allows to write that the equation below is equivalent to

$$\|f\|^2 - \sum_{n=0}^{p-1} \left| \langle R^n f, g_{\gamma n} \rangle \right|^2 \leqslant \varepsilon^2 \|f\| \tag{6.25}$$

the residue $R^n f$ was not calculated at each iteration, therefore we have to check if the equation $\|R^p f\| = \left\| f - \sum_{n=0}^{p-1} \langle R^n f, g_{\gamma n} \rangle \right\| \leqslant \varepsilon \|f\|$ is verified.

6.1.5 Mallat and Zhang Version of Matching Pursuit with Dictionaries of Time–Frequency Atoms g_γ

The objective is to proceed to *an adaptive time–frequency transformation*. It provides a particular *decomposition in a sum of complex time–frequency atoms*. The particularity is that "the distribution of time–frequency energy results from the sum of the Wigner distribution of each time–frequency atom". The preceding Mallat and Zhang theorem proves[4] that the "Matching Pursuit" decomposes any function of L^2 into:

$$f = \sum_{n=0}^{+\infty} \langle R^n f, g_{\gamma n} \rangle g_{\gamma n}, \qquad (6.26)$$

with:[5]

$$g_{\gamma n}(t) = \frac{1}{\sqrt{s_n}} g\left(\frac{t - \mu_n}{s_n}\right) e^{i\xi_n t}, \qquad (6.27)$$

where $\gamma_n = (s_n, u_n, \xi_n)$. In the discrete mode, (for a signal length N), we have: $\gamma = (s, u, \xi) = (2^j, 2^{j-1}p, k\pi 2^{-j})$, with $0 \leqslant j \leqslant \log_2 N$, $0 \leqslant p \leqslant N\, 2^{-j-1}$, $0 \leqslant k \leqslant N\, 2^{j+1}$. Thus, for an atom n, we have:

$$g_{\gamma n}(t) = \frac{1}{\sqrt{2^{j_n}}} g\left(\frac{t - 2^{j-1}p_n}{2^{j_n}}\right) e^{ik_n \pi 2^{-j_n} t} \qquad (6.28)$$

$(j_n, p_n, \xi_n$ are the values for the atom n). From this decomposition we can obtain for each atom a Wigner distribution. Let us recall that initially the definition of the Wigner distribution (with cross terms) for two functions $f(t)$ and $h(t)$:

$$\mathbf{D}[f, h](t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(t + \tau/2)\bar{h}(t - \tau/2)e^{-i\omega\tau} d\tau, \qquad (6.29)$$

and the Wigner distribution of $f(t)$ is $\mathbf{D}f(t, \omega) = \mathbf{D}[f, f](t, \omega)$:

$$\mathbf{D}[f, f](t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(t + \tau/2)\bar{f}(t - \tau/2)e^{-i\omega\tau} d\tau. \qquad (6.30)$$

Since the *Wigner distribution is quadratic*, we obtain from *the atomic decomposition of $f(t)$* the following expression:

$$\mathbf{D}f(t, \omega) = \sum_{n=0}^{\infty} \left|\langle R^n f, g_{\gamma n} \rangle\right|^2 \mathbf{D}g_{\gamma n}(t, \omega) \qquad (6.31)$$

$$+ \sum_{n=0}^{+\infty} \sum_{n=0,\, m \neq n}^{+\infty} \langle R^n f, g_{\gamma n} \rangle \overline{\langle R^m f, g_{\gamma m} \rangle} \mathbf{D}[g_{\gamma n}, g_{\gamma m}](t, \omega).$$

---

[4] A dictionary of time–frequency atom is complete.

[5] ["a_n" in the equation of the "Matching Pursuit" $\tilde{f} = \sum_{n=1}^{N} a_n g_{\gamma n}$, corresponds to $\langle R^n f, g_{\gamma n} \rangle$ in the expression above of $f = \sum_{n=0}^{+\infty} \langle R^n f, g_{\gamma n} \rangle g_{\gamma n}$].

The double sum corresponds to the cross terms of the Wigner distribution, which generally are removed to improve the image of the energy distribution (see sections about Cohen's class and Wigner distributions). This is also done here and it makes it possible to obtain *an expression which is interpreted as a new energy density*

$$\mathbf{E}f(t,\omega) = \sum_{n=0}^{+\infty} \left| \langle R^n f, g_{\gamma n} \rangle \right|^2 \mathbf{D}g_{\gamma n}(t,\omega). \tag{6.32}$$

From the dilation and translation properties of the Wigner distribution, one can write for $\gamma = (s,\mu,\xi)$:

$$\mathbf{D}g_\gamma(t,\omega) = \mathbf{D}g\left(\frac{t-\mu}{s}, s(\omega - \xi)\right) \tag{6.33}$$

thus, for $\gamma_n = (s_n, u_n, \xi_n)$:

$$\mathbf{E}f(t,\omega) = \sum_{n=0}^{+\infty} \left| \langle R^n f, g_{\gamma n} \rangle \right|^2 \mathbf{D}g\left(\frac{t-\mu_n}{s_n}, s_n(\omega - \xi_n)\right). \tag{6.34}$$

Since the Wigner distribution also checks: $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{D}g(t,\omega) dt d\omega = \|g\|^2 = 1$, the energy conservation equation: $f = \sum_{n=0}^{+\infty} \langle R^n f, g_{\gamma n} \rangle g_{\gamma n}$, makes it possible to write:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbf{E}f(t,\omega) dt d\omega = \|f\|^2. \tag{6.35}$$

$\mathbf{E}f(t,\omega)$ is analyzed as an *Energy Density* of f in a time–frequency plane (t,ω). If $g(t)$ is, for example, the following *Gaussian window*:

$$g(t) = 2^{1/4} e^{-\pi t^2}, \tag{6.36}$$

it comes:

$$\mathbf{D}g(t,\omega) = 2e^{-2\pi\left(t^2 + (\omega/2\pi)^2\right)}. \tag{6.37}$$

Mallat and Zhang explain that *the time–frequency atoms $g_\gamma(t)$ are then called Gabor functions* and *the energy distribution can be understood as "a sum of Gaussian pulses"* whose positions and spreads in the time–frequency plane depend on $\gamma_n = (s_n, \mu_n, \xi_n)$.

Illustration for an artificial signal: Figure 6.5 provides an elementary illustration of what has just been presented previously. It is the application of the Matching Pursuit on a simple signal and the representation of the Wigner distribution in the time–frequency plane of the atomic decomposition of this signal. This signal is composed of a simple sinusoid, two Gabor atoms and a Dirac function and a noise along the entire signal;

Signal = [a short sinusoid + two Gabor atoms + a dirac + noise].
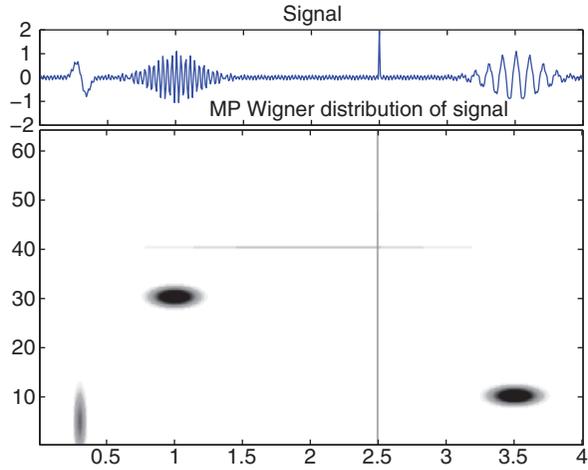
Fig. 6.5 Energy distribution of the signal in the time–frequency plane
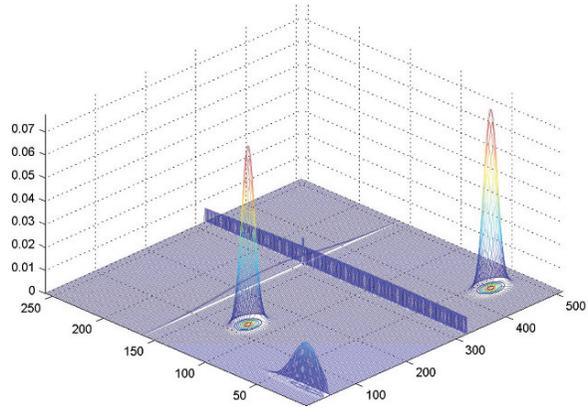
Fig. 6.6 Meshgrid plot
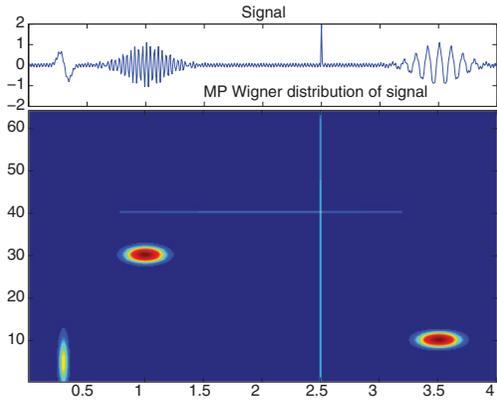
Fig. 6.7 Energy distribution of the signal (jet version)
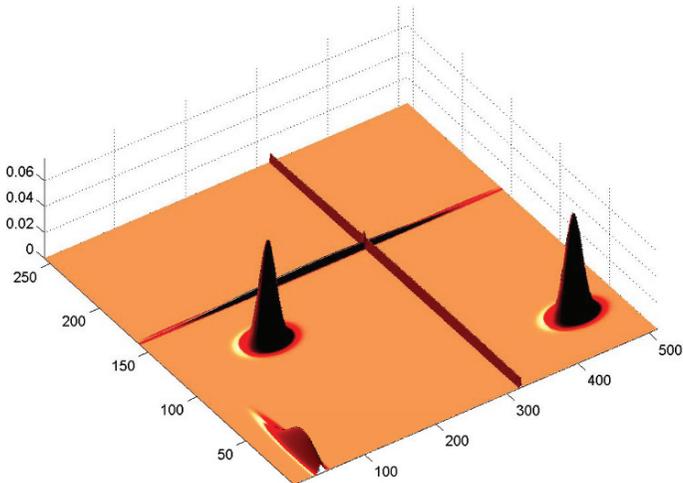
Fig. 6.8 Energy distribution of the signal in the time–frequency-(amplitude) space

We observe clearly the elements from the atomic decomposition of the MP in the time frequency-(amplitude) space. We identify the images of the two Gabor atoms which have different frequencies, the vertical line of the Dirac function, the very localized sinusoid in the lower left part and the horizontal line in the upper half which represents the noise along the signal (Figs. 6.6–6.8).

Application to a standard signal of transitory behavior (i.e. transient): This sample made of 512 points is well-known and is used in many works relating to signal processing.
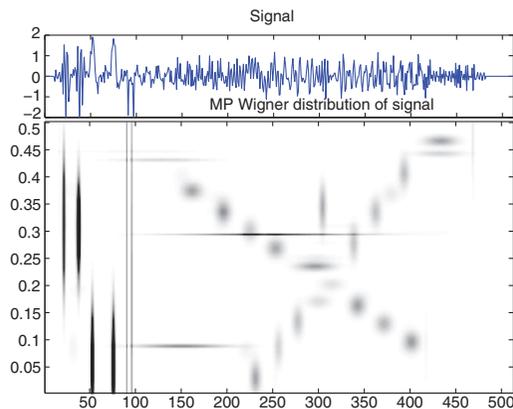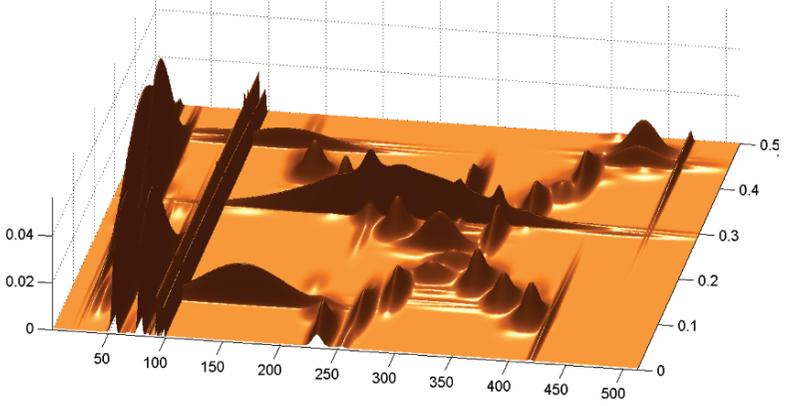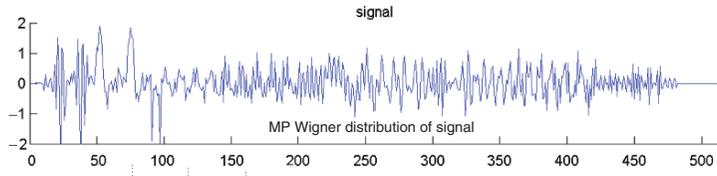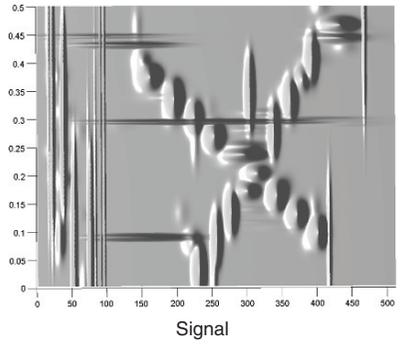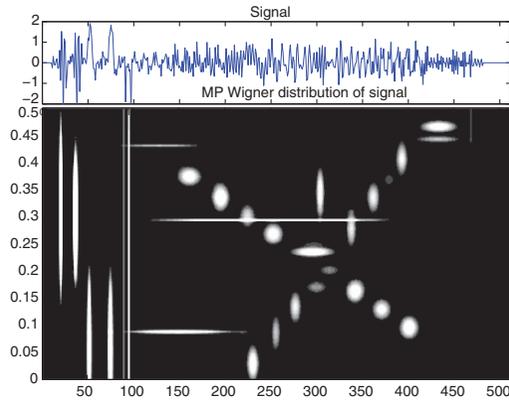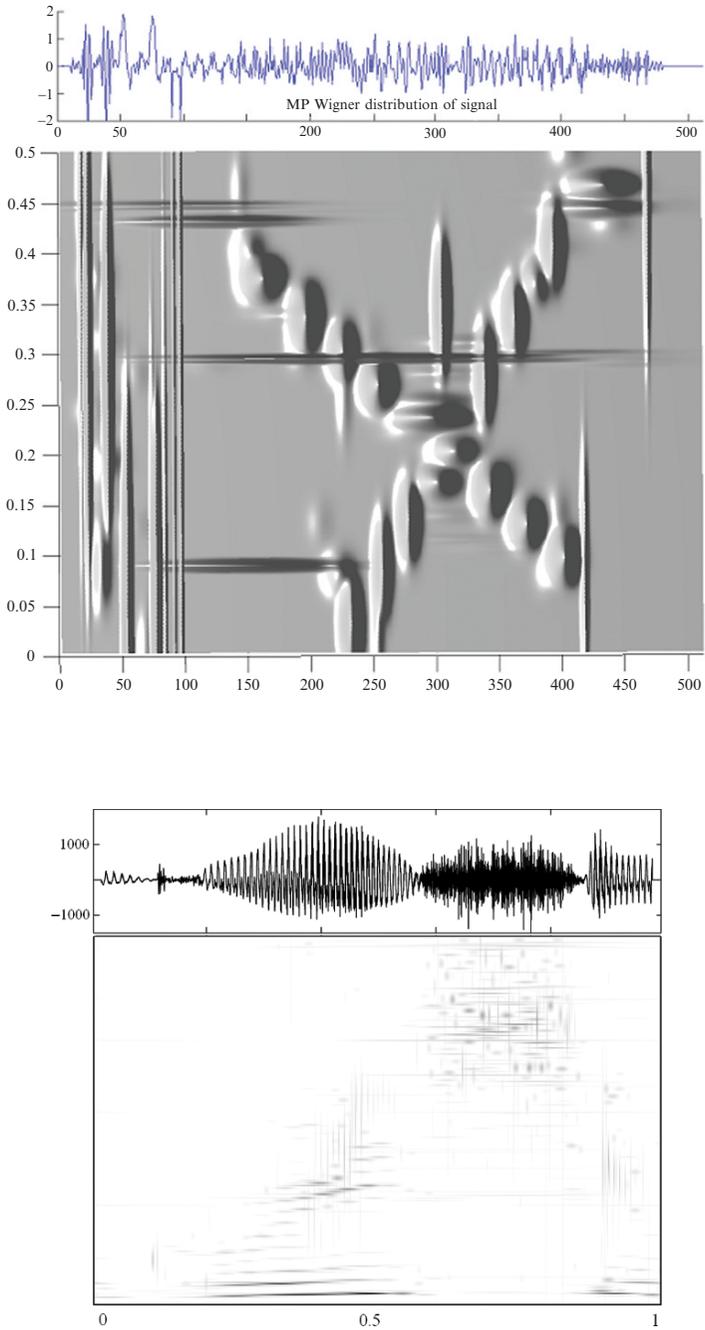
Fig. 6.9 Wigner–Ville distribution of Gabor atoms

Fig. 6.10 Signal and energy density (B&W)
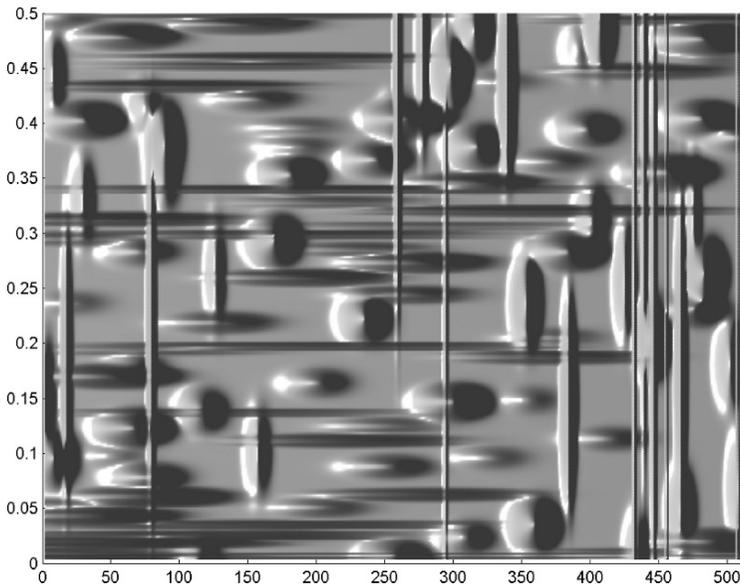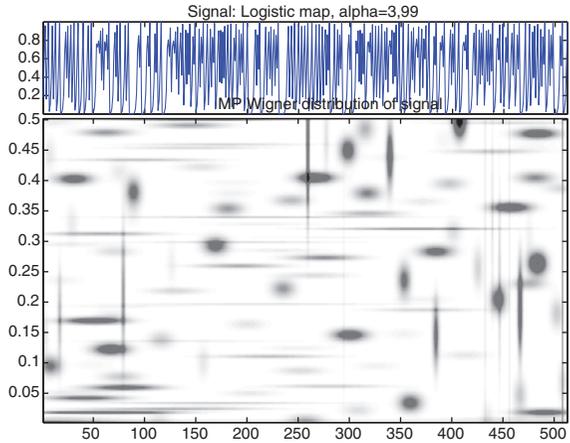
Fig. 6.11 Energy density or Wigner density of the atoms

Application to a sound recording: Recording of the word "greasy" with a sample of 8,192 points. The dark blobs are Wigner–Ville distributions of Gabor atoms extracted by the matching pursuit and shown in the time–frequency plane (Fig. 6.9).

Application to the logistic equation in the chaotic field: Illustration by the logistic equation $x_{t+1} = \alpha x_t (1 - x_t)$ for $\alpha \simeq 4$ with a sample of 512 points. We observe the *different atoms* with *their respective shape* and *spread* (Figs. 6.10–6.15).

Fig. 6.12 Wigner density (elevation 70°)
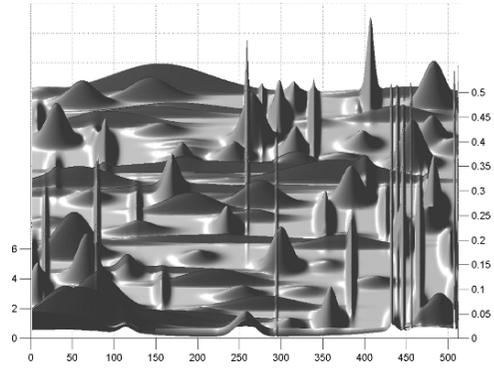
Fig. 6.13 Perspective of Wigner density of the attractor
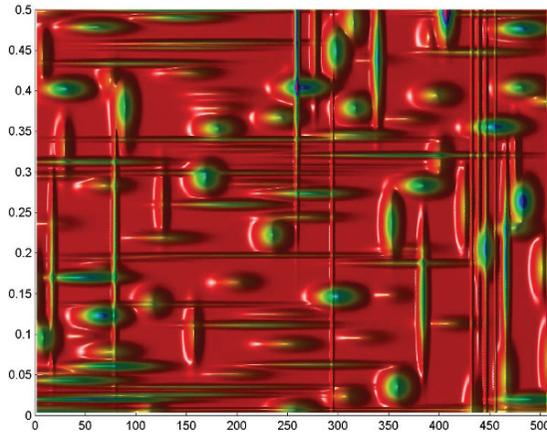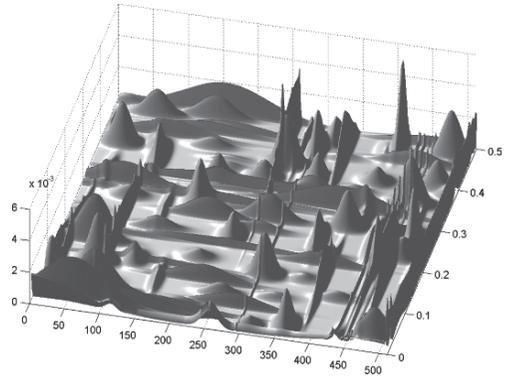
Fig. 6.14 Wigner density of the logistic attractor (HSV colormap version)

Fig. 6.15 Perspective (HSV)
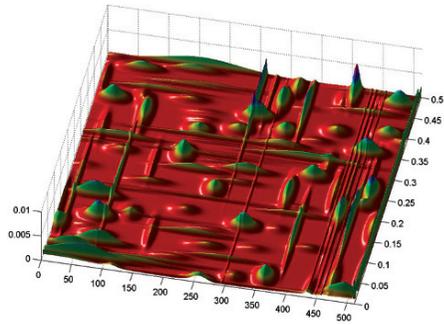
6.2 Applications of the Different Versions of the "Matching Pursuit" to a Stock-Exchange Index: *Cac40*

In this example we successively apply the following algorithms to the French stock-exchange index:

- The atomic decomposition by the MP with Fourier dictionaries
- The atomic decomposition by the MP with Wavelet dictionaries
- The atomic decomposition by the MP developed by Mallat and Zhang with time–frequency atom dictionaries

The first and second decompositions are applied to a sample of 2,048 daily values of the *Cac40 growth rate* corresponding to more than 8 years from January 1988 (Figs. 6.16 and 6.17).

The *third decomposition* that applies the Mallat and Zhang version of the Matching Pursuit (with Gabor, Fourier and Dirac atoms) is tried out on a longer sample, i.e. 11 years starting from January 1, 1988.

6.2.1 Matching Pursuit: Atomic Decomposition with Fourier Dictionary

We apply the atomic decomposition of the matching pursuit by means of Fourier dictionary, that means sine and cosine packets (i.e. sinusoid packets). As a preliminary and by way of example, with an aim of better visualizing the Fourier atom repertories, we present the extraction by means of the Matching Pursuit of 50 vectors of sinusoids packets from the 256 values (taken randomly) of the growth rate of the stock-exchange index: Cac40, we obtain the following repertories (Fig. 6.18).
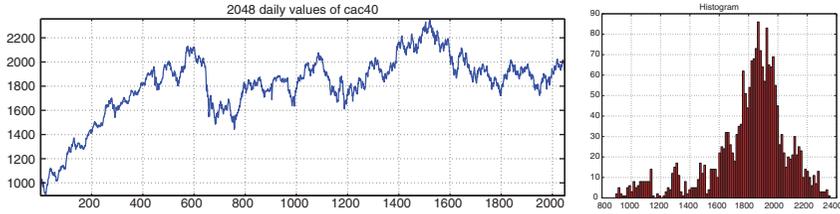
Fig. 6.16 *Left:* 2,048 daily values of the stock index. *Right:* Distribution
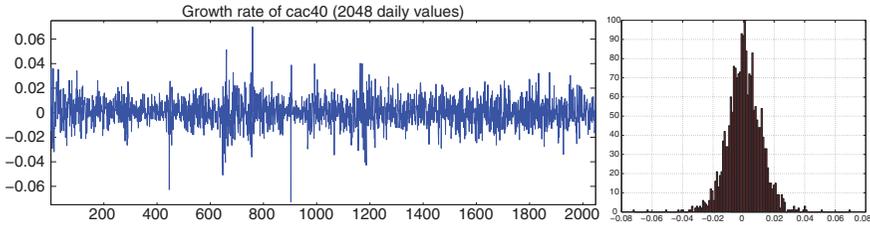
Fig. 6.17 *Left:* 2,048 daily growth rates. *Right:* Distribution

The following figure shows the Fourier dictionaries (i.e. sinusoid packets) that decompose the signal by depth levels.
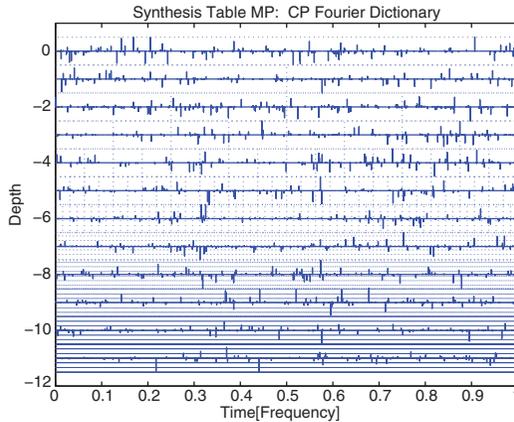
Observe in Fig. 6.19 the representation in the time–frequency plane of the Heisenberg boxes of the Fourier atoms resulting from the decomposition by the MP of 2,048 daily values of the stock-exchange index. The different types of symbol (circle, cross, point) inside the rectangles represent the intensity or the amplitude. Then, in order to better representing the tiling of the plane by the Heisenberg boxes, we provide various images, either through a gray scale or a jet color scale.

In order to better visualize the boxes, we provide an illustration for a shorter sample corresponding to 512 daily values (Fig. 6.20).
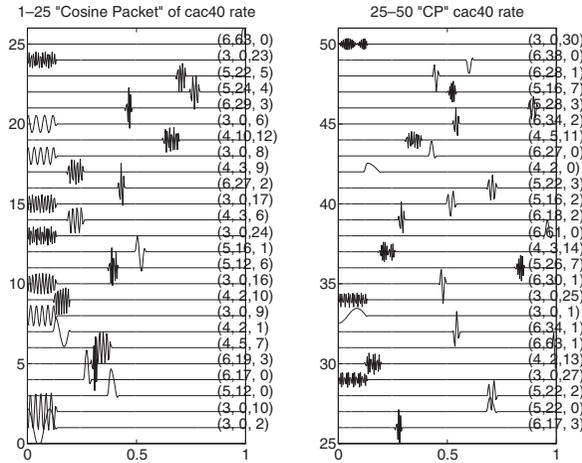
Fig. 6.18 Sample of the dictionary repertories of Fourier waveforms
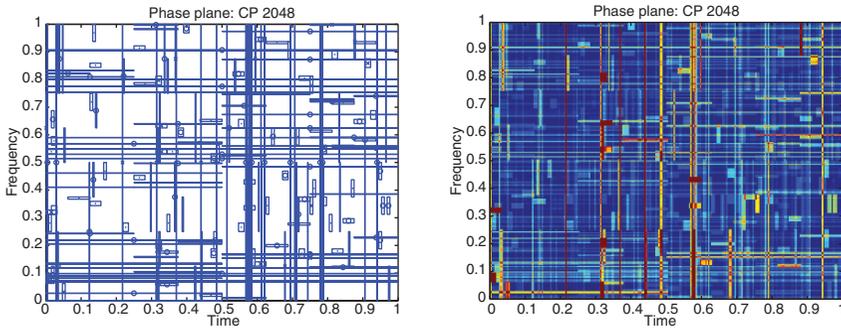
Fig. 6.19 Heisenberg boxes of Fourier atoms (*left*), image of Heisenberg boxes (*right*)

6.2.2 Matching Pursuit: Atomic Decomposition with Wavelet Dictionary

We apply the decomposition of the matching pursuit by means of wavelet packet dictionary. We will show neither the repertories, nor the wavelet dictionaries whose pictures have a weak interest. On the other hand, the Heisenberg boxes are shown in Fig. 6.21.

In order to better visualize the boxes, one provides an illustration for a shorter sample, corresponding to 512 daily values (Fig. 6.22).
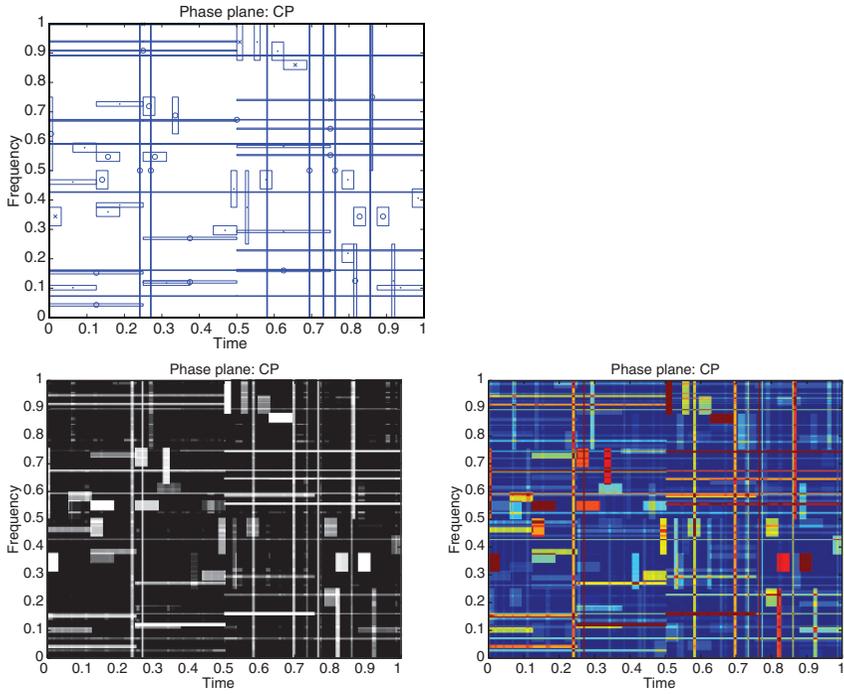
Fig. 6.20 Heisenberg boxes of Fourier atoms for a shorter sample (512 values) (*top left*), image of Heisenberg boxes (B&W) (*bottom left*), Jet version (*bottom right*)
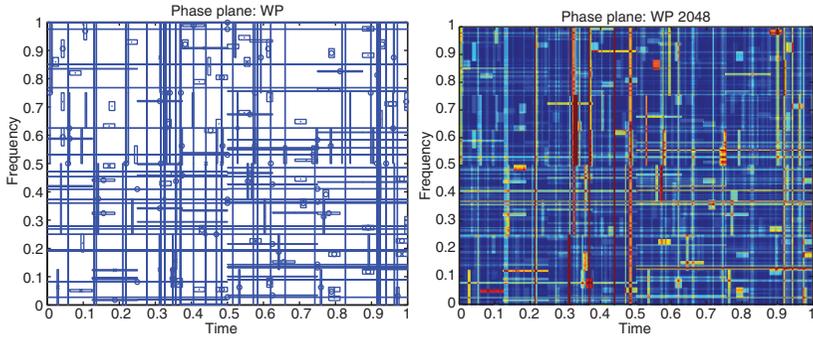
Fig. 6.21 Boxes of wavelet atoms (*left*), image of the atoms (*right*)
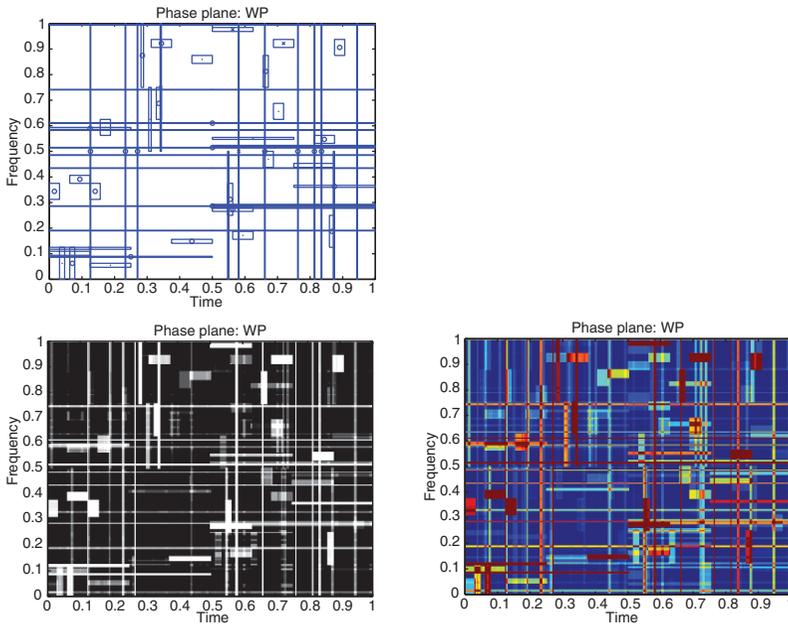
Fig. 6.22 Heisenberg boxes of wavelet atoms for a shorter sample (*top left*), image of boxes (B&W) (*bottom left*), Jet version (*bottom right*)

6.2.3 An Application of the Mallat and Zhang "Matching Pursuit" Version: An Adaptive Atomic Decomposition of a Stock-Exchange Index with Dictionaries of Time–Frequency Atoms g_γ

Consider again the expression of the new energy density suggested by Mallat and Zhang and described in one of the preceding sections:

$$\mathbf{E}f(t,\omega) = \sum_{n=0}^{+\infty} |\langle R^n f, g_{\gamma n}\rangle|^2 \, \mathbf{D}g\left(\frac{t-\mu_n}{s_n}, s_n(\omega-\xi_n)\right), \qquad (6.38)$$

where $\mathbf{E}f(t,\omega)$ is analyzed as an energy density of a signal f in the time–frequency plane (t,ω). We apply this MP version to the growth rate of 2,847 daily values of the Cac40 (see Fig. 6.24) from January 01, 1988 to June 30, 1999 (Figs. 6.25–6.30).

It is possible to observe in Figs. 6.25–6.30 the great number of atoms due to the length of the signal which shows more than 11 years of the French stock-exchange index. We can also observe for those which are visible in this image, the Dirac functions, in particular, at the beginning of the signal around $t = 100$, then little before the point $t = 500$, then around $t = 1,000$, then at $t = 1,400$, $t = 1,990$, $t = 2,550$ and especially at $t = 2,710$.

Fig. 6.23 *Left:* 2,847 daily values. *Right:* Distribution
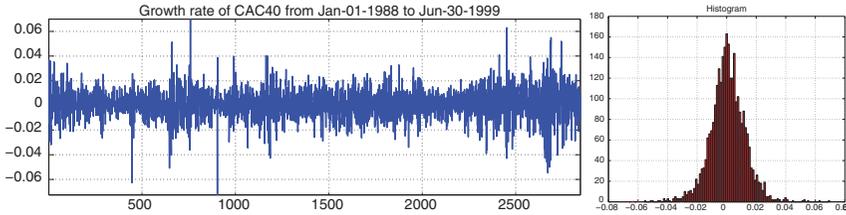
Fig. 6.24 *Left:* 2,847 daily growth rate. *Right:* Distribution
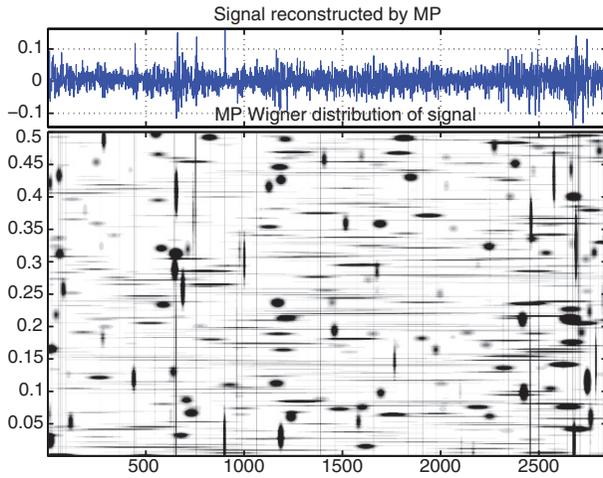
Fig. 6.25 Distribution of the atoms

Hereafter, in order to better visualize the energy density (i.e. the Wigner density) in this time–frequency plane, we present it in perspective through three dimensions: i.e. time–frequency-amplitude.
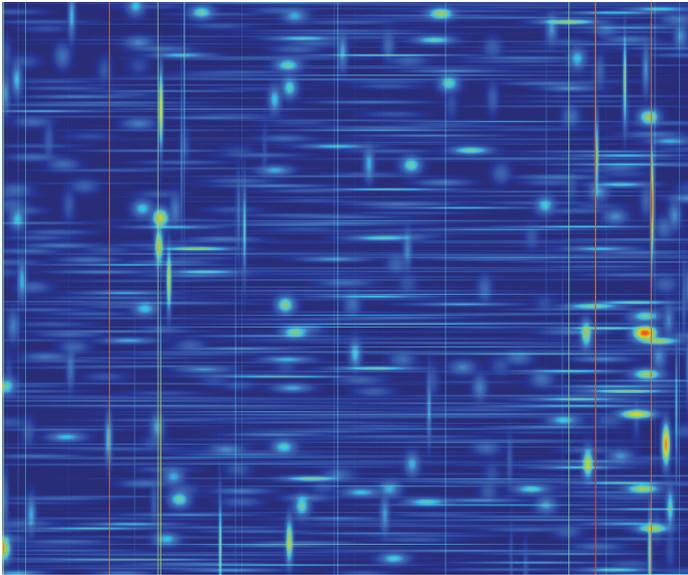
Fig. 6.26 Wigner density (Jet)
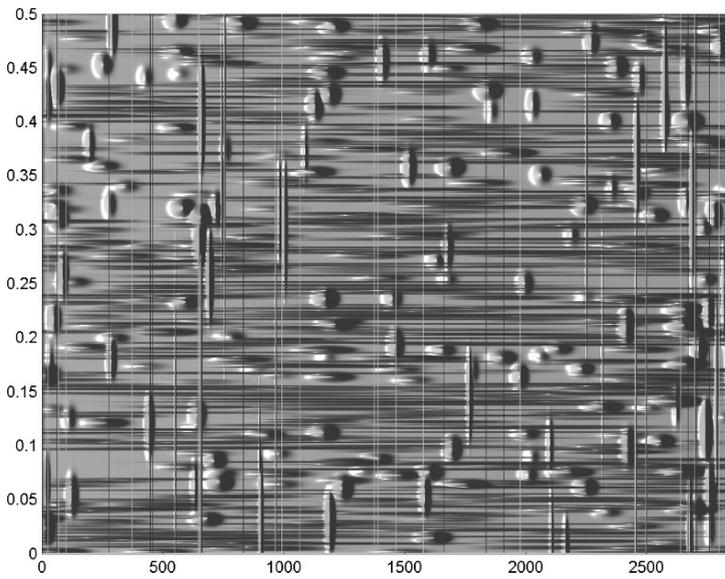
Fig. 6.27 Wigner density
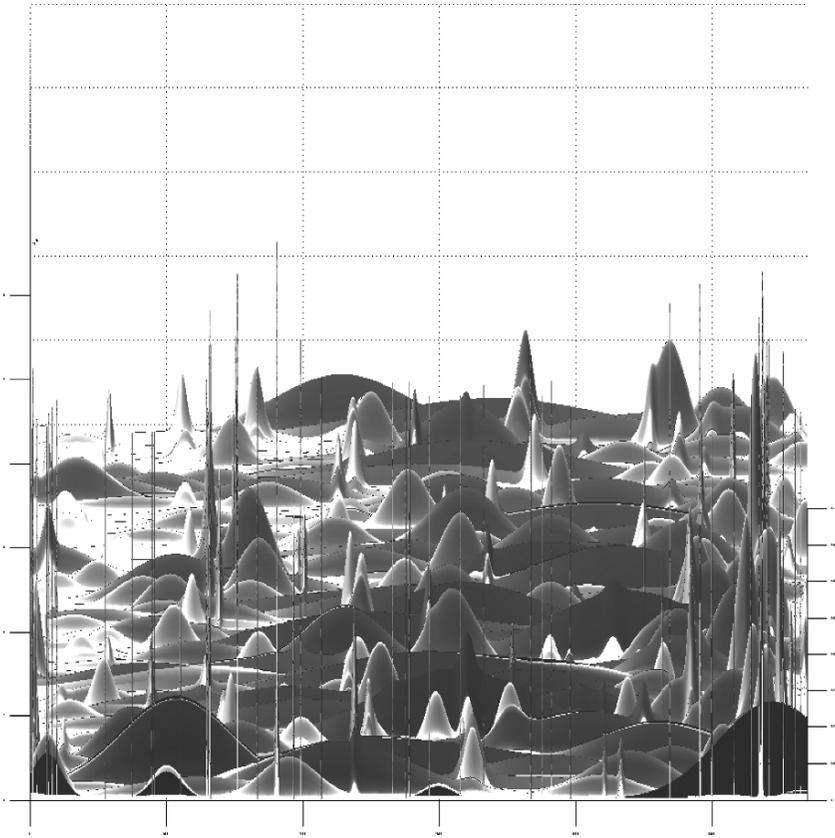
Fig. 6.28 Density of energy [Wigner distribution] (Elevation 35°, azimuth 0°)
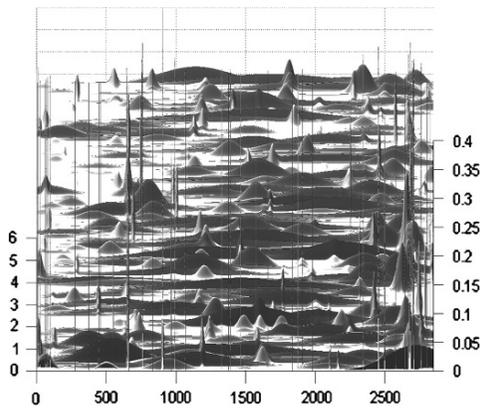
Fig. 6.29 Density of energy [Wigner distribution] (Elevation 75°, azimuth 0°)
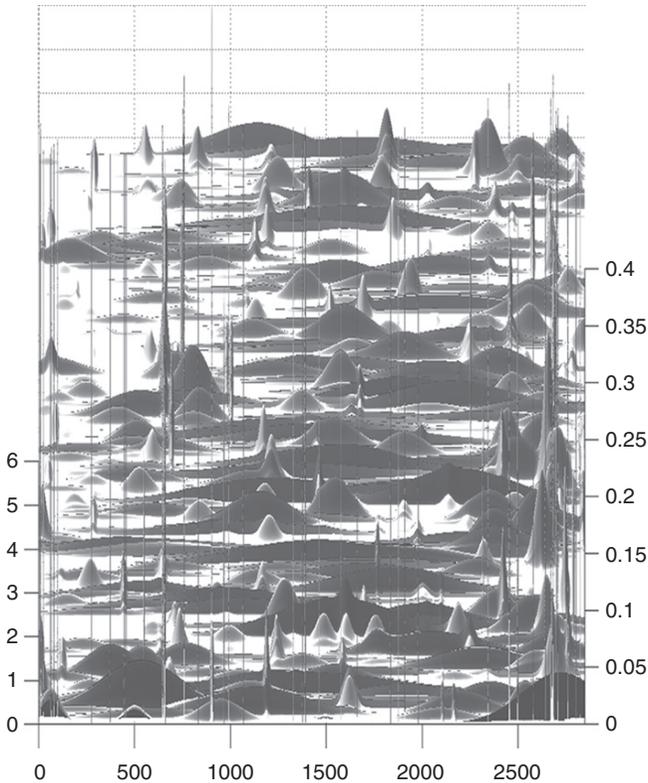
Fig. 6.30 Distribution of Time-Frequency atoms and Dirac functions.

6.3 Ramsey and Zhang Approach of Stock Market Crises by Matching Pursuit with Time–Frequency Atom Dictionaries: High Intensity Energy Periods

It would be interesting to test this method on a stock market index offering a *continuity* throughout the twentieth century and compare this series and its decomposition with the historical events which impacted this financial index and the economy. The French index does not allow probably this experiment, indeed, it would be necessary to study the continuity between the French index before and after 1988, since in 1988 a new index has been created (i.e. Cac40). The Ramsey and Zhang work (using the Mallat and Zhang algorithm) is based on the *S&P500 index* of the United States *which seems to have a better structural continuity*. If one evokes the growth rate of the index, it is possible to observe that the very dense signal is traversed by fast oscillations with localized bursts of very large amplitude. This type of behavior favors the use of analysis tools such as the *waveform* dictionaries, i.e. the time–frequency

atoms dictionaries. The *Dirac Delta function* have a particular interest in this signal decomposition method to understand these types of behavior.

6.3.1 The Dirac Function Would Allow to Distinguish Isolated and Intense Explosions: Internal Shocks and External Shocks

J.B. Ramsey and Z. Zhang distinguish for a signal between "isolated impulses (i.e. isolated bursts or explosions)" and "intense impulses" (i.e. explosions of high intensity energy). The first one being able to be represented by *Dirac delta functions which indeed involve all the frequencies of the spectrum*, and the second one involving *almost all the frequencies* but relatively less at the very high and very low frequencies. The distinction enunciated above comes from the observation of stock market crises. At the observation, the stock exchange crises are periods of high intensity energy which appear "without apparent warning", such that Ramsey and Zhang explain it, and go up very quickly and die out equally abruptly. Nevertheless, these bursts are not the result of isolated impulses represented by Dirac delta functions. In fact, there is many cases of isolated impulses in a stock exchange index but their occurrence is separated in general from intense bursts.

An interesting lighting is that the crisis of 1929 and that of 1987 do not show, by means of their decomposition, significant Dirac delta functions, even if the two periods of crises are the center of narrow zones very localized in time with many energy bursts. This indicates that the major crises of the stock exchange and even the smaller crises are not the result of isolated impulses. Thus, by means of the observation, the intense bursts cannot be associated with isolated impulses, or external shocks. *Ramsey and Zhang explain that the intense bursts of high intensity energy which are prominent in the stock-exchange times series seem to be internal to the systems*. Thus, according to their own terms, it is possible "reasonably to infer from these results that bursts of intense activity are not the result of isolated, unanticipated, external shocks, but more likely the result of the operation of a dynamical system with some form of intermittency in dynamics". *The Ramsey and Zhang analysis conclude that the data of financial markets do not follow a random walk*, *even if their structure is very complex*. This observation is based on the fact *that the number of "waveforms" necessary to decompose a financial series is smaller than to decompose a random series*. This result is stronger when the period of the financial index observed is rather stable but *is traversed by some very strong very localized bursts*. That means that when these strong energy explosions occur without possibility to anticipate them by means of the period *which precedes* and when the following period shows also a strong stability. Nevertheless, these bursts are not the result of isolated impulses which would be represented by Dirac Delta functions. There exist however many cases of isolated impulses, but their occurrence generally are separated from the intense energy explosions that we mentioned

previously. The occurrence of Dirac delta functions would represent impulses or shocks of the system from external sources, whereas on the other hand the bursts of high intensity energy, which are preeminent, seem to be internal to the system. These explosions contain high energy at most frequencies, although relatively less at very low and very high frequencies.

Even if this analysis poses many questions, the interesting result shows that the isolated impulses which can be formalized by Dirac Delta functions could be the expression of external shocks which cannot be anticipated. On the other hand the explosions of intense activity are not the result of isolated, unanticipated, external shocks, but according to the own sentence of the authors, "but more likely the result of the operation of a dynamical system with some form of intermittency in dynamics".

6.4 Comments About Time–Frequency Analysis

The spectral analysis identifies the similarities of a signal and it is commonly said that the spectrum of a series is the equivalent in the frequencies of the autocorrelation function in the field of time. *Thus, the periodic phenomena are read by the spectral analysis. Whereas the non-periodic, non-recurrent and localized phenomena, as the shocks, the intermittencies, the turbulences, the transitory behaviors, the abrupt variations or the chaotic regimes are better read by the wavelet analysis, although there is unfortunately no perfect tools.* The spectral analysis is not very effective to treat the abrupt variations because of in particular the error diffusion in the decomposition of this type of phenomenon. *However, this is also in these localized (punctual) and singular behaviors that a dynamics (or a possible nonlinear dynamical system) expresses its own fundamentals.* Information is quite as important there as in the periodic and recurrent aspects. Both methods, spectral analysis and wavelet analysis are thus complementary. Hence, *the hybrid techniques of signal processing in the time–frequency planes prove their utility*. In particular, it is the algorithm of the "best basis" choice and the "Pursuit" algorithm. The "best basis" poses a problem, because it is not adapted to the non-stationary signals composed of disparate elements. On the other hand, *in such a case the "Pursuit" algorithm appears more adapted*. The last algorithm finds at each part of the signal the "wave" which resembles to it the most. It does not seek a decomposition or optimal representation on the entire signal, but seeks it for each characteristic of the signal.

As explained in the general introduction, the conclusions of traditional statistical analysis about the random walk of stock markets is contested today. In particular, we think about the Lo and MacKinlay statistical works already mentioned, but we think also, within the atypical framework for the economists of time–frequency analysis, about the Mallat–Ramsey–Zhang works concerning the decomposition of the stock-exchange series by the "Pursuit" algorithm with time–frequency atom dictionaries.

Furthermore, we know that the wavelet analysis (used among others by the "Matching Pursuit" algorithm) has the ability to *treat* the *transitory behaviors*, *turbulent* or the *non-recurrent* bursts in signals. However, there is certainly an analytical work to carry out *to identify in the stock indexes*, the *behaviors of this nature* which *potentially concern nonlinear mechanics*. Regarding what precedes, the methods of the nonlinear signal processing and the statistics of nonlinear processes presented in the first two parts of this book, have certainly an important role to play.

Chapter 7
Evolution of Economic Growth Models

The economic growth models based on linear structures masked for a long time the instability phenomena. The nonlinearity concept seems relevant to explain many phenomena and in particular the economic fluctuations.

In this chapter, firstly, we endeavor to describe the evolution of growth models: The linear models, then, models that highlight unstable behaviors (which do not necessarily result from nonlinear systems[1]) and finally models whose structure are of nonlinear type.

After an evocation of neo-classical theory of the distribution and after a reminder about the notion of macroeconomic aggregates, we endeavor to describe the evolution of growth models. This description allows to understand the progressive introduction of instability phenomena into the structure of models. Instability appears for example in the optimal growth models, such as the models of Cass–Koopmans–Ramsey, Tobin and Benhabib–Nishimura. A basin of instability characterizes these models where the trajectories of stable equilibrium appear as "singularities", immersed in a set of explosive trajectories. The instability also appears when we introduce the externalities into the endogenous growth models. Indeed, if a lack of externality leads to stationary trajectories, on the other hand, too much externality leads to explosive trajectories. Consequently, the trajectory along the razor's edge (the "wire of the razor") of the Romer endogenous growth model appears also as a singularity.

The nonlinearities are gradually introduced into growth models. In this respect, the Day model in 1982 is a kind of paradigm. In Economics, it is representative of the epistemological revolution appeared during the 1970s, that of chaos theory. The periodic and chaotic behaviors are highlighted in it, through dynamics of capital accumulation which is written in the same way as the logistic equation. The dynamics of accumulation of the Day–Lin model in 1992 concerns a construction more elaborate. It is based on the concepts of *imperfect information* and *adaptive expectations*. In such a model, the dynamics is also of nonlinear nature and

---

[1] *Saddle point:* We think about the *linear systems of order 2*, having in particular *a saddle point* as solution. However, the *saddle point* can result also from *bilinear nonlinear systems* as presented in the part I of this book.

the behaviors highlight not only a subharmonic cascade but also the presence of a strange attractor. In the framework of overlapping generations, the structure of models developed by Benhabib and Day (1982) and in particular by Grandmont (1985) also allows to exhibit cyclic or chaotic dynamics. Complex dynamics and nonlinearities are also present in the construction of the Benhabib–Nishimura optimal growth model (1979). The dynamics of this model allows to admit the existence of a limit-cycle at the equilibrium, immersed in the basin of instability of the saddle-point mentioned previously. This leads to observe the place of chaos theory in cycle theory (and in economic shocks) and in particular in the Real Business Cycle theory (RBC).

One of the characteristics of financial markets is their instability. This instability finds an expression in the model of portfolio choice, such as the Tobin model. The trajectories of the saddle point are essentially explosive, and the stable equilibrium of the axis of the saddle depends on restrictive initial conditions. The trajectories of stock markets often are described as random walks by the statisticians. This observation analyzed by Paul Samuelson in 1965 led him to say that the more a market is efficient, the more the series of prices produced by such a market is random. This remark, which can appear contradictory, immediately raises the question of the predictability of price trajectories of stock markets. Grossmann and Stiglitz however considered that this efficient markets assumption was academic and in so doing such markets considered as an unrealizable economic idealization. Deprive the stock markets of their characteristic of efficiency would be perhaps a way of restoring their predictability. However, some recent statistical works tend to show that the New York Stock Exchange does not follow a random walk. In particular we think about Lo and MacKinlay works in 1988.

Another aim of this chapter is also to present the evolution of economic growth factors and provide arguments in order to explain the growth rates of per capita products. The recent endogenous growth models and the Romer model in particular, develop new arguments to characterize the fundamental "engine" of the growth. These growth factors, conceived as exogenous in the different evolutions of the Solow model, found a new legitimacy while becoming endogenous in the models. These factors can correspond (as in one of the first endogeneisation attempts of the Ramsey model) to an intertemporal optimization mechanism of consumption levels, or, much more recently to positive externalities (as in the Romer model), they have directly or indirectly incorporated the heart of macroeconomic growth models. Lastly, the Von Neumann model (1946) although anterior to the majority of them has a particular place in all of these models. While avoiding the neoclassical production function, it privileges the "technology" although conceived without technical (or technological) progress in its initial version and it points out the importance of the choice of production processes and that of the organization of technology in the perspective to maximize the global growth of the economy.

7.1 Growth and Distribution in the Neoclassical Framework

7.1.1 Aggregates and National Income

The main part of the arguments was developed through the concepts of distribution and allowance of the national wealth between the actors and the factors which take part of this wealth. The GDP (Gross Domestic Product) remains the principal aggregate which is supposed to measure the production of this national wealth; it is expressed, at the national level, to some extent, as the cumulation of the value of consumer goods[2] and gross investments. Instead of GDP, one can approach the domestic product or more precisely the domestic income (or domestic revenue), through the Net Domestic Product (NDP) which is expressed as the cumulation of the value of consumer goods and net investments; the purpose here is to *eliminate the partial depreciation of production tools and equipments still contained in the value added* which by summation provides the GDP. Thus, after *the elimination of the amortization*, we obtain the NDP.[3] Then, these aggregates can be written as follows:

$GDP = \sum values\ added$
$GDP = values\ of\ consumer\ goods\ and\ gross\ investments$
$GDP = amortization + national\ income + indirect\ taxes$
$NDP = value\ of\ consumer\ goods + value\ of\ net\ investments$

$$NDP = C + I. \qquad (7.1)$$

After deduction of the amortization contained in the GDP, we obtain: $NDP = national\ income + indirects\ taxes$. If the State is omitted and if we note Y as the national income:

$$NDP = \ national\ income \ = \ Y. \qquad (7.2)$$

Thus, it seems that we can refer to the NDP to represent the national income[4] Y of macroeconomic models. As explained previously, this national income is expressed as the sum of the consumption C and net investment I:

$$Y = C + I. \qquad (7.3)$$

Moreover, the portion of this national income which is not spent by means of the purchase of consumer goods is saved by the individuals, and if the savings is denoted S we write $Y = C + S$. The neo-classical analysis gives to the savings a prevalent role, then the consumption is regarded as a residue or a remainder. Let us recall *the*

---

[2] Analyzed as the final goods, and sometimes called the consumption goods.

[3] The national accounting distinguishes Net Domestic Product (NDP) and the Net National Product (NNP), due to the contribution of the nonresident agents of the considered country.

[4] Let bet Y the nominal national income, the real national income $\hat{Y}$ is: $\frac{Y}{P} = \frac{\text{Nominal national income}}{\text{General index of prices}}$.

equality between investment and savings developed in the neo-classical theory of the distribution:[5]

$$I = S. \tag{7.4}$$

This equality expresses that there exists an equilibrium point between the savings offer and the investment demand (arbitrated by the interest rate). In the neo-classical models, the construction of curves of global savings offer and global investment demand result from the "aggregation" of individual behaviors and individual choices. We will not endeavor to describe the *Supply* and *Demand* construction in the neo-classical model, however it will be said that *by convention for the neo-classics any savings which would not be intended for the investment is eliminated.* Indeed, the purchase on credit of a good by a private individual, considered as an individual offer of negative savings, is *entered negatively in the global savings offer*; on the other hand, for a company which contracts an investment loan, considered as an individual offer of negative savings, this investment loan is *entered positively in the global investment demand.* Consequently, by this convention, the global savings offer is the aggregation of all the individual savings offers coming from the *consumers.* And if we then consider that the individual offers of savings are increasing according to the interest rate, the global offer is also an increasing function of the interest rate.

7.1.2 Neo-Classical Production Function and Diminishing Returns

The macroeconomic production function is written from the stock of capital and labor, which are generally the only factors of production, because the others appear only temporarily through the intermediate consumption, according to the national accounting. The macroeconomic production function is written:

$$Y = F(K, L). \tag{7.5}$$

This function links the stock of capital and the quantity of labor available with the national income.[6] With this production function, one generally associates the substitutability of production factors, the decreasing returns and the property of "constant returns to scale".[7] Such a macroeconomic production function is built as if an economy were a kind of immense enterprise or firm. One of the questions that preoccupies the neo-classical macroeconomists is to find the conditions of the optimal combination of factors of production of capital and labor?

---

[5] However, in the Keynesian assumptions, the consumption has the prevalence and the savings is a "remainder".

[6] Y and K are expressed in monetary units, whereas L is often expressed in a number of workers.

[7] The macroeconomic production function generally supposes *constant returns to scale*, whereas the production function in microeconomics supposed *decreasing returns to scale*.

7.1.3 Conditions of the Optimal Combination of the Factors K and L

The optimal combination is expressed by referring to the analysis mode of micro-economics. The method consists in the construction of a national "isorevenue" curve. The points of this curve are composed of the capital K and the labor L and the curve is written $Y_0 = F(K, L)$. Considering the "decreasing returns" hypothesis, it is said that there exists on this curve only one point $A_0 = (K_0, L_0)$ which minimizes the costs in capital and labor. This point verifies the equality between the technical substitution rate at the macroeconomic level and the price ratio of the production factors (i for the interest rate and w for the wages):

$$\frac{\partial L}{\partial K} = -\frac{i}{w}. \tag{7.6}$$

If we consider that i and w are given, then an economy which verifies the equality above uses "rationally" (in the neo-classical sense) its labor and capital resources. One can wonder how this "rule", which requires the optimal use of resources, finds its realization in our real economies. Because, if in the microeconomic sphere, the company can insure an optimal combination of factors of production by maximizing its profit or minimizing its expenditure, it is not possible to say there exists, at the level of this immense function of macroeconomic production, an "optimisator" who rationally will use capital and labor resources of the economy; in addition, such an approach, in relation with a form of planning, would not be really admitted in the neo-classical field, for which it is at the microeconomic level of individuals and enterprises that the dynamics is engendered.

Consequently, there is no reason to insure that the equality $\partial L / \partial K = -i/w$ is verified. Thus, the economy will be able to carry out the income of the curve of national isorevenue $Y_0 = F(K, L)$ from any combination of the production factors K and L, and then it will not be possible to say there is the optimality. All the points of the isorevenue curve, except for A_0, are not optimal points, i.e. are not optimal combinations of factors. In the framework of the Walrasian equilibrium at the origin of neo-classical models, which supposes that the production functions are given, described a rigid framework that does not favor the optimization of the model. Indeed, when the production functions are given to the entrepreneurs, i.e. when they are installed in given activity branches, within the framework of the Walrasian equilibrium, they cannot reorientate their resources of these branches towards the other branches where the distribution of the factors K and L would be different. In other words, the transfer of resources by creation and closing of enterprises are not allowed. A "slackness" of the Walrasian rule is brought through the Adam Smith invisible hand which allows the producer to reorientate the resources towards branches of activities which will enable to maximize its profit; thus, it will be possible to manufacture products on these new branches and to cease manufacturing the other products on the old branches.

The game of the competition which is at the center of the neo-classical basic model must allow, freely, the entry and the exit of enterprises, but this is an attack to Walrasian equilibrium. To resolve this contradiction, the neo-classical theory appealed to *the assumption of the tendency to zero profit in the long term*. This assumption, resulting from *microeconomics, will make it possible to obtain the optimal combination of the factors K and L in the macroeconomic level.* We will have to face, in fact, a kind of *succession of Walrasian equilibrium*, under the effect of competition.

7.1.4 Optimal Combination of Factors, and Tendency Towards a Zero Profit in the Long Term

In micro-economics, to determine *the long run equilibrium* of the enterprise, it is to determine *the optimal volume of production* by combining all the factors. In a competitive economy, the price of outputs is fixed by the market and the producer takes it as being a given element. He will attempt to maximize his profit and to determine the *level of optimal production*. The calculation of this optimum is carried out commonly from the given price P of the output and from the curve of *marginal-cost* C_m which allows to provide the optimum level of the output.[8] When the price comes from Walrasian equilibrium, the enterprise will produce in the long-run this optimal quantity and will obtain a positive profit, because the price is higher than the *average cost* in the long run: $P > C_M$. If one introduces new enterprises into the activity branch considered, the Walrasian equilibrium will be modified. The offer of these enterprises will come to increase the total offer and mechanically will cause a drop in the price. As long as the *price* remains higher than the *average cost*, there will be always an interest for the enterprises to enter into the market until the price reaches *the minimum of the average-cost curve*, *point from which the profit is equal to zero. From this point, the enterprises are not attracted any more by this branch of activity. And one can consider that the profit will remain fixed into zero.*

However, this propensity to the zero profit of the entire economy has to be relativized; indeed, the technical innovation can lead the costs and the cost curves to decrease and then the profit becomes again positive. Thus, the tendency towards zero profit will depend on the capacity of innovation of the branch considered. This is this tendency towards the zero profit, resulting from micro-economics, which allows to conceive an optimal combination of the capital and labor factors at the macroeconomic level.

---

[8] Let us recall the shape of the costs curves as follows: $C_m < C_M \Leftrightarrow C_M$ is decreasing; $C_m > C_M \Leftrightarrow C_M$ is increasing.

Product Exhaustion Rule

A zero-profit means that the national income will be entirely shared between the *interests* and the *wages*. When the profit is equal to zero, firstly, it is possible to write at the level of the enterprise:

$$P \times Q = \textit{values of intermediate consumptions} + \textit{amortization} + iK + wL$$

$$P \times Q - \textit{values of intermediate consumptions} = \textit{amortization} + iK + wL,$$

(where Q can be basically understood as the quantity of national product and P its price). The equation above is equivalent to:

$$\textit{Value Added} = \textit{amortization} + iK + wL. \tag{7.7}$$

The value added is denoted *VA*. Moreover, at the *national level,* the $GDP = \sum VA$, thus:

$$GDP = \textit{amortization} + iK + wL, \tag{7.8}$$
$$GDP - \textit{amortization} = iK + wL, \tag{7.9}$$
$$NDP = iK + wL. \tag{7.10}$$

In accordance with what was introduced previously, if one omits the State and the Exterior, the NDP can be regarded as the *national income Y*:

$$Y = iK + wL. \tag{7.11}$$

This equation represents the product exhaustion rule, introduced by Clark and Wicksteed (see Clark–Wicksteed theorem[9]), which is also called the national income exhaustion rule. This means that *the national income is equal to the interests plus the wages paid in 1 year*. From this observation, the optimal combination of the factors K and L will be essential. Indeed, if we use again our macroeconomic production function by imposing this product exhaustion rule, we obtain:

$$F(K,L) = iK + wL. \tag{7.12}$$

---

[9] **Clark–Wicksteed theorem:** The Clark–Wicksteed theorem asserts that output (Q) is exactly equal to income (Y) if each unit of a resource generates a payment to its owner precisely equal to the value of its marginal product (MPP): Such payments are an equilibrium result if all resource and product markets are perfectly competitive, and if production functions are linearly homogeneous. If these conditions are verified, then the Clark–Wicksteed theorem holds and $Q = Y = (MPPL \times L) + (MPPK \times K) + (MPPN \times N) = wL + iK + rN$, where L = labor, w = the wage rate, K = capital, I = interest payment to each unit of capital, and N = land, and r = rental rate per unit of land. The Clark–Wicksteed theorem is named after the American economist John Bates Clark (1847–1938) and the English economist Phillip Wicksteed (1844–1927). See also marginal productivity theory of income distribution.

Since by assumption *the production function had constant returns to scale*, *this function is also a homogeneous function of degree 1 which verifies Euler identity:*

$$F(K,L) = K \times F_K'(K,L) + L \times F_L'(K,L), \tag{7.13}$$

where F_K' and F_L' are the partial derivatives of F. Consequently, after commutation, we obtain the following equality:

$$iK + wL = F_K'(K,L) \times K + F_L'(K,L) \times L. \tag{7.14}$$

However, this equation is true for any value of K and L; thus there are $F_K' = i$ and $F_L' = w$, which is written also:

$$\frac{\partial Y}{\partial K} = i \quad \text{and} \quad \frac{\partial Y}{\partial L} = w. \tag{7.15}$$

The first equality means that the rate of interest is equal to the marginal productivity of the capital factor, and the second equality means that the wages are equal to the marginal productivity of the labor factor.[10] These equalities are verified at the points $A_0 = (K_0, L_0)$. However, near this point, if we simulate a small variation of the capital factor ∂K, we will obtain a variation of the national income ∂Y_1, and if we simulate a small variation of the labor factor ∂L we will obtain a variation of the national income ∂Y_2 and it is possible to write $\frac{\partial Y_1}{\partial K} = i$ and $\frac{\partial Y_2}{\partial L} = w$. From these equalities one obtains $\frac{\partial Y_1}{\partial K} \times \frac{\partial L}{\partial Y_2} = \frac{i}{w}$; however, the variations ∂Y_1 and ∂Y_2 of the national income must necessarily be compensated to preserve the same level of national income, i.e. the same isorevenue curve. We must therefore have $\partial Y_1 + \partial Y_2 = 0$ or $\partial Y_2 = -\partial Y_1$. Consequently, it is possible to write:

$$\frac{\partial Y_1}{\partial K} \times \frac{\partial L}{-\partial Y_1} = \frac{i}{w}, \tag{7.16}$$

$$\frac{\partial L}{\partial K} = -\frac{i}{w}. \tag{7.17}$$

They represent the condition of the optimal combination of the factors K and L. Thus, the tendency towards the zero profit in the long term has as consequence the product (or income) exhaustion rule which makes it possible to verify the definition of the optimal combination of factors of production K and L. In other words, the free entries and exits of enterprises in the branches where the profit is positive involves a zero-profit in the long term. The profit disappearing, the "national income becomes the unique resultant of the only incomes of the factors K and L", which is written $Y = iK + wL$. And since the returns to scale is constant, the "remuneration" of production factors is carried out in proportion to their marginal productivity. In relation to what

---

[10] These nominal equalities have a transcription into the national real income $\hat{Y}$: $\frac{\partial \hat{Y}}{\partial K}P = i$ and $\frac{\partial \hat{Y}}{\partial L}P = w$.

precedes, a distinction is emphasized between macro-economic optimum and micro-economic optimum. Indeed, in micro-economics it is the producer who, adopting a rational behavior, determines the optimal combination of factors, whereas in the macro-economic level this combination is the result of the competition and the result of the transfer of resources between branches.

7.1.5 The Ground Rent and the Ricardo Growth Model

An extension of the basic model consists in introducing a variable which represents *the primary production factor*, as the ground or the natural resources. This factor, noted T, is supposed to remain invariant and *the neo-classical function of macroeconomic production* is written then:

$$Y = F(K, L, T). \tag{7.18}$$

This configuration allows to evoke the Ricardo growth model. One of the effects of the invariance of T is that *the increasing use of the capital and labor factors is decreasingly effective.* If it is admitted that the factors K and L are used in an optimal way such as presented in the previous sections, one can gather these variable factors in a new variable V and the production function is written as follows:

$$Y = F(V, T). \tag{7.19}$$

If we note P_V the price associated with the new variable V and P_V^m the *marginal productivity* of the production function, we write:

$$P_V^m = \frac{\partial Y}{\partial V}. \tag{7.20}$$

This marginal productivity is decreasing from its extremum.

7.1.5.1 Tendency to Zero Profit, Surplus and Ground Rent

Unlike the case of the production function with the two factors K and L, when one introduces a fixed factor T, the enterprises progressively with their entry have productivities which increasingly worsen, and the last enterprise will not have positive profit. On the other hand, all the enterprises which are arrived before and which are more productive, will have a positive profit, i.e. a "ground rent" or a "surplus". On the curve of marginal productivity of the factor V, the critical point "c" from which the profit is not positive any more, is written $\partial Y / \partial V = P_V$ and one can write $\partial Y = P_V \partial V$.
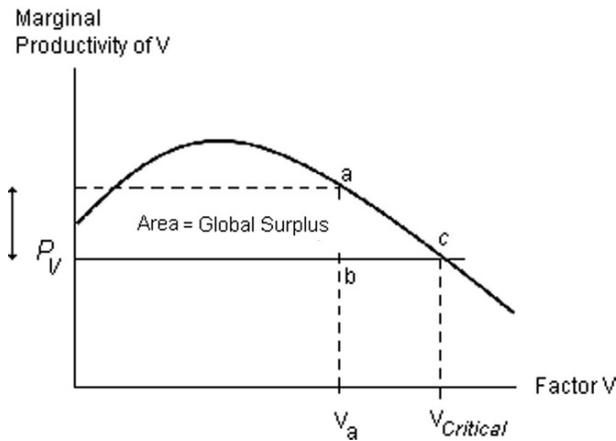
On this curve, at the point "*a*", a small increase in V causes a small increase in income ∂Y. The cost of this increase being equal to P_V, and the surplus which results from this corresponds to the segment [a,b]. If one carries out the sum of all these units of used factors V, one obtains the "global surplus of the economy": it corresponds to the surface ranging between the productivity curve and the line of price, and this surface also represents the ground rent of the economy. It is a neo-classical presentation of the concept of ground rent that can be connected to the Ricardian approach.[11]

Ricardian Rent, Technical Progress and Stationary State

The notion of ground rent results from the fertility disparities between the grounds. The least fertile ground does not produce a surplus, it covers only the costs. On the other hand a more fertile ground, for a same production cost, will make it possible to obtain a higher income and this supplement of income is the ground rent. The difference in fertility between the grounds constitutes for Ricardo and Marx the explanation of the ground rent. In other words, it comes from the differences in productivity resulting from fixed factors and from the decreasing returns which result from it. If we consider again the presentation through the preceding graph, the displacement of the point "c", towards the straight line, on the curve of marginal productivity of the variable factor, shows a growth trajectory of the national income in time. However, this growth Y can be realizable only if the price P_V of the variable

---

[11] *Ricardo:* In 1815, Ricardo published his Essay and introduced the differential theory of rent and the *law of diminishing returns* to land cultivation. (This principle was discovered independently in the same time by Malthus, Torrens and West.) In the 1815 Essay, Ricardo formulated his theory of distribution in a one-commodity ("corn") economy. With wages at their "natural" level, Ricardo argued that rate of profit and rents were determined residually in the agricultural sector. Then, he used the concept of arbitrage to claim that the agricultural profit and wage rates would be equal to the counterparts in industrial sectors. With this theory, Ricardo could show that a rise in wages did not lead to higher prices, but lowered profits.

factor decreases, due to the decreasing form of the marginal productivity. But, the decrease in price cannot occur indefinitely, and this price which includes wages and interest rate will meet a critical point. From a critical interest rate the preference for the *consumption of "households"* (i.e. the final consumption) will be stronger and the savings equal to zero. Zero savings means that the investment is equal to zero and thus the capital K remains constant. K and L thus will remain identical, it is the beginning of a stationary state of the economy, that Ricardo had highlighted. Facing this stationary state, the technical progress provides the solution in order to counter this type of process, the marginal productivity rises with the technological level and the limits of the stationary state are pushed back at the same time.

7.1.6 The Expansion Path and the Limitation of the Nominal National Income Growth

Thus, according to the rule of the optimal combination of production factors (presented previously), for a given level of national income Y_0, there is only one optimal combination of factors denoted:

$$A_0 = (K_0, L_0). \qquad (7.21)$$

From this observation, one can introduce the notion of *expansion path* of an economy, which consists in a "trajectory" which passes by zero the origin of the axes (K, L) and by the points:

$$A_i = (K_i, L_i) \qquad \text{where } i = 0, \ldots, n, \qquad (7.22)$$

which are the optima of each level of national income Y_i. In microeconomics, this notion of expansion path for the enterprise and its production function exists and from this notion, one reaches a final concept which is that of *the optimal size* of the enterprise. This concept of optimal size does not have a transposition for our path of macroeconomic expansion, because in this case the returns to scale are regarded as constant, *thus, here is a potential cause of limitation of the national income which disappears*. In addition, the "law of markets" (attributed to Jean-Baptiste Say [1767–1832], in French "loi des débouchés" literally "law of outlets", in English "law of markets") for which "the offer creates its own demand", excludes a limitation of the latter and the national income. It thus seems that the only limitation of the growth of *the "nominal" national income can result only from a limitation of available resources in L and K production factors*. With this couple *(K,L)* usually one associates the growth of the population and the working population in particular which contributes to the displacement of the national income along the expansion path. *The path constructed with the set of level points of different incomes is also called "a model of growth"* and within this neo-classical framework, the paradigm is *the Solow growth model*.

GDP	Price and exchange rate of 1995					%	Price and current rate			
(OECD)	1995	1997	1998	1999	2000	volume	1997	1998	1999	2000
Billions $						1999/2000				
USA	7,338	7,943	8,293	8,645	9,077	5	8,257	8,729	9,237	9,896
France	1,553	1,600	1,655	1,703	1,756	3.1	1,406	1,452	1,438	1,294
Germany	2,458	2,511	2,543	2,603	2,680	3	2,114	2,151	2,112	1,873

However, within the neo-classical framework, this growth of the nominal national income contains problems related to the downward trend of the growth rate of the per capita national income presented in the model of Solow.

Robert Solow: Solow is one of the major figures of the Neo-Keynesian synthesis macroeconomics. With Paul Samuelson together they constituted the core of the MIT economics department which has been viewed as the mainstream of the post-war period. Their contribution concern in particular the von Neumann growth theory (1953), the capital theory (1956), the linear programming (1958) and the Phillips Curve (1960). Solow is known in particular for his work on the neo-classical growth model (1956, 1970). His use of an aggregate production function launched the Cambridge capital controversy with pitted Solow and Samuelson against Robinson and the Cambridge Keynesians. The different papers of Solow on the issue of capital confirmed the importance of the subject (1963–1967). Solow had debates with Kaldor on the issue of growth and technological progress (1957, 1960). Solow was also one of the co-inventors of the constant elasticity of substitution (CES) production function (1961). Solow is also responsible for exploring and popularizing the long-run multiplier derived from a dynamic government budget constraint (1973). Robert Solow received Nobel prize in 1987.

7.1.7 Stationary Per Capita Income of the Solow Model in the Long Term

The main conclusion of the Solow model is that the accumulation of the "physical" capital *does not explain the long periods of strong growth of the "per capita income"*. The basic model radically ignores the externalities and in particular those generated by the capital, to justify the growth. The growth is explained from an endogenous source, which is *the capital accumulation*, and from an exogenous source, which is *the available quantity of labor*. If one supposes that the exogenous cause is constant in order to facilitate the analysis of the growth, then, one can observe the evolution of the production by means of the only accumulation process. Within such a framework, *if a fall of the growth is observed*, it will be the consequence of *a production function with constant returns to scales* and of *a decreasing marginal productivity of the capital.*[12] In contrast with what precedes, if it is supposed that the returns to scales were increasing, contrary to the neo-classical

---

[12] *Inada condition (1964):* The marginal productivity of the capital tends towards zero. If $f(\cdot)$ is the intensive form of the production function F($\cdot$), then $lim_{k\to 0} f'(k) = \infty$, $lim_{k\to\infty} f'(k) = 0$.

assumption,[13] we could have the risk to observe explosive trajectories. Thus, the Solow model seems to highlight in the long-term an interruption of the per capita income growth. Then how to explain the real growth rates in our contemporary economies?

7.2 Linear Technical Relations Outside the Neo-Classical Theory Framework of the Distribution: Von Neumann Model of Semi-Stationary Growth (1946)

7.2.1 Primacy of the Organization of "Technical Processes"

This model has been proposed by the mathematician John Von Neumann[14] in an article published in 1946 in *Review of Economic Studies* and entitled "A Model of General Equilibrium". The model is specifically oriented towards the production and is written so that all the production is used for the new productions. Moreover, the *surpluses*, if they exist, are eliminated without cost. The model poses technical relations of "linear" nature. The purpose of the model is to determine the manner of using the available techniques *in order to reach the possible highest growth rate*, i.e. *the optimal rate*. The *"Primal" operates in the space of goods*, whereas the *"Dual" operates in the space of values*. This *model associates with its optimal trajectory in terms of quantity, a dual trajectory of price*, which is interpreted as the dual of a linear program.[15] *The "Primal" consists in seeking an optimal use of the vector of the technical intensities $y(G)$* which carries out a homothetic growth of the maximum rate G: thus, the couple which is solution of the Primal problem is $(y(G), G)$. The *Dual* of the model consists in seeking a row vector of price $p(R)$ and a minimal profit rate R: thus, the couple which is solution of the dual problem is written $(p(R), R)$.

7.2.2 Presentation of the Von Neumann Model

The economy in this model consist of n goods which can be inputs or outputs, or both. The *technical possibilities* of the economy are represented by the couple of matrices (A,B), whose columns are respectively formed by the inputs a and the outputs b, and each one of these matrices consists of m columns and n goods.

---

[13] See the Product exhaustion theorem.

[14] *John von Neumann* (mathematician) has played a rather important role in post-war economic theory through two essential pieces of work: his 1937 paper on a multi-sectoral growth model and his 1944 book (with Oskar Morgenstern) on game theory and uncertainty.

[15] *Linear programming*. A linear programming problem, or LP, is a problem of optimizing a given linear objective function over some polyhedron. (The simplex method is the algorithm most commonly used to solve LP. In practice it runs in polynomial time, but the worst-case running time is exponential. Two polynomial-time algorithms for solving LP are the ellipsoid method of Khachian and the interior-point method of Karmarkar.)

The m techniques are symbolized by two column matrices with n elements, the first represents the inputs, it is $y_t a^j$, and the second represents the outputs $y_t b^j$ obtained from the inputs. For the j-th technique, one denotes by a^j *the column matrix of inputs* and b^j *that of outputs*. One thus knows that an unspecified technique consumes a certain number of inputs restricted in relation to the set of goods in the economy and will produce only some outputs or only one, these matrices are composed of 0 and the other components are "positive". y is the vector of the production technique with m elements at the period t, whose each component is positive or zero. y represents the "intensity" with which a technique is used. If y exists, i.e. if there is an intensity vector of use of available techniques, then the system of involved techniques is productive. What is produced by the By model is normally equal to, or higher than, what is consumed Ay, that is written:

$$Ay \leq By : Minimal\ condition. \tag{7.23}$$

It will be noted that the essential element is the selection by A and B of components of y (carried out by the multiplication of vectors), the resultant corresponding to the "structure" of the considered economy. This condition $Ay \leq By$ is the "minimal" condition so that there is *reproduction* of the system. If there is equality for some goods, the consumed quantity of these goods is equal to the produced quantity, *consequently the system will be reproduced exactly,* i.e. it will be duplicated and even if there is over-production (excess, surplus) of the other goods. The surpluses in the model are supposed to be eliminated without cost. Thus, *there will be a production growth if the inequality $Ay \leq By$ is strict for each good, indeed, it is because there are surpluses for all the goods that it is possible to "nourish" the growth.*

7.2.2.1 Conditions and Hypotheses of the Model

From this structure oriented towards the production, the model defines its operating framework by postulating what follows:

- The *returns* are supposed to be *constant*.
- The model admits the *joint productions*.
- The model admits the *technique changes* to produce the same good.
- The number of activities is at least equal to the number of goods, i.e. $m \geq n$
- The model *does not incorporate technical progress*.
- The model depicts a closed economy.
- The model contains only *intermediate goods*. This means that all the produced goods are reintroduced in the model as factors of production.
- The *surplus*[16] *is excluded* of the system.

---

[16] *Surplus:* The notion of surplus generally indicates two very different concepts:
- What remains in the production when we have removed all which is necessary for its implementation or execution.
- The set of the gains, in term of utility, gotten by the exchanges.

- The model excludes the *superprofit*[17] (i.e. each activity must carry out a negative or zero superprofit).

  Von Neumann makes *two central hypotheses* for the system *perpetuation*:

- Hypothesis H1: In order to perpetuate the system, *each good B_j must be produced by at least one activity*.
- Hypothesis H2: *Each activity A_i must consume at least one good*.

7.2.2.2 Semi-Stationary and Homothetic Optimal Growth

The problems, corresponding to the *primal* of the model, are *to determine the vector of production intensities which allows the largest growth possible*. If one names "g" the rate of growth, one thus seeks a vector $y(G)$ and a rate G which maximizes g under the constraint:

$$(1+g)Ay \leq By : \textit{Condition of "expanded reproduction".} \qquad (7.24)$$

We understood that this constraint allows the expanded reproduction at the rate G of the model, the quantity of the outputs being at least as important as that of the inputs used to produce. The model highlightsthat there exists a solution couple $(G, y(G))$ which satisfies the condition of the expanded reproduction and which maximizes the growth.[18] For this optimum the inequality is written:

$$(1+G)Ay(G) \leq By(G) : \textit{Optimal growth.} \qquad (7.25)$$

We observe that the inequality above is not strict for at least one of the goods, if it was the case, this would have as consequence a surplus (excess) for each good of the model. And it would be then possible by exploiting these surpluses to have a stronger growth, contrary to the optimal nature of G and $y(G)$. If Φ represents the set of indexes of goods for which there is equality in $(1+G)Ay(G) \leq By(G)$ and Ψ that of the indexes of goods for which there is inequality, one can write:

$$(1+G)\sum a_{ij}y_j(G) = \sum b_{ij}y_j(G), \quad \textit{for } i \in \Phi, \qquad (7.26)$$
$$(1+G)\sum a_{ij}y_j(G) < \sum b_{ij}y_j(G), \quad \textit{for } i \in \Psi. \qquad (7.27)$$

In the set Φ the quantities of goods increase at the optimal rate G, whereas in the set Ψ the quantities increase more quickly. In the set Ψ the goods produced beyond the rate G, i.e. the surplus is eliminated at each period, and will not be consumed by

---

[17] *Superprofit:* Extremely discussed notion for which there is many definitions. One of them consists in saying that the superprofit is the equivalent of the *economic rent*, which is what the organizations gain beyond the cost of the invested capital. The terminology is not very satisfactory. Indeed, the reference framework was worked out by David Ricardo at the beginning of the nineteenth century, at the moment when agriculture dominated the economic activity. *The economic rent was then renamed "profit", "superprofit" or "excess benefit" (or "surplus benefit").*

[18] By using *the fixed-point theorem* and by making assumptions on the A and B matrices.

the production process. Indeed, it should be associated with other inputs belonging to the set Φ, which then would be "lacking" to be able to produce a new output.

Such a system develops at the optimal rate G, and corresponds to a *semi-stationary state*, that means that the *structure* of the model is preserved exactly during its growth. It is said that a semi-stationary state is a situation where all the variables in quantity and level grow at the same rhythm. An economy in a semi-stationary state has an *evolution of the homothetic type* where the proportions between the different variables remain constant in the course of time. This type of semi-stationary state is also called "permanent regime", because the ratio of the quantity or price variables is permanent during the growth of the economy. Note that the stationary states are particular cases of semi-stationary states, for which the rate of growth is equal to zero. The semi-stationary states are known as equilibria of long term, because they are supposed to symbolize the underlying deep evolution of the economy, except for exogenous impacts on the system.

7.2.2.3 Price and Profit Associated with the Optimal Growth

The dual program of the model *associates with the optimal rate of growth, a system of price and also a weakest possible rate of profit or interest such that (provided with this couple: price and rate of profit) the profit unitarily in the production of each good is negative or zero.* One seeks therefore to determine a row vector of price $p(R)$ and a minimal profit rate R which verify the condition:

$$PB \leq (1+r)PA. \tag{7.28}$$

Or, unitarily for each activity j:

$$Pb^j \leq (1+r)Pa^j. \tag{7.29}$$

Thus, for the price P the value of the goods produced with the j-th technique is lower or equal to the value of the inputs used to produce them. One deduces from the duality theorem that the solutions of the primal $(G, y(G))$ and dual $(R, p(R))$ are in conformity with the following rules:

- The rate of maximal growth G is equal to the minimal profit rate R. Under H1 and H2 : $(A + B) > 0 \Rightarrow R = G$. (The conditions are those of the technological or economic "indecomposability".)
- In the optimal growth, the price of goods, whose production is surplus, is zero (i.e. the goods belonging to the set Ψ).

Thus, it is possible to write the two following relations which are *exclusion* relations, in accordance with linear programming:

$$(1+G)\sum a_{ij}y_j(G) < \sum b_{ij}y_j(G) \Rightarrow p_i(R) = 0 \tag{7.30}$$

the goods generating negative profit are not produced:

$$(1+R)\sum a_{ij}p_i(R) > \sum b_{ij}p_i(R) \Rightarrow y_j(G) = 0. \tag{7.31}$$

The Von Neumann model can be understood in the perspective of a planner, considering that this planner has the objective to find how to use the available techniques in order to maximize the growth of the economy. The planner can, either:

- "To organize" the production system to obtain the strongest growth, i.e. to impose the use intensities of available techniques, or
- On the contrary, "to display" the couple of price $(R, p(R))$ and to let the "production units" determine $(G, y(G))$ the corresponding quantity couple. The different units must thus find the suitable use "intensities" of techniques. There will be only production of goods for which the profit is not negative and the goods for which the profit is zero will be offered only. Generally, the Von Neumann model implies that it is the "market" which "finds" *the most suitable price couple* $(R, p(R))$, i.e. the rate of profit and vector of price.

The Von Neumann model is *a general equilibrium model* in a closed economy and oriented towards the production, whence *the demand is excluded* from the system. Only the postulate of the model, which poses the existence of constant returns to scale, makes it possible "to admit a sort of neutrality of the demand". The Von Neumann model and its regime of homothetic growth i*s not very representative of a real process of accumulation* of a real economy. The accumulation processes of "surpluses" as well as the nature of their stock (of surplus) for the concerned agents, in a real economy have an impact on the equilibria of the model, on its solutions, and on the phenomenon of reproduction of the system. Not only the distribution of surpluses between the actors but also their rhythm of accumulation and *their use* are sources of conflict, and modifies the construction of equilibria of the model.

7.2.3 The Optimal Path and the Golden Rule in the Von Neumann Model

7.2.3.1 Optimal Distribution, Remuneration of Factors and Accumulation

As explained previously the Von Neumann model could be understood in the perspective of a planner, who would aim to find how to use the available techniques in order to maximize the growth of the economy. As previously presented, the planner can either organize the production system to optimize the growth or on the contrary, "display" the couple of price and to let the production units determine the quantity couple. Generally, one considers that it is the market which finds the price couple, i.e. the profit rate and the vector of price. In the Von Neumann model and in the

stationary state of the Solow model, there is *homothetic growth* phenomenon,[19] at the same rhythm, of all the variables in level (production, consumption, investment, etc.). Thus, the ratio of two unspecified variables of the model is constant in the course of time, it is for example the case of the production per capita, or consumption per capita. In the initial model of Solow, the economy tends towards a stationary state of the production and consumption per capita, thus there is not growth, except the growth of the population, this means that it is not a model of growth strictly speaking and is not therefore able to explain the growth phenomena.

First, it is Von Neumann who approached the problem of the growth within the framework of a linear model, *with constant technical coefficients*, where all the *surplus* is reinvested at each period. Von Neumann associate with the trajectories in quantity (produced, consumed or reinvested) the trajectories of prices which are interpreted as the "dual" of a linear program of quantities. The logic of the model is that of a planner who seeks the best resource allocation, from the point of view of the optimal growth, i.e. the strongest possible growth, the considered evolutions being of the semi-stationary type. In this model, the trajectory of prices is deducted by duality from the quantities and is thus not at the origin of the coordination of individual choices, in fact, the model does not comprise individual agent. The structure of the model, oriented towards the production, postulates in particular that:

- The model is linear.
- The returns are constant.
- The model does not incorporate technical progress.
- The model contains only intermediate goods, and all the produced goods are reinvested in the model.

The model concludes that when the two basic hypotheses of the model are satisfied, *at the optimum* there is *equality* of *the growth rate (maximum)* and of *the rate of profit (minimum)*:

$$G = R. \tag{7.32}$$

However, in Growth Theory the "golden rule" corresponds to conditions on the parameters of the model characterizing an economy, so that the consumption per capita is the highest possible in semi-stationary states.[20] This means that it is

---

[19] *Homothetic:* Two figures are homothetic if they are related by an expansion or geometric contraction. This means that they lie in the same plane and corresponding sides are parallel; such figures have connectors of corresponding points which are concurrent at a point known as the homothetic center. The homothetic center divides each connector in the same ratio k, known as the similitude ratio. For figures which are similar but do not have parallel sides, a similitude center exists. Also known as radially related figures. *Homothetic transformation:* A transformation that leaves the origin of coordinates fixed and multiplies the distance between any two points by the same fixed constant. Also known as *transformation of similitude*. In connection with the homothetic transformation, although different, see also the notion of *Similarity Transformation* which transforms objects in space to similar objects. Similarity transformations and the concept of self-similarity are important foundations of fractals and iterated function systems.

[20] *Semi-stationary state:* The semi-stationary states are situations during which the variables in quantity and level grow at the same rhythm. I.e. semi-stationary economies have an evolution of a homothetic type, where the proportions between the different variables remain constant in time.

necessary to determine *the optimal sharing between consumption and savings*. The golden rule was implemented on the Solow model, Von Neumann model and optimal growth models. This step, which is of a normative nature, concludes that *the marginal productivity of the capital* (noted p_k^m) is equal to *the growth rate of the economy*, which is itself equal to *the real interest rate:*

$$p_k^m = \gamma_{growth} = \gamma_{real\ interest\ rate}. \tag{7.33}$$

Such a rule applied to the Von Neumann model leads to say that it would be sufficient to know the "marginal productivity of the capital" to know the growth rate and the rate of profit of the model (and the interest rate). But, in the Von Neumann model the marginal productivity of the capital does not appear spontaneously: there is not explicit production function but rather technical matrices expressing at a given moment the "technology" (A,B) of the model, i.e. the production techniques which use all or part of the production processes of the model. The marginal productivity of the capital, which is expressed as the derivative of the production function of a model with respect to the capital (i.e. variation of the output of the model when one causes a variation of the capital factor) can find a transcription in the Von Neumann model. Generally, *the concept of "marginal productivity" is highlighted when a company wishes to determine the level of production which enables it to optimize its profit*. The optimum is conditioned by the nature of the marginal productivities which can be as well constant, as in the case of the Von Neumann model, than decreasing, and this approach is valid for any factor of production, i.e. the ground, the labor or the capital. Within the *framework of the perfect competition*, it is said that the marginal productivity of the production factors (or inputs), evaluated in terms of "quantity of product", is equal to the price of factors. Based on such a framework, one can deduce that the remuneration of production factors (or inputs) is equal to their marginal productivity: *Remuneration of production factors = Marginal productivity of factors*.

Moreover, it is known that the equilibrium of perfect competition is a *Pareto optimum*, and it is possible to deduce that *the remuneration according to the marginal productivity is "optimal" for an economy. The remuneration of production factors according to their marginal productivities is "optimal"*. This deduction is valid as much for the remuneration of the owners of the production tool, than for the remuneration of the ground for example, or for the remuneration of labor, and we are inscribed here within the neo-classical framework of *Distribution Theory*. This theory, which is of *a normative nature*, leads implicitly to admit that the optimal distribution is defined by the conditions of a technical nature, i.e. obviously the structure of the production and various marginal productivities. *It is possible to say that it is the technical structure which decides the optimal distribution*.

In the Von Neumann model the transition from the matrices (A,B), representing the model, to the marginal productivity of the capital is not spontaneous, but we admit that it is possible to obtain the marginal productivity from these matrices. Consequently, they are the technical conditions, i.e. the *technology* of the model, through matrices (A,B), which makes it possible to determine the marginal productivity determining then the growth rate:

$$f'(k) = G. \tag{7.34}$$

The main subject is the way in which the capital factor influences the growth of the model. Moreover, the remuneration of the capital factor corresponds also to the rate of profit of the model and, if one considers again the equalities stated by the golden rule, then it is possible to write:

$$f'(k) = R = G = I_r. \tag{7.35}$$

We are in a model in which the growth is a process of *homothetic type*, i.e. the proportions between the various variables remain constant in the course of time:[21] this growth is also called a *semi-stationary process*, in which the structure of the economy is preserved exactly in the long-term.[22] The optimal path of the model is sometimes called the Von Neumann ray. The objective of the golden rule was thus to determine the highest consumption per capita in the long term, which requires implicitly *to determine the best saving rate or the interest rate, i.e. to determine the optimal distribution.* In the case of *the Von Neumann model, the golden rule corresponds to the optimal growth.*

7.2.4 Comments About Von Neumann and Solow Models

The Von Neumann model, of which the initial structure excludes the technical progress, is based on linear relations and concludes that the economy converges towards a semi-stationary state. The Solow model also concludes that there is convergence towards a semi-stationary state, but it also concludes that *the per capita product* is stationary in the long-term,[23] with furthermore, nominal variables which all grow at the same rhythm, which is generally the rhythm of the growth rate of the population. The Von Neumann model made it possible to avoid the production function and the neo-classical theory of the distribution. Such a conception proposes *the idea of the organization of production techniques* (although conceived without technical progress). By disregarding population, the model cannot conclude (like the Solow model) that the nominal variables grow according to the (exogenous) rate of population growth. Von Neumann centered its model on the production processes and their remuneration, avoiding thus to incorporate the labor factor,[24] as in a number of models posterior to that of Solow, which study the behavior of the (per capita)

---

[21] And it is also said that each quantity couple (G, y(G)) solution of the model is itself a system of Homothetic Optimal Growth.

[22] The optimal path of growth (also called the Von Neumann ray) is composed by the couple of inputs vectors x_t and outputs vectors X_t, i.e. the couple (x_t, X_t) which verifies $\forall t$, $X_t = (1 + G)x_t$.

[23] See following section.

[24] Furthermore, Neumann model revised through the Sraffa contribution, submits another approach "including the labor factor" and its remuneration.

product by using the only criteria of the capital, technology and its remuneration, "neutralizing" thus the effect of the population.

The Von Neumann model postulates *constant returns to scale, it does not incorporate technical progress*, but it admits, however, *the possibility of technique changes to produce a good*. We know that *there is an optimal production technique* which produces a *homothetic growth at a maximum rate G*.[25] The model thus seems to show an "engine of growth" linked *to the organization of the production techniques and to the used technical intensity, in the whole of the economy*.[26] However, t*he Solow model postulates the full employment of all the resources of the economy*: in the sense that in such a case the technical intensity is maximal and we cannot find a combination of processes which would increase the level of production. Thus, in the Solow model, it seems that we are immediately located at the end of the Von Neumann growth maximization process. But, that means that we forget the fact that the Von Neumann model is made of many production processes, which are, on the other hand, completely aggregated in the Solow production function. This aspect highlights the importance, not that of the technical progress since the Von Neumann model excluded it in its postulate, but that of the choice of the "technology", as if the "flexibility" of the entire economy could increase its level of production, i.e. *the "return" of all the economy*, and to maximize its growth. The model say nothing about the existence of firms and focus on the only production processes, but, unless one assimilates the economy to one large firm, or unless one assimilates the enterprise to a process, to admit such a flexibility of an economy, implies to admit that the organization of the production techniques is carries out, at the same time, inside and between the firms. Implicitly, we admit that there is a planner who aims to find the available techniques in order to maximize the growth. But, it can either directly "organize" the production system to obtain the strongest possible growth, or "to display" the vector of price and to let the production units determine the couple of quantities. Generally, the model implies that the *market* finds the suitable price couple. Moreover, one can wonder what covers this idea of the "organization of the production techniques" as regards the real economy and the more recent growth models. Because of the ability of the Von Neumann model to postulate the need for organizing technology to maximize the growth, within the framework of a matrix model, the model implies that there are "technical possibilities" which increase the effectiveness of the economy (without technical progress). Beyond the model, how to describe these global technical possibilities while the technical coefficients are fixed? The underlying idea is that "the organization or technical possibilities are *augmentative factors* of the *effectiveness of the technology* of the economy and always in the absence of technological progress". Solow model or neo-classical theory imply the optimality, but, are the modern economies as optimal as the theory suggest? Or, is it necessary to return upstream, in the optimization process and to reconsider, as in the Von Neumann model, that this is the organization of the economy which optimizes globally the use of processes and which exploits 100% of

---

[25] Optimal production technique: $y(G)$.

[26] Associated of course with the vector of price.

the production tool, in order to maximize the growth? The *neo-classical aggregate production function* is built as if the economy were a kind of immense company which is always in full employment of its resources. In the Von Neumann model, *the organization*[27] *of the economy* is *a source of growth*, which will meet at least in the theory, early or late, the optimum of full employment of the neo-classical aggregate production function, at the end of the process of maximization of the growth. Is Von Neumann model another manner to formalize the "neo-classical distribution theory"? Or, is there something other? The model says little thing about the production processes and about the global organization of processes. This *indetermination* can be a source of various interpretations. In the real economy, processes and organization of processes can cover various meanings and are necessarily coupled with the intervention of human resources which are subjacent (even if it is not always formalized in the models). Must we consider that the concept of process organization in the Von Neumann model is only the manner of deeply exploiting the production tool? Or, can we understand it as an augmentative factor of "the technology" (always without technical progress)? And thus to introduce an "additional factor", which is generator of growth in relation for example with the impact of the knowledge, know-how, or learning by doing, on the *use* and *combination* of production *processes*. And to also introduce, instead of the "indetermination" of certain components of the model, the game of interactions of production units with each other and of "hidden agents"[28] with each other. The production units are supposed to have a role: they can determine, for example, the price couple of the model, after the planner has fixed the level of production (even if it is said, generally, that it is the market which determines the price couple). In short, the production units are actors of the model and the problems of coordination are at the center of the model. To make coexist individually the production units in a model corresponds to potentially accepting what occurs in the real economy, i.e. *their interactions* and those of agents. This also means not excluding the role and the consequences of these interactions on the global economy. The concept of externality, for example, developed in the *models of endogenous growth*, can be considered as the consequence of the game of these interactions. The *externalities* indicate all situations where the *activities* of one (or several) economical agent(s) have consequences on the well-being of other agents, without having exchanges or transactions. To admit the existence of externalities, negative or positive, corresponds to accepting the necessity of "revisiting" from a different viewpoint, the *aggregate production function*. The Von Neumann model evokes, obviously, neither the concept of externality, nor the concepts of knowledge, know-how or learning by doing, but its *non-aggregated matrix structure*, where the production units coexist and where the organization of production processes is at the center of the optimization diagram, offers a favorable context, to highlight and to admit their possibility, as augmentative factors of the global effectiveness of the economy.

---

[27] *Of the production techniques.*
[28] Since they are not defined explicitly in the Von Neumann model.

7.3 Stability, Stationarity and Diminishing Returns of the Capital: The Solow Model (1956)

7.3.1 Diminishing Returns and Stationarity of the Per Capita Product

The model has been introduced by Robert Solow in 1956 in an article entitled "A contribution to the theory of economic growth", (Q J Econ 70: 65–94). It is regarded as the reference model of neo-classical "growth", although certain elements of its initial construction are borrowed from the Keynesian theory, like the consumption function. Robert Solow postulates an economy where there is the full employment of all the resources and first the human resources. The Solow purpose is to show that an economy based on a production function of the neo-classical type, i.e. where there is constantly the full employment of resources, converges towards a semi-stationary state, where all the variables grow at a same constant rate. In fact, the basic model concludes that in the long run, there is stationarity (i.e. particular case of semi-stationary states where there is no growth) of the product *per capita*, whereas the nominal variables increase all at the same rhythm, which is generally that of the exogenous growth rate of population. Let us describe the reference Solow model.

7.3.2 The Reference Model

The assumptions of the model are the following:

- The countries produce and consume only one homogeneous good Y.
- The production is carried out within the framework of *perfect competition*.
- The *technology is exogenous*.
- The technology can be represented by a production function of the neo-classical type with substitutable factors which are the capital and the labor: (K,L).
- The aggregate production is represented by a Keynesian function:

$$C = cY \Rightarrow S = (1 - c)Y = sY. \tag{7.36}$$

- The rate of participation in the employment of the population is constant. If the growth rate of the population is n, the offer of labor L grows at same rate n:

$$\frac{dLog(L)}{dt} = \frac{dL/dt}{L} = \frac{\dot{L}}{L} = n. \tag{7.37}$$

In order to present the model, we will use a production function of the Cobb Douglas type:

$$Y = F(K,L) = K^{\alpha} L^{(1-\alpha)}. \tag{7.38}$$

Let us recall that the model postulates constant[29] returns to scale. Within this framework of perfect competition, the companies maximize their profit in relation to the market prices:

$$max_{K,L} F(K,L) - rK - wL. \tag{7.39}$$

With the rate of interest r and the wages w. Applied to our function, one obtains:

$$w = \frac{\partial F}{\partial L} = (1-\alpha)\frac{Y}{L}, \tag{7.40}$$

$$r = \frac{\partial F}{\partial K} = \alpha\frac{Y}{K}. \tag{7.41}$$

Since, by assumption, the production function has constant returns to scale, this function is also *a homogeneous function of degree 1* which verifies the Euler identity,[30] then it possible to write:

$$wL + rL = Y. \tag{7.42}$$

It is possible to write the *per capita* model, by knowing that $\frac{L}{L} = 1$, $k = \frac{K}{L}$, then, the model is written (Fig. 7.1):

$$y = \frac{Y}{L} = f(k) = \frac{F(K,L)}{L} = \frac{K^\alpha L^{(1-\alpha)}}{L} = \left(\frac{K}{L}\right)^\alpha = k^\alpha, \tag{7.43}$$

$$y = f(k) = k^\alpha. \tag{7.44}$$

Such a production function with $\alpha \in [0,1]$ presents a form which highlights the decreasing returns of the *per capita* capita. With such *per capita* production function which highlights the way in which the factor K influences the production, one will highlight the dynamics of the capital accumulation:

$$\dot{K} = \frac{dK}{dt} = I - \delta K \tag{7.45}$$
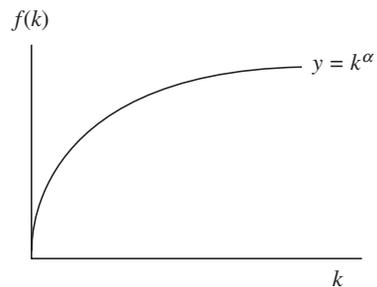
Fig. 7.1 Cobb Douglas function of production per capita

---

[29] That is, for the Cobb Douglas function: $\alpha + (1-\alpha) = 1$.
[30] Euler's identity: $F(K,L) = K \times F'_K(K,L) + L \times F'_L(K,L)$.

The way in which the capital varies is function of the difference between investment I and depreciation of the capital δK. By definition, in a closed economy, the investment is equal to the savings (s: saving rate):

$$I = S = sY. \tag{7.46}$$

Thus

$$\dot{K} = sY - \delta K. \tag{7.47}$$

Furthermore, it is possible to write:

$$k = \frac{K}{L}, \tag{7.48}$$

$$\log(k) = \log(K) - \log(L), \tag{7.49}$$

$$\frac{d\log(k)}{dt} = \frac{\dot{k}}{k} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L} = \frac{sY - \delta K}{K} - \frac{\dot{L}}{L}. \tag{7.50}$$

However, we known that $\dfrac{d\log(L)}{dt} = \dfrac{dL/dt}{L} = \dfrac{\dot{L}}{L} = n$, i.e. we know the growth rate of the labor factor, and we write:

$$\frac{d\log(L)}{dt} = n, \tag{7.51}$$

$$\log(L) = \int ndt = nt + C_0, \tag{7.52}$$

$$L(t) = e^{nt+C_0} \times L(0) = e^{C_0} = L_0, \tag{7.53}$$

$$L(t) = L_0 e^{nt}. \tag{7.54}$$

Consequently, the equation $\dfrac{\dot{k}}{k} = \dfrac{sY - \delta K}{K} - \dfrac{\dot{L}}{L}$ becomes:

$$\frac{\dot{k}}{k} = \frac{sY}{K} - \delta - n = \frac{sy}{k} - \delta - n. \tag{7.55}$$

Thus, we obtain the fundamental dynamical equation of the capital:

$$\dot{k} = sf(k) - (\delta + n)k. \tag{7.56}$$

The coefficient fitting the labor factor in this equation corresponds to the sum of the depreciation rate of the capital with the growth rate of the labor factor.

7.3.2.1 Balanced Growth Path and Solow Diagram

In the preceding section one highlighted the two fundamental following equations, the *per capita* production function and the fundamental dynamics of the capital accumulation:
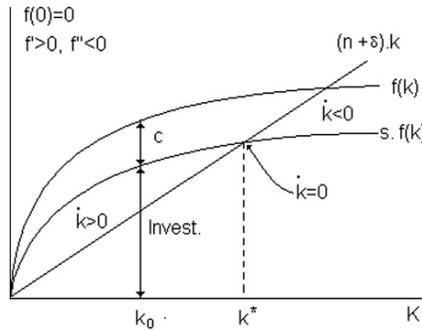
Fig. 7.2 Solow diagram and BGP in k^*

$$y = f(k) = k^\alpha, \tag{7.57}$$
$$\dot{k} = sf(k) - (\delta + n)k. \tag{7.58}$$

If one notes *the initial conditions* of the model $k_0 = K_0/L_0$, the production function provides the level of production, the investment, the savings and the fundamental dynamical equation of the capital provides the evolution of the capital. These evolutions can be represented through the following dynamic diagram:

$$
\begin{array}{ccccc}
k_0 & \xrightarrow{f(k)} & y & \xrightarrow{s,\delta} & \dot{K} \\
n\downarrow & & & & \downarrow \\
L_1 & \longrightarrow & k_1 & \longleftarrow & K_1
\end{array}
$$
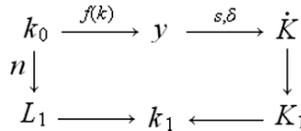
The Solow diagram is represented in Fig. 7.2.

In this diagram, the rate of variation of k is provided by the difference between the curves: $f(k)$ and $(n + \delta)k$. the intersection point of these two curves is $\dot{k} = 0$, at such a point $\dot{k}/k = 0$ and this critical value of k is noted k^*,[31] it designs the *stationary-state* of the model (also called *steady-state*) for which the *per capita* capital remains unchanged. (Inada conditions,[32] with the conditions on the first derivative of f insure the possibility of a k^* such that $k^* = 0 \Leftrightarrow sf(k^*) = (\delta + n)k^*$.) Outside this stationary state, one can have, either:

$$k_0 < k^* \Leftrightarrow \dot{k} > 0, \tag{7.59}$$

where the *per capita* capital increases, i.e. there is an intensification of the capital, or:

$$k_0 > k^* \Leftrightarrow \dot{k} < 0, \tag{7.60}$$

where, on the other hand the *per capita* capital decreases.

---

[31] In $k^* : y^* = f(k^*), c^* = (1-s)f(k^*)$.

[32] *Inada condition*: the marginal productivity of the capital tends towards zero. If $f(\cdot)$ is the intensive form of the production function $F(\cdot)$, then: $\lim_{k \to 0} f'(k) = \infty, \lim_{k \to \infty} f'(k) = 0$.
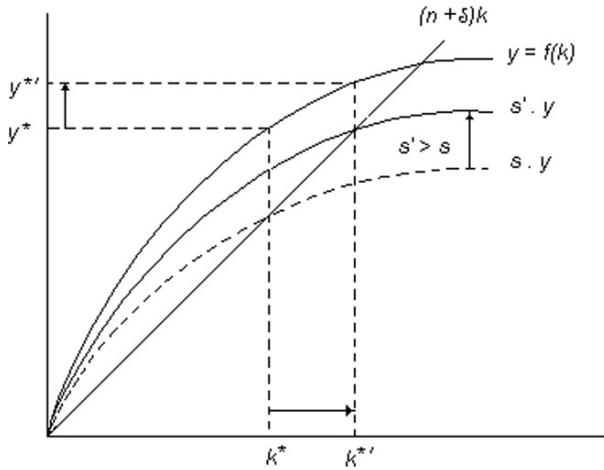
Fig. 7.3 Rise in the rate of investment

7.3.2.2 Comparative Statics and Stationary-State Properties

Rise in the Rate of Investment

Let us imagine that we are located in a stationary state and that then the individuals decide to increase their saving rate which passes from s to s' (knowing that the rate remains positive). Consecutively, it will result in an increase in the investment and the goal is to observe the impact on k and y, by means of the following graph (Fig. 7.3).

Rise in the Demographic Growth

Let us observe what occurs when the growth rate of the population increases from n towards n'. The *per capita* capital retrogrades since its denominator increases. They are represented with the stationary state in Fig. 7.4.

7.3.2.3 The Wealth Level of an Economy

The stationary state is represented by

$$\dot{k} = sk^\alpha - (n + \delta)k = 0. \tag{7.61}$$

It gives k^*:

$$k^* = \left(\frac{s}{n + \delta}\right)^{1/(1-\alpha)}. \tag{7.62}$$
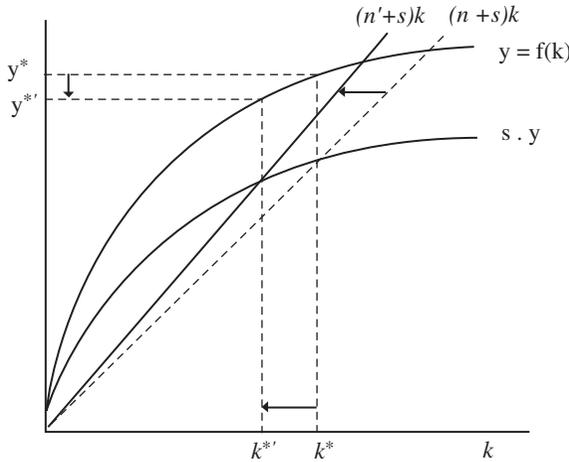
Fig. 7.4 Rise in the growth rate of the population

And it comes for the production *per capita:*

$$y^* = f(k^*) = \left(\frac{s}{n+\delta}\right)^{1/(1-\alpha)}. \tag{7.63}$$

This leads to conclude that:

- The countries which have a high rate of savings-investment, compared with other countries, will tend to being richer.
- The countries which have a high demographic growth rate, will tend comparatively to being poorer.

It is an explanation attempt of the development differences existing between rich and poor countries by the model.

7.3.2.4 Convergence and Dynamics of Transition in the Long-Term

In this model, the absolute variables (Y,C,K) increase at the same rhythm as population (i.e. n) and the *per capita* variables are constant in *a stationary state*.

$$\frac{\dot{k}}{k} = \frac{\dot{y}}{y} = 0 \Rightarrow \frac{\dot{Y}}{Y} = \frac{\dot{K}}{K} = \frac{\dot{L}}{L} = n. \tag{7.64}$$

In such case, in a long-term stationary state, one observes:

- A constant rate $\frac{K}{Y}$ (capital/product), because k and y are constant.
- A constant marginal productivity of the capital k, because k is constant.
- *A variation of the GDP/person between different countries* (corresponding to statistical observations).

*In such a model, the economies can grow short-term but not in the long run:
indeed, if a country deviates at a given time from the stationary state, it will have a
period of transition and then will reach a new stationary state. The growth will slow
down as the economy approaches the stationary state.* Indeed, let us take again the
fundamental (generic) dynamical equation of the capital:

$$\dot{k} = sf(k) - (\delta + n)k. \tag{7.65}$$

By introducing the *per capita* production function of the model, we obtain:

$$\dot{k} = sk^\alpha - (\delta + n)k. \tag{7.66}$$

Let us pose the growth rate of the capital γ_k:

$$\gamma_k = \frac{\dot{k}}{k} = s\frac{f(k)}{k} - (\delta + n), \tag{7.67}$$

$$\gamma_k = \frac{\dot{k}}{k} = sk^{\alpha-1} - (\delta + n). \tag{7.68}$$

From this equation, which depends in particular on α which is lower than 1, one
understands that the increase in the *per capita* capital factor induces a trend fall
of the growth rate of the capital γ_k; the growth rate of y which is written $\gamma_y = \dot{y}/y$
also decreases, since it is proportional to the growth rate k. Figure 7.5 presents the
behavior of γ_k and isolates the two components of the equation which are $s\frac{f(k)}{k}$ and
$(\delta + n)$.

7.3.2.5 The Golden Rule of the Capital Accumulation

In the Solow model, the semi-stationary states are defined by the equality $sf(k^*) =
(n + \delta) k^*$. The objective is to find the value of the saving rate s to which corresponds
a maximum stationary consumption *per capita*. However, since $f(k^*)$ provides the
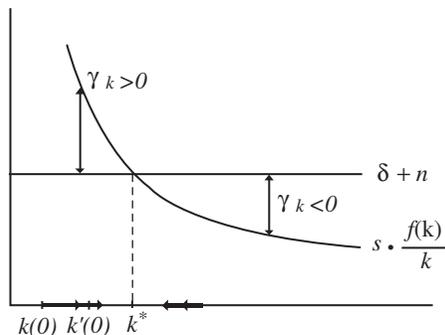
Fig. 7.5 γ_k: Growth rate of k. And convergence towards the BGP

per capita product when the *per capita* capital is k^* and since the consumed portion of the product is $(1-s)$, then, the stationary *per capita* consumption is written:

$$(1-s)f(k^*). \tag{7.69}$$

If the growth rate n of the labor factor is given, as well as the rate of depreciation δ of the capital, each value of s corresponds to a value $k^* > 0$:

$$k^*(s), \frac{dk^*(s)}{ds} > 0, \tag{7.70}$$

$$sf(k^*(s)) = (n+\delta)k^*(s), \tag{7.71}$$

$$c^*(s) = f(k^*(s)) - (n+\delta)k^*(s). \tag{7.72}$$

The balanced growth in the sense of c^*, first, is increasing with s, because s makes it possible to finance the investment and the demand; then after, it is decreasing, because s reduces the demand through reducing the consumption. And there is a maximum for c^* for a value of s. The optimal couple is (s_{GR}, c^*), s_{GR} being the value of s which maximizes c^*. This value of s_{GR} is written:

$$s_{GR} = \arg\max c^*(s), \tag{7.73}$$

$$\frac{dc^*}{ds} = \left(f'(k^*) - (n+\delta)\right)\frac{dk^*}{ds} = 0, \tag{7.74}$$

the *golden rule* (Fig. 7.6) is written:

$$\Rightarrow f'(k_{GR}) = (n+\delta) \tag{7.75}$$

while knowing that $k_{GR} = k^*(s_{GR})$ and $c_{GR} = f(k_{GR}) - (n+\delta)k_{GR}$.

The golden rule of the capital accumulation is thus written: $f'(k_{GR}) = (n+\delta)$. In such a case, *the variation of the per capita product compensates for exactly the total depreciation of the per capita capital.* The saving rate s_{GR}, which is "dynamically effective", is the saving rate making it possible to obtain the balanced growth path which maximizes the per capita consumption.
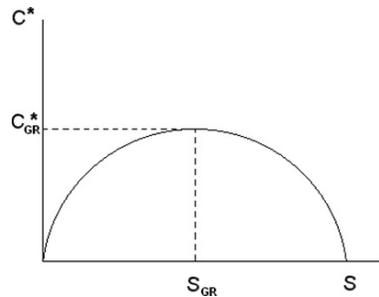
Fig. 7.6 The golden rule

7.3.3 Introduction of the Technological Progress into the Solow Model and Balanced Growth Path

In the previous model of Solow, with the *per capita* variables there is no growth. This fact which was just described above can be modified if the technological progress is included in the model of Solow. By the introduction of "A" which represents *the technology evolution through a technological progress which reinforces the labor*.[33] This means that production processes, or labor units become more productive. Consider a production function such that:[34]

$$Y = F(K, AL) = K^\alpha (AL)^{1-\alpha}. \qquad (7.76)$$

The factor "AL" is often indicated as *the effective supply of labor* (Aghion and Howitt 1999), and is also called the *effective population*. ("AL" is sometimes indicated as being *the efficiency units of labor*.) The technological progress which is *exogenous* in the model results from A which grows at a constant rate. If one notes g the growth rate of the technological progress A, one obtains:

$$\frac{\dot{A}}{A} = g \Leftrightarrow A = A_0 e^{gt}. \qquad (7.77)$$

Let us observe the behavior of such a model. Let us recall:

$$I = S = sY, \qquad (7.78)$$
$$\dot{K} = sY - \delta. \qquad (7.79)$$

Thus, the capital accumulation is written:

$$\frac{\dot{K}}{K} = s\frac{Y}{K} - \delta. \qquad (7.80)$$

The production function becomes:

$$y = k^\alpha A^{1-\alpha}. \qquad (7.81)$$

And by using the logarithmic derivative:

$$\gamma_y = \frac{\dot{y}}{y} = \alpha\frac{\dot{k}}{k} + (1-\alpha)\frac{\dot{A}}{A} = \alpha\gamma_k + (1-\alpha)g. \qquad (7.82)$$

---

[33] "Labor augmenting"; neutral in the sense of Harrod.

[34] Let us remind the various types of technical progress:
   Neutrality in the sense of Harrod: $Y = F(K, AL)$,
   Neutrality in the sense of Solow: $Y = F(AK, L)$,
   Neutrality in the sense of Hicks: $Y = F(K, L)$.

Based on the equation $\frac{\dot{K}}{K} = s\frac{Y}{K} - \delta$, the growth rate of the capital K will be constant only if the ratio $\frac{Y}{K}$ is constant, because s and δ are constant. Thus, the growth rate of the production and growth rate of the capital are equal $\gamma_Y = \gamma_K$ and the growth rates of the *per capita* production and the *per capita* capital are equal: $\gamma_y = \gamma_k$. Consequently, there is in this case a *balanced growth path* (BGP) for which the production, consumption and the population grow at the same constant rate

$$\gamma_y = \alpha\gamma_y + (1-\alpha)g \quad \Leftrightarrow \quad \gamma_y = \gamma_k = \gamma_A = g \geq 0. \tag{7.83}$$

The consequence of such a model of technological progress is that the capital and GDP per capita deploy and increase on a balanced growth path. This conclusion is different from the basic Solow model.

Remark 7.1. The analysis of Solow (1957) of the "capital accumulation" notion is founding in many respects. Indeed, the principle is that the capital accumulation will undergo the effect of decreasing returns, if there is no technical productivity. The technological progress and its continuous improvements make it possible to compensate for the effects of decreasing returns by improving the productivity.

7.3.4 Evolution of the Solow Model and the Neo-Classical Models

We are interested in the various evolutions of the Solow model and in their capacity to explain the convergences between countries.

7.3.4.1 The Solow Model with Human Capital

The model results from two articles, the first one is the paper of Mankiw et al. (1992) and the second is the paper of Lucas (1988). The subject is to note that confrontation between the model of Solow and the experimental data of the economic growth highlights explanatory qualities of the model, which are interesting in particular if one makes it evolve by introducing the notion of human capital. The approach consists in taking into consideration the different levels of skill (or qualification) and education (or learning, training) of countries. Thus, one considers the production function of Cobb Douglas and one supposes that the production results from a combination between physical capital and skilled labor H. The production function is written

$$Y = K^\alpha (AH)^{1-\alpha}. \tag{7.84}$$

A is the technological progress which "reinforces" the labor and A increases at the rate g, which is an exogenous rate. The individuals can arbitrate between the education and labor, and use a part of their time to education rather than to the

direct labor. "u" is the part of time that an individual devotes to education and L *is the global quantity of used labor* in the production. Thus, if N_t represents the population of the model, it is possible to write the equation:

$$L_t = (1 - u)N_t. \tag{7.85}$$

Education and training transform *unskilled labor* into *skilled labor*, this transformation is written:

$$H_t = e^{\psi u}L_t, \tag{7.86}$$

where ψ is a constant $\psi > 0$. The shape of this function results from the analysis of observations relating to the labor economy, which explains that every year of study increases by 10% the wages of the individual. In the particular case where $u = 0$, then $H = L$, i.e. the production is carried out with *unskilled labor*. Let us formalize the impact of a variation of u on the *skilled labor H*:

$$\log H = \psi u + \log L, \tag{7.87}$$

$$\frac{\partial \log H}{\partial u} = \frac{\partial H/\partial u}{H} = \psi. \tag{7.88}$$

Thus, the impact of u on H comes from the positive constant ψ. As regards the accumulation of the physical capital, it is possible to write it in the following way:

$$\dot{K} = s_K Y - \delta K, \tag{7.89}$$

where s_K is the exogenous rate of investment and δ the rate of depreciation of the capital. The accumulation of the physical capital results from the level of investment and savings. *Per capita*, the production function becomes:

$$y = \frac{Y}{L} = k^\alpha (Ah)^{1-\alpha}, \quad \text{with } h = e^{\psi u}. \tag{7.90}$$

If we suppose that "u" is exogenous and constant, and since s_K is exogenous, then "h" is also constant and thus we are in a similar model to the Solow model with technological progress. Inside a balanced growth path (BGP), y and k will increase at the constant rate g. And we take into account the constant variables on the BGP, we can write:

$$\tilde{y} = y/Ah, \quad \tilde{k} = k/Ah, \quad \tilde{y} = k^\alpha. \tag{7.91}$$

And we obtain:

$$\dot{\tilde{k}} = s_K\tilde{y} - (n + \delta + g)\tilde{k}. \tag{7.92}$$

Then, on the growth balanced path, we must have $\dot{\tilde{k}}/\tilde{k} = 0$, this involves that:

$$\frac{\dot{\tilde{k}}}{\tilde{k}} = \frac{s_K}{n + \delta + g} \Rightarrow \left(\frac{\dot{\tilde{k}}}{\tilde{k}}\right)^\alpha = \left(\frac{s_K}{n + \delta + g}\right)^\alpha. \tag{7.93}$$

Since $\tilde{y} = y/Ah$, $\tilde{k} = k/Ah$, $\tilde{y} = k^\alpha$, by dividing by $\tilde{y}^\alpha$, we obtain:

$$\tilde{y}^{*(1-\alpha)} = \left(\frac{\dot{\tilde{k}}}{\tilde{k}} \right)^\alpha = \left(\frac{s_K}{n + \delta + g} \right)^\alpha, \tag{7.94}$$

$$\tilde{y}^* = \left(\frac{s_K}{n + \delta + g} \right)^{\alpha/(1-\alpha)}. \tag{7.95}$$

This equation gives the value of the product per *effective skilled labor unit* on the BGP, whereas the *per capita* product on the BGP is written:

$$y_t^* = \left(\frac{s_K}{n + \delta + g} \right)^{\alpha/(1-\alpha)} hA_t. \tag{7.96}$$

Thus, the *per capita* product grows at the rate g on the BGP. *The difference of wealth between countries* is better explained by the equation above:

- The countries are richer, because they have *a high-rate of investment in physical capital.*
- The countries are richer, because they have *a lower demographic growth rate.*
- The countries are richer, because they have *a strong technological progress.*
- The countries are richer, because they give a lot of time and importance to the *education* and training: $e^{\psi u}$.

7.3.4.2 Convergence and Diversity of Growth Rates

If one observes the growth rates of the OECD (Organisation for Economic Co-operation and Development) countries between 1960 and 1990, one notes a kind of convergence between these countries: they concentrate along a (decreasing) regression line in a plane made up of the product/*per capita* and growth rates of different countries (Fig. 7.7). This observation disappears if, over the same period, one takes into account all countries constituting the world growth: the concentration disappears and the convergence also

$$\frac{\dot{\tilde{k}}}{\tilde{k}} = s_K \frac{\tilde{y}}{\tilde{k}} - (n + \delta + g), \quad with \quad \frac{\tilde{y}}{\tilde{k}} = \tilde{k}^{\alpha-1}. \tag{7.97}$$

This convergence is called "absolute convergence", which would occur between the economies which have identical stationary-states, and, among these economies, the poor countries should grow more quickly than the rich countries. In addition, the more an economy is below its stationary-state, the more its growth will be fast, and on the contrary, the more an economy is above its stationary-state, slower will be its growth.
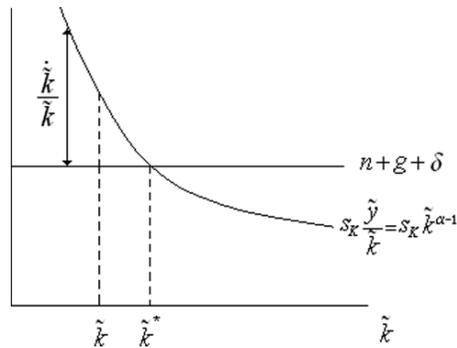
Fig. 7.7 Convergence and dynamics of transition

7.4 Introduction of Externalities, and Instability: Endogenous Growth Theory

The theory has been stated in 1986 by P. Romer (Romer 1986) and in 1988 by R. Lucas (Lucas 1988). Theory seeks *to explain the growth of the per capita production of an economy by means of the accumulation process*, without justifying it by exogenous factors. The Solow model, neo-classical paradigm, has been used for a long time to illustrate the works on the economic growth, but it did not make it possible to justify the phenomena of "prolonged growth" and the disparities of growth rates between countries. Thus, the endogenous growth theory, which claims to adhere to neo-classics, adapts certain assumptions of the Solow model.

7.4.1 Interrupted Growth in the Solow Model and Long-Term Stationarity

The conclusion of the model is that the accumulation of the physical capital does not explain the long periods of strong growth of the *per capita* income and does not explain either the different levels of *per capita* income between the different countries. In the model, the participation of the capital in the production is direct and this participation is remunerated according to its marginal product. The basic model *radically ignores the (positive) externalities*, and in particular those which are generated by the capital, to justify the growth. The Solow model explains the growth by using:

- An *endogenous* source, which is the capital accumulation, determined by the model
- An *exogenous* source, which is *the quantity of available labor*

If the exogenous source is supposed to be "constant" in order to facilitate the analysis of growth, then, one can observe how the production evolves by basing oneself on the only accumulation process. *Within such a framework, if a decrease in the growth is observed, it will be explained by:*

- A *decreasing marginal productivity of the capital*[35]
- The *constant returns to scale*

Having postulated that the *labor quantity* is constant in the course of time, its contribution to the production through the quantity is also constant and only the capital factor can contribute to the increase in the production until the impact of a decreasing marginal productivity is observable and deadens the growth of the product.

7.4.2 Introduction of Positive Externalities

7.4.2.1 Excess or Lack of Externalities: Between Explosive Trajectory and Stationary Trajectory

The endogenous growth models are mainly built on the existence of positive externalities which are the source of the growth. The introduction of these externalities into the model is carried out by admitting that "the increase in the total quantity of spare capital in the economy involves a rise of the labor productivity". It is the major assumption of the endogenous growth theory which introduces in the model the "learning by doing" factor, which explains that the "knowledge" accumulation is done in the course of time. The endogenous growth models are thus based on the existence of positive externalities. *"Too many externalities" generate explosive behaviors and not enough externalities generate stationary behaviors.* The writing of the assumption *is done by assigning a "multiplier" A to the labor factor L* in the production function:

$$A = a\left(\sum K\right), \tag{7.98}$$
$$AL = a\left(\sum K\right)L, \tag{7.99}$$

$\sum K$: corresponds to the total stock of capital of the economy,
$a(\sum K)$: is the *multiplier of L*,
$a(\cdot)$: is a strictly increasing function,
$a(\sum K) > 1$: what means that there is a positive externality.

It is possible to say that the contribution of the capital to the production is "indirect". The new production function is written:

$$Y = F\left(K, a\left(\sum K\right)L\right). \tag{7.100}$$

---

[35] *Inada condition (1964):* The marginal productivity of the capital tends towards zero. If $f(\cdot)$ is the intensive form of the production function, then $lim_{k\to 0}f'(k) = \infty$, $lim_{k\to\infty}f'(k) = 0$.

Because of the initial assumption, the production function is homogeneous of degree 1, then we can write:

$$\frac{Y}{L} = F\left(\frac{K}{L}, a(\textstyle\sum K)\right). \tag{7.101}$$

By posing y the *per capita* product and k the *per capita* capital, the equation above can be written:

$$y = F(k, a(\textstyle\sum K)). \tag{7.102}$$

With $y = Y/L$ and $k = K/L$. If one writes the temporal derivative of the equation above, one obtains the following generic writing:

$$\dot{y} = F'_K(k, a(\textstyle\sum K)) \times \dot{k} + F'_L(k, a(\textstyle\sum K)) \times a'(\textstyle\sum K) \times \textstyle\sum \dot{K}. \tag{7.103}$$

Moreover, in the Solow model, the growth is interrupted asymptotically, i.e. $\dot{y}_{t\to\infty} \to 0$. This is the absence of externality which explains this observation. One can come back to Solow model by posing:

$$a(\textstyle\sum K) = 1, \tag{7.104}$$

$$a'(\textstyle\sum K) = 0, \tag{7.105}$$

$$\dot{y} = F'_K(k, 1)\dot{k}. \tag{7.106}$$

Moreover, remember that in the Solow model $\dot{K}_{t\to\infty} \to 0$, which is the Inada condition of the decreasing marginal productivity of the capital.

   Since we had assumed, in the initial description of the Solow model, that the quantity of available labor was constant in the course of time, and since the marginal productivity of the capital is decreasing, we deduce within the Solow model framework that the capital *per capita K/L converges towards a stationary value*. Thus, the *per capita* production $y = Y/L$ *tends to a stationary value in the long term*. The only possible growth comes from the increase in the population. On the other hand, if there is externalities, the temporal derivative $\dot{y}$ preserves all its components, because $a'(\textstyle\sum K) \neq 0$:

$$\dot{y} = F'_K(k, a(\textstyle\sum K)) \times \dot{k} + F'_L(k, a(\textstyle\sum K)) \times a'(\textstyle\sum K) \times \textstyle\sum \dot{K}, \tag{7.107}$$

Furthermore, from this equation, if we assume the following condition:

$$a'(\textstyle\sum K) > 0. \tag{7.108}$$

If $\textstyle\sum K$ grows continuously, *then the per capita product $y = Y/L$ will increase indefinitely*, whereas in the Solow model, the *per capita* product tends to a stationary value. The endogenous structure of the model resides in the function $a(\textstyle\sum K)$ which represents the endogenous process of accumulation, engine of the growth. The *paradigm of the endogenous growth model described by Romer, which will be evoked more in detail in the section which follows, chooses a "Cobb Douglas" as*

production function and writes the multiplier of the labor factor which is also the "learning" factor, as follows:

$$a\left(\sum K\right) = \left(\sum K\right)^{\beta}. \tag{7.109}$$

The Romer model chooses a value β for which the *growth* is carried out at *a constant rate*.[36] The structure of the production function becomes then, after a rapid transformation of the well-known form aK

$$Y = aK : Production\ function\ at\ constant\ rate. \tag{7.110}$$

If one poses s the *saving rate*, in such a case, contrary to what occurs in the model from initial Solow, the growth depends on the saving rate, because using

$$\dot{K} = sY : Investment = Savings \tag{7.111}$$

and $Y = aK$: *Trajectory at constant rate,* one obtains: $\dot{K}/K = as$.[37] One has therefore as solution of the linear differential equation: $K(t) = K(0)e^{ast}$. In this economy, K grows at the constant rate "as", and, since the production is proportional to K, it also grows at this rate. *This generic model provides an example where the long-term growth is endogenous and depends on the saving rate.*

The determinant of the knowledge accumulation is based on the idea that while the individuals produce goods, they find ways to progressively improve the production processes. Moreover, this increase in the productivity is done without visible introduction of innovation into the production process. It is possible to say that the knowledge accumulation, which is called the *learning by doing* in such a case, is a involuntary "collective" product of the economic activity itself. The learning by doing is the fortuitous consequence of the production of new capital (Romer 1996). Indeed, such as the model was posed, the increase in knowledge depends on the increase in capital. And the stock of knowledge is a function of the stock of the capital. Generally, *the idea of externality is analyzed as inefficient in the Pareto sense*, because of the existence of interactions between the agents, which raises *the question of the optimal remuneration of these "subjacent" externalities* which do not exist obviously as an explicit factor of production in the production function and which are however the engines of the growth. They are supposed being remunerated by other production factors in an indirect way.

---

[36] The nature of the growth will depend on the β value:

- Explosive growth
- Growth $\to 0$, as in the Solow model
- Constant growth

[37] That means that the capital accumulation is controlled by $\dot{K} = asK$. The equation implies that K grows at constant rate: as.

7.4.2.2 Stability of the Romer Model: Self-Sustained Semi-Stationary Growth at Constant Rate on the "Razor's Edge"

The endogenous growth models are based on the existence of *positives externalities*[38] which are supposed to be generators of the growth. "Too many externalities" generate explosive behaviors and "not enough" externalities generate stationary behaviors, whereas the Romer model is of a semi-stationary nature. Paul Romer describes a model which is used as a base of endogenous growth models. The selected production function is a Cobb–Douglas function:[39]

$$F(K,L) = K^{1-\alpha}L^{\alpha} : Production\ function. \tag{7.112}$$

At this stage, Romer takes into account a new factor:

$$a(\sum K_i) = (\sum K_i)^{\beta} : Knowledge\ or\ learning\ by\ doing\ factor$$
$$(\sum K_i : global\ capital).$$

This learning factor applies to all the production units and increases the skill level of workers. If $j = 1,\ldots,n$ indicates the companies of the model, then it is possible to write for one company j:

$$F(K_j, a(\sum K_i)L_j) = K_j^{1-\alpha}[(\sum K_i)^{\beta}L_j^{\alpha}]^{\alpha} : Production\ function\ of\ j.$$

The global production of the economy is:

$$Aggregate\ production:$$
$$Y = F(\sum K_i, a(\sum K_i)\sum L_i) = (\sum K_i)^{1-\alpha}(\sum K_i)^{\beta\alpha}(\sum L_i)^{\alpha}$$
$$= (\sum K_i)^{1-\alpha+\beta\alpha}(\sum L_i)^{\alpha}.$$

The evolution of the model depends on the exponent of the global capital $1 - \alpha + \beta\alpha$:

(a) If $1 - \alpha + \beta\alpha < 1$: the *per capita* production increases more and more slowly as in the case of the initial Solow model.
(b) If $1 - \alpha + \beta\alpha > 1$: the *per capita* production increases more and more quickly.
(c) If $1 - \alpha + \beta\alpha = 1$: then, $\beta = 1$. Thus, the production becomes:

$$(\sum K_i)(\sum L_i)^{\alpha} : Aggregate\ production. \tag{7.113}$$

---

[38] *Positive externality:* indicates any situation where the activities of one (or several) economic agent has (have) consequences on the well-being of other agents, without exchanges or transactions. The externalities can be positive or negative.

[39] With $0 < \alpha < 1$.

By posing $\mathbf{a} = (\sum L_i)^\alpha$ and $\mathbf{K} = (\sum K_i)$, if the *labor factor* is supposed to be constant, then the aggregate production grows only in proportion of the global capital $\mathbf{K}$ and is written:

$$\mathbf{Y} = \mathbf{a} \cdot \mathbf{K} : \textit{Aggregate production:}$$
$$\textit{Self-sustained growth at constant rate.}$$

Such a case results from strict assumptions, in particular about β, *and the consequence is a self-sustained growth at a "constant rate". A β value modification will see a change of trajectories more than proportional which can then either explode or become stationary.*

The main difficulty or "critique" concerning this type of model is related to the question of the aggregation of the agents of the model. The Solow model, like many models where the growth is due to exogenous factors, is an aggregate model and the production function represents all the agents, as if it was a unique representative agent. This conception cannot be retained as regards the endogenous growth models since the *externalities,* which are (mainly[40]) the source of the endogenous growth, *can exist only if there are interactions between the agents*. Thus, these models must treat the question of the aggregation without doing it as in the Solow model for which there is only a unique agent.

7.4.3 Endogenous Growth Without Externality

Certain models of endogenous growth, always constructed from the Solow model structure, do not use the concept of externality to justify an endogenous growth. The endogenous growth, in such a case, is always caused by the learning factor, but one considers that this factor can create "labor factor" unities, which have the form of the human capital. Thus, the concept of learning appears in the model and the individuals have arbitrages to do between investment in the material capital and investment in the learning capital.

7.5 Incentive to the Research by Profit Sharing: The Romer Model (1986–1990)

In addition to the problem of the disparity of growth between countries, the endogenous theories of growth seek to explain the origin of the *technological progress,* which is economic growth factor or promoter. The endogenous growth models introduce an idea initiated by P. Romer, which consists in saying that the technological progress results from the search of profit sharing by the researchers and the inventors.

---

[40] Indeed, endogenous growth can exist without externality.

7.5.1 Basic Components of the Romer Model

The Romer model internalizes (i.e. endogenizes) the technological progress, unlike (for example) the Solow model with human capital where the growth rate of technological progress "g" is exogenous. The novation consists in saying that the inventors, who are involved in profit sharing, generate technological progress. Thus, technological progress results from the Research and Development: R&D. The purpose of the Romer model is to explain why the developed countries profit from a sustained high growth. The model is described, as previously, by two fundamental equations, i.e. the production function and the dynamics of the production factor(s). The production function is written:

$$Y = K^{\alpha}(AL_Y)^{1-\alpha}. \tag{7.114}$$

L_Y is the labor devoted to the production, and A is the level of technology of the economy. For a given level of technology "A", the production function has *constant returns in K and in L_Y*. But, if one then considers that the level of technology (or ideas) A is also a factor of production, *the technology has increasing returns*:[41]

$$F(tK, tA, tL) = (t^{\alpha}K^{\alpha}) \times (t^{1-\alpha}A^{1-\alpha}) \times (t^{1-\alpha}L_Y^{1-\alpha}), \tag{7.115}$$

$$F(tK, tA, tL) = t^{2-\alpha} \times F(K, A, L) > t \times F(K, A, L). \tag{7.116}$$

In addition, the equations of capital and labor accumulation are similar to that of the Solow model:

$$\dot{K} = s_K Y - \delta K, \tag{7.117}$$

$$\frac{\dot{L}}{L} = n. \tag{7.118}$$

Lastly, one must write the equation defining the evolution of the technological progress which concerns the new production factor (we point out that, in the neoclassical models, the growth rate of the technological progress was exogenous). The Romer model *endogenize* the evolution of the production factor A. $A(t)$ is the stock at the instant t of the production factor, it is the stock of ideas invented previously. $\dot{A}$ represents the evolution of the factor A and is written as follows:

$$\dot{A} = \tau L_A, \tag{7.119}$$

where L_A is the number of people which devotes their time in the search of new ideas and τ is the rate to which they find new ideas, i.e. also the number of new ideas. What means in particular that:

$$L = L_Y + L_A. \tag{7.120}$$

---

[41] The presence of increasing returns result from the non-rival use of the ideas.

We write also that τ depends on stock (of already found ideas):

$$\tau = A^\phi, \quad \text{with } \phi < 1. \tag{7.121}$$

The return of the research activity of new ideas will depend on the number of people involved in the R&D process. Thus, it is not possible any more to use L_A, but we will use L_A^λ where λ expresses the *duplication* of research efforts. Consequently, the evolution of A (stocks of ideas or knowledge) is written:

$$\dot{A} = \rho L_A^\lambda \times A^\phi. \tag{7.122}$$

The equations $\tau = A^\phi$ and $\dot{A} = \rho \, L_A^\lambda \times A^\phi$ are fundamental in the new economic growth models. The individual researchers consider that τ is given and they *observe* "constant returns". However, at a global level, using the last equation $\dot{A} = \rho \, L_A^\lambda \times A^\phi$, the production of new ideas have not necessarily constant returns, even if τ only varies very slightly under the action of individual researchers. It is however clear that τ varies by being subject to the influence of the global research. It is possible to give an illustration with $\lambda < 1$ which expresses a kind of saturation or congestion and is translated in terms of externalities associated with the duplication; another illustration can be given by $\phi > 0$ which is a positive externality in the research.

7.5.1.1 Growth and BGP in the Romer Model

If a constant part of the population uses its time for the production of new ideas, the model of P. Romer concludes like the neo-classical model, i.e. *"all the per capita growth is due to technological progress", and like in the Solow model with technological progress, the following equalities must be verified:*

$$\gamma_y = \gamma_k = \gamma_A. \tag{7.123}$$

The purpose is to identify the growth rate of the technological progress along the BGP and with this intention, one poses:

$$\frac{\dot{A}}{A} = \rho \frac{L_A^\lambda}{A^{1-\phi}} \tag{7.124}$$

and along the BGP, we must verify: $\dot{A}/A = \gamma_A = constant$. However the last equality is possible only if the numerator and the denominator of the equation, above, evolve at the same rhythm:

$$0 = \lambda \frac{\dot{L}_A}{L_A} - (1 - \phi) \frac{\dot{A}}{A}. \tag{7.125}$$

Along the BGP, we have $\dfrac{\dot{L}_A}{L_A} = \dfrac{\dot{L}}{L} = n$, then it is possible to write:

$$\gamma_A = \frac{\dot{A}}{A} = \frac{\lambda n}{(1 - \phi)}. \tag{7.126}$$

Thus, the growth rate of A, but also the long-term growth rate of the economy γ_y depends on the growth rate of the population and depends on the parameters of the production function of ideas:

- Let us examine the case $\lambda = 1$ (duplication of research efforts $= 1$) and $\phi = 0$: this implies $\tau = \rho$ and $\dot{A} = \rho L_A$. However, if L_A is constant, the number of new ideas at each period is constant, and the part of these new ideas, in the global stock, decreases in the course of time, i.e. $\dot{A}/A = 0$. *Thus, one will have a sustained high growth of the economy if the number of new ideas is increasing in the course of time, i.e. if the population devoted to the research increases, or, if the global population increases, which is written:*

$$\gamma_Y = \gamma_A = n. \tag{7.127}$$

With this equality, once more, one notes a similarity with the Solow model with technological progress; on the other hand, the active principles in the P.Romer model are very different, because they use the notion of endogenous creation of new ideas. A population growth must involve a growth of the number of new ideas. Thus, a stop of the population growth in this model suggests a stop of the economic growth. Furthermore, in order to have a long term growth, the effort of research must be more than constant, i.e. increasing.

- Let us examine the particular case $\lambda = 1$ and $\phi = 1$, which is connection with the production function of new ideas of the Romer model in 1990, we have:

$$\dot{A} = \rho L_A A \Rightarrow \frac{\dot{A}}{A} = \rho L_A. \tag{7.128}$$

In such a case, the growth rate of new ideas is equal to ρL_A. Unlike the preceding case, the economic growth will not stop even if the effort of research is constant, and this *because the productivity of the research $\tau = \rho A$ is increasing, even if the number of researcher remains constant*. This interesting case is rare in economic actuality. Indeed, we know that the efforts of research strongly increased these last decades and, however, the growth rates did not increase proportionately, which leads to suppose $\phi < 1$.

Lastly, it will be emphasized that the economic policies cannot have effect on the growth, because none of the usual levers of economic policies is operative in the fundamental equation of the Romer model:

$$\gamma_A = \frac{\dot{A}}{A} = \frac{\lambda n}{(1 - \phi)}. \tag{7.129}$$

7.5.1.2 Comparative Statics and Transitory Effect of a Permanent Increase in the R&D Effort in the Population

Let us examine the case where, for example, the government subsidies would aim at increasing the R&D effort in order to increase the number of researchers in the

population, the final purpose being to know its impact on the economy within the framework of this model. If $\lambda = 1$ and $\phi = 0$, the growth rate of factor A is written then:

$$\frac{\dot{A}}{A} = \rho \frac{s_R L}{A}, \qquad (7.130)$$

where s_R expresses the ratio of the population which takes part in the effort of research and development, thus we have $L_A = s_R L$.
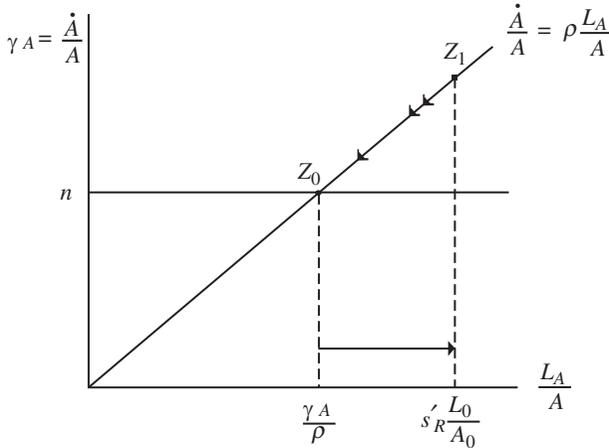
The initial balanced growth path is located in $\gamma_A/\rho = L_0/A_0 s_R$, and if s_R is increased, the part of L_A increases in relation to L, and the part of L_A also increases in relation to A. However, when we are in z_1, we have $\gamma_A = \dot{A}/A$ which is higher than n, thus the ratio L_A/A will tend to decrease until $\gamma_A = n$. In fact, the effect of a permanent increase in s_R on the economy is transitory. The transition dynamics must be connected to the Solow model after an increase in s.
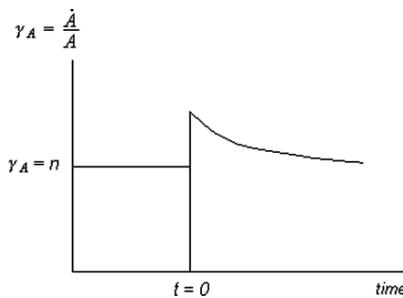
A constant growth rate implies that y/A is constant (s_K = rate of investment):

$$\left(\frac{y}{A}\right)^* = \left(\frac{s_K}{n + \gamma_A + \delta}\right)^{\alpha/(1-\alpha)} (1 - s_R). \qquad (7.131)$$

On the BGP, the equation $\frac{\dot{A}}{A} = \rho\frac{s_R L}{A}$ corresponds to $A = \rho\frac{s_R L}{\gamma_A}$ and we obtain by substitution in the preceding equation:

$$y^*(t) = \left(\frac{s_K}{n + \gamma_A + \delta}\right)^{\alpha/(1-\alpha)} (1 - s_R)\rho\frac{s_R}{\gamma_A}L(t), \tag{7.132}$$

$$y^*(t) = \theta L(t). \tag{7.133}$$

This formulation shows a *scale effect* which expresses that *an economy having a larger population is also richer*. This observation can be explained by the non-rival nature of ideas and by the fact that a larger economy has a vaster market wherein an idea can expand and thus have a larger return (see below Figure of the Level effect).
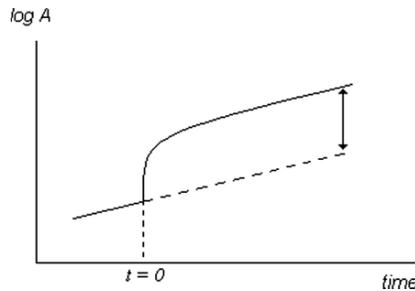
7.5.2 Imperfect Competition, Externalities and R&D Optimality: The Reconciliation in the Romer Model

The role of the *technology* is central in Solow models: From a purely exogenous initial conception in the first models, the technology was incorporated in the more recent approaches. The technology is the way in which the inputs are transformed into outputs during the process of the production. It is thus logical that technological progress must fully integrate the production function and must be modeled somehow or other (i.e. direct or indirect). Romer leads an analysis which aims at explaining the evolution of technological progress by the role of researchers and their financial incentive to innovation. By explaining the technological progress by its source (i.e. the research), we are then interested in the creation processes of "ideas" (technological innovations, process innovations). This is what occurs in the model that we have just studied, the ideas and their production are fundamental. Consequently, the ideas become a Good of which it is advisable to specify the nature. This nature is particular at least for two reasons: First, the ideas considered as goods can be regarded as having a non-exclusive use, in spite of their protection by rights of ownership; Second, the non-exclusive use of these goods has as a consequence of increasing returns. The cost index of a good of this nature can be defined as follows:

$$C(q) = F + cq, \quad \text{with } AC(q) = F/q + c, \ MC(q) = c, \tag{7.134}$$

and

$$\frac{\partial AC(q)}{\partial q} < 0, \quad \lim_{q \to \infty} AC(q) = c \quad \text{and } AC(q) > MC(q),$$

where p, q, C, F, AC, MC are respectively the price, quantity, global cost, fixed cost, average cost and marginal cost. Thus, each produced additional unit is less expensive than the unitary average cost of already produced units. Lastly, these increasing returns to scale leads to reject the possibility of a perfect competition and to admit for these goods the existence of an imperfect competition. Indeed, in perfect competition one can write briefly that, since the price of the good corresponds to its marginal cost:

$$p = MC(q) = c < AC(q), \tag{7.135}$$

the profit is negative:

$$\text{Profit} = pq - AC(q)q < 0. \tag{7.136}$$

It appears that the perfect competition cannot remunerate the production of this type of good; the absence of profit is obviously not incentive. To consider the ideas as a good requires a description of exchanges, but also of equilibria on the market of this good. These exchange mechanisms will not be presented here whereas we made it for the Romer model. We will simply note that, taking into account an imperfect competition, Romer introduces, within a general equilibrium framework, a "sector of intermediate good" for the exchanges of the good (i.e. the "ideas") and a sector of the "research" to express the production mechanisms of new ideas, while preserving the sector of the final good which is produced from several *capital good* which are the intermediate goods. In such a framework, a new idea corresponds to the production of a new *capital good*. The production of ideas and goods are dissociated and the sector of intermediate goods shows increasing returns. The existence of increasing returns in a production function was not reconcilable with the perfect competition and with the principle of the remuneration of production factors to their marginal productivity. It is from this observations that Romer introduced the concepts that we have described previously, on the one hand the imperfect competition (e.g. the monopoly[42]) and the two new sectors, that of "ideas" and that of the "research" which integrate now what could not be integrated before, i.e. the increasing returns. The monopolies sell the intermediate goods at a price higher than the marginal cost and the totality of the carried out profit remunerates the researchers and the inventors. This observation cannot lead us to accept the idea of optimality within this framework of imperfect competition. In the Solow

---

[42] *Imperfect competition:* is a situation in any market where the conditions necessary for perfect competition are not satisfied. Forms of imperfect competition include: (1) Monopoly, in which there is only one seller of a good. (2) Oligopoly, in which there is a small number of sellers. (3) Monopolistic competition, in which there are many sellers producing highly differentiated goods. (4) Monopsony, in which there is only one buyer of a good. (5) Oligopsony, in which there is a small number of buyers. There may also be imperfect competition in markets due to buyers or sellers lacking information about prices and the goods being traded, etc.

model, the price of factors is equivalent to their marginal productivity (Euler's Identity):

$$wL + rK = Y. \tag{7.137}$$

In such a framework, there does not remain any more product Y to remunerate the researchers and their activity. This absence of remuneration within the framework of competitive equilibrium makes the imperfect competition necessary, so that the capital factor is remunerated below its marginal productivity, and the rest of the product is assigned to the remuneration of the new ideas. Furthermore, a major fact led to reject the idea of optimality on the R&D market, this is the presence of positive externalities in the search of new ideas, which is not a stimulating factor, because the market, globally, "under-incites" the research. Indeed, if $\lambda < 1$, "the market does not penalize the researchers who reduce" the productivity of the other researchers, there is an over-incentive to research by the market. On the other hand, if $\phi > 0$ "the market does not take into account the fact that the productivity of the research is increasing with the stock of already discovered ideas, and the current researchers are not rewarded to have the productivity of the future researchers"; there is an under-incentive to research. There is also the question of pure innovations, i.e. the creation of new ideas producing a new capital-good in the market of final goods, which is a real source of surplus whose researchers do not receive the counterpart; there is here also an under-incentive to the research.

7.5.3 Romer Model and Transfer of Technology Between Countries

The countries produce a good Y with the labor factor and a range of capital-good x_j.[43] Since this range of capital-good is limited by the skill level of the labor h, the function is written:

$$Y = L^{1-\alpha} \int_0^h x_j^{\alpha} dj. \tag{7.138}$$

In addition, a gross capital unit is necessary to produce one unit of capital-good, thus one writes:

$$K(t) = \int_0^{h(t)} x_j(t) dj. \tag{7.139}$$

One will consider a small country which is still remote from the technological border of the world economy. The growth will depend on the "learning" of the use of the advanced production tools which are already exploited in the rest of the world (the model comprises intermediate goods: $x_j = x = K/h, \forall j$). One obtains the following Cobb–Douglas function, where *the skill level reinforces the labor:*

$$Y = K^{\alpha} (hL)^{1-\alpha}. \tag{7.140}$$

---

[43] Refer to the economic mechanism of the Romer model.

The dynamics of the capital accumulation is written:

$$\dot{K} = s_K Y - \delta K. \tag{7.141}$$

The accumulation of skills (i.e. know-how) here, will be written in the following way:

$$\dot{h} = \mu e^{\psi u} A^\lambda h^{1-\lambda}. \tag{7.142}$$

With $\mu > 0, 0 < \lambda \leq 1$; u is the time devoted to the know-how accumulation; A represents the world technological border, i.e. it corresponds to the index of the most advanced capital-good. ($e^{\psi u}$ *finds an empirical justification – Nelson and Phelps 1966*). And the variation of the skill $A^\lambda h^{1-\lambda}$ is a geometric mean of the technological border and of the skill of the country. If one writes the growth rate of the skill of the work force:

$$\gamma_h = \frac{\dot{h}}{h} = \mu e^{\psi u} \left(\frac{A}{h}\right)^\lambda \tag{7.143}$$

this equation shows that, "the more a country approaches the world technological border, the more the growth of its level of skill slows down". In this model, the technological border evolves at a rate

$$\gamma = \frac{\dot{A}}{A}, \tag{7.144}$$

which is the consequence of worldwide investments in research and development, and this rate is constant from the point of view of the considered country here. Along the BGP, the growth rate of h must be constant:

$$\gamma_h = \frac{\dot{h}}{h} = \mu e^{\psi u} \left(\frac{A}{h}\right)^\lambda. \tag{7.145}$$

However, h reinforces the labor in the production function of the model; thus, h will contribute to the determination of y and k,[44] and we can observe in the preceding equation that $\dot{h}/h$ depends on A/h, this means that $\dot{h}/h$ will be constant only if A/h is also constant. Thus, the following equality must be satisfied:

$$\gamma_y = \gamma_h = \gamma_k = \gamma_A = \gamma. \tag{7.146}$$

At this stage, one can determine:

$$\left(\frac{K}{Y}\right)^* = \frac{s_K}{n + \gamma + \delta}, \tag{7.147}$$

which, after being introduced into the production function $Y = K^\alpha (hL)^{1-\alpha}$, provides:

$$y_t^* = \left(\frac{s_K}{n + \gamma + \delta}\right)^{\alpha/(1-\alpha)} h_t^*. \tag{7.148}$$

---

[44] Reminder: $Y = K^\alpha (hL)^{1-\alpha}$ and $\dot{K} = s_K Y - \delta K$.

Starting from the equations $\gamma_h = \frac{\dot{h}}{h} = \mu e^{\psi u} (\frac{A}{h})^{\lambda}$ and $\gamma_y = \gamma_h = \gamma_h = \gamma_A = \gamma$ one obtains:

$$\gamma_h = \gamma \Rightarrow \left(\frac{h}{A} \right)^{*} = \left(\frac{\mu}{\gamma} e^{\psi u} \right)^{1/\lambda}. \tag{7.149}$$

That means that, when one devotes more time to the "know-how" accumulation, the economy is closer to the technological border on its BGP (balanced growth path). And if one introduces this value $(\frac{h}{A})^{*}$ in the equation:

$$y_t^* = \left(\frac{s_K}{n + \gamma + \delta} \right)^{\alpha/(1-\alpha)} h_t^*, \tag{7.150}$$

one obtains the trajectory of y^* on the BGP:

$$y_t^* = \left(\frac{s_K}{n + \gamma + \delta} \right)^{\alpha/(1-\alpha)} \left(\frac{\mu}{\gamma} e^{\psi u} \right)^{1/\lambda} A_t, \tag{7.151}$$

$$\gamma_y = \gamma_h = \gamma_h = \gamma_A = \gamma, \tag{7.152}$$

$$y_t^* = \left(\frac{s_K}{n + \gamma + \delta} \right)^{\alpha/(1-\alpha)} h A_t, \tag{7.153}$$

thus, the factor h of the Solow model with human capital is defined within the framework of this model. And the equation:

$$y_t^* = \left(\frac{s_K}{n + \gamma + \delta} \right)^{\alpha/(1-\alpha)} \left(\frac{\mu}{\gamma} e^{\psi u} \right)^{1/\lambda} A_t, \tag{7.154}$$

appears as *a new interpretation of the economic growth* which aims at saying that the *economies have the growth because they learn how to use the new ideas invented in the worldwide economies*. The key element is the skill level of workers in a country which will make it possible to optimize the exploitation of worldwide "technology". This exploitation is possible only if we suppose that there is a fast diffusion of technology between the countries. The facts seem to show that this diffusion has an effective reality through the "internationalization" of companies. It seems that *all information* (ideas and technologies) *is made available at the worldwide level* and the access to information does not constitute an obstacle, but that the *only obstacle* is the learning capacity of new technologies by the countries. Briefly, it is said a perfect information and an imperfect exploitation of this information. Obviously, the cost of international *patents* comes to moderate this approach, and also the R&D costs, but the set of these fixed costs is all the more quickly profitable if the market is worldwide. Another conclusion of the equation above concerns the long-term, indeed the economies "must" have the same growth rate which is the growth rate of the worldwide technological border. The technology will tend in the long term to be diffused in the countries, and the growth levels will tend to be equalized.

7.6 Nonlinearities and Effect of Economic Policies in the Endogenous Growth Models

The previous developments concern the models where the economic policies such as the increase in subsidies to research and development for example, have only level effects and do not have a long-term growth effect. The effect is not permanent, it is only temporary, because the economies come back in the long run to their initial growth rate. These models are sometimes called the "semi-endogenous" growth models. Indeed, even if technological progress is endogenous without the growth of the population, which is exogenous, or without the growth of the number of researchers, the growth of the *per capita* product will tend to stop. However, the "endogenous growth" expression had been created to define the policies which must influence the growth rates in a permanent way. This dimension of the endogenous growth theory, with permanent effect of economic policies, is presented in the following sections.

7.6.1 AK Model: The Limit Case of Solow Model for $\alpha = 1$

They are models which generate an endogenous growth with an influence of economic policies on the growth rate. It is an emanation of the Solow model without technological progress, i.e. the case $\dot{A}/A = 0$, but with $\alpha = 1$. The production function is written:

$$Y = AK. \tag{7.155}$$

This equation resulting from the Romer model (1986) and Rebelo (1991) provides a production proportional to the stock of capital. The capital accumulation is still written:

$$\dot{K} = sY - \delta K. \tag{7.156}$$

In order to simplify the model, one poses $n = 0$ and K becomes the *per capita* capital by normalizing the population to $N = 1$. One can represent the Solow's diagram as follows (Fig. 7.8).

If the *initial conditions* of the model are positioned at $sY > \delta K$, then the stock of capital will increase and this increase will persist in the course of time, i.e. the investment is higher than the depreciation in a permanent way. And we are supposed to have a permanent growth. In the Solow model, due to $\alpha < 1$, i.e. because of decreasing returns, each unit of additional capital which is added by means of the savings contributes less and less to the production. In the AK model, due to $\alpha = 1$, i.e. because of constant returns, "each unit of additional capital" added contributes in a constant way to the production. The marginal product of each unit of additional capital is always A. The growth rate of the capital, by using $\dot{K} = sY - \delta K$, is written:

$$\gamma_K = \frac{\dot{K}}{K} = s\frac{Y}{K} - \delta = sA - \delta = \text{Constant}. \tag{7.157}$$
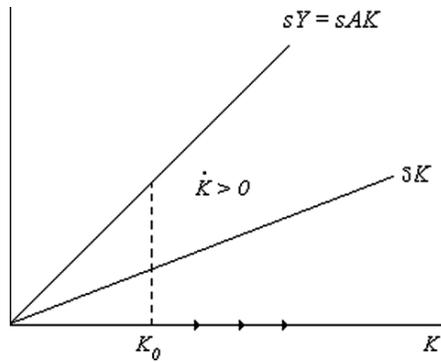
Fig. 7.8 Solow's diagram in AK model

And, by the logarithmic derivative of the production, one obtains:

$$\gamma_Y = \gamma_K = sA - \delta. \tag{7.158}$$

That means that the growth rate of the GDP is an increasing function of the rate of investment "s". Consequently, the economic policies that increase the rate of investment are supposed to increase, at the same time, the growth rate of the GDP of the economy and in a permanent way. *The "AK" model generates the endogenous growth even if the population (or the technology level) does not grow in the model.*

- The case described in the present section is a "limit-case" of the Solow model for $\alpha = 1$, and the parameter $\alpha \in [0, 1]$ indicates the bend of sY. Indeed, the more α is weak, the more the bend of the curve is high and the more the intersection between sY and δK is carried out quickly; on the other hand, the more α is close to 1, the more the form of sY approaches a straight line and the more the intersection with δK is carried out tardily. However, the intersection in K^* of sY and δK indicates the balanced growth path (BGP) (when $sY = \delta$ because $n = 0$). The more the value of α is high, the more the transition from the initial state K_0 towards K^* *takes time* and *for $\alpha = 1$ the dynamics of transition never stops* (Fig. 7.9).

7.6.2 Linearities and Endogenous Growth

7.6.2.1 AK Model

In the preceding "AK" model, the endogenous growth is generated because of a "linearity" in the differential equation of the model. If in the Solow model, the dynamics of the capital accumulation is associated with the production function, one obtains
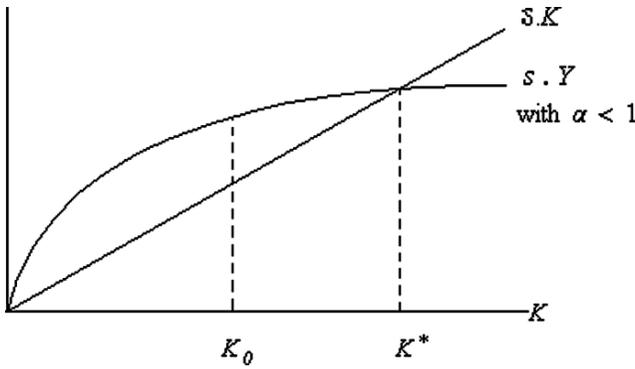
Fig. 7.9 Diagram of the Solow model with $\alpha < 1$

the following differential equation $\dot{K} = sAK^\alpha - \delta K$. Starting from this equation, two cases must be distinguished:

- Either, $\alpha = 1$ and we are in presence of a *linearity* in K which is that of the "*AK*" models,
- Or, $\alpha < 1$ (decreasing returns of the capital) and we are in presence of a *nonlinearity* in K. Then, based on this equation, we can write the growth rate of K in the following way: $\gamma_K = \frac{\dot{K}}{K} = \frac{sA}{K^{1-\alpha}} - \delta$. We can clearly identify the decreasing nature of the relation.

7.6.2.2 Solow Model with Technological Progress

In this case, the *linearity* is present in the writing of the (exogenous) growth rate of the technological progress $\dot{A}/A = g$ which allows to write the following differential equation $\dot{A} = gA$ which is of linear nature. A permanent increase in the rate g impacts, in a permanent way, the growth rate of t*he augmented Solow model*. There is an absence of decreasing returns for A.

7.6.2.3 Lucas Model (1988)

The model uses a Cobb Douglas production function combined with a factor "h" representing *the human capital by person* (i.e. *per capita*). The function is written:

$$Y = K^\alpha (hL)^{1-\alpha}. \tag{7.159}$$

Lucas assumes that the human capital h evolves in the following way, by knowing that "u" is the time devoted to labor and $(1-u)$ is the time devoted to the formation:

$$\dot{h} = (1-u)h. \tag{7.160}$$

The growth rate is written:

$$\frac{\dot{h}}{h} = (1 - u).\qquad(7.161)$$

In such a case, "h" is included in the production function, in the same way as the technological progress "A" was included in the production function of the augmented Solow model, i.e. by *the introduction of a linearity*. Any permanent increase in time devoted to the education or the learning $(1 - u)$ will thus lead to the increase in growth of the *per capita product*, in a permanent way

The three examples which precede illustrate the importance of linearities in the conception of endogenous growth models. However, the endogenous growth phenomenon can exist without linearity by using, for example, a model with two differential equations from which the respective convexities of trajectories are different. Thus, in certain configurations, in the absence of linearity, one can observe the endogenous growth.

The *linearity* assumption is simple to pose, write and exploit, this is why it is so often used. The models of the type $Y = AK$ also seem to belong to this assumption; in this expression, the output is constantly proportional to the input and A is taken as a parameter and not as a variable obviously. However, it proves that this assumption undergoes some critiques, and is not necessarily relevant. Indeed, the basic rules of the economic theory, which state the decrease of the marginal productivity or of the marginal utility, implicitly implies that the relation between input and output is rather of a nonlinear nature. Moreover, even within the framework of models whose construction is linear, the nature of the "domain",[45] in particular at its limits, can be the source of nonlinearity due to *the "edge-effects"*. One can evoke the variables which are expressed in the form of rate (or ratio), like the *per capita* variables for example. The assumption of the nonlinearity is probably more relevant than its opposite which, beyond its didactic virtues, is often replaced by more realistic concepts relating to the nonlinearity notion.

7.6.3 Externalities and AK Models

The introduction of ideas, knowledge or technology in the production function produces the increasing returns to scale. The "increasing returns to scale" seem to imply the abandon of the concept of perfect competition and the introduction of the imperfect competition notion in order to justify them. The neo-classical theory and its framework of perfect competition remunerate the factors of production by their marginal productivity, and we are led to the question of the knowledge remuneration, because *there does not remain product in order to remunerate it*.

---

[45] *Domain (of a function):* The term domain is most commonly used to describe the set of values D for which a function (or map, transformation ,...) is defined. For example, a function $f(x)$ that is defined for real values $x \in \mathbb{R}$ has domain $\mathbb{R}$, and is sometimes said to be *a function over the reals*. The set of values to which D is sent by the function is then called the *range*.

This seems to mean that the researchers cannot be rewarded for their work. *The notion which makes it possible to reconcile perfect competition and increasing returns to scale is the existence of "externalities"*, i.e. the accumulation process of knowledge is an "involuntary" consequence of other activities in the economy. And it is possible to say that the remuneration of the knowledge accumulation is "indirect". Let us consider, for example, a Cobb Douglas function of an enterprise:

$$Y = BK^{\alpha}L^{1-\alpha}. \tag{7.162}$$

It is known that the *returns* of K and L in this equation are *constant*. As regards B, which can represent the knowledge if one imagines that its accumulation process is endogenous, the returns of B would be thus increasing. On the other hand, one can suppose that the capital accumulation (by means of externalities) generates new knowledge in the economy:

$$B = AK^{1-\alpha} \tag{7.163}$$

with A a constant. The individual enterprise does not recognize that its effort of investment improves the level of knowledge of the technology in the economy, taking into consideration its size. The *knowledge* seems not identified by the companies and is thus considered by these companies as "external". But, at the global level of the economy the improvement, resulting from knowledge, is appearing. Thus, the capital is remunerated according to its marginal productivity. It is a remarkable fact, because the companies do not identify in the accumulation of this capital factor the "subjacent" accumulation of knowledge. The production function becomes:

$$Y = AKL^{1-\alpha}. \tag{7.164}$$

And if one removes the factor L by normalizing it to 1, the production function becomes:

$$Y = AK \tag{7.165}$$

which is, by definition, the writing of the production function of the "AK" model. Thus, there exist two approaches in order to endogenize the accumulation of the knowledge by justifying the existence of increasing returns:

- The *imperfect competition*
- The *externalities*

In the Romer model, both approaches (imperfect competition and externalities) were taken into account. Indeed, the externalities allowed the increasing returns of knowledge, and the imperfect competition allowed the increasing returns in the production.

7.6.4 Nonlinearities and Effect of Economic Policies in Endogenous Growth Models: Transitory or Permanent Effects

As described previously, it is possible to do a dichotomy between the endogenous growth models according to the effects of policies on these models. It is a question to distinguish between permanent effect and transitory effect (i.e. level effect).

Let us suppose an economy with a subsidy active policy of the R&D, one can wonder whether the effects of this policy will last indefinitely, or in the long-term (from 10 to 50 years), or in the mid-term (5 years for example). Furthermore, the analysis could lead to say that an effect of long term but not undefined and thus transitory, can be close to a permanent effect. Such an analysis suggests privileging the "level effect" approach (M-LT). Moreover, the measured empirical elements do not make it possible to admit the economic models as being linear, i.e. having $\alpha = 1$. On the contrary, α is close to 1/3 for the capital and even if human capital and externalities are added to it, we are close to 4/5, thus we are always lower than 1.

The linearity does not seem either to be present in the R&D, since it is known that the investment efforts in this field and the number of researchers over the 40 last years increased considerably, whereas the growth rate of economies did not increase in a significant way, because it is near 1.8% during the same period. This leads to conclude that the positive externalities ($\phi > 0$) remain however lower than 1: ($\phi < 1$).

7.7 Basin of Instability and Saddle-Point: Optimal Growth Model of Ramsey without Technological Progress

This founding model in many respects, takes into account at the same time the behavior of households and that of companies. The intertemporal choices of households consist in *arbitrating* between consumption and savings. Savings and consumption are not fixed in an exogenous way, but are the consequence of a optimization mechanism of an endogenous nature. This mechanism is the main novation, contrary to the basic Solow model, which posed an exogenous saving rate. The optimal growth model of Ramsey is often used to show that there is a unique interest rate which is used as indicator for the consumption or investment decisions.

7.7.1 Intertemporal Choices and Utility Function

The choice between consumption and savings is carried out only for one good (the corn for example). One admits that the "standard of living" of a given state of the model can be transcribed in a "measurement" by means of the utility of the *per*

capita consumption $c(t)$. This measurement is denoted $u(c)$, it is the utility function which is increasing whereas its first-derivative is decreasing $u'(c) > 0$, $u''(c) < 0$. For the needs of the model, one considers an economy of which the *lifespan* is equal to T and with *consumption trajectories* of the form:

$$(c_0, \ldots, c_t, \ldots, c_T). \tag{7.166}$$

c_t is the consumption at the moment t. The *utility function* of the model which is noted $U(\cdot)$ is considered separable, i.e. it is of the form: $U(c_0, \ldots, c_t, \ldots, c_T) = u_0(c_0) + \cdots + u_T(c_T)$. Another characteristic consists in supposing that there exists a *"discount factor"* (or a rate of discount), i.e. a "positive *rate of time preference*" which *translates a certain preference for the present*, it is denoted ρ: it is *the rate of preference for the present* in the economy. The *intertemporal utility function* of the model is written as follows:

$$U(c_0, \ldots, c_t, \ldots, c_T) = u_0(c_0) + \cdots + \frac{u_T(c_T)}{(1+\rho)^T}. \tag{7.167}$$

This rate ρ is subjective, it corresponds to the preferences of the *representative agent*. After this discrete description, it is possible to provide a continuous representation:

$$\int_0^T u(c_t)e^{-\rho t}\,dt. \tag{7.168}$$

The program will seek to maximize the intertemporal utility function, such as:

$$\max \int_0^T u(c_t)e^{-\rho t}\,dt. \tag{7.169}$$

In such a case, the *economic horizon* of the model *is limited* to a date T, but the choice of this horizon is an important question because T could be selected unlimited.

7.7.2 The Production Function

The selected production function is the neo-classical function such that:

- The firms rent the services of production factors, i.e. the capital and the labor, and they sell their production to the households and to the other firms. The technology of firms enables them to transform these factors into products (and if we wish, it is possible to make evolve this technology in the course of time, but it will not be the case in this model).
- The households possess the assets and the factors of production, and they choose the part of their income which will be consumed.
- The markets of output and inputs exist and the exchanges determine the prices of the output and inputs.

The production function can be written:

$$Y(t) = F(K(t), L(t), t), \tag{7.170}$$

$$Y_t = F(K_t, L_t). \tag{7.171}$$

Obviously this production function will be written later $Y = F(K, L)$. Technological progress is not taken into account in the model and the *returns to scale are constant*. The *working population* of the model, denoted N_t, grows in an exponential way at rate n. And, since there is *no unemployment* in the model, one can write that $N_t = L_t = N_0 e^{nt}$. All the variables of the model can be expressed *per capita* and the *intensive form* of the production function will be written: $y = f(k)$. In this model, one takes into account the depreciation of the capital which is considered proportional to the stock of capital, and the rate of depreciation is noted ω. By definition, it is admitted that the production function respects the equilibrium condition on the markets and thus the equality:

$$Y = C + I. \tag{7.172}$$

C is the consumption and I the investment. (The equation can be written *per capita* as follows: $y = f(k) = c + i$.) From this equality, one can write the equation of the capital variation in the course of time:

$$dK/dt = \dot{K} = Y - C - \omega K = F(K, L) - C - \omega K. \tag{7.173}$$

The per capita capital, which is written $k = K/L$, enables us to write by taking the logarithm: $\log k = \log K - \log L$. By taking the derivative of this expression, one obtains:[46]

$$\frac{dk}{dt} = \dot{k} = \frac{1}{L}\frac{dK}{dt} - nk. \tag{7.174}$$

Thus, one can write, by introducing the value of $\dot{K}$:

$$\dot{k} = \frac{1}{L}(F(K, L) - C - \omega K) - nk. \tag{7.175}$$

Or:[47]

$$\dot{k} = f(k) - c - \omega k - nk. \tag{7.176}$$

This equation represents the dynamics of the *per capita* capital accumulation in this economy. We can observe clearly that the level of consumption at each period plays a very important part in the dynamics of accumulation.

---

[46] Developed expression: $\frac{1}{k}\frac{dk}{dt} = \frac{1}{K}\frac{dK}{dt} - \frac{1}{L}\frac{dL}{dt}; \frac{dk}{dt} = \frac{k}{K}\frac{dK}{dt} - \frac{k}{L}\frac{dL}{dt}; \frac{dk}{dt} = \frac{1}{L}\frac{dK}{dt} - nk.$

[47] Which can be also written in the following form: $k_{t+1} = f(k_t) - c_{t+1} - \omega k_t - nk_t.$

7.7.3 *Mechanism of Optimization, and Trajectories*

In order to have an optimal growth, it is necessary to carry out an *arbitrage* at each period between consumption and savings. this is because the agents aim at consuming later on. The division between consumption and savings, at each period, depends on the form of the intertemporal utility function (in particular of the rate of preference for the present) and in addition depends on the "production techniques" which are used. To determine the optimal trajectory of consumption and thus the trajectory of the production, one maximizes $\int_0^T u(c_t)e^{-\rho t}dt$ under the constraints represented by $\dot{k} = f(k) - c - \omega k - nk$. The objective is to seek an optimal c which maximizes the utility of agents, or more exactly, of the representative agent at each moment. With this intention, one extracts "c" from the preceding expression:

$$c = f(k) - (\omega + n)k - \dot{k} \tag{7.177}$$

and one introduces it into the function of utility to maximize, which is also the objective function and which becomes:

$$U^* = \max \int_0^T u[f(k) - (\omega + n)k - \dot{k}]e^{-\rho t}dt. \tag{7.178}$$

The unknown element of this problem of maximization is k. The candidate-functions for the resolution of this problem must satisfy the Euler condition, which is written as follows:

$$g'_k - \dot{g}'_{\dot{k}} = 0. \tag{7.179}$$

The g function is defined by:

$$g(t, k(t), \dot{k}(t)) = e^{-\rho t}u(f(k(t)) - (\omega + n)k(t) - \dot{k}(t)). \tag{7.180}$$

Let us calculate g'_k the derivative with respect to k:

$$g'_k(t, k(t), \dot{k}(t)) = e^{-\rho t}u'(f(k(t)) - (\omega + n)k(t) - \dot{k}(t))(f'(k(t)) - (\omega + n))$$
$$= e^{-\rho t}u'(c(t))(f'(k(t)) - \omega - n).$$

Let us calculate the derivative with respect to $\dot{k}$:

$$g'_{\dot{k}}(t, k(t), \dot{k}(t)) = e^{-\rho t}u'(f(k(t)) - \omega - n)k(t) - \dot{k}(t))(-1)$$
$$= e^{-\rho t}u'(c(t))(-1).$$

The derivative with respect to time of the preceding expression is written $\dot{g}'_{\dot{k}}$:

$$\dot{g}'_{\dot{k}}(t, k(t), \dot{k}(t)) = -e^{-\rho t}(-\rho u'(c(t))) - e^{-\rho t}u''(c(t))\dot{c}(t).$$

Taking into account the preceding results, the Euler condition $g'_k - \dot{g}'_{\dot{k}} = 0$ is written

$$e^{-\rho t}u'(c(t))(f'(k(t)) - \omega - n) + e^{-\rho t}(-\rho u'(c(t))) + e^{-\rho t}u''(c(t))\dot{c}(t) = 0$$

then we can extract the expression of $\dot{c}$:

$$\dot{c}(t) = -\frac{u'(c(t))}{u''(c(t))}(f'(k(t)) - \omega - n - \rho). \tag{7.181}$$

By omitting t the simplified writing of $\dot{c}$ is as follows:

$$\dot{c} = -\frac{u'(c)}{u''(c)}(f'(k) - \omega - n - \rho)). \tag{7.182}$$

7.7.3.1 Optimal Trajectories of the Consumption

The following system of differential equations of order 2 *determines the optimal trajectory of consumption and investment, starting from the given initial conditions* $(k(0), c(0))$:

$$\dot{k} = f(k) - c - (\omega + n)k, \tag{7.183}$$

$$\dot{c} = -\frac{u'(c)}{u''(c)}(f'(k) - \omega - n - \rho). \tag{7.184}$$

A brief reminder:

  n: is the *growth rate* of the *population*.

  ρ: is the *rate* of *preference* for the *present (positive rate of time preference)*.

  ω: is the *rate* of *depreciation* of the *capital*.

Starting from given initial conditions, the system thus made up will provide the solutions of the model (if they exist). It is important now to study the behavior of the different trajectories of the system.

The cas $\dot{k} = 0$ *and* $\dot{c} = 0$:

From the system defined previously, one will consider the solutions of:

$$\dot{k} = 0: f(k) - c - (\omega + n)k = 0,$$
$$\dot{c} = 0: -\frac{u'(c)}{u''(c)}(f'(k) - \omega - n - \rho) = 0. \tag{7.185}$$

By admitting that $u'(c) \neq 0$, one obtains:

$$c = f(k) - (\omega + n)k, \tag{7.186}$$
$$f'(k) = \omega + n + \rho. \tag{7.187}$$

From the last equation and by postulating that the marginal productivity of the capital is strictly decreasing, and thus that the function $f(\cdot)$ is *concave*, one can write:

$$k^* = f'^{-1}(\omega + n + \rho). \tag{7.188}$$
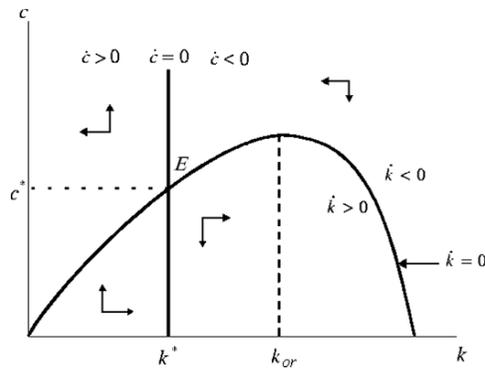
Fig. 7.10 Phase diagram

Replacing this value of k in the equation $c = f(k) - (\omega + n)k$, one obtains:

$$c^* = f(k^*) - (\omega + n)k^*. \tag{7.189}$$

$\dot{c} = 0$ corresponds to a vertical straight line which for a single value of k, noted k^*, and $\dot{k} = 0$ corresponds to a concave curve (The conditions of Inada being respected); both are represented on the phase diagram (Fig. 7.10).

The diagrammatic orientation of the flow (of solutions) is represented by the arrows which help to understand that the solutions are of the saddle-point type. Indeed, let us study the behavior of the system by analyzing the Jacobian matrix. The Jacobian matrix is written:

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial \dot{k}}{\partial k} & \dfrac{\partial \dot{k}}{\partial c} \\ \dfrac{\partial \dot{c}}{\partial k} & \dfrac{\partial \dot{k}}{\partial c} \end{pmatrix} \tag{7.190}$$

for (k^*, c^*), $\mathbf{J}$ is written:

$$\mathbf{J}^* = \begin{pmatrix} f'(k) - \omega + n & -1 \\ -\dfrac{u'}{u''}f''(k) & -\left(\dfrac{u'}{u''}\right)'(f'(k) - \omega - \rho - n) \end{pmatrix}. \tag{7.191}$$

The notations, here, are simplified. Thus, the expression[48] $(\frac{u'}{u''})'$ represents the derivative with respect to c of the term between brackets. Moreover, the component $\frac{\partial \dot{k}}{\partial c}$ of $\mathbf{J}^*$ is zero for (k^*, c^*), due to the factor $(f'(k) - \omega - \rho - n)$ which is zero itself, because $f'(k) = \omega + n + \rho$, when $\dot{c} = 0$. Thus, the determinant of $\mathbf{J}^*$ is written:

$$\det \mathbf{J}^* = [f'(k) - \omega + n] \cdot 0 - (-1)\left(-\dfrac{u'}{u''}f''(k)\right), \tag{7.192}$$

---

[48] Reminder: $\left(\frac{u'}{u''}\right)' = \frac{u'^2 - u'u''}{u''^2}$.

$$\det \mathbf{J}^* = -\frac{u'}{u''}f''(k).\qquad(7.193)$$

Since the model postulated the following assumptions $u'(\cdot) > 0$, $u''(\cdot) < 0$ and $f'(\cdot) > 0$, $f''(\cdot) < 0$, one deduces that: $\det \mathbf{J}^* < 0$. Thus, $\mathbf{J}^*$ has two eigenvalues of the opposite sign, which is symptomatic of a saddle-point. The axis of the saddle is obtained by means of the eigenvector resulting from the eigenvalue of which the real part is negative. *With an aim of obtaining convergence towards the point E (which corresponds to a balanced growth "regime" at constant rate), the capital-consumption initial values couple $(k(0), c(0))$ must belong to the axis of the saddle. The choice of initial conditions is thus crucial for the convergence of the model and leads to the problem of the coordination of agents' behaviors which could be the coordination of a planner or of a unique (representative) agent.*

7.8 Basin of Instability and Saddle-Point: Optimal Growth Model of Cass Koopmans Ramsey with Technological Progress

As in the previous case, this model[49] takes into account, at the same time, the behavior of households and that of companies, and in a competitive economy. The intertemporal choices of the model consist in arbitrating between consumption and investment (or savings). Unlike the Solow model, the savings and consumption are not fixed by an exogenous fundamental rule but are the result of an optimization mechanism of endogenous nature. Thus, unlike the basic Solow model, the saving rate is endogenous. We are thus faced with a growth model with technological progress, without externalities, and the determination of the saving rate is endogenous. The behaviors of the savings and consumption result from a maximization program of the utility.

7.8.1 Enterprises and Production Function

From the point of view of the enterprises, the production function of the model depends on the capital and labor factors, but also on the *technological progress* whose growth rate is exogenous. The function is written:

$$Y = F(K, AL).\qquad(7.194)$$

In this *aggregate production function*, the population L is replaced by the "effective" population AL. A reflects the current state of technological knowledge,[50] and A can

---

[49] Model of Ramsey (1928), Cass (1965), Koopmans (1965).

[50] A could be interpreted as *the number of efficiency units per unit of labor.*

be understood as a productivity parameter which grows at the constant rate $g > 0$. The exogenous value of g is supposed to reflect "scientific progress". The writing of such a production function makes technological progress equivalent to an increase in the *the effective supply of labor AL* which grows according to the *growth rate of the population n plus the productivity:* $(n + g)$. (It is possible to define all the variables according to the "effective supply of labor" AL. Thus, for example, K/AL would correspond to the supply of capital per "effective person".) The companies use work force and capital and A grows according to the exogenous rate g. One considers that the households are the owners of companies and therefore they are the beneficiaries of the profit which is maximized by these companies.

Taking into account the competitive framework, the factors of production are remunerated according to their marginal productivity; since the returns to scales are constant, the profits of companies are considered equal to *zero. The marginal product of the capital* $\partial F(K, AL)/\partial K$ is equal to $f'(k)$. Indeed, taking into account the assumption of the constant returns to scales, one can use the *intensive form*[51] of the production function $f(k)$ (which satisfies the conditions $f(0) = 0$, $f'(k) > 0$, $f''(k) < 0$) and one deduces that $f'(k)$ is the marginal product of the capital because $F(K, AL) = ALf(K/AL), \partial F(K/AL)/\partial K = ALf'(K/AL)(1/AL) = f'(k)$. In the absence of depreciation, one also writes that the real return rate of the capital is equal to its marginal product, consequently, at moment t, the real interest rate is equal to:

$$\textit{Real interest rate : :} r(t) = f'(k(t)),$$
$$\textit{or :} r = f'(k).$$

Moreover, the remuneration of the "effective" labor is carried out according to its marginal product, one can write that $\partial F(K,AL)/\partial AL$ corresponds to $f(k) - kf'(k)$. Therefore, the remuneration, i.e. the *real wage rate per unit of labor* is equal to:

$$w(t) = f(k(t)) - k(t)f'(k(t)), \textit{ or :} \tag{7.195}$$
$$w = f(k) - kf'(k). \tag{7.196}$$

The remuneration of an individual at the moment t is written $A(t)w(t)$, because the marginal product of the labor in this case corresponds to $A\partial F(K, AL)/\partial AL$.

7.8.2 Households and Maximization of the Utility Function Under the Budget Constraint

From the point of view of the households, let us consider a given standard of living, represented by a consumption $c(t)$ and an index of utility $u(c)$ (knowing that

---

[51] The *intensive form* is obtained as follows: $F(K/AL, 1) = (1/AL)F(K, AL)$ where K/AL is the amount of capital per "effective" labor unit. If one notes $k = K/AL$, $y = Y/AL$ and $f(k) = F(k, 1)$, then one obtains the *intensive form* of the production function $y = f(k)$.

$u'(c) > 0, u''(c) < 0$. One considers that there is a great number of identical house-
holds and the size of each household grows according to rate n. The households are
supposed to be the owners of the capital, and they loan it to the companies. *Each
member of the household supplies a unit of labor per unit of time*. If one notes H the
number of households and $K(0)$ the stock of the initial capital of the model, then,
$K(0)/H$ corresponds to the stock of initial capital of the household. In this model,
one does not take into account the depreciation of the capital to simplify the writing.
The households perceive an income for their work and for the loan of their capital,
but they can possibly perceive profits. Then, starting from the perceived income,
they arbitrate between consumption and savings in order to maximize their utility.
The utility function of each household is defined through an infinite lifespan. The
utility function is written:

$$U = \int_{t=0}^{\infty} e^{-\rho t} u(C(t)) \frac{L(t)}{H} dt. \tag{7.197}$$

$L(t)$ is the global population of the model and $L(t)/H$ is the number of person in
the household. $C(t)$ is the consumption of each member of the household, and $u(\cdot)$
is the instantaneous utility function of each member of the household. Moreover,
$U(C(t))L(t)/H$ is the instantaneous utility at the moment t and ρ is the positive rate
of time preference (i.e. the rate of preference for the present) of the function. The
more ρ is high, the more the preference for the present is high. If one notes θ the
rate of "risk aversion", the function of instantaneous utility is written[52]

$$u(C(t)) = \frac{C(t)^{1-\theta}}{1-\theta}, \quad \text{with } \theta > 0 \ \text{ and } \ \rho - n - (1-\theta)g > 0. \tag{7.198}$$

This function expresses a constant "risk aversion" and the coefficient of risk aversion
is written $-Cu''(C)/u'(C)$ and is equal to a constant θ. The utility function depends
on θ. The more θ is small, the more the utility diminishes proportionally to the
increase in $C(t)$. If $\theta \to 0$, the utility function only depend on $C(t)$ and is even equal
to $C(t)$. The *elasticity of intertemporal substitution* of the consumption is equal
to $1/\theta^2$.

The representative household maximizes its utility taking into account its budget.
Since we are in a logic of intertemporal choices, this budget constraint is stated as
follows: it is said that the *discounted value* of its global intertemporal consumption
must remain lower than the current value of its initial wealth increased by the *dis-
counted value* of the *labor incomes* of all its life. This constraint, rather heavy to
write, requires to take into account the variations of the *real return of the capital r*
which can be written $R(t): \int_0^t r(\tau)d\tau$. And $e^{R(t)}$ represents the ratio at the moment

---

[52] (a) If $\theta < 1$, then $C^{1-\theta}$ grows functions of C. If $\theta > 1$, then $C^{1-\theta}$ decreases functions of C.
Thus, while writing $u(C(t)) = C(t)^{1-\theta}/1 - \theta$ it is known that the marginal utility of consumption
is positive for all θ.

(b) And if $\theta \to 1$, then $u(C(t)) \to \log C(t)$.

(c) The constraint $\rho - n - (1 - \theta)g > 0$ ensures the convergence of the intertemporal utility
function, otherwise the maximization program can have no solution.

t of an invested unit (i.e. good or capital) at the beginning of the period. The remuneration of labor[53] of a household is equal to $A(t)w(t)L(t)/H$ and its consumption is written $C(t)L(t)/H$; thus the budget-constraint of the household is:

$$\int_{t=0}^{\infty} e^{-R(t)} C(t) \frac{L(t)}{H} dt \le \frac{K(0)}{H} + \int_{t=0}^{\infty} e^{-R(t)} A(t) w(t) \frac{L(t)}{H} dt. \qquad (7.199)$$

If we reason *per unit of effective labor*, it is possible to write the constraint as follows:

$$\int_{t=0}^{\infty} e^{-R(t)} c(t) \frac{A(t)L(t)}{H} dt \le k(0) \frac{A(0)L(0)}{H} + \int_{t=0}^{\infty} e^{-R(t)} w(t) \frac{A(t)L(t)}{H} dt$$

which can be also written (using the fact that $A(t)L(t)$ is equal to $A(0)L(0)e^{(n+g)t}$):

$$\int_{t=0}^{\infty} e^{-R(t)} c(t) e^{(n+g)t} dt \le k(0) + \int_{t=0}^{\infty} e^{-R(t)} w(t) e^{(n+g)t} dt. \qquad (7.200)$$

This expression is difficult to use but can be simplified by expressing the budget constraint according to the capital possessed by the household when $t \to \infty$. The budget constraint is written now:

$$\lim_{s \to \infty} e^{-R(s)} \frac{K(s)}{H} \ge 0 \qquad (7.201)$$

that is also written:

$$\lim_{s \to \infty} e^{-R(s)} e^{(n+g)s} k(s) \ge 0. \qquad (7.202)$$

Since the *per capita* consumption is written $A(t)c(t)$, one can write:

$$\frac{C(t)^{1-\theta}}{1-\theta} = \frac{(A(t)c(t))^{1-\theta}}{1-\theta} = \frac{(A(0)e^{gt})^{1-\theta} c(t)^{1-\theta}}{1-\theta} = A(0)^{1-\theta} e^{(1-\theta)gt} \frac{c(t)^{1-\theta}}{1-\theta}.$$

The objective-function of the household, i.e. the *utility*, is written *per capita* as follows:

$$U = \int_{t=0}^{\infty} e^{-\rho t} \frac{C(t)^{1-\theta}}{1-\theta} \frac{L(t)}{H} dt, \qquad (7.203)$$

$$U = \int_{t=0}^{\infty} e^{-\rho t} \left(A(0)^{(1-\theta)} e^{(1-\theta)gt} \frac{c(t)^{1-\theta}}{1-\theta} \right) \frac{L(0)e^{nt}}{H} dt, \qquad (7.204)$$

$$U = A(0)^{1-\theta} \frac{L(0)}{H} \int_{t=0}^{\infty} e^{-\rho t} e^{(1-\theta)gt} e^{nt} \frac{c(t)^{1-\theta}}{1-\theta} dt, \qquad (7.205)$$

which can be written in a simplified form:

---

[53] A brief reminder:

    L: total population
    H: number of households
    L/H: number of person composing a household.

$$U = B \int_{t=0}^{\infty} e^{-\beta t} \frac{c(t)^{1-\theta}}{1-\theta} dt, \tag{7.206}$$

where $\beta = \rho - n - (1-\theta)g$ is positive by definition and $B = A(0)^{1-\theta} \frac{L(0)}{H}$. The household must choose the path $c(t)$ which maximizes its total intertemporal utility under the budget constraint. The *Lagrangian* method allows the resolution of this maximization program under constraint (where λ *is the Lagrange multiplier*); it is written:

$$\pounds = U + \lambda \, (budget \; constraint): \tag{7.207}$$

$$B \int_{t=0}^{\infty} e^{-\beta t} \frac{c(t)^{1-\theta}}{1-\theta} dt$$

$$+ \lambda \left(k(0) + \int_{t=0}^{\infty} e^{-R(t)} w(t) e^{(n+g)t} dt + \int_{t=0}^{\infty} e^{-R(t)} c(t) e^{(n+g)t} dt \right).$$

The condition of the first order is:

$$B e^{-\beta t} c(t)^{-\theta} = \lambda e^{-R(t)} e^{(n+g)t}. \tag{7.208}$$

If we use the logarithm, the expression becomes:

$$\log(B) - \beta t - \theta \log(c(t)) = \log(\lambda) - R(t) + (n+g)t. \tag{7.209}$$

Then, by applying the relation $R(t) = \int_{\tau=0}^{t} r(t) d\tau$ and by calculating the derivative at t of the new expression, we have the condition:

$$-\beta - \theta \frac{\dot{c}(t)}{c(t)} = -r(t) + (n+g). \tag{7.210}$$

Then, it is possible to write the following expression:

$$\frac{\dot{c}(t)}{c(t)} = \frac{r(t) - n - g - \beta}{\theta} \tag{7.211}$$

taking into account $\beta = \rho - n - (1-\theta)g$ we write:

$$\frac{\dot{c}(t)}{c(t)} = \frac{r(t) - \rho - \theta g}{\theta}. \tag{7.212}$$

7.8.3 Dynamics and Balanced Growth Path

7.8.3.1 Dynamics of c and k

The dynamics of c for one household is also valid for all the households, since they are considered identical in the model. The following equation corresponds to the

Euler equation of the maximization problem and is also the *growth rate of c which is written*:

$$\frac{\dot{c}(t)}{c(t)} = \frac{r(t) - n - g - \beta}{\theta} = \frac{r(t) - \rho - \theta g}{\theta}, \tag{7.213}$$

since we know that $r(t) = f'(k(t))$, the expression above can be also written:

$$\frac{\dot{c}(t)}{c(t)} = \frac{f'(k(t)) - \rho - \theta g}{\theta} = \gamma_c. \tag{7.214}$$

A brief reminder:

   g: is the *exogenous growth rate* of A.

   n: is the growth rate of the size of households.

   ρ: is the *rate* of *preference* for the *present (positive rate of time preference)*.

   θ: is the *rate of "risk aversion"*.

One observes immediately that, when $f'(k(t)) = \rho + \theta g$, then $\dot{c}(t) = 0$, and one can denote k^* the value for which this equality is satisfied. If $k > k^*$, then $\dot{c} < 0$; by contrast if $k < k^*$, then $\dot{c} > 0$. c grows when $k < k^*$, c decreases when $k > k^*$, and $c = 0$ (i.e. is constant) for $k = k^*$. The *dynamics of the per capita capital* is written:

$$\dot{k}(t) = f(k(t)) - c(t) - (n + g)k(t). \tag{7.215}$$

That means that the evolution of the capital corresponds to the difference between the current investment $f(k) - c$ and the *investment* of the *break-even point* $(n + g)k$, except the depreciation of the capital which is not taken into account in this model. In the figures which follow, one shows the behavior and the trajectory of c, when the variation of the capital is zero, i.e. for $\dot{k} = 0$. One observes clearly, in Fig. 7.11 for $\dot{k} = 0$, the point corresponding to the equality $f(k) - c = (n + g)k$, i.e. the point for which the current investment is equal to the investment of the break-even point.
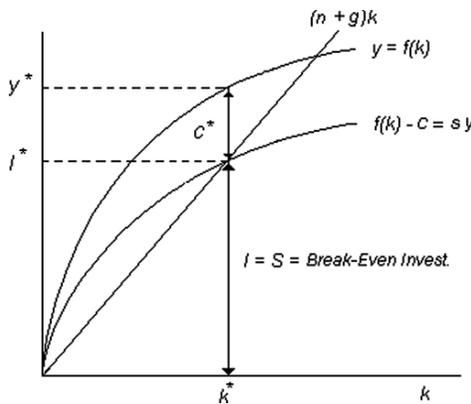
Fig. 7.11 For $\dot{k} = 0 : c = f(k) - (n + g)k$
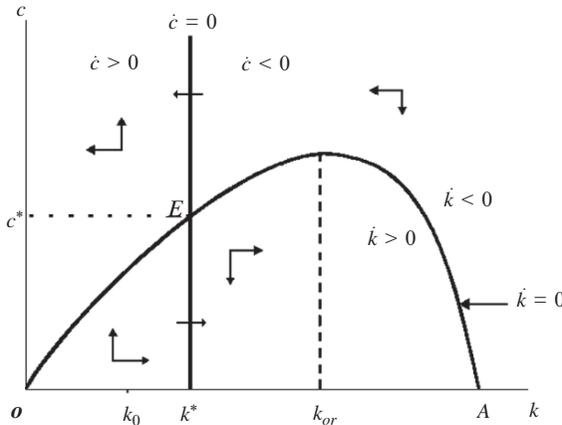
Fig. 7.12 Dynamics of the model

By contrast, outside the border $\dot{k} = 0$, i.e. when $f(k) - c \neq (n+g)k$, two cases are obviously possible: either $\dot{k} > 0$ or $\dot{k} < 0$. And the behavior of c according to these two cases is different when k varies. The dynamics of the model is represented in Fig. 7.12.

When $\dot{k} = 0$, the consumption $c = f(k) - (n+g)k$ is equal to the difference between the *per capita production* curve and the straight line of the investment of the break-even point. However, the value of this consumption increases until $f'(k) = (n+g)$, which corresponds to the golden rule of capital accumulation in the model. Indeed, if $\dot{k} = 0$, then c initially will grow until a maximum, then, it will decrease taking into account the shape of the curve of c (in relation to k) for $\dot{k} = 0$. The maximum of c, for $\dot{k} = 0$, is reached for $f'(k) = (n+g)$ (what corresponds to the golden rule of this model), then, c decreases when k grows (the golden rule is of course different from the golden rule of the Solow model which was written as follows: $f'(k_{GR}^{Solow}) = (n+\delta)$, while knowing that δ is not taken into account, here). When c is higher than the border for which $\dot{k} = 0$, then, k decreases ($\dot{k} < 0$); conversely, when c is lower than the border for which $\dot{k} = 0$, then k increases ($\dot{k} > 0$).

7.8.3.2 The Modified Golden Rule of the Model

Let us recall that k^* is the level of k for which $\dot{c} = 0$ and k_{GR} corresponds to the maximum of the curve $\dot{k} = 0$. In the different graphs, one can observe that $k^* < k_{GR}$. Let us recall that k^* corresponds to $f'(k^*) = \rho + \theta g$ and furthermore, the value k_{GR} corresponds to $f'(k^*) = n + g$. And, since $f''(k) < 0$, k^* is lower than k_{GR} if $\rho + \theta g$ is higher than $n + g$. This condition is verified, because by definition in this model (in order to prevent the divergence of the intertemporal utility), we have $\rho - n - (1 - \theta)g > 0$, i.e. $\rho + \theta g > n + g$. Thus, k^* is lower than k_{GR}. Thus, one observes that k converges towards k^* which is lower than the *per capita* capital of
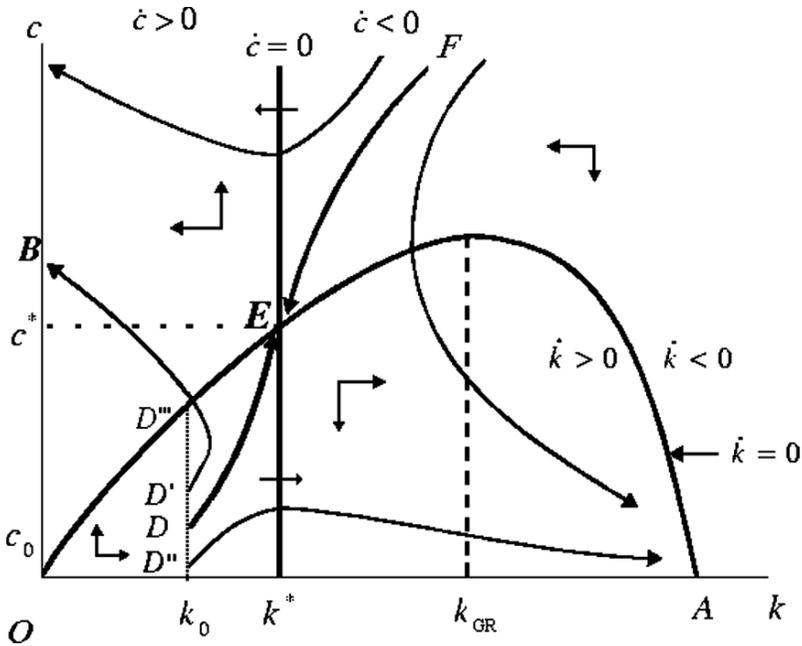
Fig. 7.13 Saddle-point and phase diagram of the model

the golden rule k_{GR}, but since k^* is the *optimal level* of k towards which the model converges, it is called *the stock of capital of the modified golden rule*.

Figure 7.13 gives a diagrammatic representation of the solutions of the model, which have the shape of a saddle-point; the orientation of the flow of solutions is also depicted, as well as the different frontiers delimiting the trajectories, except the curve of the *per capita* production function to make lighter the graph.

Dynamics:
$$\dot{k} = 0 \Leftrightarrow f(k^*) - c^* - (n+g)k^* = 0$$
$$c > c^* \Rightarrow \dot{k}(k^*,c) = f(k^*,c) = f(k^*) - c^* - (n+g)k^* < 0$$
$$\dot{c} = 0 \Leftrightarrow f'(k^*) - \rho - \theta g = 0$$
$$k < k^* \Rightarrow f'(k) > f'(k^*) = \rho + \theta g \Rightarrow \dot{c}(k) > 0.$$

Golden rule:
$$f'(k^*) = \rho + \theta g > n + g = f'(k_{GR})$$
$$k^* < k_{GR} \text{ (since } f''(k) < 0)$$
$$f(k^*) < f(k_{GR}) \Rightarrow c^* < c_{GR}$$

The point E corresponds to the couple which verifies at the same time $\dot{k} = 0$ and $\dot{c} = 0$ (i.e. c and k are constant at such a point).

7.8.3.3 Trajectories (k,c) when t Tends Towards the Infinite and the Choice of c

Conditions of optimization

$$\dot{k}(t) = f(k(t)) - c(t) - (n+g)k(t), \tag{7.216}$$

$$\frac{\dot{c}(t)}{c(t)} = \frac{f'(k(t)) - \rho - \theta g}{\theta}, \tag{7.217}$$

these equation define the behavior of the model and the trajectories will depend on the initial conditions $(k(0), c(0))$. But, if k is given, this is not the case of c which must be selected and defined. Let us study the main behaviors of the system according to the initial values of c, represented in the preceding graph:

- Let us observe the case of the trajectory starting at $c(0)$, which belongs to the curve $\dot{k} = 0$, i.e. at the point D''': this trajectory $(k(t), c(t))$ will move towards the upper part of the (k,c)-plane and towards the left in direction of the axis of c.
- When the trajectory starts in D', i.e. below the border $\dot{k} = 0$, in a zone where $\dot{k} > 0$ and $\dot{c} > 0$, then the couple (k,c) initially will move towards the top and towards the straight line, but starting from the intersection point with the curve $\dot{k} = 0$, it will move towards the point B.
- When the trajectory starts in D'', for a low value of $c(0)$, the couple (k,c) is directed asymptotically towards the axis of k, and, while coming from the zone $\dot{k} > 0$ and $\dot{c} > 0$, crosses $\dot{c} = 0$ towards the zone $\dot{k} > 0$ and $\dot{c} < 0$.
- The point E (where $\dot{c} = 0$, $\dot{k} = 0$, i.e. c et k are constant) represents the stable point of the model.
- Moreover, for a given value $k(0)$, there is a value of c for which the system converges towards the stable point. This value of $c(0)$ is represented by the point D, which is the critical point which allows the convergence towards E.
- Above the point D, the model always undergoes an increase in c stronger than that of k until the intersection of the curve $\dot{k} = 0$, and then the capital decreases whereas the consumption continues to increase.
- By contrast, if the initial consumption $c(0)$ is below D, the trajectory crosses initially $\dot{c} = 0$ towards $\dot{c} < 0$ and until values of k higher than the golden rule; thus, the *real interest rate* $r(t) = f'(k)$ is then lower than the rate of the golden rule $f'(k_{GR}) = n + g$. Consequently, the expression $e^{-R(s)}e^{(n+g)s}$ in the budgetary constraint will increase and, since k also increase, the budget constraint $\lim_{s \to \infty} e^{-R(s)}e^{(n+g)s}k(s) \geq 0$ will tend towards the infinite.
- If a trajectory starts at the critical point D, then the value of k will converge towards k^*, and thus the real interest rate $r(t)$ will converge towards $r(t) = f'(k^*) = \rho + \theta g$; consequently, the expression $e^{-R(s)}e^{(n+g)s}$ in the budget constraint will diminish until $\beta = \rho - n - (1 - \theta)g$ which was introduced in the objective-function and which will lead the budget constraint $\lim_{s \to \infty} e^{-R(s)}e^{(n+g)s}k(s)$ towards its limit value which is zero.

Above, the search of the optimal trajectory was carried out with a fixed k: the problem was to choose $c(0)$. One can extend the approach and select an initial couple $(k(0), c(0))$, then to observe the trajectories under the constraints of the model. The optimal trajectory is called the *saddle-path* (or the axis of the saddle). This saddle-path is represented by D–F branch in the graph; it does not violate the conditions of the model, in such a framework the trajectories converge towards the balanced growth path (BGP), i.e. towards the equilibrium point E: it is a stable branch.

7.8.4 Comments About the Trajectories and Maximization of the Level of Consumption

The trajectory of the model which starts at the point D''', i.e. for a low value of $c(0)$, corresponds to a behavior of the agents of the "thrifty" type (with a rather weak preference for the present and consumption). The trajectory which begins at this point must be completed (if the process is long enough) on the abscissa-axis when the level of consumption will be reduced to zero.

The trajectory of the model which begins at the point D', i.e. for a high value of $c(0)$, corresponds to a typical behavior of preference for the consumption (with a high preference for the present). The trajectory which begins at that point will meet the constraint of the production function, i.e. will have consumed all what is possible to consume, and consequently, the *per capita* capital will decrease until being on the ordinate-axis when the capital is zero. The trajectory which begins at the point D, i.e. for a single value of $c(0)$ (on the vertical k_0 of the graph), will tend towards the point E, which is the intersection point of the borders $\dot{c} = 0$ and $\dot{k} = 0$. The point E is the balanced growth "regime" at constant rate. In fact this point is *reached* only *after an infinite time*, because the more one approaches E, the more $\dot{c}$ and $\dot{k}$ become smaller and *the convergence will take an infinite number of iterations*.

In the growth theory, the golden rule corresponds to conditions on the parameters of the model characterizing an economy, so that the *per capita* consumption is the highest possible. For the maximum level of consumption, the rule explains that the *marginal productivity of the capital* is equal to the *growth rate* which is equal to the *real interest rate*:

$$p_K^m = \gamma_{growth} = \gamma_{real\ interest}. \tag{7.218}$$

The level of *per capita* consumption is thus maximum for a (solution) value of k equal to k_{GR} (which maximizes the level of consumption by definition), which means that the rate of preference for the present ρ is zero. The value of k is that of the golden rule of the model. Thus, when the rate of preference for the present ρ is zero and when the level of *per capita* consumption is maximum, then the real interest rate of the model which was written:

$$r(t) = f'(k) \tag{7.219}$$

for the value k_{GR} which maximizes the *per capita* consumption, *the interest rate becomes:*

$$r_{GR}(t) = f'(k_{GR}). \qquad (7.220)$$

By definition, it is equal to the growth rate, and this same interest rate is equal to the rate of profit in an economy with only one good, thus one can write:

Interest rate = Profit rate = Growth rate

$$\gamma_{interest} = \gamma_{profit} = \gamma_{growth}. \qquad (7.221)$$

In the Ramsey–Cass–Koopmans model, the savings corresponds to a rational arbitrage of households which is made according to the utility that they obtain from their consumption. Thus, a growth path with a consumption which would be permanently lower than other paths cannot be an equilibrium (i.e. a balance): indeed, the households, in such a case, would decrease their savings with an aim of taking a path which would increase their consumption.

In the preceding sections, we noticed that the stock of *per capita* capital k^* is lower than the stock k_{GR} of the golden rule, which means that the model does not converge towards the balanced growth path which in the absolute would maximize the consumption permanently. Trajectories beginning with values k other than those of k^*, either will diverge excluding the balanced paths or will converge towards (k^*, c^*) because positioned on the axis of the saddle. *Thus k^*, which is the optimal level of the per capita capital towards which the economy of the model converges, is called the stock of capital of the modified golden rule.*

When g the growth rate of A is equal to zero, the marginal productivities of k^* and k_{GR} is written respectively $f'(k^*) = \rho$, $f'(k_{GR}) = n$ and, if we take into account the initial constraint $\rho - n - (1 - \theta)g > 0$, we will observe that $\rho > n$. That means that the real interest rate r^*, for which there is a BGP optimum, is equal to the rate of preference for the present which is higher than the growth rate of the size of households.

7.8.5 Equilibria and Instability of Solutions

Such models highlight complex solutions in which stable and unstable trajectories but also balances and unbalances coexist. The former approaches associated stability with balance, as in the Solow model, but also associated instability with imbalance as in the Harrod model.

The possible behaviors, in the model presented here, are mainly unstable and the stability has the appearance of an exception, like the axis of the saddle reached for restricted initial values of $(k(0), c(0))$. Outside this trajectory, there are only divergence, instability and unbalance. Thus, one attaches a growing interest to this type of model and to their solutions which have the form of the saddle-point. The trajectory of the steady-equilibrium belongs to the set of the possible trajectories of the model which are mainly unstable. Nevertheless, the fact that the "agents"

have the choice, characterizes the optimal growth model presented previously, i.e. they can arbitrate between consumption and savings. This constant possibility of arbitrage makes more complex the field of the behaviors of the model. The *optimal trajectory is* "plunged" or "immersed" *in a field of instability* which dominates the model. The rareness of the optima (maybe their absence) is in opposition to the structure of many former models, such as the basic Solow model, whose conception aimed at the convergence towards stable states, which offered little choice to the agents, except the faculty of substitution of the production factors K and L.[54] Thus, the arbitrage possibility for the actors of a model, leads to accept a sort of prevalence of the instability and, at the same time, leads to accept the relation between the accuracy or fidelity of a model and the *prevalence of the instability*. Moreover, it is necessary to state the fact that, *within the framework of the optimal growth models*, one can *generate (without external shock) cyclic or aperiodic growth paths*. This is what happens in the optimal growth model presented in 1990 by M. Boldrin and M. Woodford, which *generates cyclic and chaotic endogenous fluctuations*.

7.8.6 Endogenous Growth Without Externality, and Saddle Point

The Cass–Koopmans–Ramsey model previously presented depicts an economy with *complex behaviors* whose saddle-point solutions *converge* towards a *balanced growth path* only for particular *initial values* of k and c. It is the principle of the optimal growth model, in which the individual considers that it is not sufficient to save a constant part of its income, but carries out an intertemporal choice based on its function of utility and its budget constraint. *The saddle-point solution is the consequence of the opposite signs of the eigenvalues of the Jacobian matrix (of the system).* Nevertheless, one observes also solutions of the saddle-point type for other growth model kinds. Indeed, while maintaining the framework of the Solow model, *certain models show endogenous growth modes without externality*. In such cases, the growth is primarily due to the "learning", which considers that it is possible to *create labor factor in the form of human capital*. The choice in this type of model is carried out between an investment in human capital and an investment in training capital. One associates the production function of material goods with a production function of human capital, and the model is *then represented by a system of differential equations of order 2*, whereas the Solow model is of order 1. The dynamics generated by the model is complex, and the trajectory depends on the initial conditions of the model. In such a case, one observes a solution which has the saddle-point shape for this dynamics and a major role is attached to the *axis of the saddle* which is the trajectory that the economy is supposed to choose. In this type of model, the generated dynamics is that of a saddle-point with its characteristic instability.

---

[54] Ref: M. Kurz on the stability in the Solow model (Kurz 1968).

7.9 Day Model (1982): Logistic Function, Periodic and Chaotic Behaviors

The model of R. Day allows to introduce the logistic function into a model whose structure is a continuation of the Solow model. The logistic function which is a simple nonlinear function is of a great interest in order to study the diversity of behaviors of a model, and in particular the chaotic regimes. The Day model is a capital accumulation model written *in discrete time. This characteristic makes it possible to avoid the concept which states that in order to generate chaotic behaviors in continuous time it is necessary to have three equations (ref. to Poincaré: "The three body problem"). Thus, the Day model with its single discrete equation* makes it possible to *obtain behaviors of chaotic nature.*

7.9.1 The Model

On the labor market, where the supply is equal to the demand, one can write the following equations of the model:

$$L_{t+1} = (1+n)L_t \quad \text{and} \quad L_t = (1+n)^t L_0, \tag{7.222}$$

where the growth rate of the population is noted n and L_t is the labor demand at t:

$$S_t = sY_t. \tag{7.223}$$

The investment is immediately adapted to the saving capacity and the stock of capital lasts only one period:

$$I_{t+1} = K_{t+1} = S_t = sY_t. \tag{7.224}$$

The production function[55] (without technological progress), is written:

$$Y_t = F(K_t, L_t). \tag{7.225}$$

But also:

$$Y_t = F(K_t, L_t) = L_t F(K_t/L_t, L_t) = L_t f(k_t). \tag{7.226}$$

One can write the recurrence equation of the capital:

$$k_{t+1} = \frac{K_{t+1}}{L_{t+1}} = \frac{sY_t}{(1+n)L_t} = \frac{s}{1+n} f(k_t) \tag{7.227}$$

therefore

$$k_{t+1} = \frac{s}{1+n} f(k_t). \tag{7.228}$$

---

[55] F is homogeneous of degree 1 and satisfies the Inada conditions.

From this writing, one deduces the convergence of k towards a *fixed point* k^*:

$$k^* = \frac{s}{1+n}f(k^*) \quad \text{or} \quad f(k^*) = \frac{1+n}{s}k^*. \tag{7.229}$$

The fixed-point can be located at the intersection of the curves $k_{t+1} = f(k_t)$ and $k_{t+1} = \frac{1+n}{s}k_t$. The convergence towards k^* can be carried out in two different ways, according to the initial value $k(0)$, which can be either higher or lower than k^*. If a Cobb Douglas function is selected for $F(K_t, L_t)$, we obtain:

$$Y_t = F(K_t, L_t) = BK^\beta L^{1-\beta}. \tag{7.230}$$

And it comes:

$$f(k_t) = Bk_t^\beta. \tag{7.231}$$

And by replacing in $k_{t+1} = \frac{s}{1+n}f(k_t)$ the recurrence equation for k, it comes:

$$k_{t+1} = \frac{s}{1+n}Bk_t^\beta. \tag{7.232}$$

This recurrence equation is increasing (β is positive) and it converges towards a permanent regime of which the *capital factor* or the *capital intensity* is written as follows:

$$k^* = \frac{s}{1+n}Bk^{*\beta}, \quad \text{i.e. } (k^*)^{1-\beta} = \frac{sB}{1+n}, \tag{7.233}$$

$$k^* = \left(\frac{sB}{1+n}\right)^{\frac{1}{1-\beta}}. \tag{7.234}$$

7.9.2 From Dynamics of Capital Towards Logistic Function

At this stage, one admits that the economy undergoes a "pollution effect": indeed, when the *capital intensity* increases, the production is limited by a factor $(m-k)^\gamma$ which decreases the productivity of the model. Consequently, the Cobb–Douglas function evolves:

$$g(k_t) = (m-k)^\gamma Bk_t^\beta \tag{7.235}$$

with $m > k$ and $\gamma > 0$. the recurrence equation which corresponds to the dynamics of k, is written:

$$k_{t+1} = \frac{s}{1+n}(m-k_t)^\gamma Bk_t^\beta. \tag{7.236}$$

It is increasing for $k = [0, k^* = \frac{\beta}{\beta+\gamma}m]$ and decreasing for $k =]k^*, m]$. For k^* there is thus a maximum:

$$\frac{sB}{1+n}\beta^\beta\gamma^\gamma\left(\frac{m}{\beta+\gamma}\right)^{\beta+\gamma} = g(k^*). \tag{7.237}$$

When $\beta = \gamma = m = 1$ the equation $k_{t+1} = \dfrac{s}{1+n}(m-k_t)^\gamma B k_t^\beta$ becomes:

$$k_{t+1} = \frac{sB}{1+n}k_t(1-k_t), \qquad (7.238)$$

or:

$$k_{t+1} = ak_t(1-k_t). \qquad (7.239)$$

This equation is similar to the logistic equation: it is the dynamics of k and it can generate chaotic behaviors. Thus, to know its behavior, one can refer to the logistic equation. The factor "a" in the logistic equation corresponds to the fertility rate and usually varies between zero and four, the chaotic regime occurring from the value $a \simeq 3.56$.

7.9.3 Periodic and Chaotic Solutions of the Dynamics of k

The dynamics of k depends on $a = sB/(1+n)$; the values of n and s being *exogenous*, only B in this expression is variable, whereas the values of k_t were normalized. Thus, the values of k will depend on B:

$$
\begin{aligned}
&0 < B < (1+n)/s: &&k \to 0\\
&(1+n)/s < B < 3(1+n)/s: &&k \to 1 - 1/a\\
&3(1+n)/s < B < 3.449(1+n)/s: &&k \to \textit{Period-2}\\
&3.449(1+n)/s < B < 3.544(1+n)/s: &k \to \textit{Period-4}\\
&3.544(1+n)/s < B < 3.57(1+n)/s: &k \to \textit{Period-n}\\
&3.57(1+n)/s < B < 4(1+n)/s: &&k \to \textit{Chaotic regime}
\end{aligned}
$$

One observes about B that between 0 and 3.57, k converges towards periodic values of which the number of period regularly increases. Starting from 3.57, one gets into a chaotic regime without periodicity except punctually some windows of value for B, where there are re-emergences of periodic behaviors. The dynamics of k are thus either periodic (i.e. cyclic, from period-1 to period-n) or aperiodic (i.e. chaotic).

7.10 Day–Lin Model (1992): Imperfect Information and Strange Attractor

At the end of the second world war, the planning questions and the strategies of growth and investment were in the forefront. In such a framework, the optimal growth models were significant. Today, the environment is different. The role of planner or coordinator of the State is weakened and because of the "incertitude" existing on the markets, the firms do not find necessarily the indicators which are indispensable to the coherence of their choices and arbitrages. Consequently, the accumulation processes of the capital are not, within this framework of imperfect

market, the result of perfectly rational and efficient choices and anticipations. Consequently, we have to face the temporal incoherence of decisions and to dynamical inefficiency situations. It is also in this context that we can explain the chaotic evolutions of the growth.

7.10.1 Imperfect Information, Price Uncertainty and Adaptive Expectations

In this model, the *State* has the role of a *planner* and must decide *investment* flows for different periods in order to *maximize* the *discounted sum* of the *utility* of *the consumption at each period*. In such a case, unlike what occurs in the initial optimal growth model, one considers that the *consumption choices* are *carried out* by *an infinite number of individual agents,* which succeed one another in the course of time and have a short two-periods lifespan, in relation to the cycle of the economy. Moreover, we postulate an imperfect information because of an increase in the uncertainty which is justified by resorting to the last developments as regards planning known as strategic. Within this framework, the agents consider that the price announcement or display by the State, are not "certain" and the anticipations or expectations of agents take into account this uncertainty, thus these expectations are of *adaptive* nature.

7.10.1.1 The Equations of the Model

The planner must carry out, at each period, the sharing between consumption and savings, in order to maximize a collective utility function. The maximization of the utility is written:

$$\max \sum_{i=0}^{\infty} \gamma^i u(c_{t+i}) \tag{7.240}$$

with $u'(c) > 0, u''(c) < 0$; $\gamma = 0, \ldots, 1$: the *discount factor* expresses the preference for the present of the decision maker. The maximization of the utility is carried out under constraints and in particular that of the equality between savings and investment:

$$I = S: \tag{7.241}$$
$$k_{t+1} = y_t - c_t. \tag{7.242}$$

Like in the Day model, one supposes that the capital is entirely renewed at each period. The investment at t corresponds to the stock of capital at $t+1$. And the savings corresponds to the part of the income which is not consumed at t, as we can observe it in the equation above. Furthermore the *distribution equation* is written:

$$y_{t+1} = w_{t+1}^a \times l_{t+1} + r_{t+1}^a \times k_{t+1} \quad \text{with}: l_{t+1} = 1. \tag{7.243}$$

This equation shows at $t+1$, the national distribution of income, between wages and profits. In this fundamental equation, the prices of labor and capital factors are anticipated and the principle of anticipation is the following:

$$w_{t+1}^a = w_t : \text{Wages anticipation at } t+1,$$
$$r_{t+1}^a = r_t^a + \mu(r_t - r_t^a) : \text{Interest rate anticipation at } t+1.$$

In this model the *anticipation of the interest rate*, i.e. the anticipation of the price of the capital factor, *is adaptive*, whereas the anticipation of the wages at $t+1$ is very simple: it is carried out by means of existing wages at t; This is called *naive anticipation*. The consumption can be written as follows:

$$c_{t+1} = y_{t+1} - k_{t+1}. \tag{7.244}$$

The equations $k_{t+1} = y_t - c_t$ and $c_{t+1} = y_{t+1} - k_{t+1}$ express two different constraints. Indeed, $k_{t+1} = y_t - c_t$ defines the *ex-ante*[56] equilibrium between savings and investment, it is the equilibrium on the market of goods, whereas the equation $c_{t+1} = y_{t+1} - k_{t+1}$ defines the *ex-post*[57] equilibrium between savings and investment. The coexistence of the two equations expresses an equilibrium-state on the market of goods for which the resources of the economy are entirely used. By replacing in the equation ex-post, the expression of $y_{t+1} = w_{t+1}^a l_{t+1} + r_{t+1}^a k_{t+1}$, it comes:

$$c_{t+1} = w_t + (r_{t+1}^a - 1)k_{t+1}. \tag{7.245}$$

Then, by using $k_{t+1} = y_t - c_t$, one obtains:

$$c_{t+1} = w_t + \rho_{t+1}^a (y_t - c_t) \quad \text{with } \rho_{t+1}^a = (r_{t+1}^a - 1). \tag{7.246}$$

We known that the lifespan of individual agents is short, i.e. two periods and one calls a "cohort" the individuals who live the same two periods. The optimal sharing carried out by a cohort, or by its representative agent, is carried out by maximizing $u(c_t) + \gamma u(c_{t+1})$ under the saturation of the constraint $f(k_t) - c_t \geq 0$ with k_t and k_{t+1} fixed.

7.10.1.2 The Accumulation Dynamics According to Price Expectations

The optimization program is carried out by means of a Lagrangian, which is written:

$$L(c_t) = u(c_t) + \gamma u(c_{t+1}) + \lambda_t (y_t - c_t). \tag{7.247}$$

---

[56] **Definition** *(Ex-ante):* ex-ante is Latin for "beforehand". In models where there is uncertainty that is resolved during the course of events, the ex-antes values (e.g. of expected gain) are those that are calculated in advance of the resolution of uncertainty.

[57] **Definition** *(Ex-Post):* Latin for "after the fact". In models where there is uncertainty that is resolved during the course of events, the ex-post values (e.g. of expected gain) are those that are calculated after the uncertainty has been resolved.

The purpose is to maximize the utility of the consumption of the current period but also that of the following period, under the saturated constraint $f(k_t) - c_t \geq 0$. Replacing by the value of $c_{t+1} = w_t + \rho^a_{t+1}(y_t - c_t)$, then:

$$L(c_t) = u(c_t) + \gamma u\left(w_t + \rho^a_{t+1}(y_t - c_t)\right) + \lambda_t(y_t - c_t). \qquad (7.248)$$

The first order condition is written:

$$u'(c_t) = \frac{1}{\gamma}\left(\frac{w_t}{\rho^a_{t+1}} + y_t\right). \qquad (7.249)$$

If one poses $u(c_t) = \log(c_t)$ and also

$$f(k_t) = k_t^\beta \qquad \text{where } 0 < \beta < 1, \qquad (7.250)$$

then, the expression of c_t is written:

$$c_t = \frac{1}{1+\gamma}\left(k_t^\beta + \frac{(1-\beta)k_t^\beta}{(r^a_{t+1} - 1)}\right). \qquad (7.251)$$

The following system defines the accumulation dynamics of the capital:

$$k_{t+1} = y_t - c_t = k_t^\beta - \frac{1}{1+\gamma}\left(k_t^\beta + \frac{(1-\beta)k_t^\beta}{(r^a_{t+1} - 1)}\right), \qquad (7.252)$$

$$r^a_{t+1} = r^a_t + \mu(r_t - r^a_t), \qquad \text{with } r_t = f'(k_t). \qquad (7.253)$$

We observe that the capital accumulation explicitly depends on expectations. Thus, the optimal sharing by the agents is carried out according to their own price anticipations. The couples (r^a_t, k_t) along t constitute the growth path of the model.

7.10.2 Chaotic Growth and Intertemporal Non-Optimality

7.10.2.1 Subharmonic Cascade of the Capital, and Strange Attractor

The dynamics of the capital accumulation in the long term shows the presence of "deterministic chaos", like the behaviors of the logistic function and the Day model (1982). By contrast, only the construction of the dynamics of the capital using at least two recurrence equations, like in the model of Day and Lin (1992), creates the conditions of existence of a strange attractor. Indeed, such an attractor can exist only in systems with at least three variables, but their fractal dimension can only be *non-integer* and here lower than three but higher than two. This type of attractor was defined by Ruelle and Takens and one knows that the attractor is named *strange* if:

- The attractor is of *zero volume* in the phase space.
- The dimension D of the attractor is *non-integer*, i.e. fractal $2 < D < n$, where n is the dimension of the phase space.
- There is *sensitive dependence on initial conditions (SDIC)*, i.e. two trajectories (of an attractor) initially very close will diverge.

The definition of an attractor of this type is described in the part I, one will simply point out that their existence allows to identify the *existence* of dynamic *torus* and *self-similarity* at various scales of analysis. Moreover, the bifurcation diagram of the model which shows subharmonic cascades, is constructed in the (γ, k)-plane of the parameter of the preference for the present (i.e. the time preference) and of the stock of capital. When the preference for the present increases, then the *period-doublings* occur and quickly, starting from a critical-point $\gamma_{critical}$ the behavior of the capital become chaotic. That means that when we are inside the chaotic mode, the capital accumulation by the agents can be very different from a period to another. *The growth dynamics of the model depends on the dynamics of the capital which can adopt chaotic behaviors.* And, in the case of a chaotic growth, it is not possible to say there is an intertemporal optimality. Thus, the conclusions about this model are very different from those of the Ramsey optimal growth model, which let the agents choose to consume more at certain periods to increase their standard of living than the following periods. Indeed in the last case, the *marginal rate of substitution* $(MRS_{t+1,t})$ between present consumption and future consumption was equal to the *marginal rate of technical substitution*[58] $(MRTS_{t+1,t})$, along the path of optimal consumption of the Ramsey model:

$$\frac{\frac{1}{1+\rho}u'(c_{t+1})}{u'(c_t)} = \frac{1+n}{1+f'(k)}, \tag{7.254}$$

$$MRS_{t+1,t} \Leftrightarrow MRTS_{t+1,t}. \tag{7.255}$$

These rates are independent of the period where the choice intervenes, which is not the case of the Day–Lin model, where these rates are different according to time because of expectations. One will simply state that a chaotic dynamics does not seem to allow the respect of the transversality condition of the optimal growth models. And, if this condition is necessary in a strict sense, the *chaotic growth paths* are

---

[58] *The marginal rate of technical substitution (MRTS)* usually is the increase in productivity a company experiences when it substitutes on unit of labor input – i.e. an hour worked by a factory worker – for one unit of capital (i.e. a machine). A positive MRTS indicates that it is advantageous for a company to make this substitution, and a negative MRTS implies that the company would drop in productivity if it did this. The MRTS can also been seen as the slope of an Isoquant at the point in question.

not the solution of the optimization program. The equilibrium of the optimal growth model maximized the well-being of the representative household whereas in this model (like in the Diamond model) to define a suitable measure of the well-being is more difficult: we are thus *immersed in a dynamic inefficiency context in uncertainty situations*.

7.11 The Instability of Stock Markets, and Random Processes: Model of Portfolio Choice

In such a model, the agents choose the destination of their savings, and can arbitrate between savings invested directly in the production process and savings assigned to a "portfolio", in which there is a sub-arbitrage between the "money" and "capital-goods". The construction of the model is however different.[59] Indeed, the agents must arbitrate between the holding of goods and holding of money. The quantity of money in the economy, denoted $M(t)$, grows at an exogenous constant growth rate θ. p is the price of a unit of good and n is the growth rate of the population. The set of agents is in possession of the total quantity of money in the economy. The set of assets of an agent corresponds to the sum of goods and money in possession of the agent. The "real wealth" of agents corresponds to the sum of the value of the stock of capital (i.e. the good) with the value of the stock of money expressed in terms of price of the unique good (M/p: *Real Money Balances,*[60] also called real cash balances). Thus, the real wealth of agents is written $W = K + M/p$, and *per capita:*

$$W/L = K/L + M/pL = k + m. \tag{7.256}$$

One considers that the savings in the model can not be entirely used for the investment in the production process, but can be invested in a "portfolio" inside which one can arbitrate between the money (whose nominal income is zero) and the holding of "capital goods" whose remuneration is r. This remuneration corresponds to the *interest rate* written: $r = f'(k) + E(\dot{p}/p)$, i.e. the sum of the *marginal productivity of the capital* $f'(k)$ plus the *anticipated inflation*, which is written as the *expectation* of the *growth rate* of the price of the good $E(\dot{p}/p)$. One postulates that the price expectations are perfect, as within the *neo-classical* framework of *rational expectations*; consequently, the anticipated price is equal to the effective price which is noted ω. Thus, the remuneration of the holding of "capital goods" is written:

$$r = f'(k) + \omega. \tag{7.257}$$

---

[59] Ref: Tobin (1965); and in Burmeister (1980). And Abraham-Frois and Berrebi (1995).
[60] See Pigou.

7.11.1 Dynamics of Accumulation of k and m

7.11.1.1 The Case $\dot{m}=0$

The accumulation dynamics of the per capita variables k and m will determine the trajectories of the model. One will observe the behavior of the *per capita real money balances* $(m = M/pL)$ for $\dot{m} = 0$. The dynamics of m is written:

$$\dot{m}/m = \dot{M}/M - \dot{p}/p - \dot{L}/L \tag{7.258}$$

that we can write:

$$\dot{m}/m = \theta - \omega - n. \tag{7.259}$$

The money demand can be formalized in the following way: $m = d(k,r)$. The interest rate leads the arbitrages between *per capita* capital and money,[61] it is written as a function of these two factors $r = \Phi(k,m)$. Thus, one poses:

$$\omega = r - f'(k) = \Phi(k,m) - f'(k). \tag{7.260}$$

From this equation, one can rewrite dynamics $\dot{m}/m$ as follows:

$$\dot{m}/m = (\theta - \Phi(k,m) + f'(k) - n) \tag{7.261}$$

or, also:

$$\dot{m} = (\theta - \Phi(k,m) + f'(k) - n)m. \tag{7.262}$$

If we wish that $\dot{m} = 0$, m being positive, it is necessary that $(\theta - \Phi(k,m) + f'(k) - n) = 0$. Thus, we write $H(k,m) = f'(k) - \Phi(k,m) + \theta - n = 0$. This equation defines *the increasing trajectory of the couple* (k,m).[62]

7.11.1.2 The Case $\dot{k}=0$

Let us consider the *income of agents*, it is constituted of the income coming from the production, but also by the *increase* in the *value of real money balances*. This total income is written:

$$\mathbf{Y} = Y + (\theta - \omega)M/p. \tag{7.263}$$

The impact of the rise of prices was removed from the total income of agents. Taking into account $k = K/L$, we obtain the following rough equation:

$$\dot{k}/k = \dot{K}/K - \dot{L}/L. \tag{7.264}$$

---

[61] With $\Phi'_k > 0$ and $\Phi'_m < 0$.

[62] Ref: in accordance with the "theorem of the implicit functions", we obtain $dm/dk = -H'_k/H'_m$. Since it is possible to write $H'_m = -\Phi'_m > 0$ and $H'_k = f''(k) - \Phi'_k < 0$, with the other assumptions of the model, we obtain $dm/dk > 0$.

One knows that $\dot{L}/L = n$ (according to the postulates of the model); the dynamics of the capital accumulation is defined by the investment of a fraction of the income Y which is not consumed. Then, we obtain:

$$\dot{k}/k = \frac{Y - (1-s)((\theta - \omega)M/p)}{K} - n, \tag{7.265}$$

$$\dot{k}/k = \frac{sy - (1-s)(\theta - \omega)m}{k} - n. \tag{7.266}$$

The dynamics of the *per capita capital accumulation* is thus written:

$$\dot{k} = sf(k) - (1-s)(\theta - \omega)m - nk. \tag{7.267}$$

"A balanced growth at constant rate implies a constant per capita capital, i.e. $\dot{k} = 0$", and if the rate ω is replaced by its expression given previously, one obtains the expression of m as follows:

$$m = \frac{sf(k) - nk}{(1-s)(\theta - \Phi(k,m) + f'(k))}. \tag{7.268}$$

This equation expresses the *behavior* of *real money balances* according to k for $\dot{k} = 0$. This trajectory is concave in relation to the abscissa axis and reached its maximum for:

$$f'(k) = n/s. \tag{7.269}$$

When the real money balances are equal to zero at the intersection of the curve with the *abscissa axis* in k_s, all the savings is used for the investment, and one comes back to the conditions of the Solow model.

7.11.2 The Solution is a Saddle-Point

Figure 7.14 in the (k,m)-plane represents the *two dynamic equilibrium conditions* $\dot{m} = 0$ and $\dot{k} = 0$ of this model. The *equilibrium point* $E : (k^*, m^*)$ is a *saddle point* (Abraham-Frois and Berrebi 1995, p. 35).

The dynamic equations of the model, are written:

$$\dot{m} = (\theta - \Phi(k,m) + f'(k) - n)m,$$
$$\dot{k} = sf(k) - (1-s)(\theta - \omega)m - nk. \tag{7.270}$$

The Jacobian matrix is written:

$$\mathbf{J} = \begin{bmatrix} \partial \dot{m}/\partial m & \partial \dot{m}/\partial k \\ \partial \dot{k}/\partial m & \partial \dot{k}/\partial k \end{bmatrix}. \tag{7.271}$$

Let us study the components of this matrix. Firstly, we know that $\partial \dot{m}/\partial m = \theta - n - \Phi(k,m) + f'(k) - \cdot m\Phi'_m$; then, since $\dot{m} = 0$, taking into account the dynamic
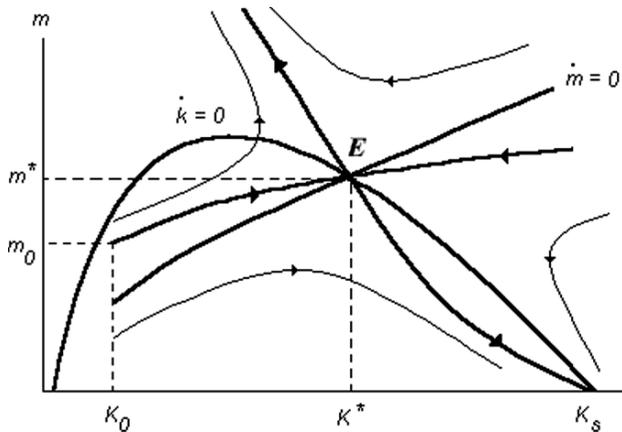
Fig. 7.14 Dynamic equilibrium in Tobin model and saddle point

equation of m, $\theta - n - \Phi(k,m) + f'(k) = 0$ and knowing that $m > 0$ and $\Phi'_m < 0$, we obtain:

$$\partial \dot{m}/\partial m = -m\Phi'_m > 0. \tag{7.272}$$

Knowing that $\Phi'_k > 0$ and $f''(k) < 0$, then:

$$\partial \dot{m}/\partial k = m(-\Phi'_k + f''(k)) < 0. \tag{7.273}$$

Then, taking into account the $\omega = \Phi(k,m) - f'(k)$, we obtain: $\partial \dot{k}/\partial m = -(1 - s)\theta + (1 - s)\omega + (1 - s)m\Phi'_m = (1 - s)(m\Phi'_m - n)$. Then, knowing that $\Phi'_m < 0$ and being at the *stationary point* (k^*, m^*) we know that $\dot{m} = 0$ and thus $\dot{m}/m = \theta - \omega - n = 0$, which is thus written also $n = \theta - \omega$; then we obtain:

$$\partial \dot{k}/\partial m < 0. \tag{7.274}$$

Lastly, from $\omega = r - f'(k) = \Phi(k,m) - f'(k)$, we can write:

$$\partial \dot{k}/\partial k = sf'(K) - n - (1 - s)m(f''(k) - \Phi'_k) - (1 - s)(\theta - \omega)dm/dk. \tag{7.275}$$

At the stationary point (k^*, m^*), we know that $\dot{k} = 0$ and the expression of m is $m = [sf(k) - nk]/[(1 - s)(\theta - \Phi(k,m) + f'(k))]$, which, due to the equation $\omega = \Phi(k,m) - f'(k)$, becomes $m = [sf(k) - nk]/[(1 - s)(\theta - \omega)]$. This expression of m makes it possible to provide the expression of dm/dk, that we replace in the initial equation of $\partial \dot{k}/\partial k$, and we obtain $\partial \dot{k}/\partial k = -(1 - s)m(f''(k) - \Phi'(k)) + [(sf(k) - nk)/(\theta - \omega)][f''(k) - \Phi'_k]$ which disappears, taking into account the value of $m = [sf(k) - nk]/[(1 - s)(\theta - \omega)]$, which was highlighted above. Thus, we obtain:

$$\partial \dot{k}/\partial k = 0. \tag{7.276}$$

The study of the sign of the components of the Jacobian matrix allows to write that:

$$\mathbf{J} = \begin{bmatrix} + & - \\ - & 0 \end{bmatrix}. \tag{7.277}$$

The determinant of $\mathbf{J}$ is negative for the *equilibrium point* (k^*, m^*), this solution couple is thus a saddle-point. In this model of arbitrage between "investment" and "portfolio", we have to face behaviors showing a saddle-point, where the only trajectory leading to the equilibrium, is the axis of the saddle, i.e. any initial couple (k_0, m_0) will tend towards the point E.

7.12 Goodwin's Cyclical Growth Model

In this model the assumption of the full employment of labor is no more assumed. The production function has fixed coefficients, constant for the capital and with technical progress for the labor (labor productivity increased at a fixed rate):

$$K(t) = (1/h)Y(t) \quad \text{and} \quad L(t)e^{\beta t} = \alpha Y(t), \ h > 0, \alpha > 0, \beta > 0. \tag{7.278}$$

Labor supply $N(t)$ is exogenous and increases at a constant rate $v : \dot{N}(t)/N(t) = v$. If the real wage rate $w(t)$ does not exceed the labor productivity $(w(t) \le e^{\beta t}/\alpha)$, the labor demand is determined by the full employment of capital: $L^d(t) = \alpha h e^{-\beta t} K(t)$, and the current employment is: $L(t) = L^d(t)$ if the labor supply is sufficient $(\alpha h e^{-\beta t} K(t) \le N(t))$. The investment $I(t) = K(t)$ is equal to profits $Y(t) - w(t)L(t)$; and the variation rate of the real wage: $\dot{w}(t)/w(t) = cx(t) - \gamma$ is an increasing affine function of the demanded employment rate: $x(t) = L^d(t)/N(t)$, with $c > 0, \gamma > 0$. Note that $z(t) = w(t)L(t)/Y(t)$ the part of wages in the national income. The evolution of the economy is studied in the case: $L(t) = L^d(t) = \alpha h e^{-\beta t} K(t)$. Therefore, we have in this case:

$$\dot{K}(t) = Y(t) - w(t)L(t) = (1 - z(t))hK(t), \tag{7.279}$$

$$x(t) = L(t)/N(t) = \alpha h e^{-\beta t} K(t)/N(t), \text{ and } z(t) = \alpha e^{-\beta t} w(t), \tag{7.280}$$

$$\frac{\dot{x}(t)}{x(t)} = \frac{\dot{L}(t)}{L(t)} - \frac{\dot{N}(t)}{N(t)} = -\beta + \frac{\dot{K}(t)}{K(t)} - v = h(1 - z(t)) - \beta - v, \tag{7.281}$$

$$\frac{\dot{z}(t)}{z(t)} = -\beta + \frac{\dot{w}(t)}{w(t)} = cx(t) - \beta - \gamma. \tag{7.282}$$

Then we obtain the *differential system* at x and z:

$$\dot{x}(t) = hx(t)(b - z(t)) \quad \text{and} \quad \dot{z}(t) = cz(t)(x(t) - a), \tag{7.283}$$

where we have posed: $\alpha = (\beta + \gamma)/c$ and $b = 1 - (\beta + v)/h$. Conversely, to any solution of the above differential system that verifies $0 < x(t) \le 1$ and $0 < z(t) \le 1$ corresponds an evolution of the economy where the wage rate $w(t) = z(t)e^{-\beta t}/\alpha$ is

lower or equal to the productivity $e^{-\beta t}/\alpha$ of the labor, and where, the demand of labor $L^d(t) = x(t)N(t)$ does not exceed the supply $N(t)$, the effective employment is $L(t) = x(t)N(t) = \alpha h e^{-\beta t}K(t)$ (this equality is verified in $t = 0$ and the derivatives are equal). (1) *Study of the equilibrium:* Consider the evolution in $E =]0,1[\times]0,1[$ resulting from the above differential system. This evolution has a unique equilibrium $(x^*, z^*) = (a,b)$ in E if and only if we have: $\beta + \gamma < c$ and $\beta + v < h$. These conditions on the parameters are verified by the empirical data and we made these assumptions. The Jacobian matrix at (a,b) of the differential system above is $\begin{pmatrix} 0 & -ha \\ cb & 0 \end{pmatrix}$. Its eigenvalues $\pm i\sqrt{habc}$ have zero as real parts, and it is not possible to conclude on the stability of the equilibrium. (2) *Study of the evolution*: For any solution of the differential system in E, we have

$$(b - z(t))\frac{\dot{z}(t)}{z(t)} = (b - z(t))cz(t)(x(t) - a) = \frac{c}{h}(x(t) - a)\frac{\dot{x}(t)}{x(t)}. \tag{7.284}$$

However, $b\dot{z}(t)/z(t) - \dot{z}(t)$ is the derivative of $f(z(t))$, where $f(z) = b\ln z - z$, and $(c/h)(\dot{x}(t) - a\dot{x}(t)/x(t))$ is the derivative of $g(x(t))$, where $g(x) = (c/h)(x - a\ln x)$. Consequently the solution of the differential system in E such that $x(0) = x_0$ and $z(0) = z_0$ verifies on any interval where the solution exists:

$$f(z(t)) = g(x(t)) + d_0, \quad \text{with} \quad d_0 = f(z_0) + g(x_0). \tag{7.285}$$

(3) *Study of the curve:* $f(z) = g(x) + d_0$. The function $g(x)$ whose derivative is: $(c/h)(1 - a/x)$ (decreasing on $]0,a]$ and increasing on $[a,+\infty]$); and $f(z)$ whose derivative is: $b/z - 1$ (increasing on $]0,b]$ and decreasing on $[b,+\infty]$). There exists two pints $\bar{x}_1$ and $\bar{x}_2$ solutions of $g(\bar{x}_1) = f(b) - d_0 = g(\bar{x}_2)$ such that $0 < \bar{x}_1 < a < \bar{x}_2$, if $(x_0, z_0) \neq (a,b)$ (because we have: $f(b) - d_0 - g(a) > f(z_0) - d_0 - g(x_0) = 0$).

If $x \notin [\bar{x}_1, \bar{x}_2]$, we have: $g(x) > f(b) - d_0 \geq f(z) - d_0$ for any $z \in \mathbb{R}_+^*$, and there is no number $z > 0$ such that $f(z) = g(x) + d_0$.
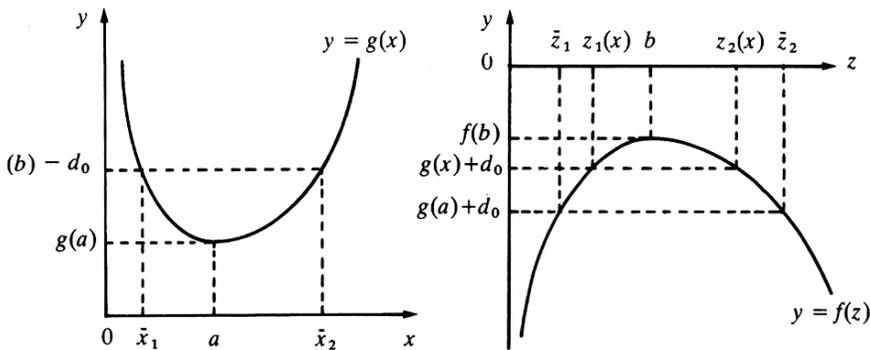
If $x \in [\bar{x}_1, \bar{x}_2]$, we have: $g(x) + d_0 < f(b)$, and there exists two solutions $z_1(x)$ and $z_2(x)$ of: $f(z) = g(x) + d_0$, such that $0 < z_1(x) < b < z_2(x)$. The minimum $\bar{z}_1$ of $z_1(x)$ and the maximum $\bar{z}_2$ of $z_2(x)$ are obtained for $x = a$. Thus the curve of the equation: $f(z_0) = g(x) + d_0$ is a closed curve of the plane (x,z), in which when x increases from $\bar{x}_1$ to $\bar{x}_2$, two arcs $z_1(x)$ and $z_2(x)$ start from b to return to b. This curve is

entirely located in E if and only if the upper bounds $\bar{x}_2$ and $\bar{z}_2$ are less than 1. These conditions equivalent to $g(1) > f(b) - d_0$ and $f(1) < g(a) + d_0$, are written

$$f(b) - f(z_0) < g(1) - g(x_0) \quad \text{and} \quad f(1) - f(z_0) < g(a) - g(x_0). \qquad (7.286)$$

They are verified by the initial conditions (x_0, z_0) not too far from the equilibrium (a, b). (4) *Cyclic evolution*: If the initial conditions (x_0, z_0) verify the equation above, the solution of the differential system (on any interval where they exists) remains in the compact set $[\bar{x}_1, \bar{x}_2] \times [\bar{z}_1, \bar{z}_2]$ that is included in E. This solution exists on $[0, +\infty]$. The solution moves along the curve: $f(z) = g(x) + d_0$ in a finite time interval, because on this curve the continuous function $hx\,|b - z| + cz\,|x - a|$ that does not cancel, remains higher than a positive constant δ, and therefore $|\dot{x}(t)| + |\dot{z}(t)|$ sum of the absolute values of speeds remains higher than $\delta > 0$. The evolution of the economy is cyclic.

Note that the equilibrium (a, b) is not *locally stable*, because if (x_0, z_0) is different from (a, b), the cyclic solution exhibits the curve: $f(z) = g(x) + d_0$, and does not tend to the equilibrium. But if (x_0, z_0) tends to (a, b), $\bar{x}_1$ and $\bar{x}_2$ tend to a, and $\bar{z}_1$ and $\bar{z}_2$ tend to b; therefore the solution remains in the neighborhood $B((a, b), \varepsilon)$ of the equilibrium if the initial conditions (x_0, y_0) are enough close to (a, b). The *equilibrium (a, b) is stable in the Lyapunov sense*.
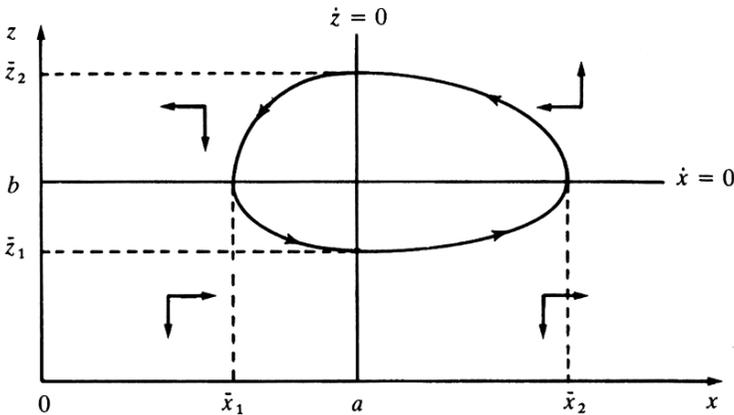
7.13 Catastrophe Theory and Kaldor Model

The first application of catastrophe theory to economics, and in particular to Kaldor model (1940) probably dates back to H.R. Varian (1979). Consider an expanded version of the Kaldor model proposed by Varian. Given the system:

$$\dot{Y} = \alpha[I(Y, K) - S(Y, W)], \qquad (7.287)$$
$$\dot{K} = I(Y, K) - D, \qquad (7.288)$$
$$\dot{W} = \Gamma(W^* - W), \qquad (7.289)$$
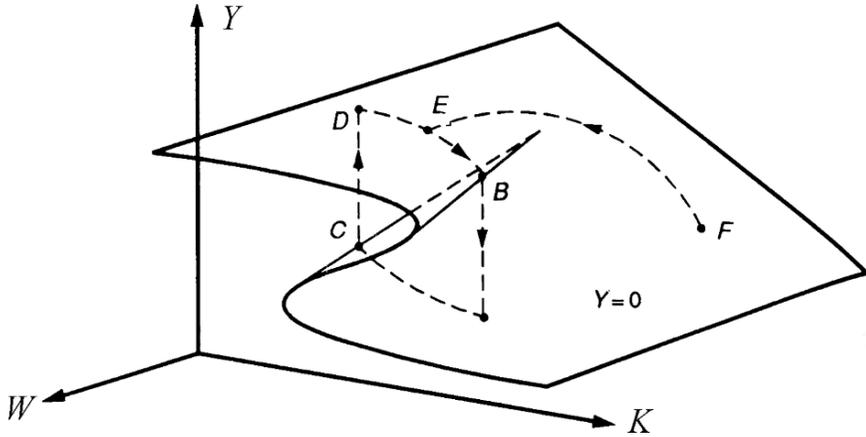
Fig. 7.15 Cusp catastrophe in the Kaldor model
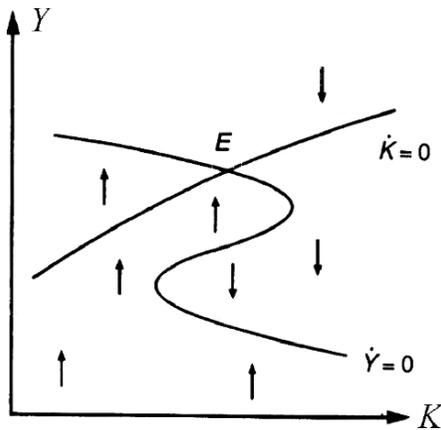
Fig. 7.16 Cusp catastrophe

where W denotes the "wealth", W^* the equilibrium value of the wealth, D a level of depreciation that is constant and independent. The investment function has usual characteristics. Suppose that savings is a decreasing function of the wealth so that when the wealth increases that leads to (1) the decreasing of the part of savings that is independent of the income level and also to (2) the decreasing of the marginal propensity to save.

According to the value of W, we can have one or three intersections between savings and investment functions. In the tridimensional space (Y, W, K), we obtain a representation of a "cusp catastrophe" (see Fig. 7.15). The equilibrium surface can be considered as the aggregation of different layers representing $\dot{Y} = 0$ for different values of W. Note that a representation of a cusp catastrophe in the bidimensional space (K, Y) is shown in Fig. 7.16.
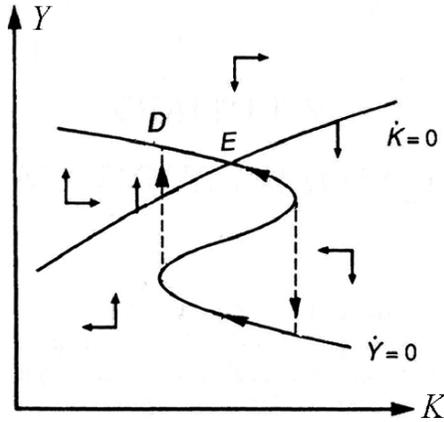
Fig. 7.17 Unique orbit
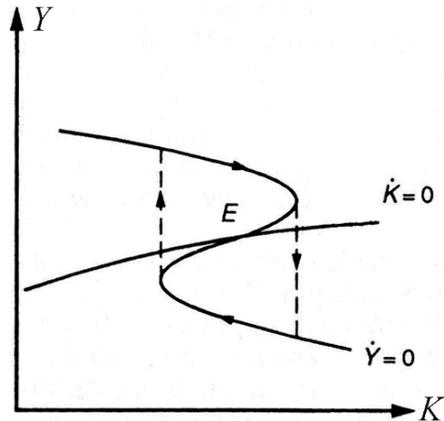
Fig. 7.18 Self-sustained cycle

Suppose an enough high level of W. If the exogenous shock is relatively weak, the system returns to E from F. However, when K is increased so that we reached the point B, a catastrophe appears and the income jumps on the low branch of $\dot{Y} = 0$. Then start a slow motion on the low branch of $\dot{Y} = 0$ until the point C where Y jumps again to the high branch. In this case, Y tends to return to the equilibrium E and the scenario stops. Therefore, in this case there is a single oscillation triggered by the initial shock (see Fig. 7.17). However another scenario is also possible (see Fig. 7.18), indeed, suppose that the long term equilibrium is locally unstable, i.e. the curve $\dot{K} = 0$ cuts the curve $\dot{Y} = 0$ in its increasing part, then it is potentially possible to observe *endogenous cycles*.

When we return to the three-dimensional representation (Y, W, K), by taking into account the third variable (i.e. the wealth), then the exogenous shock can lead to a change in the proper dynamic of the system. Suppose a stock market crash of a

small importance, then we find again the previous evolution. However, if the stock market crash is of a great importance, then there is a significant diminishing of the wealth and then it is possible to pass from a system with three equilibria to a system with only one equilibrium, and potentially to a system without equilibrium.

7.14 Overlapping Generations Models: Cycles, Chaos

The concept of an OLG model was devised by Allais (1947) and popularized by Samuelson (1958) as a way of simplifying monetary economics and macroeconomic models. Later, Diamond (1965) in his seminal contribution has examined the effect of government debt on the long-run competitive equilibrium of an economy with overlapping generations. OLG models can have varying characteristics depending on the subject of study, but most models share several key elements: individuals receive an endowment of goods at birth, goods cannot endure for more than one period, money endures for multiple periods, individuals must consume in all periods, and their lifetime utility is a function of consumption in all periods.

Hereafter, the agents are placed in an economy where they have to make an intertemporal consumption choice in order to achieve an objective subject to given or expected prices and resource constraints. In the basic model, the population is constant, there is a representative agent per generation and an exogenously specified endowment stream of the consumption-good. "y" indicates the variables pertaining to the youngs and "o" those for olds, t indicates the time. The utility function is $U(c_t^y, c_{t+1}^o)$, where c_t^y is consumption when young of an agent born at t, and c_{t+1}^o is the same agent's consumption when old. The time-invariant endowment pair is denoted (w^y, w^o), the price of the homogeneous good at time t is p_t, so that $\rho_t = p_t/p_{t+1}$ is the interest factor at time t. The representative agent maximizes the utility function $U(c_t^y, c_{t+1}^o)$ under the budget constraint:

$$c_{t+1}^o = w^o + \rho_t[w^y - c_t^y]. \tag{7.290}$$

In the context of the standard concavity hypothesis, an intertemporal competitive equilibrium will be given by the triplet of vectors (ρ_t, c_t^y, c_t^o) such that utility of each generation is maximized under the equation above and the following *balance constraint* is also satisfied: $[w^y - c_t^y] + [w^o - c_t^o] = 0$. The youngs can either save or borrow and carry claims or debts into the second period. Note that the case in which the young are impatient and borrow can be called "classical" and the opposite case can be called "Samuelson" in accordance with Gale (1973). The shape of the utility function U and the importance of w^y, w^o determine the state of the system. The no-exchange and no-money equilibrium is a possible result and will be repeated from the first period since we are in a time-invariant economy. Such an autarkic equilibrium is locally unstable in the classical cases and locally stable in the Samuelson case (ref. to Gale).

7.14.1 Benhabib–Day (1982)

Benhabib and Day (1982) were the first to consider such economies from the point of view of nonlinear analysis (and previously through a first paper in 1980). They analyzed the behaviors in the classical case. In such a model, by using the first-order condition for utility maximization with the budget constraint we obtain the *equality:*

$$\frac{U_1(c_t^y, c_{t+1}^o)}{U_2(c_t^y, c_{t+1}^o)} = \frac{w^o - c_{t+1}^o}{c_t^y - w^y}, \tag{7.291}$$

where U_1, U_2 are the partial derivatives of U. The equation above can be solved uniquely for c_{t+1}^o, then we have the function

$$c_{t+1}^o = G(c_t^y; w^y, w^o). \tag{7.292}$$

From this equation and from the previous equality, c_{t+1}^o can be eliminated from the left-hand side of the equality that must be equal to ρ_t (in equilibrium). The obtained ratio can be called the constrained marginal rate of substitution (CMRS), it is a function denoted $V(c_t^y; w^y, w^o)$ where w^y, w^o are parameters. This function combined with the balance constraint give the difference equation in the youngs consumption levels:

$$c_{t+1}^y = w^y + V(c_t^y; w^y, w^o)(c_t^y - w^y) \equiv f(c_t^y). \tag{7.293}$$

At this stage, the authors looked for conditions on $f(c_t^y)$ to make it unimodal[63] and with a degree of steepness sufficient to generate chaotic behaviors. They looked for *chaos* in the *topological sense*, i.e. existence of a period-three orbit, but also highlighted cases of utility functions and endowment pairs of stronger sort of chaos, i.e. existence of an invariant, absolutely continuous and ergodic measure. Then $V(c_t^y)$ can vary between w^y and $w^y + w^o$ and the conditions are as follows: $\exists \hat{c}, \hat{c} > w^y$ such that:

$$\begin{array}{ll} \alpha_1 = V(\hat{c}) > 1 & (<1) \\ \alpha_2 = V(\alpha_1 \hat{c} + (1 - \alpha_1) w^y) > 1 & (<1) \\ 0 < \alpha_3 = \alpha_1 \alpha_2 V(\alpha_1 \alpha_2 \hat{c} + (1 - \alpha_1 \alpha_2) w^y) \leq 1 & (\geq 1). \end{array} \tag{7.294}$$

Under these conditions, a topological chaos occurs for the system described previously $c_{t+1}^y = w^y + V(c_t^y; w^y, w^o)(c_t^y - w^y) \equiv f(c_t^y)$. The authors also considered relevant problems such as the Pareto efficiency of the chaotic trajectories and the role of an authority regulating the credit used by the young. In addition, they also mentioned the Samuelson case, highlighting that, when cyclic or chaotic

---

[63] *Unimodal function*: In mathematics, a function f(x) between two ordered sets is unimodal if for some value m (the mode), it is monotonically increasing for $x \leq m$ and monotonically decreasing for $x \geq m$. In that case, the maximum value of f(x) is f(m) and there are no other local maxima (e.g. quadratic polynomial, logistic map, tent map). In statistics, the *mode* is the value that occurs the most frequently in a data set or a probability distribution.

trajectories could be obtained, the system of the previous equation $c_{t+1}^y = w^y + V(c_t^y; w^y, w^o)(c_t^y - w^y) \equiv f(c_t^y)$ would not be well defined since for c_t^y there will exist two equilibrium levels of c_{t+1}^y. The analysis of such a system was continued by Grandmont.

7.14.2 Grandmont (1985)

Grandmont (1985) has studied such a case, i.e. Samuelson case, in a long article that is not possible to fully depict here. The model also considers overlapping generations. The utility function is supposed to be time separable and, instead of a fixed endowment of consumption-good, assume each agent has a labor-time endowment $\bar{\ell}^i$ (where i corresponds to y or o) in each period of agent life. ℓ^i denotes the amount of $\bar{\ell}^i$ that each agent supplies for work and assume that agent's utility depends both on consumption c^i and leisure time $\bar{\ell}^i - \ell^i$, according to $U_1^y(c^y, \bar{\ell}^y - \ell^y) + (c^o, \bar{\ell}^o - \ell^o)$, where the utility function U^i satisfies standard differentiability monotonicity and strict concavity hypotheses. Since we study the case where the youngs lends to the old in exchange for "money", we introduce the latter in a fixed amount M. The representative individual maximizes the utility function subject to the budget constraints:

$$p_t(c_t^y - \ell_t^y) + m^d = 0, \quad p_{t+1}^e(c_{t+1}^o - \ell_{t+1}^o) = m^d, \qquad (7.295)$$

where m^d is the nominal amount of money demanded by the young. In addition, the technology is supposed to be such that a unit of labor is transformed into a unit of consumption-good so that we are still facing a pure exchange economy. Note that p_{t+1}^e indicates the expected future price as of time t. For now, the hypothesis of perfect foresight concerning future prices used by Benhabib and Day is not made. This system has again a unique solution depending on $\rho_t^e = p_t/p_{t+1}^e$ (i.e. the expected interest factor). By using a slightly different notation, it is possible to define an excess demand for the good $z^i(\rho^e)$ as $z^i(\rho^e) = c^i - \ell^i$. The young lends to the old in exchange money, then $z^i(\rho^e)$ is always negative and $m^d = M = -p_t z^y(\rho_t^e) = p_{t+1}^e z^o(p_t)$ at each t, along an equilibrium path (note that when old, each agent will spend all of the individual money stock in exchange for goods). In the previous equalities, M indicates the fixed amount of existing bills that must all be demanded by the young in equilibrium.

Now we hypothesize that the agents have perfect foresight, that means $p_{t+1}^e = p_{t+1}$. Taking into account the balance constraint and the equilibrium condition, the competitive equilibrium is a sequence (of p_t) that results from:

$$z^y(\rho_t) + z^o(\rho_{t-1}) = 0, \quad p_{t+1}z^o(\rho_t) = M. \qquad (7.296)$$

Note that the system satisfies the Quantitative Theory of Money. If ρ_t is considered as a function of ρ_{t-1} aiming to invert z^y to obtain a "forward dynamics", this comes down to a problem mentioned above as z^y that may be backward bending. This is an essential aspect of the model and the origin of aperiodic behaviors. When ρ_t

increases, there are two opposite consequences on the demand for consumption by the young: (1) an intertemporal substitution effect as it makes consumption more expensive today, thus z^y should decrease; (2) but, it also makes the agent richer (wealth effect) as today labor is paid more. This will tend to increase the agent demand for present consumption, thus z^y should increase. However, this is different for z^o. Furthermore, note that when ρ_{t-1} increases, both substitution and wealth effect will increase the old agent demand for consumption. Thus z^o can be inverted and an *artificial* backward dynamics can be obtained from the first of both equations above $z^y(\rho_t) + z^o(\rho_{t-1}) = 0$:

$$\rho_{t-1} = \psi(\rho_t) \equiv (z^o)^{-1}(-z^y(\rho_t)). \tag{7.297}$$

The artifice of this equation does not allows to fully solve the problem, but it underlies a way along which a perfect foresight dynamics can be analyzed in an economy of the Samuelson type. Grandmont saw the necessity to fully clarify this problem and in particular the relation between the backward and the forward dynamics, but also the necessity to analyze the implications that different expectation-formation rules have for the stability of the system, see Grandmont and Laroque (1986). Suppose a periodic trajectory for the backward dynamics, then it is possible to define a forward dynamics in a neighborhood of such trajectory so that a stability analysis can be carried out by reversing the dynamic properties of the backward paths. In this context, Grandmont provides the conditions for which the equation above defines a system as an iterative map and he provides a detailed analysis of bifurcations of such a device. Grandmont refers to Collet and Eckmann (1980) outcomes concerning a vast range of utility functions generating period-doubling bifurcations and chaos. In this connection, when z^y is not monotonic, then a large enough degree of risk aversion on the part of the old trader (referring to a strongly concave function U^o) leads to chaos under certain conditions. Grandmont outcomes confirms the previous Benhabib–Day results for the classical case. The chaotic behavior results from a conflict between the wealth and intertemporal substitution effects created by a variation in the real interest rate if the first effect is strong enough. Unlike many articles pertaining to the optimality concept, such paths are not necessarily Pareto efficient and the cycles may be damped, created or amplified by appropriate monetary and fiscal policies. The Grandmont paper is regarded as a fundamental contribution on chaos in economics.

7.15 Optimal Growth Models: Convergence, Cycles, Chaos

The previously presented optimal growth models are characterized by a saddle point equilibrium. For a pair of initial values, an economy asymptotically converges towards a permanent regime. For any other pair of initial values, there is an accentuated divergence of trajectories without fluctuations. However, within the framework of optimal growth models it is possible to generate, without exterior shock, cyclic and aperiodic growth paths.

7.15.1 Boldrin–Woodford (1990)

In an article published in the "Journal of Monetary Economics" (1990), M. Boldrin and M. Woodford described an optimal growth model that shows cyclic and chaotic endogenous fluctuations. In this discrete time model, all the agents have identical behaviors and optimize the utility $u(c_t)$ of their consumption c_t (with $t = 0, 1, 2, \ldots$) on an infinite horizon where the utility function u is increasing and concave. The "state of the world" is characterized by a vector $x_t \in \mathbb{R}^n$ of stocks and by a feasible set $F = \{(x_t, c_t, x_{t+1})\}$ where x_t and x_{t+1} are vectors of $X \subset \mathbb{R}^n$ representing the today's and tomorrow stocks that are technologically compatible. Boldrin and Woodford define the function V by:

$$V(x, y) = \max_c u(c) \quad \text{such that } (x, c, y) \in F \tag{7.298}$$

and the set $D \in \mathbb{R}^{2n}$ by the projection of F along the c's coordinates. The function V, called the *short-run* or *instantaneous return function*, gives the maximum achievable utility that it is possible to obtain at time t if the state x and if we have chosen to go into state y by tomorrow. In such a context, the maximization of the discounted sum $\sum_{t=0}^{\infty} u(c_t) \delta^t$ such that $(x_t, c_t, x_{t+1}) \subset F$ is equivalent to maximize $\sum_{t=0}^{\infty} V(x_t, x_{t+1}) \delta^t$ such that $(x_t, x_{t+1}) \in X \times X \subset D$, where $\delta \subset [0, 1]$ represents a factor of time preference for the present. In other words, δ indicates the rate at which future utilities are discounted from today's standpoint (i.e. impatience). Note that for $\delta = 0$ the agent is infinitely impatient and, in a certain sense, a repeated myopic optimization of this type can represent the results of an OLG model. The classic properties of u and F imply: *Assumption* 1: $V(x, y) : D \to \mathbb{R}$ is strictly concave and smooth (if necessary), increasing in x and decreasing in y; *Assumption* 2: $D \subset X \times X \subset \mathbb{R}^{2n}$ is convex and compact, with nonempty interior.

Assume that the initial state x_0 is given. The solution of the previous maximization problem $\sum_{t=0}^{\infty} u(c_t) \delta^t$ such that $(x_t, c_t, x_{t+1}) \subset F$, is the same as that of the previous problem $\sum_{t=0}^{\infty} V(x_t, x_{t+1}) \delta^t$ such that $(x_t, x_{t+1}) \in X \times X \subset D$ which boils down to a problem of dynamic programming. For such a problem, the function W is defined by the *Bellman equation:*

$$W(x) = \max_y \{V(x, y) + \delta W(y)\} \quad \text{such that } (x, y) \in D. \tag{7.299}$$

A solution to this equation is a *map* $y = \tau_\delta(x) : X \to X$ that depends on the parameter δ and determines the optimal sequence of states $\{x_0, x_1, \ldots, x_t, x_{t+1}, \ldots\}$ as a *dynamical system* $x_{t+1} = \tau_\delta(x_t)$ on X. The time evolution depicted by τ_δ gives all the relevant information about the dynamic behavior of this model. The price vectors p_t of the stocks x_t that realizes the optimal program as a competitive equilibrium over time follows a dynamical process that is homeomorphic to the one for the stocks and is defined by the function $p_{t+1} = \theta(p_t)$ where $\theta = \delta W' \tau (W' \delta)^{-1}$ and W' is the derivative of the value function.

The study of this model comes down to analyze the asymptotic behavior of the dynamical system $x_{t+1} = \tau_\delta(x_t)$. We wonder if there is convergence towards a stable

state as in the one-sector growth model or the apparition of other behavior types. We wonder also where a stationary economy converges under competitive equilibrium and perfect foresight. The Turnpike (and anti-turnpike) theorem(s) help(s) to answer to these questions.

7.15.2 Turnpike Theorem (and Anti-Turnpike Theorem)

Turnpike theorem. Under the assumptions 1 and 2, there exists a level $\bar{\delta}$ of the factor δ such that for all the $\delta \in [\bar{\delta}, 1]$, the function τ_δ which is solution of the previous *Bellman equation* has a unique attractive fixed-point $x^* = \tau_\delta(x^*)$ with $x^* \in X$. Under additional conditions, x^* is interior to X.

The Turnpike theorem as described above dates back to Scheinkman (1976). Its continuous-time version was proved by McKenzie (1976) and Rockafellar (1976) and its generalization to the many-agents case was proved by Bewley (1982) and Yano (1984).

The theorem above means that under the general assumptions 1 and 2, if the agents are not "too impatient" under the given V and D, then it is possible to obtain a stationary state where the state repeats itself indefinitely (and no new event happens). Thus we can obtain a maximum balanced growth trajectory $\{x_0, x_1, \ldots, x_t, x_{t+1}, \ldots\}$ that converges towards a stable stationary state x^*. Such a point is a saddle point as defined in the previously described optimal growth models, and there is convergence if x^* belongs to the stationary trajectory.

Let us note that such models deal with the level of impatience of a considered economy and the difficulty of its evaluation. Important questions have to be considered pertaining to the sensitivity of Turnpike theorem to perturbations of its conditions. We have to know the value of $\bar{\delta}$ close to 1 that makes possible the convergence. We have also to know the behavior of the model when $\delta < \bar{\delta}$. The following theorem helps to answer to the latter question.

Anti-turnpike theorem. Given a function $\theta : X \to X$ of class C^2 where X is a compact and convex set of $\mathbb{R}^n$. Then there exists a technological set D, a return function V and a factor $\delta \in [0, 1]$ such that, under these assumptions, the function θ is the function τ_δ solution of the Bellman equation for V, D, δ.

This theorem allows to build a fictive economy for a given function $\theta = \tau_\delta$ that generates the recurrent sequence $\{x_0, x_1, \ldots, x_t, x_{t+1}\}$ whose complex dynamic is compatible with the assumptions of perfect competition markets, deterministic forecasts, diminishing returns, etc.

Deneckere and Pelikan (1986) have highlighted examples of one-dimensional models satisfying the same assumptions and with the function $\tau_\delta = 4x(1-x)$ as solution of the Bellman equation for specific values of δ.

Benhabib and Nishimura (1985) have studied a two-sectors model that respectively produces a consumption-good and an investment-good starting from the investment and labor goods. Once given the production functions of each good that we suppose to be concave and homogeneous of degree one, it is possible to

define the Production Possibility Frontier (PPF) $c_t = T(x_t, x_{t+1})$ that provides the consumption level c_t associated with the stock of today x_t and x_{t+1} that of tomorrow when the labor is completely and efficiently used.

When V and D are chosen as follows $V(x_t, x_{t+1}) = u[T(x_t, x_{t+1})]$ and $D = \{x_t, x_{t+1}\}$ such that $0 \leqslant x_{t+1} \leqslant F(x_t, 1)$ where F is the production function of the investment-good sector (and the labor is normalized to 1). In this case, the function τ_δ solution of the Bellman equation is not always of positive slope and if the capital intensity of the consumption-good sector is higher than that of the sector that produces the investment-good, then the function τ_δ becomes of negative slope. In this case, if the fixed point x^* of the function $\tau_\delta(x)$ verifying $\tau_\delta(x^*) = x^*$ is such that the derivative of $\tau_\delta(x)$ at the point x^* satisfies $-1 < \tau_\delta'(x^*) < 0$, then x^* is a stable fixed-point towards which the *optimal growth trajectory* tends. When the factor δ varies between 0 and 1, the value of $\tau_\delta'(x^*)$ changes and can be equal to -1 for an admissible value of δ, this gives rise to a *first bifurcation*. When (ceteris paribus) δ still evolves, then an *optimal cycle* of order 2 appears. For other values of the factor δ, it is possible to observe *stable cycles* of order $4, 8, \ldots$ and possibly *optimal aperiodic orbits*. Boldrin (1986) proved that the function $\tau_\delta(x)$ can become unimodal for values of δ included between 0 and 1 if the ratio of capital intensity of both sectors is sometimes higher that 1 and sometimes lower than 1.

Boldrin and Deneckere (1987) have studied a two-sectors model where the production function in the consumption-good sector is a *Cobb–Douglas* function and the production function in the investment-good sector is a *Leontief*. Boldrin and Deneckere have chosen a linear utility function and the function

$$V(k_t, k_{t+1}) = (1 - k_{t+1} + \mu k_t)^\alpha (k_t(1 + \gamma\mu) - \gamma k_{t+1})^{1-\alpha}, \qquad (7.300)$$

where α is the coefficient of the Cobb–Douglas function, γ is the capital intensity of the consumption-good and $(1 - \mu)$ is the rate of depreciation of the investment-good. They proved that for any factor δ, a *stable optimal cycle* of period 2^n can be obtained for values of α, γ, μ included between 0 and 1. Concerning the *aperiodic orbits*, they appear for an acceptable value of α if the value of δ is chosen between 0.2 and 0.3 that corresponds to an unrealistic discount rate of 400%. However, if $0.1 < \alpha < 0.2$, the chaos appears for $0.7 < \delta < 0.8$, i.e. for a discount rate between 25% and 40%.

7.15.3 Benhabib–Nishimura Optimal Growth Model (1979): Equilibrium Limit Cycle

The optimal growth model presents a solution which has the shape of a saddle-point (it can occur in a space of dimension higher than two). However, the convergence towards a unique stationary equilibrium-point does not make it possible to depict the phenomena of *equilibrium cycles*. The model developed by Benhabib and Nishimura can exhibit optimal trajectories which converge towards cycles

(Benhabib and Nishimura 1979). One presents the model in a simplified manner. The objective of the program is to maximize the utility function under constraint. $U(C(y,k))$ is the utility function of the consumption, n is the growth rate of the population and δ is the rate of preference for the present (rate of time preference) or the discount rate; y is the production and k represents the stocks of capital goods *per capita*. i is the good among the capital-goods and one will note y_i, k_i, p_i, w_i some components of y, k, p, w, by knowing that p_i will express the price of the good i and w_i the remuneration for this good i.

$$\underset{y}{Max} \int_0^{+\infty} e^{-(\delta-n)t} U(C(y,k))dt. \tag{7.301}$$

Under the constraint: $\dot{k_i} = \frac{dk_i}{dt} = y_i - nk_i \quad avec \quad i = 1,\dots,n$. The conditions of optimization on y and k give:

$$\frac{\partial C}{\partial y_i} = -p_i, \quad \frac{\partial C}{\partial k_i} = w_i. \tag{7.302}$$

The program is solved by using the Hamiltonian H:

$$H = e^{-(\delta-n)t} \{U(C(y,k)) + q(y - nk)\}. \tag{7.303}$$

The maximization implies $\frac{\partial H}{\partial y_i} = 0$. And we obtain:

$$q_i = U'p_i, \tag{7.304}$$

$$\dot{k_i} = \frac{\partial H}{\partial(q_i e^{-(\delta-n)t})} = y_i - nk_i, \tag{7.305}$$

$$\dot{q_i} = -U'w_i + \delta q_i. \tag{7.306}$$

In the neighborhood of the *permanent regime*, we have:

$$\dot{k_i} = y_i(k,p) - nk_i, \tag{7.307}$$

$$\dot{p_i} = -w(k,p) + \delta p_i. \tag{7.308}$$

The Jacobian matrix of this system is written:

$$\mathbf{J} = \begin{bmatrix} (\partial y/\partial k) - nI & (\partial y/\partial k) \\ -(\partial w/\partial k) & -(\partial w/\partial p) + \delta I \end{bmatrix}. \tag{7.309}$$

For such a Jacobian matrix, under certain conditions, when δ grows there is a *Hopf bifurcation* occuring for a critical value of δ (Fig. 7.19). At this stage, in the neighborhood of the stationary state (i.e. $\dot{y_i} = \dot{k} = 0$), we observe an *orbit* and this trajectory can show an *equilibrium cycle*.
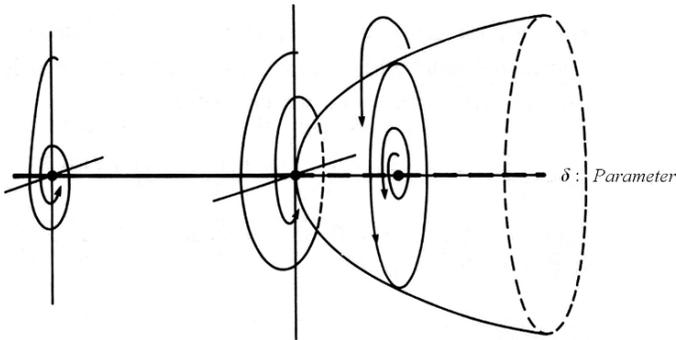
Fig. 7.19 Hopf bifurcation and limit-cycle

7.16 Nonlinearities and Cycle Theory

7.16.1 Nonlinearities and Chaos Theory

As observed previously, *Chaos theory* concerns properties of certain solutions of differential or recurrence equations of nonlinear type. The innovation concerning these differential equations resides in the ability to obtain unpredictable behaviors, whereas the equations producing them are perfectly deterministic. Indeed, the "*chaotic" behaviors* can be observed by using very simple "difference equations" of *nonlinear type*, and in this respect, the "logistic function" is one of the traditional demonstrations. The *sensitive dependence on the initial conditions* also plays a fundamental role in the definition of chaos theory. The initial state of a system is defined with more or less accuracy and one of the characteristics of this theory is that the initial states represented by two different points, but as near as possible, will produce different trajectories. In a linear system two very nearby points characterizing two different initial conditions will provide two nearby trajectories. However, when the systems are nonlinear, such an observation is not necessarily true. Indeed, the respective trajectories of two initial points will be able to move away, then, to come closer, *without any "regularity"*. It is thus impossible for such models to predict the trajectory or to find their initial condition starting from this trajectory. One touches a crucial aspect which is the numerical precision of the determination of initial conditions and the trajectory calculation. There is therefore *a sort of indeterminism inside a model whose writing is perfectly deterministic*. Thus, *Chaos theory seems to weaken certain fields of the economic modeling*. However, one of its counterpart is provided by the importance of the parameter setting. Indeed, a nonlinear "deterministic" function can show according to the value of its parameter, either monotonic behaviors (i.e. regular or periodic) or behaviors of the chaotic type. One can observe the "unpredictable" chaotic behavior of a model whereas a very weak variation of the value of its parameter could make it go back towards a monotonic or periodic behavior. Moreover, we also know that the *transition speed* from one state

to the other (understood that the speed of change of its parameter setting) influences the nature of the final state of the system and its predictability.[64]

Chaos theory also sought to study the behavior of stock markets. The statistical analysis of stock indexes seems to prove *the absence of regularity or cycle*. It is traditionally said that stock markets follow a random walk. It is also possible to say that their behavior is chaotic. However, In fact, *if the evolution of the stock market showed "regularities", it would be possible to admit that* (if one accepts for example the neo-classical concept of rational expectations) *these ones would be integrated by the different agents*. Moreover, one can wonder *how these regularities would be incorporated by the agents in the stock exchanges prices* and what would result from it on their evolution. Perhaps that any regularity would disappear and we would face again a random walk, as if the stock exchange prices could be only chaotic. *The stock markets were often regarded as the paradigm of a market of perfect competition, and we have to admit that they are "generator of chaos". By contrast, the usual economic variables do not generally show chaotic behaviors of this kind, and one voices sometimes the idea that such a fact is due to a certain "viscosity", "rigidity" or "adherence" of variables. This "adherence" seems to stabilize their behavior and act as a kind of simple attractor by showing monotonic trajectories. The justification of these adherences is difficult and vast: from the accumulations of goods, until the rigidities of actors and variables, as well as the decentralization of the establishment of equilibria.*

Chaos theory also studied overlapping generations models (OLG) in which "non-linearities" appear by selecting certain utility functions or production functions. This type of model resulting from neo-classics is thus in contradiction, in comparison with these chaotic behaviors, with the concept of rational expectations so important for the neo-classics, which cannot admit chaos as balance.

7.16.2 *Real Business Cycle Theory and Concept of Shock*

The cycle theory had vocation to explain the economic fluctuations. Furthermore, it is true that the different economic "shocks" since the second world war, as well as the post-war growing interventionism of States with in particular their *contracyclic actions*, made the analysis of these cycles more difficult because these fluctuations proved to be irregular. The analysis of these economic fluctuations took into account this established fact and thus attached importance to the occurrence of "shocks" as causality. These shocks being able to be of very different nature, i.e. from the oil crises of the 1970s until the monetary shocks. These shocks disturb the "trajectories" of our economies which move away from their original tendencies. Neo-classics consider that the shocks are due to the impact of various monetary policies which seek to influence the agents. However, one will describe it later within the framework of the *Lucas critique*, the neo-classic model of shocks, because of

---

[64] Ref: Works on the speed of transition applied to dynamical systems: *"The bifurcation paradox"*. O.Y. Butkovskii, Y.A. Kravtsov, and J.S. Brush.

the assumption of rational expectations (i.e. "perfect foresight", permanent balance, and super-neutrality of economic policies) refuses any long-run divergence of an economy in relation to its natural original tendency (i.e. before the shock).

The Real Business Cycle theory modified this approach by means of the notion of "real shocks". The argumentation is contained in the idea that the modifications of the basic parameters of the economy, such as the "households tastes, the initial endowments, or the available techniques", produce the shocks and cyclic evolution of the economy. The model which is usually used is based on the intertemporal choices of a single agent and if certain values of parameters are chosen, we obtain "a cyclic evolution of the production". The *shock,*[65] *resulting from the modification of parameters, will impact the economy in the long-term* as a *"wave" or a "group of waves".* In this type of representation, the money does not play part unlike the preceding analysis.

---

[65] *Remark:* (Shocks and DS & TS processes, see the definitions of TS and DS processes in the parts I and II). At this stage, we would like to point out a statistical notion that seems to have to be connected to the shock notion such as analyzed by economists and in particular the way in which the shocks have an impact on the economies in the long term. The statisticians state that for the non-stationary processes, as the TS processes (i.e. which become stationary by means of detrend), the shocks are regarded as transitory or "transient" because the series will recover its deterministic walk around its tendency curve. While by contrast, the shocks for the DS processes (i.e. which become stationary by means of differentiation) infinitely impact the series, even in a decreasing way. Then the TS processes appear to have to be connected to the neo-classical analysis, whereas the DS process appear to have to be connected to the real business cycle theory. In fact, the real business cycle theory justifies the business cycles by the shocks that "the fundamental structures" of an economy can undergo. This theory selects some of neo-classics hypotheses except for the fundamental hypothesis relating to money. Indeed, here the money does not have an influence on the economy, insomuch that the models of this type do not express the money in their construction, or express the money as a non-instrumental endogenous variable. Obviously, this is in opposition to the role that the neo-classics impart to the money in the theory of cycles.

Chapter 8
Efficiency and Random Walk

This chapter, which corresponds to the second subdivision of the part, attempts to characterize stock markets. The stock markets are known as efficient and are representative of the perfect competition. They are also known as advanced indicators of the global economic activity. The evolution of stock market indexes show however trajectories whose amplitudes are often considerably higher than those of gross domestic product. The concept and the models of rational expectations (arbitrage and stock price fixing) will give a first argument to try to explain these differences between trajectories. One will consider the efficiency concept, in the sense of Fama, about the stock markets and their instability.

8.1 Market Efficiency and Random Walk: Stock Market Growth and Economic Growth

We will attempt to give the theoretical definition of stock markets given by Economics using different concepts such as the perfect markets and the advanced indicator of the economic activity and through the notion of "value creation" which is transverse to Economics and Finance. Then, we will develop the economic concept of rational expectations and its self-criticism. Lastly, we will analyze the efficiency concept of financial markets.

8.1.1 Stock Exchange: Perfect Competition Market

The abstract definition of a market of perfect competition finds in the stock-exchange an *archetype* satisfying the following criteria:

- The market *atomicity* is respected. Otherwise, Securities and Exchange Commission[1] (CES) identifies the non-respect causes of the atomicity criterion, then they suspend the listings and tradings on the market. This means that the CES makes sure that no investor is *price marker*, i.e. does not exert power on market.
- The market *transparency* is respected, the listed-companies have to provide and publish information about their activities. That means that all the potential buyers of equities and shares must be equal in front of the information placed at the disposal by the listed-companies. If it were not the case, i.e. if the CES detected an *illegal "insider trading"*, it must sanction the originator of the offence. Indeed, when one or more investors have access to confidential and internal information of the company (i.e. information which is not publicized) and use the information to obtain abnormally high and illicit profits, then we are, indeed, in the presence of an illegal insider trading. It is said of this kind of investor who carries out such manipulation that it is in an *insider* position.
- The market *homogeneity* is respected. This can be summarized in the following way, whatsoever the "number" of an *outstanding stock*[2] on the market, a share having a different number but obviously of the same company (and of an equities group of comparable nature) rigorously satisfied the same needs.
- The market *fluidity* is respected. This corresponds to the *absence of barrier to entry or exit* on the market of stocks and shares or equities. Even if the cost of transactions or tradings, nowadays minor, can be regarded as a constraint less and less true, as these costs dropped considerably. The stock market is certainly *in vivo* very close to the theoretical definition of a perfect market.

8.1.2 Stock Exchange: Advanced Indicator of Economic Activity

Generally the financial system and in particular the stock exchange contribute to the optimization of the resource allowance in the economy. However, they are also places where the information is privileged. The information is the key element of decision-makings for all the actors involved in markets. During a given period, the *return* or the remuneration of *funds* and capital (that lets predict the stock exchange market, as well as the interest rates on the bond or monetary markets) leads the investment strategies of companies and also leads the behaviors of individual agents about savings and consumption. The general evolution of the prices of stock, bond, money or currency, is supposed to contain, incorporate and integrate *information about future*. The future of an economy (i.e. companies, branches of industry and

---

[1] *Securities and Exchange Commission (SEC):* commonly referred to as the SEC, is the United States governing body which has primary responsibility for overseeing the regulation of the securities industry. It enforces, among other acts, the Securities Act of 1933, the Securities Exchange Act of 1934, the Trust Indenture Act of 1939, the Investment Company Act of 1940 and the Investment Advisors Act. It removed regulatory authority from the Federal Trade Commission.

[2] *Outstanding stock:* The shares of a corporation's stock that have been issued and are in the hands of the public. Also called shares outstanding.

their markets) but also the relation with the foreign economies in terms of exchanges and influence, are represented in stock market prices. Thus, the stock exchange indexes are *advanced indicators of economy activity*. The supply and demand aggregations of the economic agents is carried out during the fixing of prices and therefore express the expectation of the future by agents. A stock-exchange index as for example the French index synthesizes at a given time all information available and its treatment by agents. Such a treatment can be rational or led by arbitrary choices. Indeed, taking into account the market fluidity (i.e. the absence of barrier to entry) all the types of investors can carry out their choices and anticipations, while including sometimes choices based on a "less rational" treatment of the information. In the stock exchange prices all information and all expectations are thus taken into account, so that it is almost possible to say that "nothing is visible any more". The role of anticipations and expectations in the evolution of the stock-exchange prices leads the economists to consider these prices as advanced indicators of the economic situation in the future. Thus, since the markets incorporate and treat all the data available about the future, the market is a source to use in order to make forecasts. Indeed, the study of stock-exchange indexes can be used to anticipate the gross domestic product (GDP). Accordingly, the study of the French index SBF250 which integrates more companies than the Cac40 (i.e. 250 against 40) can prove to be a considerable source of information to make short-run forecasts. In the United states the S&P500 index, which integrates the first 500 stock-exchange companies, is used officially in the determination of the short-term economic situation forecasts (i.e. conjuncture forecasts) carried out by the "National Office of Economic Research" (NBER).

NBER Recession	(1) Market drop	(2) Low-point of GDP cycle	Gap beetwen (1) and (2) Month	Growth of the Index beetwen (1) and (2) (%)
1960–1961	10-1960	02-1961	4	21.25
1970	06-1970	11-1970	5	21.86
1973–1975	09-1974	03-1980	6	35.60
1980	09-1974	07-1980	4	22.60
1881–1982	07-1982	11-1982	4	33.13
1990–1991	10-1990	03-1991	5	25.13
		Mean:	4.7	26.60

The strong link between economic activity and evolution of stock-exchange index seems to be so strong only in the United States. Usually, the anticipation period of the economic activity by the stock-exchange index varies between 1 and 5 months. But, this status of advanced indicator has a less strong reality for the European countries and the other industrialized countries. One of the justifications of this reality seems contained in the observed phenomenon of the *strong correlation* of the evolution of the *world-wide stock-markets with* the *American stock-market*. Consequently, if the S&P500, as advanced indicator, is adapted to the forecasts of the economic situation for the USA, the respective indexes for the other industrialized

countries are not necessarily appropriate, as advanced indicator of conjuncture for the short-term economic growth (GDP).

8.1.3 Indicators of Value Creation

Two groups of indicators are used to measure the "creation of value" in a company. These indicators are based either on the stock market performance, or on the internal economic performance of the company.

Indicators of market performance

The *Total Shareholder Return* (TSR)[3] measures the gains of the listed-company shareholders by the increase in stock-exchange price and by the cash dividend paid. This indicator, moreover very simple, provides a sort of annual average return on the *latent capital-gain* after dividends reinvestment. This indicator considers the share yield only as a ratio.

The *Market Value Added* (MVA) *compares* the *stock-exchange capitalization* of a company with its *net assets,* i.e. its *shareholder's (consolidated) Equity*. There is creation of "market value added" if the stock-exchange price evolves more quickly than the *net assets per share*, otherwise there is *destruction of value* by the company. This concept is ambiguous, indeed it is possible to imagine that the net assets is not a exact measurement, at moment t, of the value of a company, because the *latent capital-gain* and latent *capital-loss* (for assets and liabilities) are not instantaneously entered in the accounting of a company. Thus, *at moment t*, a company can have a real (or actual) value different from its *book value* resulting from the net assets.[4] Indeed, this is often the case since the values of assets (and also liabilities) can fluctuate. But we know that the accounts of a company are not permanently (understood *instantaneously*[5]) the sincere and true image of the reality, even if the accounting attempts to tend towards this purpose. However, one prefers to consider the anticipation role of this indicator.

Indicators of economic performance

The "Return on Investment" (ROI) is a measurement of a corporation's profitability, equal to a fiscal year's income divided by common stock and preferred stock equity plus long-term debt. ROI measures how effectively the company uses its capital to generate profit; the higher the ROI, the better. More generally, the income that an

---

[3] TSR proposed by the *Boston Consulting Group*.

[4] Note the different meanings of the following expressions: *actual-value, fair-value, and present-value.*

[5] In particular because of the non-instantaneous evaluation of the assets and liabilities.

investment provides in 1 year. ROI can be also expressed as the ratio of the profit (after payment of interests and taxes) to the stockholders' equity. It is *an indicator of the financial profitability built for the shareholders*.

The "Economic Value Added" (EVA)[6] is the measure of the corporate's true economic profit, i.e. after remuneration of all funded capital (debts and shareholder's equity). The objective of EVA is to understand which business units best leverage their assets to generate returns and maximize shareholder value. EVA is a way to determine the value created for the shareholders of a company. The basic formula is: "EVA = NOPAT − (NOA × WACC)", where NOPAT = Net operating profit after taxes, NOA = Net operating assets, WACC = Weighted average cost of capital. The shareholders of the company will receive a positive value added when the return from the equity employed in the business operations is greater than the cost of that capital (see Working capital management).

There is creation of value if the economic profitability is higher than the cost of the invested capital. Furthermore, *a company cannot create market value added without creating in the long run the economic value added and conversely*. The comparison of the stock market and economic indicators is certainly a way of highlighting the stock market bubbles, as well as the opposite situations. Considering only the financial theory, the Market Value Added could be taken as the sum of "discounted values" of the future Economic Added Values.

8.1.4 Corporate Governance: Market Imperfection Factors

This possibly minor component of the explanation of behaviors of the stock market plays however a role that we must not neglect. The motivations of company managers must be identified as well as the nature of their strategic choices and the incidences on the share price of their company. This approach is not recent since Adam Smith had started to analyze the motivations which led the leaders of large companies to use funds of which they are not the owners. A.Smith attributed an important responsibility to these leaders in the occurrence of the 1929 economic crisis on the stock markets. One describes three main causes for which the interests of leaders do not converge necessarily with that of companies that they manage:

- The *leaders maximize their own utility*. They allot themselves benefits in kind and prefer the investments which promote their own notoriety and their nonpecuniary interest.
- The *leaders prefer the less risky investments*, *even if they are less remunerative in order to preserve their position*.
- Aiming at a short-term effectiveness, taking into account their short-term presence in leadership of a firm, *the leaders tend to privilege the profitable investments in the short-run*.

---

[6] EVA: The underlying concept has first been introduced by Schmalenbach, but the way EVA is used today has been developed by *Stern Stewart & Co*. which also is owner of the registered trademark EVA.

These divergences of interest, as the "asymmetry" of the information between the leader and the shareholders can then be controlled only by defining what is called a set of "complete contracts", i.e. contracts which take into account all these cases and protect the shareholders against the possible abusive behaviors of leaders, to which they delegated the management of the company of which they are the owners. However, in practice these contracts seldom find their transcription. Consequently, substitute devices were set up such as mechanisms of *internal audit* for the benefit of the board of directors and shareholders, as well as *external controls which are supposed to be exerted by the financial markets*.

These divergences of interest seem to have an influence on the evolution of companies, their structures can be transformed, including the financial aspects obviously. From the influence of these divergences on the financial structures of companies to the influence on their stock-exchange prices, there is apparently only one step. This step is however skipped too quickly. But it is necessary to wonder about their impacts, specially because of the anticipation capacity of markets. Indeed, in spite of what follows, one could think that the stock-exchange prices are "affected" by the abusive decisions of the leaders of listed-companies, especially when one evokes the external control exerted by the financial markets and the possible sanctions. However, one will explain in the following section that this approach was invalidated by the conclusions of the Modigliani–Miller theorem which is supposed to have shown that the value of a firm is independent of its *financial structure*.

8.1.5 Modigliani–Miller Theorem: Neutrality of Finance on Market Perfection

The Modigliani–Miller[7] theorem shows that the *value of a company is independent of its capitalization.*[8] Thus whatsoever its *financial policy*, i.e. *whatsoever the structure of its debt and of its financing*, its intrinsic value will be unchanged. That a company uses the issue of loan rather than the new issues of capital to finance its activity, or, whatever its policy of dividend distribution to the shareholders, the intrinsic value of the company remains unchanged. It is said that the *intrinsic value* can *depend* only on *real characteristics* and on the *strategy* of the leader of firm.

A *leader* who will *use* the *debt* in order to increase profitability for his *shareholders* is likely to have to face contrary reactions which will *compensate* one another. Indeed the *market will take into account the increase in the profitability of securities*,[9] but the market also will react to this strategy which increases the risks

---

[7] *Theorem published in 1958 in the American Economic Review, entitled "the cost of capital, corporation finances and the theory of investment".*

[8] Postulate of Modigliani–Miller theorem: *The postulate on which the theorem is based is that we are in perfect and complete capital markets.*

[9] *Securities:* Commonly Securities are tradeable interests representing financial value. They are represented by a certificate. *They include shares of corporate stock* or mutual funds, bonds issued by corporations or governmental agencies, *stock options or other options, other*

related to the ownership of these securities. The investors on the market will sub-stitute less risky securities for these company securities which became more risked, consequently, their price inevitably will decrease, coming to neutralize the objective of the leader of the firm. Thus, the *expected wealth creation* will be a *failure*. It is also possible to appeal the checking of the *non-arbitrage clause* to present the the-orem and to say that if the value of a company increased with the part of shares that the company issues in relation to the part of "bonds" that the company issues, then, a capital-gain would be realizable by arbitrage by issuing additional shares and repur-chasing, by means of these funds, the bonds which had been issued previously. The theorem explains that in all the various markets (i.e. markets of goods and services, or markets of securities), "the price incorporates all information available and it plays perfectly its coordination role of individual choices". If the Modigliani–Miller theorem is generalized, the consequence of *this theorem is that the capitalizations do not have an influence on microeconomic and macroeconomic Equilibria.* The structure of the debt (and the financial leverage[10]) is neutral for the market, but it does not have influence on the investment policy of the firm which is the fundament of its strategy. Whatsoever the changes of the financial environment of agents, the resulting portfolio re-allocations are without incidence on the fundamental struc-tures of the economy. Thus, we understood why based on such a conclusion, the macroeconomic models did not take into account financial components during their conception, except for the *interest rate* obviously.

8.1.6 Role of Expectations on Equilibria and Markets: Expectation Concepts

In what precedes, we began to foresee the fundamental role of expectations in the evolution of stock market prices. *The stock market tends by anticipating the future to determine the present prices.* Thus, since the anticipation of the future, on the stock market, makes the "today's prices", it is difficult to say that the "adaptive expectations" concept is a good manners to approach the behaviors of agents on the stock market.

---

derivative securities, limited partnership units, and *various other formal "investment instruments"*. New issues of securities, including what is known as an Initial Public Offering (IPO), for new stock issues, are offered on the primary market. Securities that have already been issued may also be traded. This trading is called the aftermarket or secondary market.

[10] *Leverage:* The use of various financial instruments or borrowed capital, such as margin, to increase the potential return of an investment. (It is the amount of debt used to finance a firm's assets.) A firm with significantly more debt than equity is considered to be highly leveraged. Lever-age helps both the investor and the firm to invest or operate. However, it comes with greater risk. If an investor uses leverage to make an investment and the investment moves against the investor, his or her loss is much greater than it would've been if the investment had not been leveraged – leverage grows both gains and losses. A company can use leverage to try to generate shareholder wealth, but if the firm fails, the interest expense and credit risk of default destroys shareholder value.

8.1.6.1 Invalidated Adaptive Expectations, and Imperfect Markets

This rule which consists in envisaging the future value of a variable by using its present value and the forecast error at the former period, explains the term "adaptive". The neo-classical analysis, at variance with the adaptive expectations, explains that the individuals built their choices not according to the past but according to the present. In the method of adaptive expectations, it is clear that it is the adjustment carried out by the individual in relation to the forecast error which creates a systematic phenomenon of adaptation, because of the hypothesis that there is systematic error. In the computation formula, the former errors intervene according to a cumulative way. The method is defined by saying that *the variation of the anticipated value of a variable, between two periods, is proportional to the anticipation error made at the previous period.*[11] The formula of adaptive expectations is written by recurrence:

$$x_t^e = \alpha \sum_{i=0}^{\infty} (1 - \alpha)^i x_{t-i}. \tag{8.1}$$

Therefore, the method consists in making forecasts by using a *weighted sum* of *past values* of the variable, and the "weighting" diminishes exponentially in the course of time.

8.1.6.2 Rational Expectations and Perfect Market

The most common definition of rational expectations concept consists in saying that the agents use all available information as well as possible. The individuals are supposed to know (by definition in the neo-classical sense) the true model of the economy of which they are all the actors, and thus use information in order to "anticipate" also the possible random shocks. This concept adopted by the neo-classics seems more adapted to the perception of the behaviors of the actors of the stock market than the precedent. Unlike the preceding concept (adaptative expectations), *the rational expectations cannot admit the systematic aspect of the errors of forecasts.* Indeed, the agent would have taken into account in its expectations its systematic forecasts errors, which are expressed (according to a cumulative way)

---

[11] The difference between two periods of the expected variable is written:

$$x_t^e - x_{t-1}^e = \alpha(x_t - x_{t-1}) \quad with\ 0 < \alpha < 1.$$

It is necessary to transform this relation in order to obtain the expected value from the last values. One writes as a preliminary:
$$x_t^e = \lambda x_t + (1 - \lambda)x_{t-1}^e.$$

It is a weighted average of the observed value at t and of the expected value at $t - 1$. Thus by recurrence it is possible to write:

$$x_t^e = \lambda \sum_{i=0}^{\infty} (1 - \lambda)^i x_{t-i}.$$

inside the adaptive formula in order to correct its forecasts. The rational expectations, in fact, are analyzed as *a particular case of perfect foresight*, theory for which the random events are admitted but where the foresight is supposed to be perfect in the sense that it is made by means of the probability laws of the considered events.

The founder of the concept of rational expectations is John Muth (1960). The introduction of the concept into the macroeconomic models at the beginning of the 1970s, was initiated by Robert Lucas (1972). The neo-classics adopted the notion, which then has been broadly widespread until the 1980s among monetarists and some Keynesians. Among these economists who integrated the rational expectations concept, there are in particular those claiming to belong to the *General Equilibrium model* of Arrow–Debreu, for which the individual agents proceed to *optimal choices* guided by the *rationality principle*. Thus, the principle of rational expectations appeared adapted to their analysis and their general conceptualization of the economy, since they must integrate the concept of *expectations* in the preliminary *principle of rationality*. Then, it was necessary to introduce the "expectations" into the macroeconomic models. The major constraint of this rational expectations concept is that it supposes that the agents have *a perfect knowledge* of components and operating modes *of "the economy"*. The agents take, therefore, their decisions according to the economic environment, i.e. it is *necessary to know* the *true model of the economy*. The agents are supposed to anticipate the variables of the model, up to a random factor.

Intertemporal Equilibrium and Rational Expectations Model of J. Muth

The rational expectations suppose *the intertemporal equilibrium and the "self-fulfilling beliefs"* of agents since we are deliberately within the general framework of *the "perfect foresight"*. By definition, at the equilibrium the "actors" involved in the process do not have any reason to change their choice, behavior and expectation. Indeed, the model to which they refer is efficient since they are at the equilibrium, their beliefs are self-fulfilled and the rationality of their choice led them to the equilibrium. The principle of rational expectations require that all the agents have a representation of the whole of the economy and also requires that they formalize this representation by a model which is *the true model of the economy*. It is based on this model and the beliefs of agents that they can proceed to their rational expectations. That means that the beliefs of agents are formalized in the model, and since the model is the source of rational actions, the beliefs are thus self-fulfilled (*Beliefs* $\Rightarrow$ *Model* $\Rightarrow$ *Rational Expectations* $\Rightarrow$ Actions $\Rightarrow$ Self-fulfilling beliefs).

In such a mechanism, in a certain manner, the beliefs can be regarded as the *parameters* of the model.[12] If one gives a simple formalization of the model of rational expectations, it could be represented in the following way. One considers

---

[12] The beliefs are connected to the *probability distributions*.

a supply-demand model which are equal to each period, with however a shift (i.e. delay) of one period between demand and availability. The system is written:

$$Demand : d_t = a - bp_t + \varepsilon_t, \tag{8.2}$$

$$Supply : s_t = cp_{t-1}^e - d\sigma_{t-1}^{2\,e}, \tag{8.3}$$

$$with : d_t = s_t. \tag{8.4}$$

- *Supply is equal to demand at each period by assumption.*
- A lag of one period is necessary so that the production is available, it is a characteristic of the model called "cobweb".
- A lag of one period is necessary so that the production is available, it is a characteristic of the "cobweb" model.
- a and b are strictly positive and ε_t represents a random element, i.e. an "arbitrary" stochastic process.
- The *supply* s_t depends on the price which has been anticipated at the previous period p_{t-1}^e and depends on the anticipated variance of this price, $\sigma_{t-1}^{2\,e}$.
- The *price* p_t follows *a normal law*.
- $\sigma_{t-1}^{2\,e}$ indicate the *risk aversion*, indeed, the more the variance of the anticipated price p_t^e is high, the more they will restrict the supply.
- *The agents by definition know the Demand equation but also the probability law of ε_t which is by assumption a centered normal law $N(0,v)$, thus one can write:*

$$E(\varepsilon_t) = 0, \quad V(\varepsilon_t) = v. \tag{8.5}$$

Let us notice that the precondition $E(\varepsilon_t) = 0$ expresses the rejection of the adaptive expectations idea for which there are systematic forecasts errors. Then, from the following equality assumption $d_t = s_t$ the expression of price is deduced easily:

$$p_t = \frac{a - cp_{t-1}^e + d\sigma_{t-1}^{2\,e} + \varepsilon_t}{b} \tag{8.6}$$

and we observe that it follows a normal law with the following expectation and variance:[13]

$$E(p_t) = \frac{a - cp_{t-1}^e + d\sigma_{t-1}^{2\,e}}{b}, \tag{8.7}$$

$$V(p_t) = v/b^2. \tag{8.8}$$

On the other hand, at $(t-1)$ one does not know the value of ε_t at t, and one takes as anticipated price the expectation of the price at t:

$$p_{t-1}^e = E(p_t). \tag{8.9}$$

---

[13] Because it is supposed that $E(p_t) = 0$ and $1/b$ is the coefficient of ε_t, which is the only random term in the expression of p_t whose the variance is v.

Replacing p_t^e by $E(p_t)$ in the expression of the price expectation, it comes:

$$E(p_t) = \frac{a + d\sigma_{t-1}^{2\,e}}{b+c}.$$
(8.10)

Then, replacing the *anticipated variance* of the price $\sigma_{t-1}^{2\,e}$ by the value of the variance $V(p_t) = v/b^2$, one thus knows the law of p_t:

$$E(p_t) = \frac{a + d\frac{v}{b^2}}{b+c},$$
(8.11)

$$V(p_t) = v/b^2.$$
(8.12)

The rational expectations give rise to self-fulfilling phenomena, the supply is adapting to the demand (up to a random element ε_t). *The price and the production oscillate randomly around long-term values due to ε_t which prevents the construction of a cobweb as in the case of "perfect foresight" where ε_t is equal to zero by definition. Most remarkable in the conception of the principle of rational expectations is that, by postulating the perfect foresight (up to a random element ε_t) and thus postulating the equilibrium, the economy explained by this type of model excludes any coordination of the actions of agents, and thus excludes the existence of any coordinator. The postulates, in the conception of this model, are extremely strong, and probably excessive.* In the presupposition which consists in saying that all the agents know the true model of the economy, one admits (as described in the *Game theory*) that all the agents take into account in a rational way all the interdependences which connect their decisions between them. Thus, the models are impossible to write, this is why the *parameter setting* of models, i.e. the *beliefs* of agents, are simplified in the extreme.

Since stock markets became accessible at the same time to everybody and to any capital, seeking the maximization of profits according to the "rationality principle", the expectation concept was regarded in a new way. Consequently, *if the stock-market actors admit and integrate in their decisions the principle of rational expectations, they make of it a belief that becomes itself a self-fulfilling belief and this phenomenon influences the evolution of stock prices.*

8.1.7 The Lucas Critique of Rational Expectations and the Superneutrality of Economic Policies

The critique[14] of rational expectations was carried out by the economist Robert Lucas (1981) who himself had introduced and applied this concept in the

---

[14] The Lucas *critique* of rational expectations has been developed in Lucas (1976).

macroeconomic models, Lucas belongs to the neo-classics.[15] The critique is founded on the principle that the economic policies are ineffective if the agents act according to the rational expectations concept, because *they integrate these policies in their expectation outlines* and divert them of their objectives towards another "equilibrium". Thus, the interventionist economic policies would be not only inoperative but would be likely to disturb the organization of spontaneous balances of the economy. The economic policies, when they are incorporated into the predeterminations (coefficients and parameters) of the model and into the beliefs of agents, modify their initial projections and thus the effects. It appeared important to expose step by step how Robert Lucas, through macroeconomic model, describes the inefficiency of economic policies.

(a) *Writing of the Supply of the model after the estimate of the parameters (a,b,c).* This supply statistically estimated must be used to foresee the economic policies:[16]

$$Supply: y_t = a + bm_t + c\,t + \varepsilon_t, \tag{8.13}$$

y_t: is the logarithm of the *production* at t.
m_t: is the Money Stock (i.e. Money Mass[17]) and is the instrument of the monetary policy.
t: represents the time, this means that the estimate of parameters is carries out by means of a regression on time.
ε_t: is a random element, i.e. a stochastic process.

(b) Lucas *gives its "structural form" to the model and deduces a new "reduced form". In addition, R. Lucas considers that this supply function, which is for him only a reduced form of the model is blind.*[18] He thinks that the most adapted form of the Supply which does not appear in the function above, is its *structural form*, which is by definition the expression of the subjacent economic theory, and must take into account with their parameter setting, the following elements:

• "The forecasts of agents"
• "The variables of economic policies"

Thus, Lucas writes the structural form of the model as being constituted of:

• A *Supply function* based on the price expectations (this is the Lucas new supply)

---

[15] *Lucas Robert:* Nobel Prize in 1995 to have "developed the assumption of rational expectations and thus transformed the macroeconomic analysis and improved the understanding of the economic policy".

[16] The variables are taken as *logarithmic* in the model.

[17] That is, the equivalent of the "money supply and cash": (M1,M2,M3).

[18] *Reduced form:* the reduced form under which a model is written when each one of its *endogenous variables* is expressed according to its exogenous variables. The reduced form deduced from the "structural form", which expresses the economic theory on which the model is based. The reduced form is used for the estimate of the parameters, the forecasts or the tests of the effects of the different economic policies.

- A *Demand function* based on:

  1. The *quantitative formula* [see: paragraph (d)]
  2. "The rule of economic policy"
  3. The *rational expectations* of agents

(c) The *supply function by Lucas*. Lucas considers that the essential determinant of the supply is *"the expectation error"*, thus the *Supply is rewritten in the following way and (obviously) always fitted with a regression on time*:

$$\text{Lucas Supply: } y_t = a' + b'(p_t - p_t^e) + c't + \varepsilon_t', \tag{8.14}$$

p_t^e is the *expected price* at $t - 1$ and $(p_t - p_t^e)$ is the *expectation error* made at $t - 1$ concerning the price at t. The *price* and the *production* are the two *endogenous variables* of our model *which cannot thus be estimated since it is not a reduced form*.

(d) The Lucas *quantitative formula:* Instrumental relation of economic policy. Partisan of the Quantitative Theory of Money, Lucas then introduces the quantitative formula. Lucas thus formalizes the relations between money stock, production and price, in order to test the impact of monetary policies on the model. He writes the "instrumental" relation of the interventionist economic policy of this model. Let us note that this relation is monetary, this is the Lucas choice, but it could be of another nature, the interest rate for example in another type of model. k is the *circulation velocity* of the money, the quantitative formula is written:

$$\text{Quantitative formula: } p_t + y_t = m_t + k. \tag{8.15}$$

(e) *Instrumental rule of the economic policy.* It is a simple positive linear relation between the time and money stock, plus a random element:

$$\text{Instrumental rule: } m_t = \alpha + \beta \, t + \eta_t. \tag{8.16}$$

η_t is a random variable, i.e. a stochastic process such that: $E(\eta_t) = 0$. The equation above means that *the money stock increases at constant rate β, except for the shocks η_t*.

(f) *Rational expectations of the money stock by the agents (by Lucas)*. Lucas admits that the agents know the preceding rule perfectly, consequently they are able to anticipate the money stock, knowing that $E(\eta_t) = 0$ and the constant rate β. Moreover, $E(\cdot)$ represents the *value of the mathematical expectation of the money stock*. Then it is possible to write:

$$\text{Expected Money stock: } m_t^a = E(m_t) = \alpha + \beta t. \tag{8.17}$$

(g) Correlative expectation of prices. By using the quantitative formula and their money stock expectation, the agents can proceed (by substitution) to a price anticipation:

$$\text{Expected Money stock: } m_t^e = \alpha + \beta t, \tag{8.18}$$

$$\text{Expected Quantitative formula: } p_t^e + y_t^e = m_t^e + k, \tag{8.19}$$

$$\text{Expected Price: } p_t^e = m_t^e - y_t^e + k = \alpha + \beta t - y_t^e + k. \tag{8.20}$$

(h) *Forecast errors about prices. By subtraction of the preceding equation with the quantitative formula one deduces the forecast error:*

$$\text{Price forecast errors: } p_t - p_t^e = m_t^e - \alpha - \beta t - y_t + y_t^e. \tag{8.21}$$

In the equation above appears the forecast errors about the production which can formally be indicated by the variable e_t, whose construction now is known.

(i) The *Lucas new supply function*. In the initial supply function of Lucas, if one replaces the expression $(p_t - p_t^e)$ by its expression obtained in the equation above, it comes:

$$\text{New Supply function: } y_t = a' + b'\alpha + b'm_t + (c' - b'\beta)t + b'e_t + \varepsilon_t'. \tag{8.22}$$

R. Lucas thus obtained an "enhanced" Supply function, which is improved by the agents' anticipation plans about the money stock, price and production. It is possible to say that it is the *"true" reduced form of the Supply function*, whose approach of coefficients and parameters is more elaborate, incorporating the *quantitative formula*, the *instrumental rule* of the economic policy and the various forecasts of agents. Let us notice that the involved variables are rigorously the same one as those appearing in the equation of initial supply constructed from a statistical estimate, whence the ambiguity that Lucas wished to underline. Thus, when one proceeds to the econometric tests and to the extraction of coefficients from the observed variables, these coefficients are considered to be "rough" coefficients which will be at the origin of forecast errors. Indeed, these rough coefficients, in fact, would have to be the subject of a "sub-decomposition" to extract the estimate of the coefficients α and β which are those of "the instrumental rule" of the economic policy. If a forecast is carried out using the initial model and rough coefficients to evaluate the consequences of an arbitrary economic policy, one will not be able to forecast the anticipations that the agents will carry out concerning this economic policy, which will modify the effects of this economic policy. Only the estimate of the parameters α and β could make it possible (in a second step) to carry out good forecasts, having estimated the response of agents about the economic policies in the past, then only it will be possible to allow oneself to forecast their anticipations of the given future economic policies.

(j) The *"superneutrality" of economic policies*. Lucas extends the analysis, since he concludes that *the coefficients of the instrumental rule do not have an influence on the supply, because they disappear from its writing*. Indeed, we know that

these coefficients are used in the equations of the instrumental rule, quantitative formula, expected money stock, expected prices, price forecast errors and in the *"true" reduced form of the Lucas supply function* (i.e. the new supply function). In fact, *these coefficients* can disappear by substitution of the "price forecast errors" expression which are one of the components of the "new supply function".[19] The coefficients disappearing completely, they thus do not have any more impact, this means that the instrumental rule does not act any more on supply. Consequently, it is said that there is *superneutrality* of the money stock on the Good-supply. The supply is thus written[20]

$$\text{Supply function: } y_t = a' + c't + \varepsilon'_t + b'e_t + b'\eta_t. \tag{8.23}$$

Within this specific framework, the effectiveness of monetary policies is contested by Lucas. The Lucas conclusion is valid within the precise framework of this model. But beyond the developed example here, which aims at a model where the "leverage" is the money, in a more general way, it is the belief of agents which plays a fundamental role. And in particular the belief in the observed economic policy which will determine its efficiency.

8.1.8 Rational Bubbles and Sunspots Models

The Equilibrium is a situation where the agents are not tempted to change behavior because they achieve their projects (at the equilibrium) in the manner that they had envisaged. Their forecasts based on beliefs are completely confirmed (at the equilibrium), thus the agents do not have any reason to modify their choices. This *self-fulfilling prophecy*[21] concept had already been highlighted by Keynes in its "General theory". But it is the rational expectation concept, which integrating the idea of self-fulfilling belief, which allowed its full accomplishment in economics. The game theory reinforced the importance of the belief of agents in their choices,

---

[19] The price forecast errors are written:
$p_t - p^e_t = m^e_t - \alpha - \beta t - y_t + y^e_t$
and replacing m^e_t by m_t in the equation above, it comes:
$m_t = \alpha + \beta t + \eta_t.$
The price forecast errors are written:
$p_t - p^e_t = \eta_t + e_t.$
Moreover, substituting the equation above in the following quantitative formula:
$p_t + y_t = m_t + k,$
one obtains:
$y_t = a' + c't + \varepsilon'_t + b'e_t + b'\eta_t.$

[20] The random variable η_t which appeared in the instrumental rule still appears in the supply function and is not predictable by definition.

[21] *Self-fulfilling prophecy:* The notion dates back to ancient Greece, but it is Robert K. Merton (sociologist) who created the expression "self-fulfilling prophecy". Merton gives the following definition: "The self-fulfilling prophecy is, in the beginning, a false definition of the situation evoking a new behavior which makes the original false conception come true".

and the *self-fulfilling nature of the equilibrium* reinforces furthermore the agents in their beliefs. A typical case of phenomenon of self-fulfilling belief is pointed out in the rational bubbles models. In such a case, the beliefs of agents on the stock market lead to the rises or falls almost instantaneously after the purchase or sale decisions because there is not (price) adherence on markets.

The most symptomatic construction of what can be a model of rational bubble is symbolized through the models of "sunspots". In these significant simulations, the agents believe (it is a postulate) in the influence on the economy of the evolution of *sunspots*. This belief of agents will predetermine their choices and their rational expectations, thus they will act according to this belief and therefore *will create* systematically, by their behaviors, *correlations* in the economy *between sunspots and economic variables*. Thus, we can observe how the self-fulfilling belief phenomenon happens, whereas there is no relation in the fact between the economy and the sunspots. The most symptomatic is that the action of agents creates a relation between two events which at the beginning do not have any reason to be connected a priori. In short, it is possible to say that "the agents act in order to make true their belief". Moreover, such phenomena are identifiable on the financial markets. A bubble on a stock market appears only because the agents believe that it represents a reality. The agents act in order to make real the beliefs, even if the phenomenon itself does not exist. When the *divergence between* the value of *stock-exchange prices* (i.e. stock-exchange capitalization) and the *economic value of listed-companies* is *large*, if there is no likely element to justify the future value-creation, then indeed we are faced with a financial rational bubble. The word "rational" does not seem to be suitable, taking into account the weakness of concrete bases of the *overvaluation* of prices. But in such a case we must refer to the "*Perfect Foresight* theory" based on the "rationality principle".

As in the case of preceding rational expectations and that of the Lucas critique, we will use the same type of presentation to identify the various interactions in the model. The model of rational bubbles is based on the principle of the *permanent self-fulfilling belief* which induces the permanent equilibrium (*Permanent self-fulfilling beliefs* ⇒ *Permanent Equilibrium*).

Let us write the first variables of the model:

p_t is the *price* of a share at moment t.

d_t is the *dividend* for a share at t.

r_t is the *rate of return* at moment t.

(a) *The "non-arbitrage" condition of a share.* Since the permanent equilibrium is supposed, the arbitrage between the shares is permanent, and the rate r_t is the same for all shares (according to the homogeneity principle of a perfect competition market). Thus, the considered share must verify the non-arbitrage condition:

$$\textit{Non-arbitrage condition: } r_t = \frac{p_t^e - p_t + d_t}{p_t}. \tag{8.24}$$

The share return corresponds to ratio of the dividend plus the expected capital-gain (i.e. appreciation) to the purchase price of the share.[22]

(b) *The "rationality" assumption transforms the expectation into perfect foresight.* The *rationality assumption* in the sense of the partisans of the rational expectation, is to suppose that the anticipated and expected price p_t^e is the "true" price at $t+1$ *(according to perfect foresight)*:

$$\text{Price perfect foresight}: p_t^e = p_{t+1}. \tag{8.25}$$

(c) This same *assumption* (through perfect foresight) implies that the future dividends and returns are known. Thus, due to the equation above, the non-arbitrage condition can be written:

$$r_t = \frac{p_{t+1} - p_t + d_t}{p_t} \tag{8.26}$$

or in an equivalent way, resolving in p_t, one can extract the expression of the share price:

$$\text{Share price}: p_t = \frac{p_{t+1} + d_t}{1 + r_t}. \tag{8.27}$$

This is a typical equation of rational expectations. We will notice that the price of the share at t is expressed according to the price of the share at $t+1$, what is a symptomatic mechanism of the perfect foresight, the "expected future" determining perfectly the present. Therefore, theoretically it is possible to *extend* the reasoning to an *infinite horizon*, what is carried out during the two following steps and in particular to avoid to have to calculate the ultimate value of the model.

(d) *Extension of the price expectation* by recurrence. At the step $t+1$ the price is written:

$$\text{Future share price at } t+1: p_{t+1} = \frac{p_{t+2} + d_{t+1}}{1 + r_{t+1}}. \tag{8.28}$$

As described above in (c) the price of the share at t was expressed according to the price of the share at $t+1$, but if one replaces p_{t+1} in the equation highlighted in (c) by the second member of the equation above of the price in $t+1$, it comes:

$$p_t = \frac{p_{t+2} + d_{t+1}}{(1 + r_t)(1 + r_{t+1})} + \frac{d_t}{1 + r_t} + \frac{d_{t+1}}{(1 + r_t)(1 + r_{t+1})}. \tag{8.29}$$

Which is a new expression of the price at t. During the following step, one pushes the term of the horizon until $t+T$.

(e) The *general expression* of the *price* at the $(t+T)$ *horizon,* also called the *Fundamental value*, is written:

$$p_t = \frac{p_{t+T}}{(1 + r_t)\cdots(1 + r_{t+T-1})} + \frac{d_t}{1 + r_t} + \cdots + \frac{d_{t+T-1}}{(1 + r_t)\cdots(1 + r_{t+T-1})}. \tag{8.30}$$

---

[22] *Gross return*: If one considers a *gross* return in which the dividend is not reinvested, which is not necessarily the case.

The price at t is expressed in terms of *dividends flows* from t to $t + T$ and in terms of price at $t + T$, i.e. at the horizon. *Moreover, since one considers that "at infinite horizon the present value of the price is zero", then the price at the period T is expressed only according to the "discounted (future) dividend flow".*

$$p_t \simeq \frac{d_t}{1 + r_t} + \cdots + \frac{d_{t+T-1}}{(1 + r_t) \cdots (1 + r_{t+T-1})}. \tag{8.31}$$

(f) *An infinity of solutions.* If the initial price expressed in (c) is selected and if a constant k is associated, that is written:

$$p_t + k = \frac{p_{t+1} + d_t}{1 + r_t} + k. \tag{8.32}$$

By giving at the second part of this equation the same common denominator, then it comes:

$$p_t + k = \frac{p_{t+1} + k(1 + r_t) + d_t}{1 + r_t}. \tag{8.33}$$

Moreover, it is possible to add to "fundamental value" a suitable term such as $k(1 + r)^{T-1}$ with $r_t = r$ and $T \geqslant t$, (and by posing $r_{t-1} = 0$), which represents the *rational bubble*. The fundamental value plus this additional term makes it possible to represent the solutions of the equation above. The "rational bubble" expression is used when k is positive and different from zero, because *the agents consider that the value of the share will grow continuously and they anticipate this belief.* There will be *absence of rational bubble when k is equal to zero, i.e. when the agents do not envisage increase in prices.* Thus, we observe that the model highlights the multiplicity of possible solutions, i.e. balances of which each one depends on the beliefs of agents that predetermine the model. There will be as many balances as beliefs. This analysis can be connected to (even if the concepts used are of different nature obviously) the concept (described in the section about the indicators of value-creation) which concluded that, in the theory, *the Market Value Added (MVA) is the expression of the discounted value of the (future) Economic Value Added (EVA).* The discounted value of the Economic Value Added is not obviously the discounted value of future dividends flows, and the Market Value Added is not either the simple expression of the share price at the current period, but it is however possible to say that a present stock-market value is expressed as the *discounting* of (future) *income* flows. What tends to show that the analysis of *rational expectations* finds *transcriptions* in the *financial sphere*. Furthermore, the most recent methods of firms evaluation emphasize the *discounting* of (*future*) *profit flows* in the mid or long-term from which the present value of a firm is calculated.[23] The Business plans

---

[23] *Intrinsic value:* It is possible to present the way in which the *traditional financial analysis* calculates the share prices. It is based on the future performances and on the "rate of discount" which makes it possible to calculate the present value of these flows. The *intrinsic value* of a share is equal to the discounted value at the market interest rate of the anticipated future dividends.

$$VI_0 = \Sigma_{t=1}^{\infty} \frac{E(d_t)}{(1 + E^a(r))^t}$$

of the economic activity of a firm are built on the *income statements* based on the evolution of the structure of the company, then the profit flows are discounted. By knowing that a company in a given sector must provide a given "Return On Equities" (ROE) to show its competitiveness, we inevitably return to the price of the company since the investors request a refund of their purchase on a fixed duration (10 years for example depending on industrial branches). Otherwise, obviously, the investors do not buy and the shareholders can sell the shares for the same reason.

If we go back to the model described in this section, it concludes that there is an infinity of possible solutions which depend obviously all on the nature of the beliefs of agents which predetermine the solutions and predetermine different balances which can result from the application (in the individual projects of agents) of their beliefs and rational expectations. There is another type of model which highlights the role of the self-fulfilling prophecies, in an exemplary way, this is the model of "sunspots". The model postulates that, without scientific fundament, the sunspots apparition are the cause of business cycles. In this type of models, one admits that "the agents believe that the evolution of sunspots has an impact on business cycles". Since the agents act based on this postulate and this belief, the model by the actions of agents gives a reality to the "relation" between sunspots and economic activity. Consequently, the agents "observe" the existence of this link between the sunspots and the economic activity and then conclude that their beliefs are valid and therefore "confirmed", whereas there is no tangible bond. It is the same principle as in the model of rational bubbles, where the belief has an impact on the share prices, i.e. on their market values, but not on their "(real) economic values", i.e. the "fundamental values".

The concept of rational bubbles which is a development of the former theory of the perfect foresight, makes it possible to highlight the way in which a "gap" can be created between "market values" and "economic values" (i.e. between trading values and economic values based on the fundamental economic analysis), then to grow according to rational principles whereas the fundament rests on the belief. The rationality of choices is obviously postulated by the model, but if the choices are rational the question to be posed is that of the "rationality of beliefs". The absence of rationality in the beliefs which will predetermine the model, during the apparition of rational bubbles, will create a gap, or a shift, between real economy and exchange values, and will increase the explosion risk of bubbles which is also a way for the economy of bringing back the prices to their fundamental values. However, it is necessary to emphasize that the fundamental values are not always able to take into account the value creation of a company and an economy by extension. Here is the main difficulty, the value creation of today (technological innovation, new markets, . . .) is rarely quantified instantaneously for its real value, i.e. for the value

---

The intrinsic value today of the share is noted VI_0.

The discounted value of the market rate $E^a(r)$.

The anticipated future dividends are noted $E(d_t)$.

The value of the share today is expressed as the sum of the flows which will be generated by this one in the future, but after having discounted them in order to give of it a value in monetary units of today.

that it will represent in the future. The current innovations will make the companies of future. This is why the evaluation techniques which tend to measure the value by using the past, i.e. the sum of "past incomes" risk to be blind about the future and its innovations.[24] In addition, *the fundamental value in the sense of rational expectations theory must incorporate all available information*. In a symmetrical way, the reasoning which consists in observing how a company in the past could show adaptation and innovation on its market is also a key indicator of its capacity to create value, such indicators must be taken into account obviously.

The construction of an evaluation, before the exchange happens and before the price fixation, must thus *combine* techniques of *discounting* of the *past* and its flows *with techniques* of *discounting* of the *future* and its flows (belonging to the rational expectations framework). But anticipations of future flows will result from postulates and thus from beliefs, which will generate a multiplicity of projections. That is also during the "choice" of the belief that a rational act must be done which first consists in taking into account the history of the company and its evolution must themselves be evaluated and incorporated in the "belief". But there are obviously other less mastered elements which will incorporate the belief and will induce the anticipations. Therefore, it is also in the elaboration of the belief that the past must join the future in a *conviction* which will first take shape in the establishment of the belief itself and in all the decisions which will result from this. If an analyst has the conviction that a company which always proved in the past its ability to anticipate the technological developments and to incorporate in its R&D investments in a continuous way the cost of the innovation, the analyst will integrate this past reality in its conviction and can provide favorable anticipations, whatever the context of technological change. In the previous paragraphs a confusion is present (obviously) between the pure concept of "rational expectations" within the perfect foresight framework and the "rational anticipations" within an empirical framework where the concept lost its essence.

8.1.9 Efficiency and Instability of Financial Markets: A Non-Probabilisable Universe

From the concept of *rational bubble* to the empirical notion of *speculative bubble* there is obviously a *conceptual abyss*. However the concept have to face the empirical reality. This reality finds one of its expressions through financial crises which weaken the efficiency concept of financial markets, which at the same time can lose some of the prerogatives that one had been able to grant to them. The *efficiency concept* invested many works concerning the financial markets. By definition, it is said that a financial market is efficient "if the prices (of assets) instantaneously and fully reflect all available information". In such a market the price will be always

---

[24] Except for the initial vested capital. A company at the moment t (represented by its balance sheet structure) is also the expression of the sum of past incomes, except for its potential dividend distributions.

equal to the fundamental value, up to negligible elements (of friction) as the cost of entry or exit for shares. Thus, by definition "the share price must take into account, at any moment, all information concerning the current state of a listed-company and its future evolution perspectives". The efficiency concept (Fama E[25]) was decomposed into three forms according to the *information nature* used in the *mechanism of price setting of a* share. The velocity of incorporation by the market of the information is emphasized in all the cases, whatsoever it is. Fama proposed *three types of Efficiency*:

(1) *Weak form.* In such a case, "the prices of assets reflect instantaneously and fully all *information contained in the past of price". It is said commonly that the prices incorporate so quickly the information contained in the past that it is impossible to define profitable rules of negotiation for the agents*. Thus, all the rules of negotiation and arbitrage which would rest on the past cannot be profitable and would be without fundament.

(2) *Semi-strong form.* In such a case, the prices of assets reflect instantaneously and fully all *information available publicly*. This type of information is incorporated in the price by the agents so quickly that it is impossible for an operator to make profitable arbitrages. One can even wonder whether the effects of the publication of results of a company are not already incorporated in the price even before their publications, by anticipation, so that, the day of the publication, the market prices can remain unchanged.

(3) *Strong form.* In such a case, the prices of assets reflect instantaneously and fully all *information as much private as public*. The strong assumption in this case, consists in considering that, when an agent benefits from information of a private nature, and taking into account the reaction velocity of markets, *this information cannot generate capital-gains or "profits" which would be regarded as "abnormally high"*.

Such an academic case, often contradicted by the facts, seems to exclude all the assumptions where there is an (illegal) *insider trading,* by considering that the "Securities and Exchange Commission" makes sure to prevent them. Furthermore, this is a way of emphasizing the *conceptual limit* of the efficiency such as posed by Fama, because one considers in the case where an agent benefits from information of a private nature, accomplishing "abnormal" profits, that it is the rejection condition of the hypothesis of market efficiency. Consequently, in the last case,

---

[25] *Fama E* is known for his work on portfolio theory and asset pricing. Fama thought that stock price movements are unpredictable and follow a random walk (ref: Ph.D published in the Journal of business, 1965, entitled "The behavior of stock market prices", and in the 1966 article, "Random walks in stock market prices", published in Financial Analysts Journal and Institutional Investor in 1968. The last article inaugurated a new era in financial research: the *empirical financial research.* Fama is considered as the father of efficient market theory (ref: 1970, Journal of Finance, "Efficient Capital Markets": A Review of Theory and Empirical Work). Fama enunciated two fundamental concepts, the first one, enunciated three types of efficiency: (1) strong-form; (2) semi-strong form; and (3) weak efficiency. the second one, Fama proved that the notion of efficient market was rejected only if the model of market equilibrium was rejected too (for example, the price setting mechanism). This concept is known as the "joint hypothesis problem". The fundamental issue is to explain the differences in stock returns between "market capitalization" and "value".

the "speculative exchanges" notion appears. They seem to be able to happen only if some agents have different information, *which are representative of situations where there is uncertainty*. Consequently, one leaves the framework of *NeoClassical economics* (NCE) which tends to conceptualize a universe where the risk is "probabilisable" and leaves aside the situations of uncertainty. *It is said that there is uncertainty when the agents do not assign any probability to the possible future events,* and this *because they are unable to do it, or because they are indifferent or because they ignore them.*

The uncertainty is a concept which does not belong to the universe of Neoclassical economics but belongs to a context taken into account first by Keynes, then, developed by the Keynesian school and which continued to be described by *Neo-Keynesian economics* (NKE). The speculative exchanges on the financial markets can exist only if *there is not perfect information*. Thus, the "speculative exchanges" and "speculative bubbles" concepts would concern the universe described in Neo-Keynesian economics rather than the universe of Neoclassical economics. The exclusion of the inefficiency of markets involves the rejection of the speculation by the framework of the classical school, whereas the Keynesian framework would incorporate it better.

Things seem to be as if Neo-Keynesian economics could take over to Neoclassical economics in the comprehension of financial markets to pass from a *probabilisable universe,* where the concepts are defined, towards a *non-probabilisable space* where the *uncertainty prevails*. (In a speculative way, we could regard this approach as a synthesis of both schools apprehending thus any aspects of behaviors of financial markets. But this would imply the existence of a third school which would combine these two incompatible schools.) The speculative bubble is often analyzed as a "drift" in the long term, or a deviation of stock exchange prices in relation to the fundamental values (as if we could assert that there are "true prices", what would suppress any polemic and facilitate the work of economists and financiers).

From the point of view of the increase of these speculative bubbles, it is possible to consider that the agents do not systematically adopt a behavior dictated by the principle of rational expectations. But would be directed for example by short-term choices where notably the accentuation of tendency (rise or fall) are the resultant of a lack of discernment which reproduces the decisions of the greatest number of agents. One can observe frequently these phenomena during the *euphoria* or *depression periods* of stock markets, even if these tendency *exaggeration* phenomena by the market have not been analyzed in a significant way by the statisticians (to our knowledge). We would like to point out a distinction for better apprehending the efficiency concept. Indeed, usually, one suggests that the rejection of the efficiency hypothesis could be consecutive to the existence of speculative rises, or consecutive to the explosion of speculative bubbles. In fact, as described before, the efficiency is to be rejected only if "anomalies" are observed on markets, which would make it possible to obtain profits or capital-gains "abnormally" high. Consequently (and this represent a contradiction in relation to what was stated above), the speculative bubble notions would not be a demonstration of the rejection of the market efficiency assumption, except if one provides the proof that they result from

abnormal practices, which make it possible to obtain abnormally high profits. (And we could also wonder if the capital-gain of some agents obtained during the increasing of the speculative bubble would not be compensated, in the long term, by the capital-losses realized by other agents during the explosion of the speculative bubble. As if *the equilibrium of the market itself* were *an intertemporal equilibrium gravitating around the tendency curve of fundamental values*.) In other words the *insider trading* is *a proof of non-efficiency but not a proof of speculative bubbles*. It seems to us that there is a sort of contradiction, these speculative bubbles being able to exist only if the financial markets do not have perfect information. In any event, except for the historical shocks, many stock market crises occur due to the bursting of speculative bubbles, during which the prices collapse.

This type of bubble occurs if the assets prices move away (upward) from their intrinsic values. The operators which are wrongly informed about the tendency of fundamental values are going to disrupt, by their purchase acts, the choices of the other agents, which are supposed to use better the available information. Consequently, many agents can lead the prices upwards and lead to a situation where the rise happens not because the agents anticipate a rise rationally, but because the different actors invested the market with buying orders which imply a rise of prices. The crisis intervenes when the risk related to the difference between fundamental values and exchange values is realized through a collapse of prices.

8.1.10 The Question of the Imperfection, Inefficiency and Non-Random Walk of Stock Markets

Paul Samuelson in an article entitled "Proof that Properly Anticipated Prices Fluctuate Randomly (1965)" explains that in an efficient market, about information, the price changes are unpredictable if they are properly anticipated, i.e. if they fully incorporate the expectations and information of all market participants. In 1970, Fama summarizes what precedes in a rather explicit sentence: "the prices fully reflect all available information".

In contrast with many applications of the *random walk hypothesis* in the natural phenomena for which the randomness is assumed almost by default, because of the absence of any natural alternatives, Samuelson argues that the randomness is achieved only through the active participation of many investors seeking the increase in their wealth. They try to take advantage from the smallest information at their disposal. And while doing so, "they incorporate their information into market prices and quickly eliminate the capital-gain and profit opportunities". If we imagine an "ideal" market without friction and without trading cost, then "the prices must always reflect all available information" and no profits can be garnered from information-based trading, because "such profits have already been captured". Thus, in a contradictory way, the more efficient the market, the more random the time-series of price changes generated by such a market, and "the most efficient market of all is one in which the price changes are completely random and unpredictable".

Thus, the random walk hypothesis and the efficient markets hypothesis became emblematic in Economics and Finance, although more recently, in 1980, Grossman and Stiglitz considered that the *efficient markets hypothesis* is an *idealization* economically unrealizable. Moreover, more than 15 years ago, a collection of new statistical articles (Lo 1991; Lo and MacKinlay 1989, 1990) appeared and emphasized a new approach aiming to reject this random walk hypothesis. The statistical studies conducted by Lo and MacKinlay (since 1988) about the US stock-exchange, lead to reject the random walk hypothesis for weekly values courses of the New York Stock Exchange (NYSE), using a simple test based on the volatility of courses (Lo and MacKinlay 1988). By specifying however that the rejection of the random walk hypothesis does not imply necessarily the inefficiency of the price formation.

Conclusion

The introduction of nonlinearities is an infinite source of behavior diversity for the comprehension of natural or economic realities. Many phenomena previously adverse to any modeling can find an algebraic formalization today through the epistemological rupture of nonlinear theory.

The economic growth models built on linear structures masked the phenomena of instability, persistence, hysteresis[1] and asymmetry, which are symptomatic of economic fluctuations. The concept of nonlinearity seems relevant to model business cycles. The first initiatives prior to this epistemological breaking that explained the cycles in an endogenous way date back to Kaldor in 1940, Hicks in 1950, Goodwin in 1951 and Rose in 1967. Afterwards, the approach was extended and applied to the growth paths of macroeconomic models: And this is what is called the Equilibrium cycle.

In the endogenous growth theory, linearities do not seem to be often present either, except in AK models. These AK models are constructed around a linearity in the differential equation of the model and around a constant return of capital. In spite of didactic qualities of this construction type, the models described as semi-endogenous, built on nonlinearities from decreasing returns of the capital, seem more realistic. A dichotomy is established within endogenous growth models according to the effects of their policy. The purpose is to distinguish between permanent, transitory or level effects. The stake is to know if for example an active policy of *R&D subsidy* in an economy will have durable and *permanent* effects on *growth*, or if its effects will be only *transitory*. A temporary effect depicts an economy that will revert to its original growth rate. But to *dissociate* a *long-term* effect and a *permanent effect* is a *difficult* objective. Such an observation suggests to favor the level effect or transitory effect approach and therefore to favor semi-endogenous models. In this type of model, even if technological progress is endogenous, without the growth of the population which is exogenous, or without the growth of the number of researchers, the growth of the *per capita* product will be interrupted.

---

[1] *Hysteresis:* We can say that there is hysteresis when an effect persists whereas its cause has disappeared.

Moreover, in the economies of developed countries the available empirical elements do not allow us to admit the economic models as linear, i.e. with constant *returns* of *capital* for α equal to 1. On the contrary, α seems close to $\frac{1}{3}$ concerning the capital and even if we add human capital and externalities, the value is close to $\frac{4}{5}$, which is always less than 1. Linearities do not seem to be present in research and development either, since we know that investment efforts in this field and the number of researchers has considerably increased over the last 40 years, whereas the growth rate of economies has not increase proportionally in a significant way, as we know that this growth rate is close to 1.8% over the same period. This tends to show that positive externalities remain lower than 1.

Finally we would like to mention a phenomenon that has been highlighted in the mid-1990s by O.Y. Butkovkii, Y.A. Kravtov and J.S. Brush,[2] and could be generalized to a number of nonlinear models in macroeconomics. This phenomenon shows a new perspective about state transition phases during the occurrence of bifurcations. Called "the bifurcation paradox" by the above mentioned authors, it could reach an interesting dimension, in economic policy in particular. In short, we can say that the final state of a system is predictable if the transition phase is sufficiently fast. In fact, the occurrence probability of a final state depends on two components: the transition speed and the noise level from the considered system. A fast change of the control parameter of a system is particularly recommended in order to get from a stable state to another stable state. Otherwise, if the parameter evolves too slowly, i.e. for a speed that is less than a critical value, the final state of the system corresponds to a chaotic regime. In fact, the probability to forecast the final state of a system after a bifurcation point depends on a relation between the speed of change of the control parameter and the "noise level" of the system. Within this framework, the concept of background noise level emphasizes the relevance of methods brought together under the name of *nonlinear signal processing* as well as the relevance of *spectral analysis methods* and derived methods.

The transition from the linear to the nonlinear field can also find a transcription in signal processing at the time of the transition from traditional spectral analysis to time-frequency analysis, notably through wavelet analysis and hybrid atomic decompositions (i.e. time-frequency atoms and waveforms dictionaries). In such a particular time-frequency analysis framework, the *localized nonrecurring phenomena* such as *singularities*, *shocks*, *intermittencies*, *turbulences* escape traditional spectral analysis that looks for *similarity*. Spectral analysis deals unsatisfactorily with abrupt variations because of error diffusion in the decomposition of this type of phenomenon. It is interesting to notice that *singularities* can be the *expression of fundamental elements* of a dynamics. *Simple and sudden impulses in a dynamics can contain crucial information, as an anomaly in an electrocardiogram for example*. Time-frequency analysis has evolved in order to identify within a complex signal these singularities which make it possible to anticipate risks and, potentially, the future. *To push back the limits of the "unpredictable" requires to deal both with the signal and with the modeling simultaneously*, i.e. *from an empirical point of*

---

[2] "The bifurcation paradox". O.Y. Butkovkii, Y.A. Kravtvov and J.S. Brush. NAG300. 1996.

view and *a theoretical point of view through* the *dynamic modeling of these complex phenomena*, preferably described as being *nonlinear* but which unfortunately tend to escape dynamic modeling. It is also in these specific and *singular behaviors* that *a dynamics expresses its own "foundations"*. Information is as important in these singularities as in periodic and recurrent aspects. Abrupt or sudden variations can for example correspond to "level effects" in economics, often described as "transitory effects" in the endogenous growth theory.

Thus, in Signal Processing (and more generally in information transformation methods), *hybrid techniques make sense*; the algorithm of the *best basis* choice is in this respect very interesting.[3] In order to lose nothing of the original signal, *periodicities* must be identified, but *singular nonrecurring events* must be identified also, *as the many elements that characterize a "deterministic" dynamics*. However, the best basis algorithm will not be adapted to non-stationary signals composed of disparate elements. In such a case, the "Matching Pursuit" algorithm seems more adapted. For each segment of the signal, it finds the *wave which resembles it the most*. It does not seek an optimal representation on the entire signal, but it seeks an optimal representation of each characteristic of the signal. S. Mallat explains that this is a question of "finding the best correspondence for each word of the signal, while the best basis algorithm gives the best homologous of the entire sentence". Yves Meyer explains that, whereas only one algorithm (i.e. Fourier analysis) is appropriate for all stationary signals, *transitory signals* (or transient signals) *represent such a rich and complex universe that a unique analysis method cannot win through*. The hybrid methods are currently very advanced, but they will probably be able to evolve in the future towards new tools to improve approximation and reconstruction methods and signal forecast. Construction of an imaging documentary base in time-frequency planes, calibrated on standard academic signals, could be elaborated in such an objective; for example, in order to isolate standard form structures to obtain the reference frameworks that appears so direly absent, in particular in Economics and Finance.

Obviously, we can wonder about the relevance and advisability of signal *representations* in time-frequency planes or via imaging methods. However, we have to explain that direct reading of a signal or an arbitrary time series shows seldom its content and its "secrets", except for simple signals. This is particularly due to the facts that *complex signals generally are non-gaussian* and traditional statistics looks for new probability laws beyond the well-known Gauss law. We think in particular about the *U-shape (arc-sine) law* which characterizes more likely the densities of chaotic nonlinear dynamics. But the *statistical or econometric approach does not show all the information contained in a time series either*. If we choose the examples of *radar signals or medical imaging (magnetic resonance imaging: MRI)*, indeed, it is only after their decomposition in time-frequency planes or via *imaging methods* that these signals make sense, whereas they were illegible in their original states. And we could multiply examples in geophysics and even in vocal analysis where the objective is not to decompose an image. Consequently, systematically testing

---

[3] See Meyer or Coifman bases.

this type of method on financial or economic original time series, whose characteristics permit it (i.e. length of the series, discretization mode,…), seems to be necessary, also to relativize their relevance. Moreover, we can also note that a statistical theory of time-frequency analysis must be completed. This aspect certainly hinders the use of these methods in economics that look for necessary academic validations in order to carry out competence transfers from other scientific fields.

Postface

Why the "Nonlinear"? This postface attempts to propose a non-exhaustive set of reasons to be involved in the nonlinear field:

(1) From the point of view of the *Modeling of Dynamical Systems*
(2) From the point of view of the *Time-Series* or *Signal*

(1) to analyze and model Economics as a complex dynamical system of nonlinear nature, the *"qualitative approach" feeds the analysis* and the *model*; (2) to analyze Economics in accordance with signal theory and mathematics or statistics of complex nonlinear series, where one of the stakes is *Qualitative Extraction* and even reconstruction of Dynamics.

The irruption of Nonlinear theory has led to a profound transformation of many scientific fields. Economics takes part in this revolution. The diversity of possible dynamical behaviors coming from nonlinearities makes it possible to better understand natural and complex phenomena. Henceforth, we better understand the mechanisms at work when the trajectories of dynamics become more and more complex, thus making it possible to better reproduce them, model them and explain them. The items of this set are:

1. The "Nonlinear" is adapted to the description of *natural*, *complex* or *chaotic* phenomena.
2. The "Nonlinear" *depicts* possible periodic and quasi-periodic structures, i.e. *oscillatory structures*, an essential characteristic of the context. It also depicts chaotic behaviors resulting from the loss, disappearance or transformation of these oscillatory behaviors.
3. In the framework of "Stability Boundaries", the "Nonlinear" makes it possible to describe the different known *transitions towards chaotic regimes* (i.e. canonical "routes to chaos"):

   - *Saddle-connections* (i.e. blue-sky catastrophes)
   - *Period-doubling* and cascade of bifurcations, (ddl $\geqslant 1$)
   - *Intermittency* (i.e. explosion)

with as a corollary, the *Coupling of frequency,* sometimes called the *quasiperi-odic route to chaos* (i.e. resonance) [see KAM, Landau, Ruelle–Takens], (ddl ⩾ 3).

4. The possible *oscillatory structures* of the "Nonlinear" and the methods of *time-series decomposition* allow and lead to the use of the following "Representation Spaces".[1]

   - *Complex space* (complex Field, complex and Gauss planes)
   - *Fourier space* (Frequency-Amplitude)
   - *Hilbert space* (Banach space, …) *conceived at the origin for the "oscillating functions"* (Complex exponentials, circular functions, trigonometric functions, etc.). Space where we treat questions of *convergence*, *development into series* (i.e. Taylor series, Laurent series), *analytic or holomorphic functions*, *norms*, *differentiability*, …
   - *Time-frequency space* (and time-scale plane)

5. *Fourier* analysis shows that "*any function (or signal), even discontinuous, can be approximated by Fourier series, i.e. by trigonometric* functions". (It is an evolution of the Weierstrass theorem of function polynomial approximation.)

   - What *allows the reconstruction methods* of signals and leads to their representation in certain *Spaces* quoted above
   - However, classical Fourier analysis *hides time* and *phases*

6. In the same way, methods of signal reconstruction by wavelets are stable and complete representations and allow representations in certain spaces mentioned above. Wavelets *do not hide* time and *phases*.

7. In a similar way, "*Hybrid methods*" *of reconstruction* which use different "*wave-forms*" (*Fourier atoms*, *Gabor atoms*, *wavelet atoms, Dirac functions and time-frequency atoms, …*) *allow* representations in certain *Spaces* mentioned above (i.e. time-frequency planes).

8. *Polyspectra* (e.g. *Bicoherence*) are also *more adapted to the study of (non-Gaussian) nonlinear signals and turbulence phenomena than traditional spectra.* There is also the "wavelet Bicoherence" of which we know (because it does not hide the phases) that it *highlights* "*Coupling of Phases*" *in the turbulence phenomena* and associated "Coherent Structures" (ref. Milligen).

9. The nonlinear field concerns *Topological Invariants* (and concepts associated with the differentiable topologies, differentiable manifolds, Riemannian manifolds [i.e. non-Euclidean], etc.).

10. In the nonlinear field, the *(isolated or essential**) singularities* and the *exceptional points* (in this *oscillatory and complex nonlinear context*), which concern the analytical functions and the fields of series convergence, make sense and *are a major question*. They are better described in such a framework, i.e. within the framework of *complex continuation* rather than the framework of

---

[1] The framework used in such a case is more general than the framework of the Representation theory in Mathematics. *Representation theory*: this theory is composed of three subcategories: (C) Clifford algebras, (I) Invariant theory, (R) Representation theory of groups.

functions of the real variable (see Cauchy, Laurent, Weierstrass, Riemann). Singularities: There is an equivalence between the holomorphic functions and the functions expandable in entire series. The study of the fields of convergence of the entire series has led (among other things) to the concept of "singularity", which corresponds to exceptional points where the holomorphic continuation is not possible. Among the singularities of a function, some can disappear if one extends in a suitable way the concept of function, in contrast with the singularities where this is not possible, they are called **Essential Singularity*.

11. The concept of *Singularity* is in connection with the following concepts: The (Lipschitz) *Regularity*, Self-similarity, MBf, LRD, Process with memory, Multi-fractals, ….

12. This leads to the notion of *Measurement*, which is more *relevant* and open within the preceding framework (i.e. From the *Euclidean point* towards the *complex point* or towards the *point of a differentiable Manifold*, etc. See also Hilbert space, Time-frequency space).

13. *In Economics,* we often meet *systems with very high degrees of freedom* (even infinite) but *with attractors of limited dimension.* Thus, while working in *reconstructed systems of low dimensions,* we can obtain relevant information by observing *geometrical objects of low dimension* (see Takens, Poincaré, …).

   - Equilibrium Cycles, Overlapping Models, Optimal Growth, …

14. *In Finance*, in an analogous way, *stock market signals have non-integer "capacity dimensions" lower than the (probably infinite) degree of freedom of the system*. Thus, we are often faced with an attractor (see Hausdorff, Poincaré, Mandelbrot, …).

The reader will be able to improve and develop this incomplete open list, which, moreover, highlights the importance of certain disciplines to promote them in Economics and Finance.

Appendix A
Mathematics

A.1 Relations, Metrics, Topological Structures

In this section, devoted to the topological relations and structures, are presented in particular: Homeomorphisms, diffeomorphisms, adherent points, accumulation points, connected sets, collineations (or correlations), projective planes, orthogonal and orthonormal systems, metric spaces, Hausdorff measure, Hausdorff–Besivitch dimension (i.e. fractal dimension), etc.

The Bourbakist conception of mathematics made it possible to highlight the compartmentation which existed between the different mathematical disciplines. The study of axiomatic constructions of the different theories shows common fundamental structures. The traditional disciplines develop then from these *fundamental structures* and from the *mixed structures* (also called *multiple structures*) themselves built from the fundamental structures.

Fundamental structures: The three fundamental structures are the *algebraic structure*, the *order structure* and the *topological structure*. The *multiple structures* (also called *mixed structures*) are made up of several of the three fundamental structures. Examples of *multiple structures* are the *topological groups*, *topological vector spaces*, *ordered fields*.

(1) *Algebraic structure:* A set can be provided with an *algebraic structure* if one or several composition laws (internal or external) are defined on this set, such as addition and multiplication in the number set, or the multiplication of a vector with a scalar. The most important algebraic structures are the *semigroup*, the *group*, the *ring*, the *field*, the *module* and the *vector space*. *A module is a vector space in which the scalars are a ring rather than a field.*

(2) *Order structure:* A set can be provided with an *order structure* if an order relation is defined on this set. This means that in this set there are comparable elements in accordance with a predefined rule, as for example in the real number set by means of the relation "$\leqslant$". Examples of order structures are: *partially ordered sets (ordered sets), totally ordered sets, well ordered sets, inductive sets.*
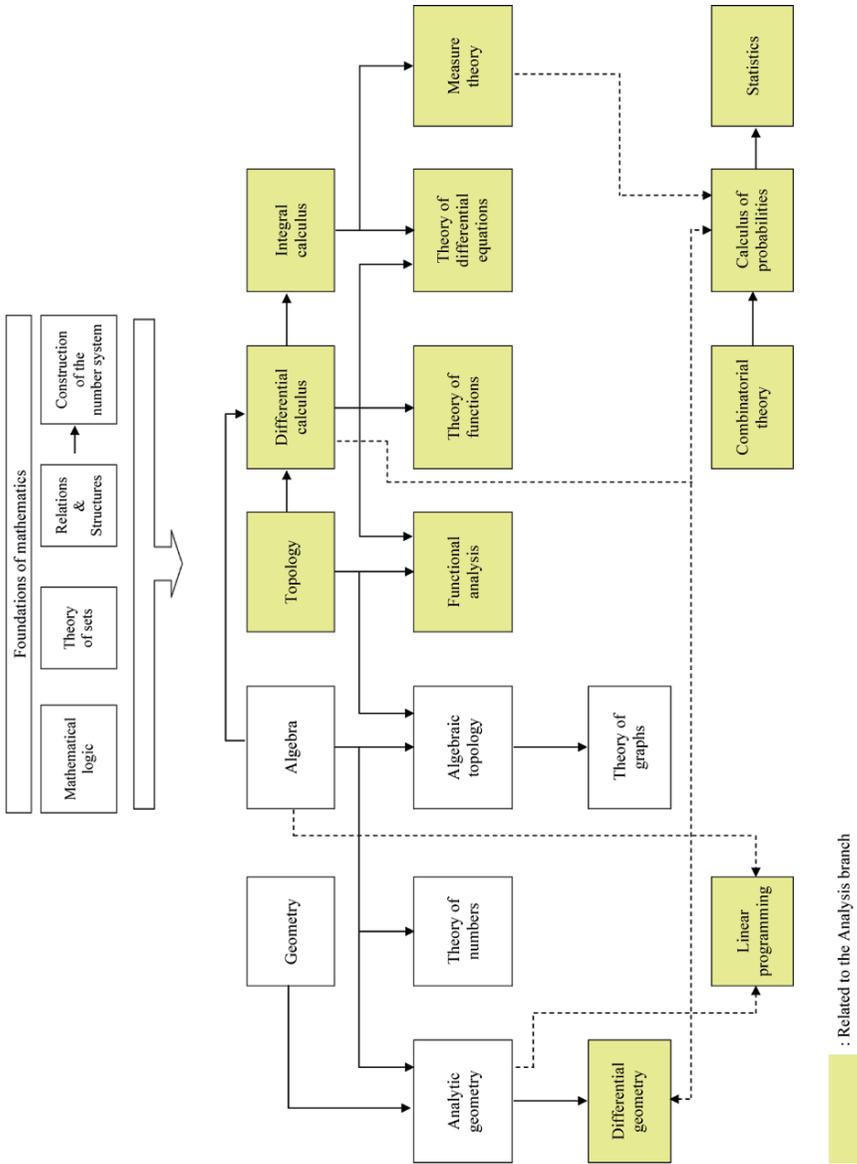
Fig. A.1 Branches of Mathematics

(3) *Topological structures:* A set can be provided with a *topological structure* if we have chosen in this set a system $\mathfrak{F}$ of subsets verifying different properties. The topological structure is very important to define the concept of *convergence*. A set endowed with a topological structure is called *topological space*.

System of relations: The fundamental structures can be understood by means of *relations* which produce these structures. Then, a structured set is a set S on which is defined a relations family $\{R_i\}$. If S is provided with relations $R_1, \ldots, R_n$, then the associated space is written $\{S; R_1, \ldots, R_n\}$. And the couple made with S and $R_1, \ldots, R_n$ is called a *system of relations*. The relations which provide the structures, when $n \geqslant 2$, must be compatible between them. The conditions of compatibility must be carefully chosen to construct a theory (see distributivity to construct the rings).

A.1.1 Relations and Diffeomorphisms

Before giving the definitions of maps until the diffeomorphism, we recall an extremely useful elementary notion of the set theory which is the partition.

Definition A.1 (Partition). If $\bigcup_{i \in I} A_i = F$ and if all the A_i are assumed $\neq \varnothing$ and pairwise disjoint (i.e. mutually disjoint), the set $\{A_i\}$ is called a partition of F.

The *main partitions* in mathematics are obtained in the form of *equivalence classes*[1] according to an *equivalence relation*.[2] Thus, the vectors or the rational numbers and many others mathematical objects are defined in the form of equivalence classes that we will describe in a next section. Thus, in the geometrical constructions, it is simply essential to put in place a representative of a class of given congruent figures.

The algebraic operations on sets exist in all the mathematical fields. In algebra, the set of solutions of an equation system is the intersection of the set of solutions of each equation. There exists a graphical method which makes it possible to visualize this principle. Indeed, *the graphical method visualizes the intersection of solutions sets of each equation as a common section of the graphs of functions associated with each equation.* A simple way to define a map can be given by the following definition:

Map (or Mapping): A way of associating single objects to every point in a given set. So a map $f : A \to B$ from A to B is a function f such that for every $a \in A$, there is

---

[1] *Equivalence classes:* The equivalence classes are the collection of pairwise disjoint subsets determined by an *equivalence relation* on a set. Thus, two elements are in the same equivalence class if and only if they are equivalent under the given relation. **Definition (Equivalence class).** The set of all elements of the set S related to $x \in S$ is called equivalence class $[[x]]$ of $x : [[x]] = \{z | x \in S \wedge z \in S \wedge xRz\}$.

[2] *Equivalence relation:* A relation is called an equivalence relation, if and only if, this relation is reflexive, symmetric and transitive.

a single object $f(a) \in B$. The terms function (sometimes improperly) and mapping are taken as synonymous for map.

Definition A.2 (Map). A relation $f \subseteq E \times F$ left-surjective and univocal is called a map.

Definition A.3 (Surjective map). $f : E \to F$ is said surjective (i.e. f is a surjection) if $f[E] = F$.

Definition A.4 (Injective map). $f : E \to F$ is said injective (i.e. f is an injection) if for any $y \in F$, $f^{-1}[\{y\}] = \{x\}$, i.e. is a un singleton or $f^{-1}[\{y\}] = \emptyset$.

Definition A.5 (Bijective map). $f : E \to F$ is said bijective, if f is surjective and injective. The isomorphisms are bijective maps particularly important.
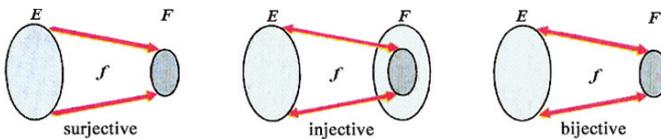
Isomorphisms: Two spaces of similar structures are regarded as equivalents, relative to the considered structures, if there is between them a bijective morphism, whose reciprocal map is also a morphism. Then we are faced with an *isomorphism*, and two spaces are known as *isomorphic*.

Automorphisms: The isomorphisms of a structured set on itself is called an automorphism. The set of all automorphisms of a structured set, provided with the composition of maps, is a Group (i.e. group of automorphisms). For example in Galois theory this type of group is important.

Topological Invariants: In geometry the objective of the topology is to study the classes of topologically equivalent points sets (see figures hereafter giving some basic examples) and to describe their representatives using properties which remain unchanged by *homeomorphism* (*topological invariants*). The invariants by continuous map, i.e. the continuous invariants are also very important.
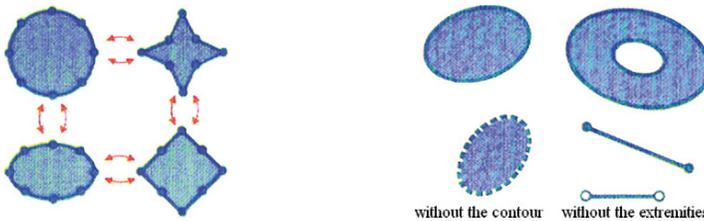
Set of topologically equivalent points   Set of topologically non-equivalent points

A function $f(x,y,z,\ldots)$ of n variables with real values can be taken as a map of a part of the space with n dimensions noted $\mathbb{R}^n$, in the set of Real numbers $\mathbb{R}$, $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. One can consider for example j functions of n variables, thus the elaborated map would be from $\mathbb{R}^n$ to $\mathbb{R}^j$.

Definition A.6 (Homeomorphism). A map f from a part A of $\mathbb{R}^n$ to a part B of $\mathbb{R}^j$ is a *homeomorphism* from A to B, if it is invertible and bicontinuous i.e. continuous for f and for the f^{-1} inverse function from B to A.

Another manner of providing a homeomorphism definition is as follows:

Definition A.7 (Homeomorphism). A map f is a homeomorphism from A to B if it is injective and continuous and if its inverse function is itself continuous.

In short, the idea of continuity concerning the topology of a space with n dimensions (defined from the Euclidean distance) requires that the parts A and B include a neighborhood of each one of their points. Thus, let us imagine a circle inside an equilateral triangle with C a center of gravity which is confused with the center of the circle. It is logical of saying that any half-line starting from C cuts the circle in M and the triangle in M'. Thus the map $f : M \longrightarrow M'$ is such that *at each point of the circle corresponds a point of the triangle* and *reciprocally* (see Fig. A.2a). Two neighbor points of a figure will correspond to two neighbor points of the other figure. The map f is said to be a *homeomorphism*. Thus the *circle* and *triangle are homeomorphic. Then a homeomorphism is defined as an invertible map and bicontinuous.* The function $g : \{(t,0)|x \in \mathbb{R}^1\} \to \{(x,y)|x^2 + (y-1)^2 = 1 \wedge y \neq 2\}$ defined by $(t,0) \mapsto \left(\frac{4t}{t^2+4}, \frac{2t^2}{t^2+4} \right)$, verifies: (1) g is bijective (2) g is continuous (3) g^{-1} is continuous, then this function is a *homeomorphism*. A straight-line in $\mathbb{R}^2$ is topologically equivalent to a circumference deprived of one of its points (see Fig. A.2b).

A *diffeomorphism is more restrictive than a homeomorphism (A diffeomorphism is an abbreviation for transformation which is one-to-one, unique, C^1 and of which the inverse exists and is also C^1).*

Definition A.8 (Diffeomorphism). The f map is a diffeomorphism from A to B if it is invertible and admits on A continuous partial-derivatives of order 1 (and the same on B for the reciprocal map f^{-1}).

Briefly, it is possible to say that a diffeomorphism is a map between manifolds which is differentiable and has a differentiable inverse. Another manner of giving a definition is as follows:
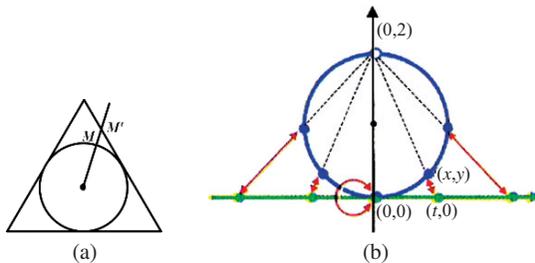
Fig. A.2 **(a)** Homeomorphism f, **(b)** Homeomorphism g

Definition A.9 (Diffeomorphism). The map f is a *diffeomorphism* from A to B, if it is a *homeomorphism and if f and the inverse function f^{-1} are differentiable*. Any diffeomorphism is a homeomorphism, but the converse is false.

A function f of a real variable defined on an interval I, for which it admits a continuous derivative and of a constant sign, is a diffeomorphism of I on the image interval $f(I)$. We notice that the *map* from a *circle $\longmapsto$ a triangle* $(f : M \longmapsto M')$, mentioned above is not a *diffeomorphism because of discontinuities in the derivative at the passing of the vertices of the triangle*. The map from a *circle towards* an *ellipse* is by contrast a *diffeomorphism*. The bijective mapping of the circle towards a segment [0;1[is a diffeomorphism.

A.1.2 Metric Spaces and Topological Spaces

A.1.2.1 Preamble: From Adherent Points Towards Compact Sets

If a subset $E \subset \mathbb{R}^p$ is provided with the *topology induced* by $\mathbb{R}^p$ and if A is a part of E, the topological situation of a point P of E in relation to A can be specified in the following way:

Definition A.10 (Adherent point). P is an *adherent point* of A, if any neighborhood of P meets A; in particular any point of A adheres to A, but there can be different ones.[3]

Definition A.11 (Exterior point). A point P, which is not an adherent point of A, is an exterior point of A.

Definition A.12 (Interior point). A point P is an *interior point of A* if P has a neighborhood included in A.

Definition A.13 (Isolated point). A point P is an isolated point of A if there exists a neighborhood of P which meets A only in P.

Definition A.14 (Accumulation point, or *cluster point,* or *limit point*). A point P is an accumulation point of A if any neighborhood of P meets A elsewhere than in P.

Definition A.15 (Frontier point, or boundary point). A point P is a *frontier point* of A if any neighborhood of P meets A and $E \backslash A$.[4]

Definition A.16 (Open set). A set only constituted of interior points is called an open set (e.g. a disc without its circumference).

---

[3] An *adherence* is taken and called more usually as a *closure*.

[4] *In mathematical logic:* $A \backslash B : \{x \mid x \in A \wedge x \notin B\}$, knowing that $\wedge, \vee$ mean: "and", "or". (e.g. $E \wedge d(x,m) < \varepsilon$, means that E is provided with a distance $d(x,m) < \varepsilon$).

Definition A.17 (Closed set). A set is called a closed set if its complementary set in E is an open set.[5] (By contrast "closed" is not equivalent to "non-opened".)

Definition A.18 (Frontier set). The set of frontier points of A, called frontier ∂A of A, is exactly the set of points of A which are neither interior nor exterior of A. i.e. there is $\partial A = \partial(E \backslash A)$. A frontier is closed.

Set M

Complement of M

Set of exterior points

P_1, P_2, P_3, P_4 Adherent points

P_5 Exterior point

P_1 Interior point.

P_2, P_3, P_4 Frontier points

P_4 Isolated point

P_1, P_2, P_3 Accumulation points

Interior M^0
Set of interior points

Closure $\overline{M}$
Set of adherent points

Frontier ∂M
Set of frontier points

Set of accumulation points
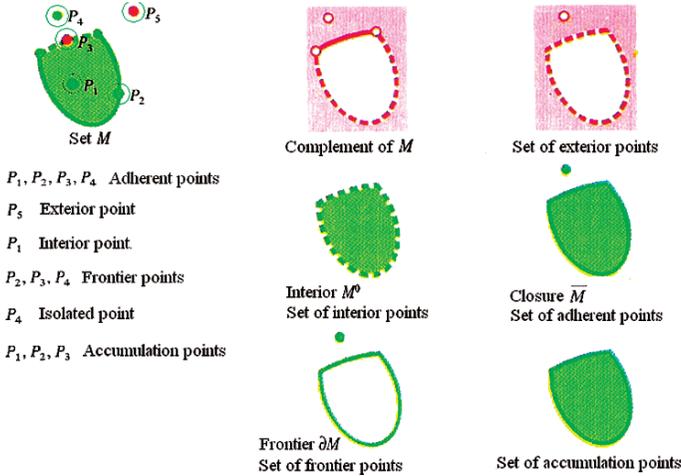
Definition A.19 (Connected set).

1. Two parts A and B of $\mathbb{R}^p$ are separated parts if there is simultaneously $\overline{A} \cap B = \emptyset$ and $A \cap \overline{B} = \emptyset$.
2. A part N of $\mathbb{R}^p$ is a *non-connected* part if we can write it under the following form $N = A \cup B$, where A and B are separated non-empty parts.
3. A part C of $\mathbb{R}^p$ is connected in the contrary case.

On $\mathbb{R}$ the only *connected (set)* are the intervals (bounded or non-bounded) and the singletons. The image of an interval (e.g. $I = [0,1]$) by the continuous map f will be still a connected set. The image $f[I]$ constitutes a *connected link* between the points $f(0)$ and $f(1)$. It is possible to say there is an *arc* between the points $f(0)$ and $f(1)$ and a *closed arc* if $f(0) = f(1)$.

Remark A.1 (Set Closure). The closure of a set A is the smallest closed set containing A. Closed sets are closed under arbitrary intersection, so it is also the intersection of all closed sets containing A. In particular, it is just A with all of its accumulation points. The term "closure" is also used to refer to a "closed" version of a given set. The closure of a set can be defined in several equivalent ways, including: (1) The set plus its "limit points", also called "boundary points", the union of which is also

---

[5] *Complementary set* of A in E, i.e. the *complement* of A in E. Since A is a subset of the set E, the *complementary set* of A is also a subset of E. Therefore, the *complementary set* of A is the *complement* of A in E.
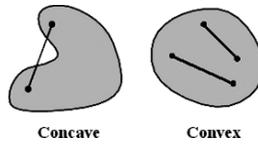
Fig. A.3 Convexity
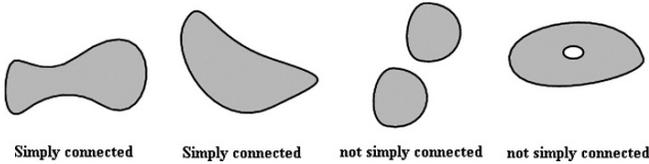
Fig. A.4 Simply and not simply connected

called the "frontier". (2) The single smallest "closed set" containing the given set. (3) The complement of the interior of the complement of the set. (4) The collection of all points such that every neighborhood of these points intersects the original set in a nonempty set Figs. A.3, A.4 show resp. convex and simply connected sets.

Connected: Connectivity properties respect the following hierarchy: *(a) Convex* $\Rightarrow$ *(b) Star convex* $\Rightarrow$ *(c) Pathwise-connected* $\Rightarrow$ *(d) Connected.*

Definition A.20 (Compact subset). The subsets of $\mathbb{R}^p$ which are at the same time closed and bounded are called compact.

Continuous invariant: The property "closed-bounded" in $\mathbb{R}^p$ is a *continuous invariant*. If we consider the image by a continuous map f of a segment I in $\mathbb{R}$ or more generally in $\mathbb{R}^p$, this image is connected but is also closed bounded; A part of $\mathbb{R}^p$ is bounded if it is contained in an open ball.

Theorem A.1 (Compact of $\mathbb{R}^n$). *A part of* $(\mathbb{R}^n, \mathfrak{R}^n)$ *is a compact if and only if it is closed bounded.*

Jordan curve: Also called *closed Jordan curve*. A Jordan curve is a plane curve which is topologically equivalent to a homeomorphic image of the unit circle, i.e. it is simple[6] and closed[7] curve.

A.1.2.2 Topological Spaces

Definition A.21 (Topological spaces). $(E, \mathfrak{T})$ is a topological space if $\mathfrak{T}$ is a subset of $\mathfrak{B}(E)$ having the following properties:

---

[6] *Simple curve:* A curve is simple if it does not cross itself.

[7] *Closed curve:* In the plane, a closed curve is a curve with no endpoints and which completely encloses an area.

$$(1): \emptyset \in \mathfrak{T}, E \in \mathfrak{T}, \tag{A.1}$$

$$(2): O_1, O_2 \in \mathfrak{T} \Rightarrow O_1 \cap O_2 \in \mathfrak{T}, \tag{A.2}$$

$$(3): \mathfrak{T}' \subseteq \mathfrak{T} \Rightarrow \bigcup_{O \in \mathfrak{T}'} O \in \mathfrak{T}. \tag{A.3}$$

$\mathfrak{T}$ is called a topology on E. The elements of $\mathfrak{T}$ are called "open", the elements of E are called "points".

Definition A.22 (Neighborhood). V is called neighborhood of x if $V \subseteq E$ and if there exists one $O \in \mathfrak{T}$ such that $x \in O \subseteq V$. One indicates by $\mathfrak{B}(x)$ the set of all neighborhoods of x. Moreover it is known that $\mathfrak{B}(x) \neq \emptyset$, because $E \in \mathfrak{B}(x)$.

A topological space can be also defined using axioms on the neighborhood, and the concept of open set results from this approach. With the neighborhood concept we can define the convergence of a sequence:

Definition A.23 (Convergence of a sequence). Let $(a_0, a_1, \ldots)$ be a sequence in a topological space $(E, \mathfrak{T})$. It is said that the sequence converges towards $a \in E$ (noted $\lim_{n \to \infty} a_n = a$) if for any neighborhood $V \in \mathfrak{B}(x)$ there exists one $n_0 \in \mathbb{N}$ such that $n \geqslant n_0 \Rightarrow a_n \in V$.

In a topological space the limit is not necessarily unique. However if for any couple of distinct points of E it is possible to find two disjoint "open sets" containing them (separated space in the Hausdorff sense), then the limit is single. Metric spaces are separated spaces.

Metric spaces: The metric spaces are a generalization of Euclidean spaces. Like Euclidean spaces they admit a "topology" defined by a metric.

A.1.2.3 Metric Spaces: Distance, Ball, Convergence and Haussdorff Measure

Definition A.24 (Metric space). A set E is a metric space if one defined a distance on E, i.e. if there is a map $d : E \times E \to \mathbb{R}_+$ verifying the properties:

$$(1): d(x,y) = 0 \Leftrightarrow x = y, \tag{A.4}$$

$$(2): d(x,y) = d(y,x), \tag{A.5}$$

$$(3): d(x,y) + d(y,z) \geq d(x,z). \tag{A.6}$$

The spaces $\mathbb{Q}$ and $\mathbb{R}$ are metric spaces for the distance defined by the following absolute value:

$$d(x,y) := |x - y|. \tag{A.7}$$

Definition A.25 (Transformation of a metric space into topological space). Any metric space can become a topological space. To this end does, one defines an open ball of center m and radius $\varepsilon > 0$ by:[8]

---

[8] *Logic:* $\wedge, \vee$: "and", "or". (e.g. $E \wedge d(x,m) < \varepsilon$, means that E is provided with a distance $d(x,m) < \varepsilon$).

$$B(m,\varepsilon) := \{x \mid x \in E \wedge d(x,m) < \varepsilon\}, \tag{A.8}$$

and it is said that a non-empty subset of E is open if it contains at least a centered open ball at each one of its points.

Remark A.2 (Topological spaces, Balls and neighborhoods). The open balls have a particular property in such a space, indeed they are themselves open sets, and thus it is possible to say that: "Any non-empty open set is a union of open balls". Consequently, rather than *the set of neighborhoods of a point*, it suffices to consider *the set of open-balls centered at this point*. Taking into account what was just stated, we can rewrite the definition of the convergence of a sequence by means of balls.

Definition A.26 (Convergence of a sequence by means of balls). Let $(a_0, a_1, \ldots)$ be a sequence in a topological space $(E, \mathfrak{T})$. It is said that sequence *converges* towards $a \in E$, as soon as for any open ball $B(a, \varepsilon)$ there exists a $n_0 \in \mathbb{N}$ such that $n \geqslant n_0 \Rightarrow a_n \in B(a, \varepsilon)$, i.e. if we have:

$$\forall \varepsilon (\varepsilon \in \mathbb{R}_+^* \Rightarrow \exists n_0 \forall n (n \geqslant n_0 \Rightarrow d(a_n, a) < \varepsilon)). \tag{A.9}$$

Any *metric space* is *separated*.

The oriented plane as a topological space: The *set of intersections of a finite number of open half-planes* constitutes a sub-basis in the topological sense, which makes it possible to provide any oriented plane with a topological structure called a *natural topology*. (Any natural topology satisfies the Hausdorff axiom of separation.)

It is possible to *use* this simple topological space, i.e. the *oriented plane*, to present some fundamental concepts that we can then transpose in more complex spaces such as the vector spaces.

The metric plane: The *point* and the *straight line* correspond to elementary geometrical concepts appearing in the axioms where they are implicitly defined. Let Π be the set of points and Γ the set of straight-lines (and respectively the points themselves are noted P and the straight-lines noted g).

We define the relation $\mathbf{I}$ (i.e. *incident to*) in $\Pi \times \Gamma$, and the relation $\perp$ (i.e. *orthogonal* or *perpendicular to*) in $\Gamma \times \Gamma$. And if we note g an arbitrary line, we can write $P\mathbf{I}g$ which means that P is on g or that p passes by P.

Definition A.27 (Projective plane). A set of points and lines are called "incidence projective plane" when it satisfies:

1. For any pair of distinct points A and B, there exists one and only one line g incident to A and B.
2. For any pair of distinct lines g and h, there exists one and only one point A incident to g and h.
3. There are four points whose three unspecified which are not incident to a same line.

Theorem A.2 (Single projective transformation). *There is one and only one projective transformation of a rectilinear division into itself or into another rectilinear*

division, transforming three given points pairwise distinct into three given points pairwise distinct.

Definition A.28 (Orthogonal collineation). A bijective map σ from $\Pi \cup \Gamma$ on itself such that $\sigma[\Pi] = \Pi$ and $\sigma[\Gamma] = \Gamma$ is called *orthogonal collineation* when it is compatible with the relations $\mathbf{I}$ and $\perp$, i.e. $P\mathbf{I}g \Rightarrow \sigma[P]\mathbf{I}\sigma[g]$ and $a \perp b \Rightarrow \sigma[a] \perp \sigma[b]$.

Definition A.29 (Collineation). A bijective map σ is called a *collineation* when it is compatible with the relations $\mathbf{I}$ and verifies only $\sigma[\Pi] = \Pi$ and $\sigma[\Gamma] = \Gamma$.

Definition A.30 (Collineation, or correlation, of a projective plane). If Π and Γ are respectively the set of points and the set of lines of a projective plane, a collineation (or correlation) of this one is a transformation which applies in a bijective way Π on Π and Γ on Γ by preserving the incidence.

Orientation: Given an *Ordered Field* K of co-ordinates, there are two *order relations* $\leq$,[9] and only two, ordering the points of an unspecified line noted g. Consequently it is possible to say that the line is oriented, it is noted $(g, \leq)$. The orientation of lines makes possible *to orient the plane*.

The oriented plane as a topological space: The set of intersections of a finite number of open half-planes constitutes a subbasis in the topological sense, which makes it possible to provide any oriented plane with a topological structure called *natural topology*.

A.1.2.4 Elements of Analytical Geometry

Canonical Scalar Product in the Vector Space $\mathbb{R}^n$

Given $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ elements of $\mathbb{R}^n$, the map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$:

$$(x, y) \rightarrow \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i \qquad (A.10)$$

is *bilinear symmetrical*. $\langle x, y \rangle$ corresponds to the canonical scalar product of two vectors x, y. It is possible to pose $|x| = \sqrt{\langle x, x \rangle}$, a *vector x* is called *unit* if $|x| = 1$.

Theorem A.3. *The absolute value of the product of the two unit vectors is* ≤ 1. *From* $(|x_i| - |y_i|)^2 \geq 0 \Leftrightarrow x_i^2 - y_i^2 \geq 2|x_i y_i|$ *we obtain:* $2 = \sum_{i=1}^{n}(x_i^2 - y_i^2) \geq 2\sum_{i=1}^{n}|x_i y_i| \geq 2|x \cdot y|$, *and we obtain:* $\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n |\langle x, y \rangle| \leq \sqrt{\langle x, x \rangle \langle y, y \rangle}$ *which corresponds to the Cauchy–Schwartz inequality.*

---

[9] *Order relation* is equivalent to *ordering*.

Canonical Euclidean Norm

It is said that the map $x \rightarrow |x|$ is a (Euclidean) norm on the *vector space* $\mathbb{R}^n$ and the map $(x, y) \rightarrow |x - y|$ is thus a (Euclidean) distance on *affine space* $\mathbb{R}^n : \mathbb{R}^n$ is a *metric space*. From $|x \cdot y| \leq |x| \cdot |y|$ we obtain $x^2 + y^2 + 2xy \leq x^2 + y^2 + 2|x||y|$, and then $(x + y)^2 \leq (|x| + |y|)^2$.

Angle Measures and Orthogonal Vectors

The inequality of Cauchy–Schwarz $\dfrac{|x \cdot y|}{|x||y|} \leq 1$ for non-null values of x and y makes it possible to define one and only one angle $\theta \in [0, \pi]$ such that:

$$cos(\theta) = \frac{|x \cdot y|}{|x||y|}, \tag{A.11}$$

where θ is the angle of two non-null vectors x and y. And the relation $x \cdot y = 0$ means that either one of the two vectors is null or that the angle $(x, y) = \pi/2$, thus (x, y) are *orthogonal*.

A.1.2.5 Similitudes, Homotheties and Orthogonal Systems

Definition A.31 (Homothety). A *homothety* is a collineation π which verifies $\pi(g) \parallel g$ for any line g and having at least a fixed-point Z, called center of the homothety.[10] From $\pi(g) \parallel g$ it is deduced that π is an orthogonal collineation, and that a point and its image are aligned with Z.

Definition A.32 (Similitudes and homothetic figures). Two figures are known as homothetic when one is the image of the other by at least a homothety.

Definition A.33 (Orthogonal system). A family of vectors $(v_1, v_2, \ldots, v_s)$ of $\mathbb{R}^n$ is known as an orthogonal system if for all (i, j), $i \neq j$, $v_i \cdot v_j = 0$.

Definition A.34 (Orthonormal system). A family of vectors $(v_1, v_2, \ldots, v_s)$ of $\mathbb{R}^n$ is known as an orthonormal system if $\forall i, j$, $v_i \cdot v_j = \delta_{ij}$ (The Kronecker symbol: $\delta_{ij} = 0$ if $i \neq j$, $\delta_{ij} = 1$ if $i = j$).

The *canonical basis* of $\mathbb{R}^n$ is orthonormal. Any orthogonal system of non-null vectors is *free*. It contains at most n vectors. If $(v_1, v_2, \ldots, v_s)$ is a basis of a vectorial subspace of $\mathbb{R}^n$, we can by means of successive operations carried out on the v_i to construct an orthonormal basis of this subspace (i.e. Schmidt Orthonormalization).

---

[10] $\parallel$: Parallel.

A.1.2.6 Metric Spaces, Hausdorff Measure and Fractal Dimension

Metric spaces constitute an important class of topological spaces:

Definition A.35 (Metric space). Let M be a non-empty set, (M,d) is metric *space* if d is a map of $M \times M$ in $\mathbb{R}^+$ (called a metric on M) verifying the properties:

$$(1): \quad d(x,y) = 0 \Leftrightarrow x = y, \qquad (A.12)$$
$$(2): \quad \forall(x,y), d(x,y) = d(y,x), \qquad (A.13)$$
$$(3): \quad \forall(x,y,z), d(x,y) + d(y,z) \geq d(x,z), \qquad (A.14)$$

where $d(x,y)$ is the distance of x and y.

- It is possible to provide $\mathbb{R}^n$ with the *distance:*

$$d_E(x,y) = \sqrt{\sum_{v=1}^{n} (x_v - y_v)^2}, \qquad (A.15)$$

between two points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$: d_E is the *Euclidean canonical metric*. And in $\mathbb{R}$ one finds $d_E = |x - y|$.
- Other *distances* on $\mathbb{R}^n$:

$$d_1(x,y) = \sum_{v=1}^{n} |x_v - y_v|^2, \qquad (A.16)$$

$$d_\infty(x,y) = \sup\{|x_v - y_v| \mid v = 1, \ldots, n\}. \qquad (A.17)$$

As explained previously, *any metric space can become a topological space.*

Definition A.36 (Separation). Let $(E,\mathfrak{T})$ be a *topological space*. The parts A and B of E are known as separated if $\overline{A} \cap B = \emptyset$ and $A \cap \overline{B} = \emptyset$.

Definition A.37 (Non-connection). Let $(E,\mathfrak{T})$ be a topological space, this space is known as non-connected if it is the union of two separated non-empty parts.

Definition A.38 (Connection). A subset of a topological space is known as connected if it is connected for the induced topology.

Theorem A.4 (Connected topological space). *A topological space $(E,\mathfrak{T})$ is connected if and only if it contains exactly two subsets which are at the same time open and closed $(E$ and $\emptyset)$.*

Theorem A.5 (Connected image). *The image of a connected space by a continuous map is connected.*

Any *metric space* provided *with* a *metric topology* is known as a *Hausdorff space*.

Definition A.39 (Internal homothety). A subset F of an Euclidean space is known as a internal homothety if there is a partition of F of which all the parts are similar to F.

Definition A.40 (Hausdorff measure). Let F be a non-empty bounded part of a metric space and a real number $\alpha > 0$. One designates by "Hausdorff exterior α-measure" of F the real $m_\alpha(F) = \lim\limits_{\varepsilon \to 0} \left(\inf \sum\limits_{i \in I} d_i^\alpha \right)$, where "inf" is taken on all the (indexed) coverings of F by the balls of diameter $d_i \leq \varepsilon$.

Hausdorff–Besicovitch dimension (or fractal dimension): For F the *real* $m_\alpha(F) = +\infty$ if $\alpha < \alpha_0$ and $m_\alpha(F) = 0$ si $\alpha > \alpha_0$. It is Besicovitch which showed the existence of such a *real number* α_0 for any bounded part F of an Euclidean space. (When F is a regular injective *nappe*[11] in an Euclidian space, the fractal dimension coincides with the topological dimension and is equal to 2. When F is a *regular Jordan arc*, the fractal dimension coincides with the topological dimension and is equal to 1.)

Any bounded part of a metric space whose fractal dimension is strictly higher than its topological dimension is a *fractal set*.

Theorem A.6 (Fractal dimension). *If F is a bounded subset of an Euclidean space of finite size with internal homothety for a regular finite partition, then the following expression is equal to its fractal dimension:* $\frac{-\log n}{\log r}$.

The fractal dimension of a *rectangular solid* (i.e. a rectangular parallelepiped or a *cuboid*) for example is equal to its topological dimension, whereas the fractal dimension of the Von Koch curve is equal to 1.26. which is higher than its topological dimension. We are thus faced with a *fractal set*. The *Cantor set* is also a *fractal set* and its dimension is equal to: $\frac{\log 2}{\log 3} = 0.63 \ldots$.

Fractal Geometry

In fractal geometry the Mandelbrot set is compact and connected, and the *fractal dimension* of its *frontier* is 2 (see Shishikura). This Mandelbrot set corresponds to a filled Julia set. It is written by the following complex sequence: $z_{n+1} = z_n^2 + c, z_0 = 0$, c is fixed, element of $\mathbb{C}$.

---

[11] *Nappe:* one of the two pieces of a double cone, i.e. two cones placed apex to apex:
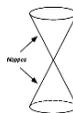
(*Remark: Hyperbolic* plane is a *differentiable manifold* which can be provided with a *Riemannian structure*).

A.2 PreHilbert, Hilbert and Banach Spaces

A.2.1 Normed Spaces

In order to use the methods of the *functional analysis* in a *vector space*, it is necessary *to provide it with a topological structure*. The normed spaces are a major illustration of such vector spaces.

Definition A.41 (Normed vector space). $(E, \| \|)$ is called normed vector space if E is vector space on $\mathbb{K}$ (with $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, i.e. $\mathbb{K}$ is either the field $\mathbb{R}$ or the field $\mathbb{C}$) and if one defined on E a norm, i.e. a function $\| \| : E \to \mathbb{R}$, verifying the following properties:

$$\|\mathbf{x}\| = 0 \Rightarrow \mathbf{x} = 0, \tag{A.18}$$

$$\forall \alpha \in \mathbb{K}, \forall \mathbf{x} \in E, \|\alpha \mathbf{x}\| = |\alpha| \, \|\mathbf{x}\|, \tag{A.19}$$

$$\forall \mathbf{x}, \mathbf{y} \in E, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|. \tag{A.20}$$

Provided with this norm, we can define a distance $d_N(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|$. The normed spaces are thus *metrics* in which we can define continuity and "compactness" by means of sequences.

A.2.2 PreHilbert Spaces

In $\mathbb{R}^n$, it is possible to introduce the *Euclidean norm* $\| \|_2$ by means of the scalar product $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. And more generally, it is defined:

$$\forall \mathbf{x}, \mathbf{y} \in E, \langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}, \tag{A.21}$$

$$\forall \mathbf{x}_1, \mathbf{x}_2, \mathbf{y} \in E, \langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{y} \rangle = \overline{\langle \mathbf{x}_1, \mathbf{y} \rangle} + \overline{\langle \mathbf{x}_2, \mathbf{y} \rangle}, \tag{A.22}$$

$$\forall \mathbf{x}, \mathbf{y} \in E, \forall \alpha \in \mathbb{K}, \langle \alpha \mathbf{x}, \mathbf{y} \rangle = \bar{\alpha} \, \langle \mathbf{x}, \mathbf{y} \rangle, \tag{A.23}$$

$$\forall \mathbf{x} \neq 0, \langle \mathbf{x}, \mathbf{x} \rangle \in \mathbb{R}_+^*, \tag{A.24}$$

where $\bar{\alpha}$ is the *complex conjugate* of α if $\mathbb{K} = \mathbb{C}$. ($\mathbb{R}_+^*$ is the set of real numbers > 0, whereas $\mathbb{R}_+$ is the set of real numbers ≥ 0 or $\geqslant 0$.)

Theorem A.7 (Norm of a vector space). *In any vector space provided with a scalar product, it is possible to define a norm by:*

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}. \tag{A.25}$$

Definition A.42 (PreHilbert Space). A space provided with a scalar product is called PreHilbert space. It is normed space whose norm is defined using the scalar product.

It is possible to give the following example: $\mathbb{C}^n$ with $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{\nu=1}^{\infty} \bar{x}_\nu y_\nu$.

A.2.3 Banach Spaces and Hilbert Spaces

In a normed space, it is possible to define the *convergence of a sequence*. We indicate $\mathbf{x} \in E$ as the limit of the sequence $(\mathbf{x})$ if $\forall \varepsilon \in \mathbb{R}^*_+, \exists n_0 \in \mathbb{N}$ such that $\|\mathbf{x}_m - \mathbf{x}_n\| < \varepsilon$ as soon as $n \geq n_0$. It is possible also to define *Cauchy sequences*, which are given by the following properties: $\forall \varepsilon \in \mathbb{R}^*_+, \exists n_1 \in \mathbb{N}$ such that $\|\mathbf{x}_m - \mathbf{x}_n\| < \varepsilon$ as soon as $n \geq n_1$ and $m \geq n_1$.

Any convergent sequence is a Cauchy sequence, by contrast the converse is not true in any normed space.

Cantor process and complete spaces: In spite of what precedes, by means of a process which is called the process of Cantor, it is possible from any normed space to build a *complete space* which is its *closure*[12] (in the topological sense).

Definition A.43 (Banach space and Hilbert space). A complete normed vector space (i.e. a normed space in which any Cauchy sequence converges) is called Banach space; a complete PreHilbert space (with the associated norm with the scalar product) is called Hilbert space.

For example, the space $\mathbf{C}^0[a,b]$ can be normed by means of:

$$\|f\|_\infty := \sup \left(\{ |f(t)| \ t \in [a,b] \} \right) \tag{A.26}$$

which is *the uniform convergence norm*. This space $\mathbf{C}^0[a,b]$ provided with this norm is *complete*, it is thus *a Banach space*.

A.2.3.1 $\mathbf{C}^n[a,b]$ Space

It is possible to build Banach spaces whose elements are functions on $[a,b]$ which are continuous and differentiable.

Definition A.44. The space of functions f on $[a,b] \subset \mathbb{R}$, with real values, n continuously differentiable times ($n \in \mathbb{N}\backslash\{0\}$), provided with the internal addition and with the external multiplication by an element of $\mathbb{R}$, is a vector space on $\mathbb{R}$ which is possible to norm by means of: $\|f\|_n := \sup \left(\left\{ \left\| f^{(k)} \right\|_0 \ k \in \{0, 1, \ldots, n\} \right\} \right)$. This space is noted $\mathbf{C}^n[a,b]$, where $f^{(k)}$ indicates the k-th derivative of f.

Theorem A.8. *The spaces $C^n[a,b]$ are Banach spaces.*

A.2.3.2 $\mathbf{L}^p[a,b]$ Space

Definition A.45. The space of the functions f on $[a,b] \subset \mathbb{R}$, with value in $\mathbb{K}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$), whose Lebesgue integral exists for a given $p \in \mathbb{R}^*_+$, i.e. *the*

---

[12] (adherence).

Lebesgue integral:

$$\int_a^b |f(t)|^p \, dt \tag{A.27}$$

exists for $p \in R_+^*$, and is noted $L^p[a,b]$.

A.2.4 Differentiable Operators

In Banach spaces, since they are vector spaces, it is possible to define a linear map from a Banach space to another (with same field $\mathbb{R}$ and $\mathbb{C}$).

Definition A.46 (Linear operator). A linear $F : B_1 \rightarrow B_2$ (B_1, B_2 are two Banach spaces) is called linear operator. F is known as bounded if there exists $c \in \mathbb{R}_+$ such that $\forall x \in B_1$, $\|F(\overrightarrow{x})\| \leq c\|\overrightarrow{x}\|$.

Definition A.47 (Operator). Any map $F : D_F \rightarrow B_2$, where $D_F \subseteq B_1$ (B_1, B_2 being two Banach spaces) is called *operator*. (This definition contains obviously the linear operators.)

Definition A.48 (Differentiable operator). An operator $F : D_F \rightarrow B_2$, where $D_F \subseteq B_1$ is known as *differentiable* at the point $\mathbf{a} \subseteq D_F$ if $\mathbf{a}$ belongs to the interior of D_F and if there exists a bounded linear operator $\dfrac{\delta F}{\delta \mathbf{x}}(\mathbf{a}) \in [B_1, B_2]$ such that we have, if $\mathbf{a} + \mathbf{s} \in D_F$:

$$\lim_{\|\mathbf{s}\| \to 0} \frac{\left\| F(\mathbf{a}+\mathbf{s}) - F(\mathbf{a}) - \dfrac{\delta F}{\delta \mathbf{x}}(\mathbf{a})(\mathbf{s}) \right\|}{\|\mathbf{s}\|} = 0. \tag{A.28}$$

The numerator corresponds to the norm of B_2, *and the denominator corresponds to the norm of B_1.*

A.2.5 Banach Fixed-Point Theorem

Let F be an operator applying a Banach space B on itself. For the functional analysis and in particular for the approximation problems, the existence of at least a fixed-point $\mathbf{x}_0$ under F ($F(\mathbf{x}_0) = \mathbf{x}_0$) is important. The theorems of the fixed-point insure under conditions its existence. We have in particular:

Theorem A.9 (Banach fixed-point). *Given $F : B \rightarrow B$ (B is a Banach space), if there exists $c \in \,]0,1[$ such that $\forall \mathbf{x}_1, \mathbf{x}_2 \in B, \|F(\mathbf{x}_2) - F(\mathbf{x}_1)\| \leq c\|\mathbf{x}_2 - \mathbf{x}_1\|$, F have one and only one fixed point $\mathbf{x}_0$.*

To show it we select a fixed $\mathbf{x}_1 \in D$ and we construct a sequence $\mathbf{x}_{n+1} = F(\mathbf{x}_n)$ ($n \in \mathbb{N}$), then we show that for $n \geq 2$:

$$\|\mathbf{x}_{n+1} - \mathbf{x}_n\| = \|F(\mathbf{x}_n) - F(\mathbf{x}_{n-1})\| \leq c^{n-1} \cdot \|\mathbf{x}_2 - \mathbf{x}_1\|, \tag{A.29}$$

then by applying several times *the triangular inequality*, and by comparing with a geometrical sequence

$$\|\mathbf{x}_{n+1} - \mathbf{x}_n\| \leq c^{n-1} \cdot \frac{\|\mathbf{x}_2 - \mathbf{x}_1\|}{1 - c} \tag{A.30}$$

for any $m > n$. Thus $\{\mathbf{x}_n\}$ form a *Cauchy sequence,* i.e. it *converges* in the Banach space B and the limit is a *fixed-point*. (It is then only necessary to demonstrate the uniqueness of this fixed-point.)

A.2.6 Differential and Integral Operators

Many applications of the *functional analysis* are based on the fact that it is possible to interpret the *differentiation* and *integration* of a function as *linear operators* applied to this function:

- One defines the *differential operator* noted $\mathbf{D} : \mathbf{C}_1[a,b] \rightarrow \mathbf{C}_0[a,b]$ by $\mathbf{D}(f) = f'$. The *linearity* of $\mathbf{D}$ results from the linearity of the derivation.
- One defines the *integral operator* noted $\mathbf{K}$, which transforms a function f into a function $\mathbf{K}(f)$ by integration. Let $\Phi : [a,b]^2 \rightarrow \mathbb{R}$ be a continuous function: it is possible consequently to define an *integral operator* $\mathbf{K}: \mathbf{C}_0[a,b] \rightarrow \mathbf{C}_1[a,b]$ by:

$$\mathbf{K}(f)h = \int_a^b \Phi(h,t) \cdot f(t)dt. \tag{A.31}$$

Such an operator is linear and bounded. Function Φ is called *kernel* of $\mathbf{K}$.

- *Hermitian integral operators:* The set $\mathbf{C}_0([a,b],\mathbb{C})$ of *continuous functions* on [a,b] with *complex values* is vectorial $\mathbb{C}$-space that one can provide with *Hermitian scalar product:*

$$\langle f, g \rangle = \int_a^b \overline{f(t)} g(t)dt. \tag{A.32}$$

The *Hermitian norm* for a continuous function is $\sqrt{\langle f, \bar{f} \rangle}$.

A.3 Complex Number Field, Holomorphic Functions and Singularities

The branch of mathematics called *theory of Analytic functions* is interested in the functions with complex values and complex arguments. The property of linear approximation for the real functions of the real variable, makes it possible to

define the class of differentiable functions, the integrability, the classes of Riemann-integrable functions (respectively Lebesgue-integrable).

By passing from $\mathbb{R}$ to $\mathbb{C}$, in the same way, classes of functions were researched. Thus were introduced for the *complex functions of the complex variable:* The *entire series*[13] (Weierstrass), the *property of integrability* (Cauchy), the *topological properties and complex derivation* (Riemann), and the definition of particular geometrical transformations (Abel).

The set of *all these constructions led to* only one class of functions, that of *holomorphic functions*. The theory of holomorphic functions is remarkable, indeed this theory appeared complete, and provides solutions to many problems that the theory of the functions of the real variable did not make it possible to solve. Often only a *complex continuation*[14] of a function of real variable can give satisfactory information on its behavior.

The *holomorphy* is in particular defined as a *complex derivability*. Since the functions from $\mathbb{C}$ to $\mathbb{C}$ are identifiable to functions from $\mathbb{R}^2$ to $\mathbb{R}^2$, a function whose real components are continuously derivable can be holomorphic (Cauchy–Riemann derivates conditions).

Singularities: The study of the *convergence domains* of *entire series* led (amongst other things) to the *singularity* concept, which corresponds to *exceptional points* where the *holomorphic continuation* is not possible. *Among the singularities of a function, some can disappear if one extends in a suitable way the function notion.* For those where it is not possible, then, they are called *essential singularities*.

The *meromorphic function* concept in the compactified complex plane and the extension of the *domain* to the sets of points of dimension 2, superposed to the complex plane, provide powerful tools for these continuations (i.e. Riemann surface). Only the concept of *Analytic function* on a *Riemann surface* really makes it possible to define the functions which are the object of the theory of functions. Based on this approach, then the *generalized series* appear, such as the *Laurent series*. Within this framework, through examples about the set of *periodic and algebraic functions*, the importance of the *theory of functions* is emphasized. In particular, we emphasize the remarkable properties of the *complex analytic continuation* of the *real exponential function* which proves to be *periodic* and *in close relationship to the circular functions*. Lastly, the *doubling periodic functions (i.e. period doubling*[15]*) make it possible to discover fundamental new properties which cannot not appear (in a classical way) in the case of the real variable functions.*

---

[13] *Entire series:* (i.e. a finite polynomial function). A power series which converges for all values of its variable; a power series with an infinite radius of convergence.

[14] The term "Continuation" is used as "Extension". More generally, the term *continuation* is used in the framework of the "Analytic functions". Analytic continuation (which is sometimes called simply *continuation*) provides a way of extending the domain over which a complex function is defined. The most common application is a *complex analytic function* determined near a point z_c by a *power series continuation:* $\varphi(z) = \sum_{i=0}^{\infty} \alpha_i (z - z_c)^i$. *Power series:* a power series in a variable z is an infinite sum of the form: $\sum_{i=0}^{\infty} \alpha_i z^i$, where α_i are integers, real numbers, complex numbers, or any other quantities.

[15] See also nonlinear functions.

A.3.1 Complex Number

The complex number Field $\mathbb{C}$ was built as an *extension field* of $\mathbb{R}$. The numbers $z \in \mathbb{C}$ are written in the form $z = x + iy$ where x and $y \in \mathbb{R}$ and $i^2 = -1$. x is the real part and y the imaginary part of the number z. This complex number $z = x + iy$ is represented in a *Gauss* plane by the point (x, y) in $\mathbb{R}^2$.

Moivre formula: The complex numbers make it possible to avoid using the sines or cosine and to have one coefficient for each frequency. The Moivre formula is written:

$$e^{i\theta} = \cos\theta + i\sin\theta. \tag{A.33}$$

When the value of the angle θ increases, the number $e^{i\theta}$ moves along the circumference of a circle of radius 1. One defines a *modulus* or the *absolute value* of $z = x + iy$ by posing:

$$|z| = \sqrt{x^2 + y^2}. \tag{A.34}$$

$\mathbb{C}$ is defined as a *metric space* by means of the *norm:*

$$d(z, z') = |z - z'| \tag{A.35}$$

by knowing that: $|z|^2 = z\bar{z}$. In the *Gauss plane*, the ε-neighborhoods of z are *open disks* with center z and radius ε. $\mathbb{C}$ provided with a metric associated with d is a *locally compact* space (but *non-compact*). Moreover $\mathbb{C}$ is a *complete space*, i.e. in $\mathbb{C}$ *any Cauchy sequence converges*. (In the following paragraph it is explained how to compactified $\mathbb{C}$.)

Compactification of $\mathbb{C}$:[16] If one adds a *complementary point* to $\mathbb{C}$, which is called *a point ad infinitum* (i.e. *at infinity*) and is noted ∞, it is possible to *compactify* $\mathbb{C}$ the *locally compact topological space*. It is noted generally: $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. The metric topology of $\widehat{\mathbb{C}}$ is defined by the distance:

$$\hat{d}(\infty, z) = \frac{1}{\sqrt{1 + |z|^2}}, \tag{A.36}$$

where the distance $\hat{d}$ is called the *chordal distance*. All parts of $\widehat{\mathbb{C}}$ which contain ∞ and the exterior of a circle with center 0 in $\mathbb{C}$ belong to the set of neighborhoods of ∞ (which is noted $\uparrow\downarrow(\infty)$), because:

$$|z| > r \Leftrightarrow \hat{d}(z, \infty) < \frac{1}{\sqrt{1 + r^2}}. \tag{A.37}$$

When it is said that the sequence (z_n) defined in $\mathbb{C}$, tends to ∞ in $\widehat{\mathbb{C}}$, i.e. $|z_n|$ is strongly divergent in $\mathbb{R}^+$, this *explains the choice of this term "point at infinity" or*

---

[16] *Compactification:* A compactification of a topological space E is a larger space F containing E which is also compact. The smallest compactification is the one-point compactification (e.g. the real line is not compact. It is contained in the circle, which is obtained by adding a point at infinity. In a same way, the plane is compactified by adding one point at infinity, providing the sphere).

point "ad infinitum", or in an equivalent way $\hat{d}(z_n, \infty) \to 0$. (It is important to say that ∞ cannot be to handle as a complex number.) *An open part of (C, d) respectively $(\widehat{\mathbb{C}}, d)$ is connected if it is not the union of two disjoint non-empty "open sets".*

Definition A.49 (Domain). Such a part is called *Domain*.

A.3.2 Construction of the Field $\mathbb{C}$ of Complex Numbers

The extension of the algebraic structure of $\mathbb{R}$ highlights a crucial property of algebraic closure.

Algebraic closure: The field $\overline{F}$ is called an algebraic closure of F if $\overline{F}$ is algebraic over F and if every polynomial $f(x) \in F[x]$ splits completely over $\overline{F}$, so that $\overline{F}$ can be said to contain all the elements that are algebraic over F. For example, the field of complex numbers $\mathbb{C}$ is the algebraic closure of the field of reals $\mathbb{R}$.

There are expressions which have no meaning in $\mathbb{R}$, e.g. $\sqrt{-1}$ cannot be defined as a real number, i.e. the polynomial $X^2 + 1$ of the *ring of polynomials* which is denoted $\mathbb{R}[X]$ (refer to "Ring of polynomials" in Algebra), does not admit real root, it is irreducible. In Algebra, It is possible to show how starting from an irreducible polynomial of the *ring of polynomials $R[X]$* over an arbitrary commutative field R, it is possible to build an *extension field* in which the aforesaid polynomial admits a root. By implementing this result to the present particular case, it is possible to form the *quotient ring* of $\mathbb{R}[X]$ by the equivalence relation "modulo $X^2 + 1$". To this end, first we define the equivalence of two polynomials $f(X)$ and $g(X)$ of $\mathbb{R}[X]$ by:

Definition A.50 (Equivalence of two polynomials). $f(X)\mathfrak{R}g(X) :\Leftrightarrow \exists h(X), h(X) \in \mathbb{R}[X] \wedge f(X) - g(X) = h(X)(X^2 + 1)$.

Definition A.51 ($\mathbb{C}$ by Ring of polynomials). $\mathbb{C} := \mathbb{R}[X]/\mathfrak{R} = \{[[f(X)]] \,|\, f(X) \in \mathbb{R}[X]\}$.

Remark A.3. Instead of $\mathbb{R}[X]/\mathfrak{R}$ it is also written: $\mathbb{R}[X]/(X^2 + 1)$.

It is possible to define in $\mathbb{C}$ an addition and a multiplication of equivalence classes independently of chosen representatives:

Definition A.52 (Addition, multiplication in the field $\mathbb{C}$). $[[f(X)]] + [[g(X)]] = [[f(X) + g(X)]]$. $[[f(X)]] \cdot [[g(X)]] = [[f(X) \cdot g(X)]]$.

$(\mathbb{C}, +, \cdot)$ is a *commutative field*. The existence of an inverse can be demonstrated for the multiplication $[[f(X)]] \neq [[0]]$ in the following way: From the irreducibility of $X^2 + 1$, one deduces that the Greatest Common Denominator (GCD) of $f(X)$ and $X^2 + 1$ is 1. Since $\mathbb{R}[X]$ is an *Euclidean entire ring,* also called Euclidean integral domain (refer to Number theory and basic concept of the divisibility), there exist polynomials $h(X)$ and $k(X)$ verifying: $f(X)h(X) + (X^2 + 1)k(X) = 1$, and therefore, $[[f(X) \cdot h(X)]] = [[1]]$. Hence $[[f(X)]]^{-1} = [[h(X)]]$.

Each class contains exactly one polynomial of the form $a + bX$ with $a, b \in \mathbb{R}$, and is consequently determined by a given couple of real numbers, with: $(a + bX)\Re(c + dX) \Leftrightarrow a = c \wedge b = d$ and $f(X) = q(X)(X^2 + 1) + r(X)$ with a degree of $(r(X)) \leq 1$ and $f(X)\Re r(X)$. One deduces that the set of polynomials of the form $a + bX$ with $a, b \in \mathbb{R}$ is a system of representatives of $\mathbb{C}$, with the *laws:* $[[a + bX]] + [[c + dX]] = [[(a+c) + (b+d)X]]$, $[[a + bX]] \cdot [[c + dX]] = [[(ac + bdX^2) + (ad + bc)X]] = [[(ac - bd) + (ad + bc)X]]$.

In $\mathbb{C}^*$ we have: $[[a + bX]]^{-1} = \left[\left[\frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}X\right]\right]$. We embed $\mathbb{R}$ in $\mathbb{C}$ by means of the map $f : \mathbb{R} \to \mathbb{C}$ defined by: $a \mapsto [[a]]$. $[[a]]$ can be identified to a. Let us pose $i := [[X]]$. Then i is root of the polynomial $X^2 + 1$, and we obtain the following representation of an element of $\mathbb{C}$:

$$[[a + bX]] = a + b[[X]] = a + bi \quad \text{with } a, b \in \mathbb{R}.$$

It follows: $i^2 = -1$. The elements of $\mathbb{C}$ are called *complex numbers*. The construction process implies that $\mathbb{C}$ is the smallest extension field of $\mathbb{R}$ in which $X^2 + 1$ has roots. The complex numbers are often taken as couples of real numbers with the corresponding operations of the addition and multiplication described above. Furthermore, the numbers bi are called *pure imaginary number* (with $b \in \mathbb{R}^*$) and i is the *imaginary unit* equal to the square of $-1 : \sqrt{-1}$. When a single letter $z = a + bi$ is used to denote a complex number, it is sometimes called an "affix". In component notation, $z = a + bi$ can be written (a, b). The field of complex numbers includes the field of real numbers as a *subfield*.

A.3.3 Geometrical Representation of Complex Numbers

The *exponentiation* is the process of taking a quantity b (i.e. the base) to the power of another quantity e (i.e. the exponent); this operation most commonly denoted b^e. Before studying in $\mathbb{C}$ the exponentiation and its reciprocal function, we are going to see the geometrical representation of complex numbers. The field $\mathbb{C}$ can be regarded as a two-dimensional vector space over $\mathbb{R}$, of base $\{1, i\}$. In the writing $z = a + bi$ (i.e. Euler formula) we interpret the real numbers "a" (i.e. *real part*) and "b" (i.e. *imaginary part*) as coordinates in a coordinate system (i.e. Cartesian coordinates). In the Descartes treatise (1637), which introduced the use of coordinates for describing plane curves, the axes were omitted, and only positive values of the x- and y-coordinates were considered, since they were defined as distances between points. For an ellipse this meant that, instead of the full picture which we would plot nowadays, Descartes drew only the upper half.

To each complex number corresponds a point of the plane and conversely. The Abscissa represents the reals and the Ordinate represents the pure imaginary numbers; Such a representation system is called *Gauss complex plane* (see Fig. A.5 left).
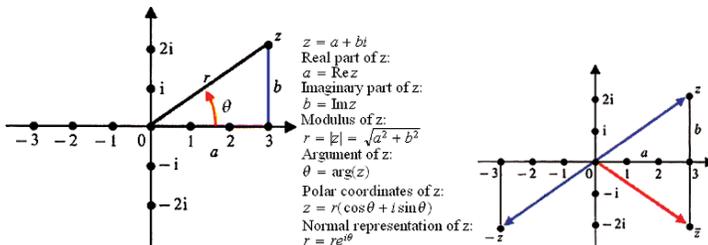
Fig. A.5 Gauss complex plane (*left*), Complex conjugate and opposite (*right*)

Modulus of a complex number: In $\mathbb{C}$ it is possible to define an absolute value which is called a *modulus:* $|z| = |a + bi| := \sqrt{a^2 + b^2}$; $(z_1, z_2) \mapsto |z_1 - z_2|$ is then a *distance* on $\mathbb{C}$.

Complex conjugate: $\overline{z} = a - bi$ is the complex conjugate of the complex number $z = a + bi$ (see Fig. A.5 right).

We obtain then $z\overline{z} = |z|^2$ and $z^{-1} = \overline{z}/|z|^2$ for $z \neq 0$. Calculation rules are written:

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}, \ \overline{z_1 - z_2} = \overline{z_1} - \overline{z_2}, \ \overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}, \ \overline{\left(\frac{z_1}{z_2}\right)} = \frac{\overline{z_1}}{\overline{z_2}}.$$

Polar coordinates: In the Gauss complex plane one uses also the polar coordinates r and θ with $0 \leq \theta < 2\pi$, so that, for $z \neq 0$, it is possible to write $z = r(\cos\theta + i\sin\theta)$. We deduce then $r = |z|$. θ is called the *principal argument* of z and is denoted $\arg(z)$. Sometimes, it is said that the principal argument corresponds to $\theta \in]-\pi, \pi[$. Furthermore, for any $k \in \mathbb{Z}$, $\theta + 2k\pi$ is an argument of z which is denoted $\arg(z) = \theta + 2k\pi$. In the theory of functions it can be demonstrated that $\cos\theta + i\sin\theta = e^{i\theta}$; we obtain then that the *normal representation* of the nonzero complex numbers: $z = re^{i\theta}$. And $\forall k \in \mathbb{Z}$, $e^{2ki\pi} = 1$.

A.3.4 Operations in the Gauss Complex Plane

The addition and subtraction are carried out in the Gauss complex plane in a vectorial manner (Fig. A.6). The multiplication and division are carried out geometrically by means of similar triangles (Fig. A.7). These operations are made by using a normal representation, i.e. representations of the form $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$, $z_1, z_2 \neq 0$; then we have $z_1 z_2 = r_1 r_2 e^{i(\theta_1 + \theta_2)}$ and $\frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$. The sum and subtraction of arguments must be reduced (modulo 2π) to a value include between 0 and 2π to obtain a *normal representation,* i.e. a representation of the form "$z = re^{i\theta}$".
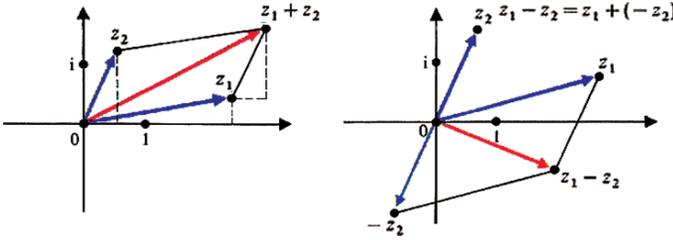
Fig. A.6 Addition and subtraction of complex numbers
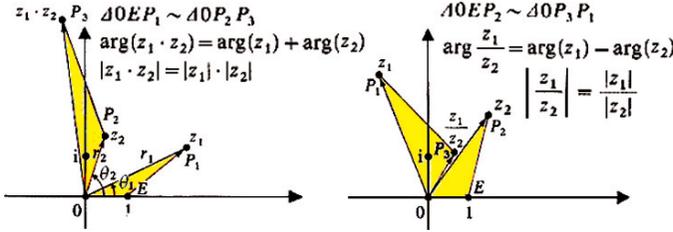
Fig. A.7 Multiplication and division of complex numbers
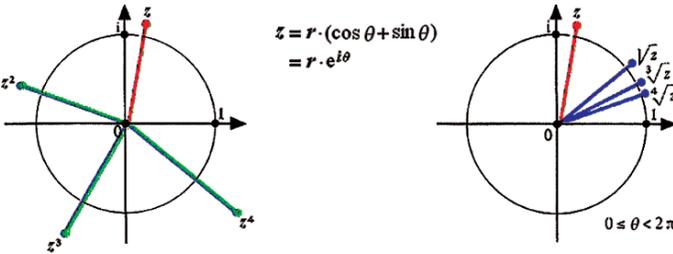
Fig. A.8 *Left:* Rise to a power. *Right:* Extraction of a root nth

A.3.5 Algebraic Closure of $\mathbb{C}$

We know that the rise to a power is equivalent to multiplications, then it is possible to carry out the rise of a complex number to a whole power: $z = re^{i\theta} \wedge n \in \mathbb{N} \Rightarrow z^n = r^n \cdot e^{i\theta n}$ $(z \neq 0)$. We obtain also $|z^n| = |z|^n$ and $\arg(z^n) = n \cdot \arg(z) \bmod(2\pi)$ with $0 \leq \arg(z^n) < 2\pi$. In a similar way the extraction of the root of $z \neq 0$ can be reduced to a calculation in $\mathbb{R}$: for $z = re^{i\theta} \wedge n \in \mathbb{N}\setminus\{0\}$ it is possible to define: $\sqrt[n]{z} = \sqrt[n]{z} \cdot e^{i\frac{\theta}{n}}$ $(\theta = \arg(z))$. We have here $|\sqrt[n]{z}| = \sqrt[n]{|z|}$ and $\arg(\sqrt[n]{z}) = \frac{1}{n}\arg(z)$; see Fig. A.8 depicting the rise to a whole power and the extraction of the root nth of a complex number, with n belonging to $\mathbb{N}\setminus\{0\}$.

Figure A.8 (left) corresponds to $z^n = r^n \cdot e^{i\theta n} = r^n(\cos(n\theta) + i\sin(n\theta))$; Fig. A.8 (right) corresponds to $\sqrt[n]{z} = \sqrt[n]{z} \cdot e^{i\frac{\theta}{n}} = \sqrt[n]{z} \cdot (\cos\frac{\theta}{n} + i\sin\frac{\theta}{n})$. $\sqrt[n]{z}$ is a zero of the polynomial $X^n - z$; such a polynomial have n distinct roots which are written $x_k = \sqrt[n]{z} \cdot e^{i\frac{2\pi}{n}k}$ with $k \in \{0, 1, 2, \ldots, n-1\}$. Furthermore, the symbol $\sqrt[n]{z}$ indicates, among the n roots, the smallest arguments. The roots of $X^n - 1$ is called the root nth of

the unity. *They correspond in the Gauss complex plane to the vertex of a regular polygon included in the unit circle*; a polygon whose one of the vertices is placed on the half-axis of positive reals. It is possible to deduce from the previous properties that (beyond $X^2 - 1$) any polynomial $X^2 - z$ with $z \in \mathbb{C}$ admits roots. Following theorem is a fundamental generalization.

A.3.6 Alembert–Gauss Theorem

Theorem A.10 (Alembert–Gauss). *Every polynomial of $\mathbb{C}[X]$ of degree $n > 0$ admits at least one root in $\mathbb{C}$ (refer to Polynomial Ring in Algebra).*

This theorem is also called the *Fundamental theorem of Algebra*, It was first proven by Gauss. It is equivalent to the statement that a polynomial $P(z)$ of degree n has n values z_i (some of them possibly degenerate) for which $P(z_i) = 0$. Such values are called polynomial roots. An example of a polynomial with a single root of multiplicity >1 is $z^2 - 2z + 1 = (z - 1)(z + 1)$, which has $z = 1$ as a root of multiplicity 2.

There are several demonstrations of this theorem, however it is not possible to demonstrate it by using purely algebraic means. Nevertheless one obtains a very simple proof as consequence of the *Liouville theorem* in *theory of functions*. The field $\mathbb{C}$ is then *algebraically closed*. (A field F is said to be algebraically closed if every polynomial with coefficients in F has a root in F.) Indeed, one deduces from the theorem that every polynomial of $\mathbb{C}[X]$ of degree $n > 0$ is decomposed into a product of n factors of the first degree. When all coefficients of the polynomial are real numbers, if x_i is root, then $\overline{x_i}$ is also root, and the polynomial is decomposed in $\mathbb{R}[X]$ into product of linear and quadratic factors. If n is odd, then we must have $x_i = \overline{x_i}$ for at least one root, and the polynomial admits a real root.

A.3.7 Exponential, Logarithm in $\mathbb{C}$

By using the *normal representation* of complex numbers, it is possible to *extend the calculation of powers* to complex exponents by conserving the calculation rules used until now. For $z = re^{i\theta}$, $r > 0$, $0 \le \theta < 2\pi$, and $w = x + iy$, we have

$$z^w = (re^{i\theta})^{x+iy} := r^x r^{iy} e^{i\theta x - \theta y} = (r^x e^{-\theta y}) e^{i(y\ln r + \theta x)}. \tag{A.38}$$

The *extraction of root* corresponds to a calculation of power like for the real numbers, one poses: $\sqrt[w]{z} := z^{\frac{1}{w}}$ for $z \in \mathbb{C}$ and $w \in \mathbb{C}\backslash\{0\}$. The normal representation is used to search for the logarithm, in particular for the logarithm of base e. It is possible to write $\ln z := \ln r + i\theta$ for $z = re^{i\theta}$, $z \ne 0$, $r > 0$, $\theta = \arg(z)$. $\ln z$ is a solution of the equation $e^x = z$. Since $e^{2i\pi} = 1$, there is an infinity of solutions for this equation, i.e. $\ln r + i\theta + 2i\pi k$, $k \in \mathbb{Z}$. However $\ln z$ is the only one verifying $0 \le \theta + 2k\pi < 2\pi$.

A particular solution of the equation $w^x = z$ is $\frac{\ln z}{\ln w}$ (for $z \neq 0$ $w \neq 0$ $w \neq 1$), this solution will be denoted $\log_w z$ and is called the logarithm of base w of z. Like for the inverse operation of the multiplication where the division by zero is excluded, it is not possible to resolve the exclusions indicated by the root extraction and the logarithm search.

A.3.8 Others Properties of $\mathbb{C}$, and Topology Theorem of $\mathbb{C}$

In the extension of the algebraic structure of $\mathbb{R}$ to the algebraic structure of $\mathbb{C}$, we lose the order structure. It is possible to provide $\mathbb{C}$ with an order structure, e.g. a strict order like: $z_1 < z_2 :\Leftrightarrow \{|z_1| < |z_2| \vee (|z_1| = |z_2| \neq 0 \wedge \arg z_1 < \arg z_2)\}$. However there is no order on $\mathbb{C}$ compatible with the algebraic structure, i.e. verifying the monotony laws for the addition and multiplication, e.g. $a < 0 \vee a > 0 \Rightarrow a^2 > 0 \Rightarrow a^2 + 1 > 0$. By contrast, it is possible to extend the real topology to a complex topology by means of the modulus, which extends the absolute value on $\mathbb{R}$. We deduce then the following theorem:

Theorem A.11 (Topology of complex numbers). *Every Cauchy sequence of complex numbers is convergent in $\mathbb{C}$. ($\mathbb{C}$ is thus complete.)*

A.3.9 Riemann Sphere (Compactification)

Let us pose the sphere whose diameter is equal to 1 in the Gauss complex plane at the origin, it is possible by using a *stereographic projection* from the North pole to make the complex plane homeomorphic to the sphere deprived of North pole (see Fig. A.9a of Riemann sphere and Gauss complex plane). Figure A.9b shows the distance between two points z_1 and z_2 of Gauss complex plane ($\mathbb{C}$) which is $|z_1 - z_2|$, i.e. $d(z_1, z_2) = |z_1 - z_2|$; the distance of their homologous points P_1 and P_2 is given by $\widehat{d}(z_1, z_2) = \frac{|z_1 - z_2|}{\sqrt{(1+|z_1|^2)\cdot(1+|z_2|^2)}}$. The equator of this sphere (see Fig. A.9a) which is called the Riemann sphere corresponds by this projection to the circle of the plane whose center is the origin and the radius is equal to 1. The North pole corresponds to a *new point* which is denoted ∞, we obtain then a *compact space* by the *adjunction of this point to the complex plane*. The exterior domains of this circle centered at the origin form a fundamental system of neighborhoods of this point ∞. The adjunction of a special point is not the only manner to compactify the plane (refer to Projective geometry), but this *adjunction* is *fundamental in the theory of complex functions*. In the *compact space* which is then formed, the *behavior of holomorphic and meromorphic functions* is particularly clear, because *one obtains a "Riemann surface" by means of this process of compactification*.

Riemann sphere: A one-dimensional complex manifold denoted $\widehat{\mathbb{C}}$, which is the one-point compactification of the complex numbers $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, together with
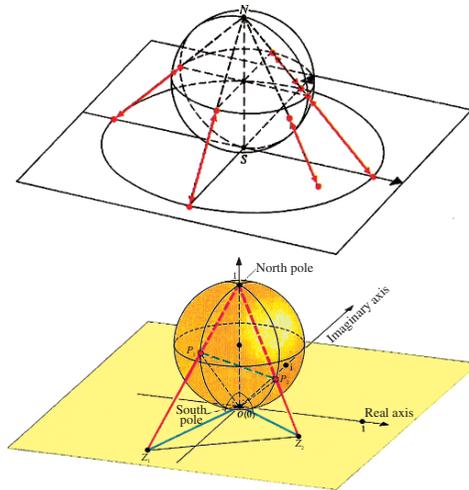
Fig. A.9 Riemann sphere

two charts (here ∞ complex infinity). For all points in the complex plane, the chart is the identity map from the sphere (with infinity removed) to the complex plane. For the point at infinity, the chart neighborhood is the sphere (with the origin removed), and the chart is given by sending infinity to 0 and all other points to $1/z$. (Instead of the notation $\widehat{\mathbb{C}}$, the following notation $\mathbb{C}^*$ is also used, but this notation is also used to denote the punctured plane $\mathbb{C}\text{-}\{0\}$.) Furthermore $\widehat{\mathbb{C}}$ is a Riemann surface. A Riemann surface is a surface-like configuration that covers the complex plane with several, and in general infinitely many, "sheets". These sheets can have very complicated structures and interconnections. Riemann surfaces are one way of representing multiple-valued functions; another is branch cuts. Riemann sphere is also known as *complex sphere*.

A.3.10 Holomorphic Function, Cauchy–Riemann Conditions and Harmonic Function

Definition A.53 ($\mathbb{C}$-differentiable). The $\mathbb{C} - \mathbb{C}$ function $f : D_f \to \mathbb{C}$ is known as $\mathbb{C}$-*differentiable* at a point $a \in \mathbb{C}$ if a is an *accumulation point* of D_f and if $\lim\limits_{z \to a} \frac{f(z)-f(a)}{z-a}$ exists.

Definition A.54 (Holomorphic function). [17] The $\mathbb{C} - \mathbb{C}$ function $f : D_f \to \mathbb{C}$ is known as *holomorphic* on the domain $D \subseteq D_f$ (resp. at a point a interior to D_f) if f

---

[17] *Holomorphic function* is a synonym for *analytic function*, regular function, differentiable function, complex differentiable function, and holomorphic map. The word "holomorphic" is derived from the Greek word "holos" which means "whole", and from the word "morphe" which means

is $\mathbb{C}$-differentiable at any point of D (resp. at a point a). The limit value $\lim\limits_{z \to a} \frac{f(z) - f(a)}{z - a}$ is called derivative of f at a and is noted $f'(a)$.

Cauchy–Riemann conditions: If f is holomorphic at a point a, $\lim\limits_{z \to a} \frac{f(z_n) - f(a)}{z_n - a} = f'(a)$ for any sequence $z_n \to a = a_1 + ia_2$, $z_n \neq a$. we take successively $z_n = a + h_n$, $z_n = a + ih_n$, where $h_n \in \mathbb{R}^*$, $h_n \to 0$.

(1) $\dfrac{f(a + h_n) - f(a)}{h_n}$

$$= \frac{[u(a_1 + h_n, a_2) - u(a_1, a_2)] + i[v(a_1 + h_n, a_2) - v(a_1, a_2)]}{h_n} \to f'(a).$$

Thus, $\frac{\partial u}{\partial x_1}$ and $\frac{\partial v}{\partial x_1}$ exist at a point a and $f'(a) = \frac{\partial u}{\partial x_1}(a_1, a_2) + i\frac{\partial v}{\partial x_1}(a_1, a_2)$.

(2) $\dfrac{f(a + ih_n) - f(a)}{ih_n}$

$$= \frac{[u(a_1, a_2 + h_n) - u(a_1, a_2)] + i[v(a_1, a_2 + h_n) - v(a_1, a_2)]}{ih_n} \to f'(a).$$

Thus, $\frac{\partial u}{\partial x_2}$ and $\frac{\partial v}{\partial x_2}$ exist at a point a and $f'(a) = -i\frac{\partial u}{\partial x_2}(a_1, a_2) + \frac{\partial v}{\partial x_2}(a_1, a_2)$. And by bringing closer the two preceding writings, we obtain the *Cauchy–Riemann conditions* which are written:

$$\frac{\partial u}{\partial x_1}(a_1, a_2) = \frac{\partial v}{\partial x_2}(a_1, a_2), \tag{A.39}$$

$$\frac{\partial u}{\partial x_2}(a_1, a_2) = -\frac{\partial v}{\partial x_2}(a_1, a_2). \tag{A.40}$$

Thus, it is possible to write that $f : D_f \to \mathbb{C}$ is *holomorphic* at a point a interior to D_f if and only if u and v are *differentiable* at a point a and satisfy the Cauchy–Riemann conditions.

Harmonic functions: Since the real and imaginary parts of the holomorphic functions are *twice continuously differentiable*, the *Cauchy–Riemann equalities* imply that:

$$\frac{\partial^2 u}{\partial x_1^2} = \frac{\partial^2 v}{\partial x_1 \partial x_2} \quad \text{and} \quad \frac{\partial^2 u}{\partial x_2^2} = -\frac{\partial^2 v}{\partial x_1 \partial x_2}, \tag{A.41}$$

hence:

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = 0 \quad \text{and} \quad \frac{\partial^2 v}{\partial x_1^2} + \frac{\partial^2 v}{\partial x_2^2} = 0. \tag{A.42}$$

"form" or "appearance". Like the analytic functions, the holomorphic functions are functions which c*an be represented by a convergent Taylor series.*

Definition A.55 (Harmonic functions). A $\mathbb{R}^2 - \mathbb{R}$-function ϕ twice continuously differentiable on a domain D is known as harmonic function on D if it verifies in any point of D the condition $\frac{\partial^2 \phi}{\partial x_1^2} + \frac{\partial^2 \phi}{\partial x_2^2} = 0$.

Proposition A.1. *The real and imaginary parts of a holomorphic function are harmonic. Any harmonic function u on a simply connected domain D corresponds to a harmonic function v defined except for an additive constant, such that u and v can be regarded as the holomorphic parts respectively real and imaginary of a holomorphic function.*

Proposition A.2 (Morera). *Let be $D \subseteq \mathbb{C}$ a simply connected domain and $f : D \to \mathbb{C}$ continuous. If along of any rectifiable oriented Jordan curve $k \subset D$ we have $\int_k f(z)dz = 0$, then f is holomorphic on D.*

Series expansion: The derivability at a point "a" for any order is a necessary condition but non-sufficient so that a real function admits a Taylor series expansion around a. Thus, one can wonder whether a holomorphic function is expandable in a series similar to the Taylor series. To this end it is necessary to study properties of complex entire series.

Proposition A.3. *If the series $\sum_{v=0}^{\infty} a_v (z_1 - a)^v$ converges for the points $z_1 \neq a$, the same applies to the series $\sum_{v=0}^{\infty} a_v (z - a)^v$ for any z such that $|z - a| < |z_1 - a|$.*

Proposition A.4. *If the series $\sum_{v=0}^{\infty} a_v (z - a)^v$ converges for points z different from a, but does not converge on any $\mathbb{C}$, there exists $r \in \mathbb{R}_+$ such that this series converges for any z such that $|z - a| > r$; r is defined by:*

$$r = \frac{1}{\lim\limits_{v \to \infty} \sqrt[v]{|a_v|}}. \tag{A.43}$$

r is the radius of convergence of the series.

Proposition A.5. *The function defined by $z \mapsto f(z) = \sum_{v=0}^{\infty} a_v (z - a)^v$ is holomorphic on the interior of the disk of convergence. By derivation term by term, it is obtained the derivative $f'(z) = \sum_{v=1}^{\infty} v a_v (z - a)^{v-1}$ (same open disk of convergence).*

Proposition A.6. *Let be $f : D \to \mathbb{C}$ a holomorphic function on the domain $D \subseteq \mathbb{C}$. f is developable in entire series around any point $a \in D$ and we have:*

$$f(z) = \sum_{v=0}^{\infty} \frac{f^{(v)}(a)}{v!}(z - a)^v. \tag{A.44}$$

The series necessarily converges towards f on the largest open disk centered at a on which f is holomorphic.

Holomorphic continuation: A real analytic function[18] is expandable in entire series around a. The definition of the radius of convergence of a entire series is the same

---

[18] *Analytic function (i.e. Holomorphic function): A function which can be represented by a convergent Taylor series. Also known as holomorphic function.*

one in the two real and complex cases. The expansion of a real analytic function in entire series, taken as an expansion in series on $\mathbb{C}$, is thus convergent on the interior of a disk centered at a and have as radius the radius of convergence of the real series. *This series defines on this disk a holomorphic function*, the *holomorphic continuation* of the real function. For entire series converging everywhere on $\mathbb{R}$, as it is the case for the functions sine, cosine, exponential,... we *obtain, by means of such a continuation,* holomorphic functions on any $\mathbb{C}$. If we compares the expansions to series, we find the following relations: $e^{iz} = \cos z + i \sin z$, $e^{-iz} = \cos z - i \sin z$, $\cos z = \frac{1}{2}(e^{iz} + e^{-iz})$, $\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$,.... This shows that *in the complex case, the exponential and hyperbolic functions* are *periodic*, and *their period is complex and is equal to $2 i\pi$.*

Definition A.56 (Analytic continuation). [19] Given $f : D_f \to \mathbb{C}$ and $g : D_g \to \mathbb{C}$ two holomorphic functions respectively on the D_f and D_g domains. If $D_f \cap D_g \neq \varnothing$ and $f(z) = g(z)$ for any $z \in D_f \cap D_g$ then g is called an analytic continuation of f and f is an analytic continuation of g.

Theorem A.12 (Monodromy theorem). *If a complex function f is analytic in a disk contained in a simply connected domain D and f can be analytically continued along every polygonal arc in D, then f can be analytically continued to a single-valued analytic function on all of D!*

Definition A.57 (Monodromy group). A group (technically defined) characterizing a system of linear differential equations $q'_j = \sum_{k=1}^{n} \alpha_{jk}(x)q_k$ for $j = 1,\ldots,n$, where α_{jk} are "complex analytic function" of x in a given complex domain.

Definition A.58 (Complex differentiable). Given $z = x + iy$ and $f(z) = u(x,y) + iv(x,y)$ on some region S containing the point z_0. If $f(z)$ satisfies the Cauchy–Riemann equations and has continuous first partial derivatives in the neighborhood of z_0, then $f'(z_0)$ exists and is given by $f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$, and the function is said to be complex differentiable (or, equivalently, "analytic" or "holomorphic").

A.3.11 Singularity of Holomorphic Functions, Laurent Series and Meromorphic Function

Given a holomorphic function on a domain D of $\mathbb{C}$. If $D = \mathbb{C}$, f is an *entire function*, i.e. f is the sum of an entire series $\sum_{v=0}^{\infty} a_v z^v$ of *infinite radius*. If $D \neq \mathbb{C}$, D have at least a frontier point. Let be such a point a. If there exists an entire series in $(z - a)$ with non-zero radius, of the sum $S(z) = \sum_{v=0}^{\infty} a_v(z - a)^v$, and an open neighborhood Φ of a such that f and S coincide on $(\Phi \backslash \{a\}) \cap D$, then f can have a *continuation* to a holomorphic function on $D \cup \Phi$, which is a domain larger than D.

---

[19] *Analytic continuation:* The process of extending an analytic function to a domain larger than the one on which it was originally defined.

Definition A.59 (Singular point and Isolated singular point). A frontier point "*a*" which does not have the property which precedes is known as a Singular point. It is "isolated" if it is the center of an open disk Ω such that $\forall b \in \Omega \backslash \{a\}$, there exists a continuation of f to a domain D_b, containing b. It is "simple isolated" if it is the center of an open disk Ω such that f admits a holomorphic "continuation" to $D \cup (\Omega \backslash \{a\})$.

Around a *simple isolated singular point* there exists a series expansion which generalizes an expansion in Taylor series.

Laurent series: Let f be the holomorphic function on a open disk Ω where we have suppressed the center "*a*" (i.e. "blunted", such a center "*a*" being a "simple isolated singular point" of f. Given K_1 and K_2 two circles centered at a contained in Ω. K_1 indicating the smallest of both, we are located at a point z of the open disk exterior to K_1, interior to K_2 (see figure below). It is possible to write the following formula:

$$f(z) = \frac{1}{2i\pi} \int_{K_2^+} \frac{f(\zeta)}{\zeta - z} dz - \frac{1}{2i\pi} \int_{K_1^+} \frac{f(\zeta)}{\zeta - z} d\zeta. \tag{A.45}$$

This expression can be transformed, in particular in the first integral $\frac{1}{\zeta - z} = \sum_{\nu=0}^{\infty} \frac{(z-a)^\nu}{(\zeta-a)^{\nu+1}}$. Since z is exterior to K_1, in the second integral it can be written $\frac{1}{\zeta-z} = -\sum_{\nu=1}^{\infty} \frac{(\zeta-a)^{\nu-1}}{(z-a)^\nu} = -\sum_{\nu=-1}^{-\infty} \frac{(z-a)^\nu}{(\zeta-a)^{\nu+1}}$. After an integration term by term, it is possible to write a series of the form:

$$f(z) = \sum_{\nu=-\infty}^{-\infty} a_\nu (z-a)^\nu \quad \text{with} \quad a_\nu = \frac{1}{2i\pi} \int_{K^+} \frac{f(\zeta)}{(\zeta-a)^{\nu+1}} d\zeta, \tag{A.46}$$

$\nu \in \mathbb{Z}$, where K is an arbitrary circle centered at "*a*" located in the disk. Such a series is called *Laurent series* admitting "*a*" as point of expansion. Figure A.10 corresponds to $f(z) = \frac{1}{2i\pi} \int_{K_3} \frac{f(\zeta)}{\zeta-z} d\zeta = \frac{1}{2i\pi} \int_{K_2} \frac{f(\zeta)}{\zeta-z} d\zeta - \frac{1}{2i\pi} \int_{K_1} \frac{f(\zeta)}{\zeta-z} d\zeta$ because of $\int_{-K_1-K_3+K_2} \frac{f(\zeta)}{\zeta-z} d\zeta = 0$.
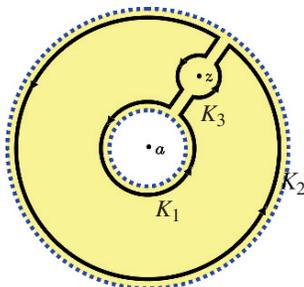
Fig. A.10 Integral representation of a holomorphic function in a Disk

Singularities of analytic functions: All points around which a function can be expanded in a *Laurent series* according to a suitable parameter of uniformisation, containing only one finite number of terms having a negative power, will be integrated in the domain of the corresponding *Riemann surface* (points of holomorphy, poles,...). The following examples give points which appear as *frontier points* of the *Riemann surface:*

	Function	Point
(1) Non-isolated singularities	: $f(z) = \sum_{\nu=0}^{\infty} z^{\nu!}$	$\{z\,|\,|z| = 1\}$
(2) Simple isolated singularities	: $f(z) = e^{1/z^2}$	0
(3) Isolated singularities on k- sheets	: $f(z) = \sqrt[k]{z}$	∞
(4) Isolated logarithmic singularities	: $f(z) = z^{\sqrt{2}} = (e^{\sqrt{2}})^{\ln z}$	0
	$f(z) = e^{\frac{1}{z}} \ln z$	0

Proposition A.7 (Picard). *Let be $f : D_f \to \mathbb{C}$ a holomorphic function on a "blunted" neighborhood of $U(z) \backslash \{z\}$ where z is an "essential isolated singular point" of f: then f take, in any blunted neighborhood of z, any value of $\mathbb{C}$ except at most only one value.*

A *pole:* This term "pole" is used prominently in a number of very different branches of mathematics. Certainly, the most important and widespread usage is to denote *a singularity of a complex function*.

In complex analysis, an *analytic function (i.e. an holomorphic function) f* is said to have a *pole* of order n at a point $z = z_0$ if, in the *Laurent series*, $a_m = 0$ for $m < -n$ and $a_{-n} \neq 0$. Equivalently, f has a pole of order n at z_0 if n is the smallest positive integer for which $(z - z_0)^n f(x)$ is holomorphic at z_0. A analytic function f has a pole at infinity if $\lim_{z \to \infty} f(z) = \infty$. (A non-constant polynomial $P(z)$ has a pole *ad infinitum* of order $\deg P$, i.e. the polynomial degree of P.) The basic example of a pole is $f(z) = 1/z^n$, which has a single pole of order n at $z = 0$ (Fig. A.11).

Meromorphic function: A holomorphic function whose only singularities are "poles" is called a meromorphic function. (The word derives from the Greek $\mu\varepsilon\rho o\varsigma$ (meros), meaning "part", and $\mu o\rho\varphi\eta$ (morphe), meaning "form" or appearance.)
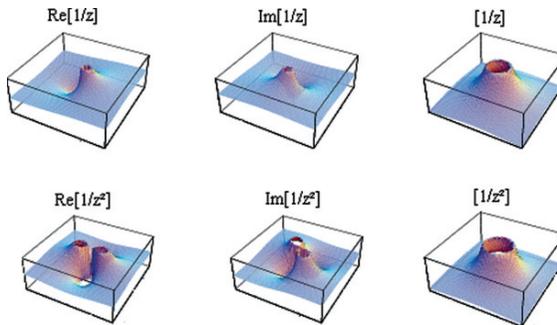
Fig. A.11 Plots of $1/z$ and $1/z^2$ are shown above in the complex plane

Definition A.60 (Meromorphic function). A holomorphic function f on a domain D of $\mathbb{C}$, except at isolated points of D which are poles, is known as "meromorphic" on D. If E is the set of poles, we can pose for any $a \in E$, $f(a) = \infty$: then f is a continuous map from D to $\mathbb{C}$, the restriction of this meromorphic function on the $D \backslash E$ domain being holomorphic.

Theorem A.13 (Residue theorem). *Given a domain D of $\mathbb{C}$ of which the frontier ∂D is a finite union of rectifiable Jordan curves. Given a holomorphic function f on D except for a finite number of points z_v, which are "poles" or "essential isolated singular points" and of which the set is noted E. If f is continuous on $(D \backslash E) \cup \partial D$ we have:*

$$\frac{1}{2i\pi} \int_{\partial D^+} f(z)dz = \sum_{v=1}^{n} (res\, f)(z_v). \tag{A.47}$$

A.4 Surfaces and Manifolds

A.4.1 Closed Surfaces, Surfaces with Boundary

Before describing closed surfaces and surfaces with boundary let us define some important notions (orientable surface, ..., simplicial complex, ...):

Orientable surface: A surface for which an object resting on one side of it cannot be moved continuously over it to get to the other side without going around an edge.

Non-orientable surface (one-sided surface): A surface such that an object resting on one side of it cannot be moved continuously over the surface to reach the other side without going around an edge; the *Moëbius strip* and the *Klein bottle* are examples.

Triangulation: Triangulation is the division of a surface or plane polygon into a set of triangles, usually with the restriction that each triangle side is entirely shared by two adjacent triangles. It was proved that every surface has a triangulation, but it might require an infinite number of triangles and the proof is difficult. A surface with a finite number of triangles in its triangulation is called compact.

Span (see also Hull): (1) For a set A, the intersection of all sets that contain A and have some specified property. Also known as "hull". (2) for a set of vectors, the set of all possible linear combinations of those vectors. Also known as linear span. (3) The difference between the highest value and the lowest value in a range of value.

Convex Hull: The smallest convex set containing a given collection of points in a real linear space. Also known as convex linear hull. The convex hull of a set of points in dimensions is the intersection of all convex sets containing.

Polytope: This word is used to mean a number of related, but slightly different mathematical objects. A convex polytope may be defined as the convex hull of a finite set of points (which are always bounded), or as a bounded intersection of a

finite set of half-spaces. A polytope can also be defined as the general term of the sequence "point, line segment, polygon, polyhedron,...", or more specifically as a finite region of n-dimensional space enclosed by a finite number of hyperplanes. The special name polychoron is sometimes given to a four-dimensional polytope. However, in algebraic topology, the underlying space of a simplicial complex is sometimes called a polytope.

Polyhedron: (1) A solid bounded by planar polygons. (2) The set of points that belongs to the simplexes of a simplicial complex. (3) See triangular space. The word polyhedron has slightly different meanings in geometry and algebraic geometry. In geometry, a polyhedron is simply a three-dimensional solid which consists of a collection of polygons, usually joined at their edges. A polyhedron is the three-dimensional version of the more general polytope (in the geometric sense), which can be defined in arbitrary dimension. The plural of polyhedron is polyhedra or polyhedrons. The term "polyhedron" is used differently in algebraic topology, where it is a space that can be constructed from such building blocks as line segments, triangles, tetrahedra, and their higher dimensional analogs by gluing them together along their faces. More specifically, it can be defined as the underlying space of a *simplicial complex* (with the additional constraint sometimes imposed that the complex be finite). In the usual definition, a polyhedron can be viewed as an intersection of half-spaces, while a polytope is a bounded polyhedron.

Tetrahedron: The regular tetrahedron, often simply called "the tetrahedron", is the Platonic or regular solid with four polyhedron vertices, six polyhedron edges, and four equivalent equilateral triangular faces. It is also *uniform polyhedron*. The tetrahedron has seven axes of symmetry.

Simplex (Simplexes, Simplices): An n-dimensional in an Euclidean space consists of $n+1$ linearly independent points $x_0, x_1, \ldots, x_n$ together with all the points given by $a_0 x_0 + a_1 x_1 + \cdots + a_n x_n$ where the $a_i \geq 0$ and $a_0 + a_1 + \cdots + a_n = 1$; a triangle with its interior and a tetrahedron with its interior are examples. A simplex, sometimes called a hypertetrahedron, is the generalization of a tetrahedral region of space to n dimensions. The boundary of a k-simplex has $k+1$ 0-faces (polytope vertices), $k(k+1)/2$ 1-faces (polytope edges), and $\binom{k+1}{i+1}$ i-faces, where $\binom{n}{k}$ is a binomial coefficient. An n-dimensional simplex can be denoted using the Schläfli symbol as follows $\frac{\{3,\ldots,3\}}{n-1}$. *The simplex represents the simplest possible polytope in any given space.* The content (i.e. hypervolume) of a simplex can be computed using the Cayley–Menger determinant. In one dimension, the simplex is the line segment. In two dimensions, the simplex $\{3\}$ is the convex hull of the equilateral triangle. In three dimensions, the simplex $\{3,3\}$ is the convex hull of the tetrahedron.

Simplicial complex: A simplicial complex is a space with a *triangulation* (Fig. A.12). Formally, a *simplicial complex* K in $\mathbb{R}^n$ is a collection of simplices in $\mathbb{R}^n$ such that: (1) Every face of a simplex of K is in K, and (2) The intersection of any two simplices of K is a face of each of them. In other words, a set consisting of finitely many simplices where either two simplices are disjoint or intersect in a simplex which is a face common to each. Also known as *geometric complex*.

Fig. A.12 Not a simplicial complex (*left*), Simplicial complexes (*right*)

Objects in the space made up of only the simplices in the triangulation of the space are called simplicial subcomplexes. When only simplicial complexes and simplicial subcomplexes are considered, defining homology is particularly easy and, in fact, combinatorial because of its finite-counting nature. This type of homology is called *simplicial homology*.

Homology group: Associated to a topological space X, one of a sequence of "Abelian groups"[20] $H_n(X)$ that reflect how n-dimensional simplicial complexes can be used to fill up X and also help determine the presence of n-dimensional holes appearing in X. Also known as *Betti group*.

Homology theory: Theory attempting to compare topological spaces and investigate their structures by determining the algebraic nature and interrelationships appearing in the various *homology groups*.

Simplicial homology: A homology for a topological space where the nth group reflects how the space may be filled out by n-dimensional simplicial complexes and detects the presence of analogs of n-dimensional holes.

Simplicial Mapping: A map of one simplicial complex into another in which the images of the simplexes of one complex are simplexes of the other complex.

The term "surface" has several interpretations. The definition which is used here, within the framework of the *algebraic topology* is the one which makes it possible to construct a homeomorphism of the surface on a polyhedron of dimension 2 simplified as much as possible, i.e. connected, with a triangulation of which all faces are included in triangular surfaces. A surface will be therefore a compact of dimension 2, this double dimension having to be perceptible at the neighborhood of any point.

Definition A.61 (Closed surface). A connected compact part of $(\mathbb{R}^n, \mathfrak{R}^n)$ is called a "closed surface" if any point of this one possesses for the induced topology an open neighborhood homeomorphic to an open disk; It is called "surface with boundary" if it possesses, on the one hand, points having an open neighborhood homeomorphic to an open disk, on the other hand, points which are often called "bordering points": A bordering point is a point which, for the induced topology, possesses an open neighborhood homeomorphic to the union of an open disk and a Jordan arc (deprived of its extremities) included in its circumference, the homeomorphic image of the bordering point is inside the arc. ($(\mathbb{R}^n, \mathfrak{R}^n)$ denotes $\mathbb{R}^n$ provided with its natural topology defined by d_E; $\mathfrak{R}^n$ is its topology.)

---

[20] **Definition (Abelian group, or Commutative group).** A group $(E; \top)$ is called an Abelian group if $\top$ is commutative, i.e. if $\top$ verifies $\forall a,b \in E$, $a \top b = b \top a$. ($\top$: internal composition law).

Remark A.4 (Closed surfaces are particular manifolds). The closed surfaces are particular two-dimensional manifolds. The boundary (or edge) of a surface with boundary is the set of its bordering points. The edge can be constituted of a Jordan Curve (e.g. closed disk; Moëbius strip; see figure hereafter) or two Jordan curves (e.g. part of revolution cylinder included between two distinct planes perpendicular to its axis), etc.
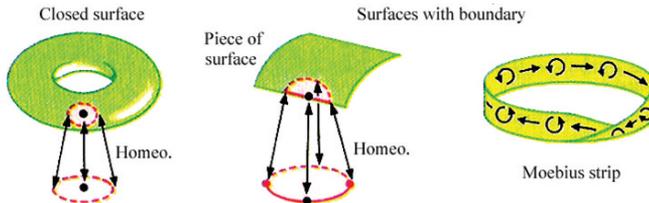
A closed surface, respectively a surface with boundary, is arcwise connected, because it is connected and each one of its points possesses a neighborhood arcwise-connected (i.e. locally arcwise-connected or locally connected by arc). It is then possible to show that a closed surface, respectively a surface with boundary, is homeomorphic to a polyhedron denoted $|K|$ where K is "simplicial complex" whose faces are of dimension ≤ 2, comprising at least a triangular simplex, the edges and the vertices being the ones of the triangular simplexes. In the case of a closed surface ($n \geq 3$) an edge belongs to two triangular simplexes (and only 2): such a property helps to explain the term "closed". It is possible to describe the fundamental group of a closed surface, respectively a surface with boundary. It is that of a homeomorphic polyhedron.

A.4.2 Classification of Closed Surfaces

The homeomorphic relation between closed surfaces of $\mathbb{R}^n$ ($n \geq 3$) is an equivalence relation, and each class admits a polyhedral representative. The problem of this classification is completely resolved. If we represent a class by a polyhedron, it is interesting to choose a polyhedron whose number of triangular faces is minimum. Then in $\mathbb{R}^3$, the skeleton of dimension 2 of the tetrahedron is convenient to represent the class of $\mathbb{S}^2$ ($\mathbb{S}^2$ corresponds to a sphere of $\mathbb{R}^3$, subspace of $(\mathbb{R}^3, \mathfrak{R}^3)$, i.e. in such a case a surface). This is the only one to have four faces, and it is the necessary minimum (a cube, which makes possible to define the same class, requires twelve triangular faces). About the class of the *torus*, it requires a polyhedron with eighteen triangular faces. Therefore the polyhedral representation is not necessarily the most simple. In spite of this aspect, the construction of classes can be carried out in a remarkable way. The involved *identification processes* are associated with the *quotient topology* notion.[21]

---

[21] *Quotient topology:* If X is a topological space, X/R the quotient space by some equivalence relation on X, the quotient topology on X/R is the smallest topology which makes the function which assigns to each element of X its equivalence class in X/R a continuous function.
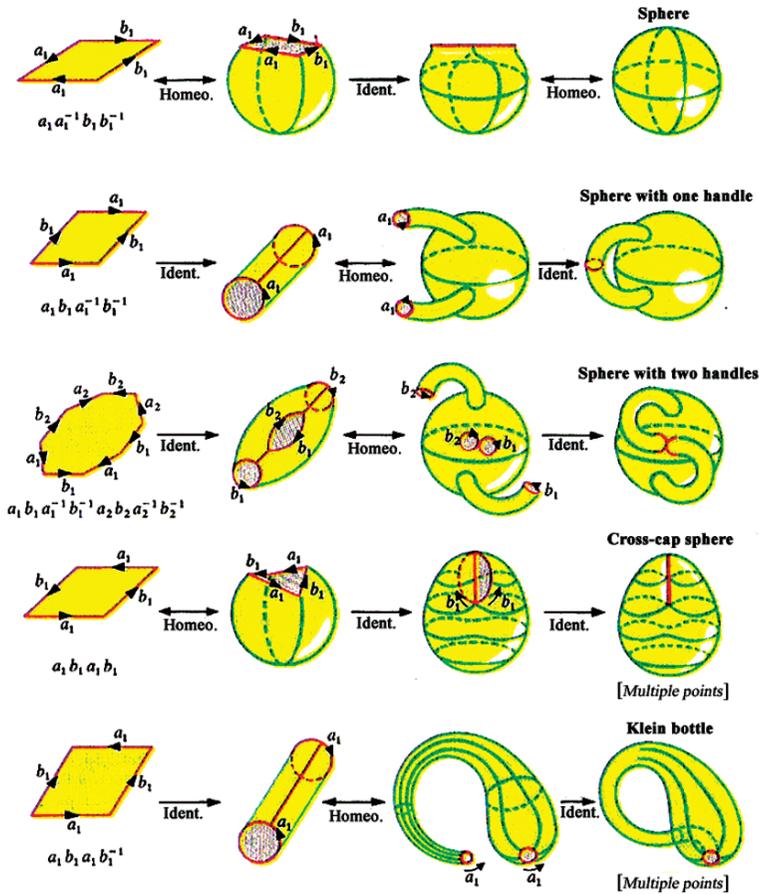
Fig. A.13 Classification of closed surfaces

First we show that *any closed surface* on $\mathbb{R}^n$ with $n \geq 3$ is *homeomorphic* to a *topological space* obtained from a *convex polygon* of $\mathbb{R}^2$ with $2\,p$ sides ($p > 2$), by identifying these pairwise in an appropriate way (see Fig. A.13). If we mark by the same letter, two sides which have to be identified, and if we indicate by an arrow each one of them, in such a way to indicate the sense of their super-position, it is possible to represent the succession of oriented segments met by carrying out a complete turn of the frontier by means of a sequence of the form: $a \ldots b^{-1} \ldots c \ldots a \ldots c^{-1} \ldots b^{-1} \ldots$, the exponents -1 mean that the corresponding segment was traversed in opposite direction (such a sequence obviously must be compatible with the vertices: Then the sequence $abb^{-1}a$ is not possible as we can see by drawing the process). Closed surface of $\mathbb{R}^n$ ($n \geq 3$) can be classified by the following primitive forms:

(1) $a_1 a_1^{-1} b_1 b_1^{-1}$ ($\mathbb{S}^2 \subset \mathbb{R}^n$, $n \geq 3$)

(2) $a_1 b_1 a_1^{-1} b_1^{-1} \ldots a_g b_g a_g^{-1} b_g^{-1}$, with $g \in \mathbb{N} \setminus \{0\}$ ($\mathbb{S}^2$ endowed with g *handles* $\subset \mathbb{R}^n$, $n \geq 3$, in the figure $g = 1, 2$)

(3) $a_1 b_1 a_1 b_1 \ldots a_k b_k a_k b_k$, with $k \in \mathbb{N} \setminus \{0\}$ ($\mathbb{S}^2$ endowed with k *cross-caps* $\subset \mathbb{R}^n$, $n \geq 4$)

Fundamental *homotopy groups* obtained from these primitive forms are pairwise non isomorphic, then the surfaces are pairwise non homeomorphic (refer to one of the topological invariant theorems).

Remark A.5 (Genus). When $g = 1$ in the form (2), we have a surface homeomorphic to a torus; The surface $\mathbb{S}^2$ endowed with g handles is also called a torus of "genus g" (or a torus with g holes). An ordinary torus is a surface having genus one, and therefore possessing a single "hole". The single-holed "ring" torus can be constructed from a rectangle by gluing both pairs of opposite edges together with no twists. In general, tori can also have multiple holes, with the term n-torus used for a torus with n holes. One of the more common uses of n-dimensional tori is in "dynamical systems". A fundamental result states that the phase space trajectories of a "Hamiltonian system" with n degrees of freedom and possessing n integrals of motion lie on an n-dimensional manifold which is topologically equivalent to an n-torus.

The form (3) is carried out in $\mathbb{R}^4$ and not in $\mathbb{R}^3$. But its representation can be done only in $\mathbb{R}^3$ or $\mathbb{R}^2$ (e.g. an affine projection from the real affine space of dimension 3 to an affine plane is a representation of the visual space). Then if we represent the form (3) for $k = 1$ starting from the sequence a_1, b_1, a_1, b_1 we obtain a surface which is called a "surface with multiple points": Two nappes of the surface cross each other at any point interior of the segment which we can observe on the last surface on the right of the figure. Relative to each nappe, considered independently of the other, a point interior of the segment possesses an open neighborhood homeomorphic to an open disk. About the extremities of the segment, it is more complicated. In the same way, it is possible to show that the Klein bottle, introduced by the sequence a_1, b_1, a_1, b_1^{-1}, which is also a *surface with multiple points*, is the representation of the form (3) for $k = 2$.

A.4.3 Orientability and Topological Invariance

Orientability can be defined by using the polyhedra. The *Moëbius strip* (see previous figure) is a *non-orientable surface*, it is possible to draw on this strip a *Jordan curve J*, whose "route" is symbolized by the almost rectilinear arrows of the drawing, along which we move continuously on the surface a curve homeomorphic to an oriented circle symbolized by a circular arrow: a complete rotation on J changes the orientation of the circular arrow to its opposite. *Any surface containing a Jordan curve for which we have this property is said "non-orientable". A surface is orientable if it contains no Jordan curve J for which this orientation change occurs.*

The *orientability*, respectively the *non orientability*, is a *topological invariant*.

Example A.1. $\mathbb{S}^2$, all tori of genus $g \in \mathbb{N}\backslash\{0\}$ are orientable. By contrast the spheres $\mathbb{S}^2$ endowed with k cross-caps, $k \in \mathbb{N}\backslash\{0\}$, are non-orientable. Moëbius strip belongs to the set of the most important non-orientable surfaces with boundary.

By knowing what precedes it is possible to show that a surface is non orientable if and only if it contains a subspace homeomorphic to a Moëbius strip.

A.4.4 Connectivity Number

Another important *topological invariant* about the surfaces is the *connectivity number Z*. It is the maximal number of Jordan curves (which can cut each other moreover) that we can draw on a closed surface without dividing this one into two disjoint parts. For an orientable closed surface $Z = 2g$, where g denotes the genus ($g = 0$ for $\mathbb{S}^2$). For a non-orientable closed surface $Z = k$ (number of cross-caps).

A.4.5 Riemann Surfaces

The passage from $\mathbb{C}$ to $\widehat{\mathbb{C}}$ makes it possible to eliminate a certain number of exceptional points of a function, like the *poles* with the proviso that we pass from the *holomorphy* concept to that of "*meromorphy*". Other singularities, appearing in the analytic continuation, are sometimes suppressed by an extension different from the domain.

(1) *First example:* The expression $\sqrt{z}$ has no precise meaning on $\mathbb{C}^*$ without additional convention, because a nonzero complex number has two square roots. If x is a real ≥ 0, $\sqrt{x}$ is (by definition) the positive square root of x. By writing for $0 \leq x \leq 2$, $\sqrt{x} = \sqrt{1+(x-1)} = 1 + \sum\limits_{1}^{\infty} \binom{\frac{1}{2}}{n}(x-1)^n$, it is possible to define the *complex analytic continuation:*

$$f(z) = 1 + \sum_{1}^{\infty} \binom{\frac{1}{2}}{n}(z-1)^n \tag{A.48}$$

which is a square root function of the complex number z on the closed disk $|z-1| \leq 1$ (normal convergence on this disk). Along the circle $|z| = 1$ it is possible to extend the restriction of f to the open disk $|z-1| < 1$ which is holomorphic, by means of entire series of the form:

$$e^{\frac{\varphi}{i2}}\left[1 + \sum_{1}^{\infty} \binom{\frac{1}{2}}{n}\left(ze^{-i\varphi} - 1\right)^n\right] \tag{A.49}$$

by varying continuously φ from the value 0: for $\varphi = \pi$ we obtain

$$i \left[1 + \sum_1^\infty \binom{\frac{1}{2}}{n} (-1)^n (z+1)^n \right] \tag{A.50}$$

and the opposite series for $\varphi = -\pi$. It is easy to observe that there are two possible roots. Such an ambiguity is not compatible with the usual function notion. It is possible to resolve this problem if (according to the Riemann process) we take as domain (no more a domain of $\mathbb{C}$ or $\widehat{\mathbb{C}}$) but a *surface* which *covers* the *plane* or *Riemann sphere*. In the example of the square root, where we come back to the same function after two turns around 0, we *place* two *planes* called *sheets on the complex plane*. Then we have to imagine that we carried out a cut of these two planes by following the axis of negative reals and (that) we identified the edges defined then crosswise. We pass from the higher plane towards the lower plane and conversely. We could have carried out the cut along any half-line going from 0 to ∞. The identification is possible in $\mathbb{R}^3$ only by the introduction of *multiple points* (for a general framework refer to "closed surfaces, surfaces with boundary"). Any point of $\widehat{\mathbb{C}}$, except for 0 and ∞, is covered by two points of the new "surface": *The points 0 and ∞, called "branch points", are covered by only one point. The surface that we have then constructed is a Riemann surface*. We can now define $z \mapsto \sqrt{z}$ on this "surface": $\sqrt{0} = 0$, $\sqrt{\infty} = \infty$ and for two points, located one above the other, the values are taken as opposite by respecting the continuity on each sheet. The function then defined is everywhere continuous and, except for the "branch points", is holomorphic. To define the function $f(z) = \sqrt{(z-a)(z-b)}$, we have to consider a surface with two sheets like the precedent but whose branch points are a and b. If we interest in roots of higher order, we have to introduce a Riemann surface having a number of sheets in relation with this order whose edges are identified in an appropriate way. Branch points of order higher than two occur (e.g. in order to define $z \mapsto \sqrt[n]{z}$ we need n sheets connecting along the axis of negative reals, 1 to 2, 2 to 3,..., $n-1$ to n, n to 1). In the Fig. A.14, topologically the surface corresponds to the surface engendered by two spheres in which we would have carried out a cut along an arc of circle and of which we would have identified the edges 2 by 2, i.e. to a sphere or a complete plane (branch points 0 and ∞ included). The Riemann surface of the square root function is shown in the Fig. A.14.

(2) *Second example:* We have to show a Riemann surface of a sphere with handle. Let f be the function defined by $f(z) = \sqrt{(z-a)(z-b)(z-c)(z-d)}$, where a,b,c,d are pairwise distinct. Except for the points a,b,c,d where the function will be 0 and except for the point *ad infinitum* (i.e. *at infinity)* where the function will be ∞, there will exist two values for f. We introduce then a surface with two sheets, which possesses four branch points. If we cut both sheets along the supposed segments $[a,b]$ $[c,d]$, what is permissible, without common points, and if we identify crosswise the edges thus created, we obtain a Riemann surface which covers $\widehat{\mathbb{C}}$ (see Fig. A.15) on which f is holomorphic except for the branch points. At branch points the function is continuous. Figure A.15 shows a *Riemann surface* corresponding topologically to a *sphere* with a *handle*, i.e. a *torus* (branch points included).
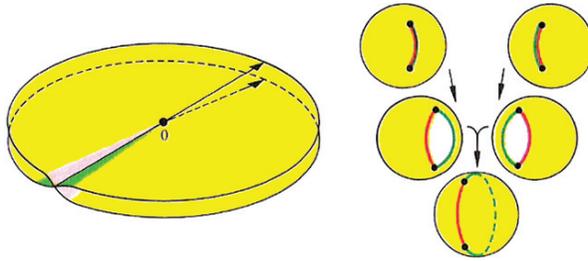
Fig. A.14 *Riemann surface* of the *square root function*
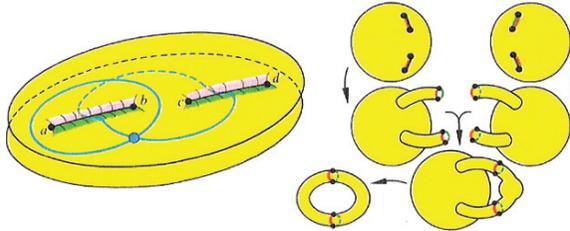
Fig. A.15 *Riemann surface* for a *sphere* with a *handle* (i.e. a *torus*)

Topological structure of Riemann surfaces of both previous examples: The Riemann surfaces constructed until now have, except for the branch points, locally a structure identical to that of the part of the complex plane that they cover. Locally these surfaces, reduced of their branch points, are then everywhere bidimensional but their global topological structure is very different that of a space with two dimensions. For example there exists on the surface, defined in the *second example* above, closed simple curves which do not separate the surface in two parts. By means of specific methods (used for the closed surfaces, surfaces with boundary and their classification) we observe that the introduced surfaces are respectively homeomorphic to a sphere (or a complete plane) with its two branch points, or to a real plane (or to a sphere deprived of a point), or to a torus with its four branch points.

A.4.6 Manifolds and Differentiable Topology

Atlas: An atlas for a *manifold* is a *collection of coordinate patches*[22] that *covers*[23] the *manifold*.

---

[22] *Patch (or local surface):* A "patch", also called a "local surface", is a differentiable mapping $\mathbf{x}: U \to \mathbb{R}^n$, where U is an open subset of $\mathbb{R}^2$. More generally, if A is any subset of $\mathbb{R}^2$, then a map $\mathbf{x}: A \to \mathbb{R}^n$ is a patch provided that $\mathbf{x}$ can be extended to a differentiable map from U into $\mathbb{R}^n$, where U is an open set containing A. Here, $\mathbf{x}(U)$ (or more generally, $\mathbf{x}(A)$) is called the map trace of $\mathbf{x}$.

[23] *Cover:* An element x of a partially ordered set covers another element y if x is greater than y, and the only elements that are both greater than or equal to y and less than or equal to x are x and y themselves. *Covering:* For a set E, a collection of sets whose union contains E.

Differentiable atlas: A family of *embeddings* $h_i : E^n \rightarrow M$ of Euclidian space into a topological space M with the property that $h_i^{-1} h_j : E^n \rightarrow E^n$ is a differentiable map for each pair of indices i, j.

Differentiable manifold: Topological space with a maximal differentiable atlas; roughly speaking, a *smooth surface*.

Manifold: A *topological space* which is *locally Euclidean; there are four types: topological, piecewise linear, differentiable, or complex analytic functions of those in Euclidean space*; *intuitively*, *a surface*. (Furthermore, a manifold can have global topological properties, such as non-contractible loops, that distinguish it from the topologically trivial $\mathbb{R}^n$.)

Differential topology: The branch of mathematics dealing with *differential manifolds*. The motivating force of topology, consisting of the study of smooth (differentiable) manifolds. *Differential topology deals with non-metrical notions of manifolds, while differential geometry deals with metrical notions of manifolds.*

Differential geometry: Differential geometry is the study of *Riemannian manifolds.* Differential geometry deals with metrical notions on manifolds, while differential topology deals with those nonmetrical notions of manifolds.

Differential operator: is an operator on a space of functions which maps a function f into a linear combination of higher-order derivatives of f;

Riemannian manifold: A differentiable manifold where the tangent vectors about each point have inner product so defined as to allow a generalized study of distance and orthogonality.

Riemann space: A Riemannian manifold or subset of a Euclidean space where tensors[24] can be defined to allow a general study of distance, angle, and curvature.[25]

Riemann surfaces: A Riemann surface is a surface-like configuration that covers the complex plane with several, and in general infinitely many, "sheets". These sheets can have very complicated structures and interconnections. Riemann surfaces are one way of representing multiple-valued functions; another is branch cuts.

Riemann sphere: The two-sphere whose points are identified with all complex numbers by a *stereographic projection*. Also known as *complex sphere*.

Riemann tensors: Various types of tensors used in the study of curvature for a Riemann space.[26]

Riemann geometry (i.e. elliptic geometry): The geometry obtained from Euclidean geometry by replacing the parallel line postulate with the postulate that no line may

---

[24] *Tensor: (1)* A tensor is an object relative to a locally Euclidean space which possesses a specified system of components for every coordinate system and which changes under a transformation of coordinates. *(2)* Or a tensor is a multilinear function on the Cartesian product of several copies of a vector space and the dual of the vector space to the field of scalars on the vector space.

[25] *Curvature:* The reciprocal of the radius of the circle which most nearly approximates a curve at a given point; the rate of change of the unit tangent vector to a curve with respect to arc length of the curve.

[26] *Riemann curvature:* A general notion of space curvature at a point of Riemann space which is directly obtained from orthonormal vectors there.

be drawn through a given point, parallel to a given line. Also known as elliptic geometry.

Hyperbolic Riemann surface (i.e. Hyperbolic type): A type of simply connected Riemann surface that can be mapped conformally on the interior of the unit circle.

A.5 Topology

Here is a brief overview. In topology (Fig. A.16), one distinguishes the *point-set topology* (also known as *set-theoretic topology*, *general topology or set topology*) and the *algebraic topology*. The algebraic topology uses algebraic processes to resolve topological problems. The point-set topology originates in *real analysis*. Note that it is possible to purely develop the theory of the convergence by means of the *properties of point sets*, without resorting to an algebraic structure or an ordered

Fig. A.16 Topology

structure. Then a third type of structure appears, i.e. the topological structure. It is based on the notions of neighborhood, closed set, open set, adherent point, accumulation point, convergence, connectedness, compactness, etc. Its purpose is in particular the study and the classification of point sets via such concepts.

Initially, the difficulty was to extract a simple axiomatic system. In fact, two starting points (incidentally equivalent) are possible. As in geometry, one distinguishes certain sets of points by a suitable choice of axioms. For example, starting from the "*neighborhoods axiom*" (sometimes also called "*axiom of neighborhoods*"), it is possible to associate with each point of a set, a system of subsets, the *system of its neighborhoods*. This construction of a topological space is mainly due to Haussdorff. By using the essential notion of *neighborhood*, we build the notion of "open set". It is this notion of open set that (axiomatically reconsidered) leads to the other introduction of topological spaces, equivalent to that of neighborhoods. It simply allows to consider a topology on a set S as a subset of $\mathfrak{B}(S)$ that has certain properties independent of the points of S: it is possible to axiomatically define certain parts of S as "open sets", and the "neighborhoods" are then defined by means of "open sets". This *second definition of a topological space* by its open sets is often better suited to conduct demonstrations. The set of all the open sets (whatever their construction) is called *topology*, and the set S is the *underlying set of the topology*.

It is possible to compare two topological spaces by means of a *map* that links the open sets of both topologies. If there exists a bijective map between the underlying sets, that induces a bijection between the sets of open sets of both spaces, then we say that both topological spaces are *homeomorphic* (it is not possible to distinguish them by topological means). And the map is called *topological map* or *homeomorphism*.

Let us point out that the main properties of topological spaces are those that are invariant by homeomorphism. Then it is possible to speak of *topological invariants* (that we can compare for example with the *invariants* by displacements in geometry).

In addition, there are also a group of topological notions that are preserved by a *class* of more extended maps, i.e. the *continuous maps*. All the homeomorphisms are included in this class, but any continuous map is not a homeomorphism: a homeomorphism is a continuous map whose reciprocal exists and is also continuous. Since any invariant by a continuous map is also a topological invariant, then the continuous maps in topology (like in real analysis) are particularly important. Thus the topology of an Euclidean space is named *continuous geometry*.

Any set that has at least two elements admits more than one topology. The choice of a topology is based on the theory to develop. Therefore, the Euclidean space for example will be in general endowed with its *natural topology*. For the definition of the open sets of this topology, we use the property of the "distance" between two points, i.e. the fact that the space is *metric*. The measure of this distance is described by the properties of a map called *Euclidean metric* in the case of an Euclidean space.

The *metric spaces* are a generalization of Euclidean spaces. Like the Euclidean spaces, they admit a topology defined by a metric. The class of metric spaces can

be studied only by means of its topological properties, i.e. without using the metric (the *measurability*).

A topology is often defined on a set by means of a known topology defined on another given set. A particular map that links both sets allows to "transport" the topology of a set towards the other set, e.g. this is the case for the topology of subsets, quotient sets, cartesian products and sum spaces.

One purpose of the topology is to extend the notion of convergence in classical analysis to the notion of convergence in a topological space. To this end, one generalizes the notion of sequence, and one introduces bases of filters. Nevertheless, this theory is satisfactory only for particular spaces, as for example Hausdorff spaces.

Hausdorff spaces are representatives of spaces verifying the "separation axiom".[27] The most important of these spaces are (in particular for *metrization* questions) the *regular spaces*, *completely regular* and *normal spaces.* The metric spaces are examples of such spaces.

Via the property of "open cover", it is possible to define the *paracompact* and *compact spaces.* The compact spaces (like all the metric spaces) verify the separation axiom. Nevertheless any metric space is not necessarily compact. However, any metric space will be *paracompact* that is an intermediate notion between compactness and normality, essential notion for metrization questions. The topology provides the bases of many disciplines of the analysis, e.g. differential geometry, theory of functions and functional analysis.

A.6 Geometry and Axioms

The construction of the geometry by means of axioms is shown in Fig. 1.17.

A.6.1 Absolute Geometry

Euclid's postulates: (1) A straight line may be drawn joining any point to any other point. (2) Any straight line segment can be extended indefinitely in a straight line. (3) Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as center. (4) All right angles are equal to one another (also written, all right angles are congruent). (5) If a straight line (the transversal) meets two other straight lines so that the sum of the two interior angles on one side of the transversal is less than two right angles, then the straight lines, extended indefinitely if necessary, will meet on that side of the transversal. (i.e. parallel postulate). [Euclid's fifth postulate cannot be proven as a theorem. Euclid himself used only the first four postulates (called "Absolute geometry") for the first 28 propositions of the Elements but was forced to invoke the parallel postulate on the 29th.]

---

[27] *Separation axioms:* Properties of topological spaces such as Haussdorff, regular, and normal which reflect how points and closed sets may be enclosed in disjoint neighborhoods.
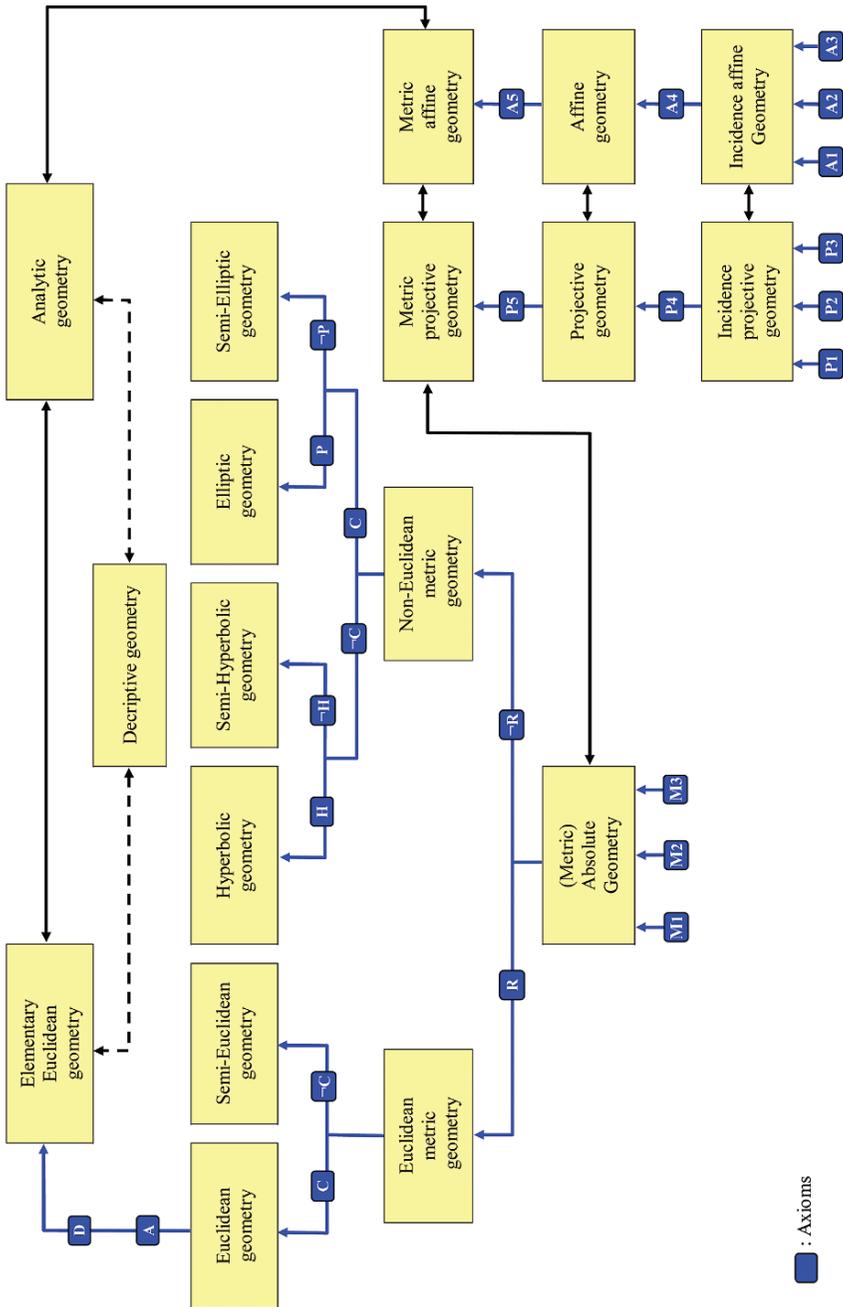
Fig. A.17 Construction of the Geometry

A.6.1.1 Metric Plane

The *point* and the *straight line* are the elementary geometrical concepts appearing in the axioms. Let Π be the *set of points* and Γ the *set of straight-lines* (and respectively the points are denoted p and the straight-lines denoted l). We define the relation **I** (i.e. *incident to*) in $\Pi \times \Gamma$, and the relation $\perp$ (i.e. *orthogonal* or *perpendicular to*) in $\Gamma \times \Gamma$. And if we note l an arbitrary line, it is possible to write $p\mathbf{I}l$ which means that "p is on l" or "l passes by p".

Definition A.62 (Orthogonal collineation). A bijective map g from $\Pi \cup \Gamma$ onto itself such that $g[\Pi] = \Pi$ and $g[\Gamma] = \Gamma$ is called *orthogonal collineation* when it is compatible with the relations **I** and $\perp$, i.e. $p\mathbf{I}l \Rightarrow g(p)\mathbf{I}g(l)$ and $a \perp b \Rightarrow g(a) \perp g(b)$.

Definition A.63 (Collineation). A bijective map g is called a *collineation* when it is compatible with the relations **I** and verifies simply $g[\Pi] = \Pi$ and $g[\Gamma] = \Gamma$.

Definition A.64 (Reflection). A reflection g_L with respect to a straight line L (or reflection of axis L) is an orthogonal collineation different from the identity map $Id_{\Pi \cup \Gamma}$, which verifies $g_L^2 = g_L \circ g_L = Id_{\Pi \cup \Gamma}$ (i.e. an involution, or involutive transformation), and makes L invariant point by point.

Definition A.65 (Isometry). A transformation composed of reflections is an isometry.

An isometry is regarded as a bijective map between two metric spaces that preserves distances, i.e. $d(f(x), f(y)) = d(x, y)$ where f is a map and $d(\cdot, \cdot)$ is the distance function. *Isometries* are sometimes also called *congruence transformations* and two figures that can be transformed into each other by an isometry are said to be congruent.

Definition A.66 (Metric plane). A metric plane is the set of points and straight-lines such that: (**M**1) There exists at least one straight line. Any straight line contains at least three points. One and only one straight line denoted $l(p_1, p_2)$ passes by two distinct points p_1 and p_2. (**M**2) If a is orthogonal to b, then b is orthogonal to a and there exists one and only one common point to both straight lines. For any point p and (straight) line l, a line perpendicular to l passes by p and is unique if p is on l. (**M**3) Any line is the axis of at least one reflection. The composition of three reflections with respect to three lines having in common a point or a perpendicular is a reflection with respect to a line.

The set of theorems derived from these definitions is called the *absolute geometry* in the sense of Bachmann. A generalization of the definitions above allows to define the *geometric metric spaces* of dimension higher than that of metric planes.

A.6.2 Euclidean and Non-Euclidean Metrics

A.6.2.1 Rectangle Axiom

The introduction of new axioms are needed to classify metric planes.

Rectangle axiom (**R**): There exist two distinct lines, which have two distinct perpendiculars in common. *Axiom ¬***R**: This axiom means that any pair of distinct lines has at most one common perpendicular.

A *quadrilateral* also said quadrangle (tetragon) is a four-sided polygon. In a quadrilateral with three right angles the fourth angle is also a right angle.

Definition A.67 (Euclidean or non-Euclidean metric plane). A metric plane verifying **R** is said to be Euclidean metric plane. A metric plane verifying ¬**R** is said to be non-Euclidean metric plane.

Theorem A.14. *In an Euclidean metric plane, two lines having a common perpendicular have all their perpendiculars in common.*

Definition A.68 (Parallel lines). Two lines l_1, l_2 of an Euclidean metric plane which have a common perpendicular are said to be parallel, denoted $l_1 \parallel l_2$.

Definition A.69 (Translation). If $l_1 \parallel l_2$, the composition $g_{l_2} \circ g_{l_1}$ is called a translation.

Theorem A.15. *In an Euclidean metric plane, the composition of three central symmetries is a central symmetry: Given three "points" A, B, C there exists D such that $g_C \circ g_B \circ g_A = g_D$.*

A.6.2.2 Axiom of Connection

Two parallel lines in a plane are seen as two non-concurrent (or non-concur) lines.

Axiom of connection (**C**): Two lines have always one point or one perpendicular in common.

Definition A.70 (Euclidean plane, non-Euclidean plane). An Euclidean metric plane where **C** is satisfied is said to be Euclidean plane. If ¬**C** is satisfied the plane is said to be non-Euclidean.

Remark: There exists an infinity of non-isomorphic Euclidean metric planes whose metric planes having a finite number of points. Let us recall that the term "isomorphic" means "having the same form". Objects which may be represented (or embedded) differently but which have the same essential structure are often said to be "identical up to an isomorphism".

A.6.2.3 Axiom of Polar Trilateral

The Axiom of connection (**C**) helps to specify the non-Euclidean metric plane. If (**C**) is satisfied it is possible to find a *triangle* with *three right angles*.

Axiom of the polar trilateral (**P**): There exist l_1, l_2, l_3 three lines orthogonal two by two (i.e. pairwise orthogonal).

The lines l_1, l_2, l_3 are pairwise orthogonal if and only if $g_{l_3} \circ g_{l_2} \circ g_{l_1} = Id$. Then the central symmetry $g_{l_2} \circ g_{l_1} = g_{l_3}$ is a reflection with respect to a line. In metric planes where **P** is satisfied, **C** and ¬**R** are valid and every central symmetry is also a reflection.

Definition A.71 (Elliptic metric plane). A metric plane where **P** is satisfied is said elliptic. When ¬**R**, **C** and ¬**P** are satisfied, the plane is said semi-elliptic.

An *elliptic plane* can be described as the projective plane with elliptic metric where the distance between two points is defined as the radian angle between the projection of the points on the surface of a sphere (which is tangent to the plane at a point T) from the antipode of the tangent point.

Theorem A.16 (Polar triangle). *If the lines* l_1, l_2, l_3 *are pairwise perpendicular, then* $l_1 l_2 l_3 = 1$, *and conversely.*

A.6.2.4 Hyperbolic Axiom

We have to study the case of non-Euclidean metric planes where ¬**C** is satisfied, that means that there exist lines which are *non connectable*. Let l and l_1 be two lines in accordance with the previous assumption. If the line l_2 passing by the point $P \in l_1$ is perpendicular to l, then $g_{l_2}(l_1) \neq l_1$ and both lines $g_{l_2}(l_1)$ and l are non connectable. However, it is not possible to deduce from the previous axioms that there do not exist more than two lines passing by P and non connectable with l, whence the need to introduce a new axiom:

Hyperbolic Axiom (**H**): It passes through a point at most two lines non-connectable with a given line.

Definition A.72 (Hyperbolic metric plane). A metric plane where ¬**C** and **H** are satisfied is hyperbolic. (In addition, from ¬**C** and **H** it results ¬**R**).

A.6.2.5 Triangle in Different Spaces

Elliptic & Hyperbolic geometries (non-Euclidean geometries): (1) In plane *elliptic geometry* there are no parallels to a given line through a given point. This can be regarded as the geometry of a spherical surface on which antipodal points have been identified and all lines are *great circles*. (2) The *hyperbolic geometry* is a geometry in which there is more than one parallel to a given line through a given point. A helpful model to visualize this geometry presents the *hyperbolic plane* as the *interior of a circle*, in which straight lines take the form of *arcs of circles* perpendicular to the circumference. In Fig. A.18 triangles are shown in Euclidean, elliptic and hyperbolic spaces.
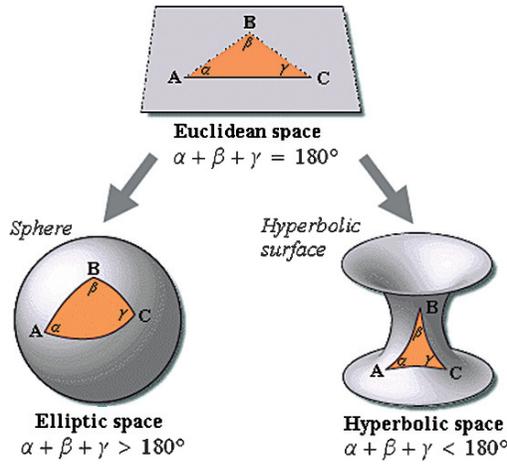
Fig. A.18 Representations of triangles in Euclidean, Elliptic, Hyperbolic spaces
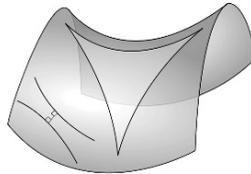
Fig. A.19 A triangle in a saddle-shape plane (a hyperbolic paraboloid), and "two diverging parallel lines" (sometimes called "hyperparallel")

Example (Triangle in hyperbolic plane). It is possible to immerse a triangle in a hyperbolic space, which has the form of a saddle-shape plane (Fig. A.19). The two limiting lines can be said asymptotic and lines sharing a common perpendicular are sometimes called "hyperparallel".

Definition A.73 (Hyperparallel). Two lines in hyperbolic geometry which diverge from each other in both directions.

Definition A.74 (Ideal point). An ideal point is a type of point at infinity in which parallel lines in the hyperbolic plane intersect at infinity in one direction, while diverging from one another in the other.

Remark (Non-Euclidean geometry). After unsuccessful attempts had been made at proving that the parallel postulate could be deduced from the other Euclid's postulates, the matter was settled by the discovery of non-Euclidean by Gauss, Lobachevsky and Bolyai. In these, all Euclid's axioms hold except the parallel postulate. (1) In the so-called hyperbolic geometry, given a point not on a given line, there are at least two lines through the point parallel to the line. (2) In elliptic geometry, given a point not on a given line, there are no lines through the point parallel to the line.

A.6.2.6 Classification of Metric Planes

The metric planes are classified through the axioms: $\mathbf{R}, \mathbf{C}, \mathbf{P}, \mathbf{H}, \neg\mathbf{R}, \neg\mathbf{C}, \neg\mathbf{P}, \neg\mathbf{H}$:

				Satisfied axioms				Plane type
$\mathbf{R}$	$\neg\mathbf{R}$	$\mathbf{C}$	$\neg\mathbf{C}$	$\mathbf{P}$	$\neg\mathbf{P}$	$\mathbf{H}$	$\neg\mathbf{H}$	
■								Euclidean metric
	■							Non-Euclid. metric
■			■					Euclidean
■					■			Semi-Euclidean
	■	■			■			Elliptic
	■	■				■		Semi-Elliptic
	■		■			■		Hyperbolic
	■		■				■	Semi-Hyperbolic

It is possible to write that $\mathbf{P}$ implies $(\neg\mathbf{R}$ and $\mathbf{C})$; $(\neg\mathbf{C}$ and $\mathbf{H})$ implies $\neg\mathbf{R}$; $\neg\mathbf{H}$ implies $\neg\mathbf{C}$.

A.6.3 Affine and Projective Planes

A.6.3.1 Affine Planes

Definition A.75 (Affine plane of incidence). An affine plane of incidence is any set of points and lines which verifies: (**A1**) For any pair of distinct points A and B, there exists one and only one line l incident to A and B. (**A2**) For any line l there exists at least one point A which is not incident to l. (**A3**) Given a line l and a point D which is non incident to l, there exists one and only one line m which is incident to l, (m is parallel to l passing through D).

To introduce the coordinates, consider the *axioms* described in *Desargues and Pappus–Pascal theorems* (also called "closure theorems"), which are not necessarily verified in an affine plane of incidence:

Theorem A.17 ((A4′) Desargues theorem – Affine version). *Given l_1, l_2, l_3 three distinct lines concurrent at the point O or parallel, and the point P_1 and Q_1 of l_1, P_2 and Q_2 of l_2, P_3 and Q_3 of l_3, all distinct from O; if the lines $\ell(P_1, P_2)$ and $\ell(Q_1, Q_2)$ are parallel, as well as $\ell(P_2, P_3)$ and $\ell(Q_2, Q_3)$, then $\ell(P_3, P_1)$ and $\ell(Q_3, Q_1)$ are also parallel.*

Theorem A.18 ((A4) Pappus–Pascal theorem – Affine version). *Given l and m two distinct lines and the points P_1, P_3, P_5 on l, and P_2, P_4, P_6 on m; if the lines $\ell(P_1, P_2)$ and $\ell(P_5, P_4)$ are parallel, as well as $\ell(P_2, P_3)$ and $\ell(P_6, P_5)$, then $\ell(P_3, P_4)$ and $\ell(P_1, P_6)$ are also parallel.*

Hessenberg stated that the Pappus–Pascal theorem leads to the Desargues theorem, whereas the converse is false, indeed there are affine planes of incidence satisfying only the Desargues theorem. When Desargues theorem is verified, then the plane is said "Arguesian plane" (of incidence), i.e. a Desargues's plane.

Definition A.76 (Affine plane). An affine plane is an affine plane of incidence verifying (**A**4).

A.6.3.2 Projective Planes

An affine geometry is a geometry in which properties are preserved by parallel projection from one plane to another. However, others are not, indeed in an affine geometry, the third and fourth of Euclid's axioms do not hold or become meaningless. Affine geometry is regarded as the study of geometry using the methods of linear algebra. The affine transformations preserve collinearity and therefore the straightness and parallel nature of lines and the ratios of distances. In affine geometry, the distinction between secants and parallels is a problem. But by an appropriate completion of the set of points and lines, it is possible to ensure that two lines have always a common point. Any sheaf of parallel lines is regarded as an "improper point" incident to all lines of the sheaf. Any line of the (completed) affine plane possesses one and only one improper point (see also point at infinity, line at infinity). Two parallel lines of the affine plane possess in common a unique point, which is their improper point. An affine plane completed in such a way is said projective in accordance with the definition:

Definition A.77 (Projective plane). A set of points and lines is called a projective plane of incidence if it verifies: (**P**1) For any pair of distinct points A and B, there exists one and only one line l incident to A and B. (**P**2) For any pair of distinct lines l and m, there exists one and only one point A incident to l and m. (**P**3) There exists four points of which three arbitrary ones are not incident to a same line.

If we replace the pair of parallel aside by the pairs aside whose intersections are collinear, that means on a same line, it is possible to write, for the projective plane, Desargues and Pappus–Pascal theorems already written for the affine plane.

Theorem A.19 ((P4′) Desargues theorem – Projective version). *Given l_1, l_2, l_3 three distinct lines concurrent at the point O or parallel, and the point P_1 and Q_1 of l_1, P_2 and Q_2 of l_2, P_3 and Q_3 of l_3, all distinct from O; then the intersection points of the lines $\ell(P_1, P_2)$ and $\ell(Q_1, Q_2)$, $\ell(P_1, P_3)$ and $\ell(Q_1, Q_3)$, $\ell(P_3, P_2)$ and $\ell(Q_3, Q_2)$, are collinear.*

Theorem A.20 ((P4) Pappus–Pascal theorem – Projective version). *Given l and m two distinct lines and the points P_1, P_3, P_5 on l, and P_2, P_4, P_6 on m: then the lines of intersection of the lines $\ell(P_1, P_2)$ and $\ell(P_5, P_4)$, $\ell(P_5, P_6)$ and $\ell(P_3, P_2)$, $\ell(P_3, P_4)$ and $\ell(P_1, P_6)$ are collinear.*

Definition A.78 (Projective plane). A projective plane is a projective plane of incidence verifying (**P**4).

By introducing a new line and adding a point to each line, we can complete an *affine plane* (of incidence) in order to obtain a *projective plane* (of incidence). This can be reversed to transform a projective plane into an affine plane.

A.6.3.3 Metric, Affine, Projec. Planes

These three concepts can be connected:

- An Euclidean plane is always affine.
- An Euclidean plane can always be completed.
- An Elliptic plane is always projective.
- A Hyperbolic plane is not affine nor projective.

A.6.4 Projective Metric

A.6.4.1 Fano's Axiom

Definition A.79 (Quadrangle). A plane figure consisting of four points, each of which is joined to two other points by a line segment (where the line segments may intersect). A quadrangle can therefore be concave or convex; if it is convex, it is called a quadrilateral.

Definition A.80 (Complete quadrangle). A complete quadrangle is formed of four points (the vertices) of which three arbitrary points are not aligned (i.e. not collinear). These vertices define pairwise six lines (the sides). Two sides which do not have a vertex in common are said opposite and their intersection point is said diagonal; there is therefore three diagonals. (An equivalent definition can be provided by using sides.)

The ideal plane of a metric plane is a projective plane. Let us present the following *axiom* or property that is not necessarily satisfied in a projective plane:

(**P5**) (*Fano's axiom*): *The diagonal points of a complete quadrangle are not collinear (i.e. non aligned).*

Definition A.81 (Metric projective plane). A projective plane satisfying the Fano's axiom (**P5**) is called metric projective plane.

An axiom equivalent to (**P5**) is written: (**P5′**) *The diagonal points of a complete quadrangle are not concurrent.* A metric affine plane can be defined by means of (**P5**), and it is possible to write the proposition: (**A5**) *The diagonals of a parallelogram intersect.*

A.6.5 Order and Orientation

A.6.5.1 Bisector Axiom

Euclidean planes can be endowed with a metric. Given u,v two vectors where $u = \binom{u_1}{u_2}$, $v = \binom{v_1}{v_2}$. The vectors u and v are said orthogonal when they cancel the bilinear

form $u_1 v_1 + \rho u_2 v_2$ where $-\rho$ is not a square in the *field K of coordinates*. Note that the factor ρ is named *orthogonality constant*. In such planes, an angle does not have necessarily a bisector (even if the angle is a right angle). K must possess some properties so that the following axioms are satisfied: $(\mathbf{A}^*)$ regarding the right angles and $(\mathbf{A})$ regarding all type of angles (see below).

Axiom (A). *For any pair of lines l and l′ incident to a point M there exists a line d such that l is transformed into l′ by the reflection of axis d.*

Axiom (A*). *For any pair of orthogonal lines l and l′, there exists a line d such that l is transformed into l′ by the reflection of axis d.*

The plane is therefore endowed with an *orthogonal coordinate system*. Given the set of points that we get starting from the point (1,0) by all reflections of axis passing through the origin. Such a set is called a *unit circle* and is made of the points which have as coordinates the couple (x,y) satisfying the equation $x^2 + \rho y^2 = 1$. Then the axiom $(\mathbf{A}^*)$ is equivalent to consider that the unit circle meets the axis of coordinates, i.e. ρ is a square in the field K. Then it is possible to place in a new coordinate system in which the intersection of the unit circle and the axis of ordinates possesses as coordinates $(0,1)$ and in which the unit circle is defined by the equation $x^2 + y^2 = 1$ $(\rho = 1)$. We can say that $(\mathbf{A}^*)$ is satisfied if and only if -1 is not a square in K.

The axiom $(\mathbf{A})$ requires that any line passing through the origin meets the unit circle. That means: $\forall m \in K$, the set of solutions of $x^2 + y^2 = 1$ and $y = mx$ must be a nonempty set, i.e. the axiom $(\mathbf{A})$ is equivalent to notice that the field K is Pythagorean. And because any *Pythagorean field* can be ordered, it results that remarkable properties for the Euclidean planes satisfy this axiom.

A.6.5.2 Orientation

In an ordered field K of coordinates there are two and only two order relations $\leq$ which order the points of an arbitrary line l. Then it is an *oriented line*, called *axis*, and can be denoted $(l, \leq)$. A point split a line into two half-lines. Furthermore, when on an oriented line l the points A,B,C are as follows $A < B$ and $B < C$ then B is said to be between A and C. (Such a statement is also valid for the inverse order on the line l).

The orientation of lines makes possible to orient the plane. Given an axis $(l, \leq)$ it is possible to split all points (of the plane) non-incident to l into two classes called *open half-planes*. Two distinct points A,B belong to the same class if the line l does not meet the line $\ell(A,B)$ between A and B. (The union of l and one of both open half-planes defines an half-plane.) One of both half-planes (separated by l) is arbitrarily defined as positive, and the other is negative. So if the axis $(l', \leq)$ is the image of $(l, \leq)$ by an isometry (i.e. by the composition of a pair number of reflections), then the image of the positive half-plane determined by $(l, \leq)$ is the positive half-plane determined by $(l', \leq)$ (ibid for the negative half-plane). So all axes possess a unique

positive (resp. negative) "side" and the plane is said to be *oriented*. There are two ways to orient a plane. At this stage, let us recall the definition of a flag:

Definition A.82 (Flag). A *flag* is a triplet (a, A, P) where a is a point, A is an *axis* passing through the *point a*, and P is an *oriented plane* containing this axis.

If the axiom (**A**) is verified in Euclidean planes, by an isometry two flags are always reciprocally images (which is the definition of the *free mobility*). Flags also exist in three-dimensional spaces and in higher dimensions. In multidimensional geometry, a flag is a collection of faces of an n-dimensional polytope or simplicial complex, one of each dimension $0, 1, \ldots, n-1$, which all have a common nonempty intersection. In normal three dimensions, the flag consists of a half-plane, its bounding ray, and the ray's endpoint. In geometry, a ray is usually taken as a half-infinite line (also said half-line) with one of the two points and taken to be at infinity.

Definition A.83 (Ray). A ray is a straight-line segment emanating from a point. Also known as half-line.

Definition A.84 (Orientation-preserving). A nonsingular linear map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is orientation-preserving if $\det(\Phi) > 0$.

A.6.5.3 Oriented Plane as Topological Space

The set of intersections of a finite number of open half-planes constitutes a sub-basis in the topological sense, which makes possible to provide any oriented plane with a topological structure called a *natural topology* (such a natural topology satisfies the Hausdorff axiom of separation).

A.6.5.4 Completion

(1) *From a geometrical point of view*, in Euclidean geometry the *field of coordinates* corresponds to the *field of real numbers* $\mathbb{R}$. Note that $(\mathbb{R}, \leq)$ is *archimedean* and is also *complete* (see *Algebraic structure of* $\mathbb{R}$). In geometry the completion is made possible by the Dedekind axiom:

Axiom (D). When a set of points of an axis $(l, \leq)$ is bounded from above, then it has a smallest upper bound.

When the axioms (**A**) and (**D**) are verified by the *field of coordinates* of an *Euclidean plane*, this field possesses an *archimedean order* and is also isomorphic to $\mathbb{R}$. Then we have an Euclidean plane $\mathbb{R}^2$ In such a plane $\mathbb{R}^2$ it is possible to associate with any line segment $[PQ]$ its *length PQ* (see *distance* in an Euclidean plane) and also to associate with any *polygon* the *area of its surface* (or lamina), since length, distance and area are positive reals. (2) *From a more general point of view*, the completion is defined as follows:

Definition A.85 (Completion). For a metric space X, a complete metric space obtained from X by formally adding limits to Cauchy sequences.

A.7 Series Expansions

A.7.1 Taylor Polynomials and Remainders

To approach a function f differentiable at the neighborhood of a point a, we can use the expression $f(a) + (x - a)f'(a)$. In fact one improves the precision of the approximation if one knows the values of derivatives of higher order at a. If such a function is defined by $f : x \mapsto f(x) = \sum_{v=0}^{n} a_v x^v$ then $f(0) = a_0$, $f'(0) = a_1$, $f''(0) = 2!a_2, \ldots, f^{(n)}(0) = n!a_n$ and one can write:

$$f(x) = \sum_{v=0}^{n} \frac{f^{(v)}(0)}{v!} x^v. \tag{A.51}$$

More generally we can write for an unspecified point a:

$$f(x) = \sum_{v=0}^{n} \frac{f^{(v)}(a)}{v!} (x - a)^v. \tag{A.52}$$

The error which is made during the approximation can be canceled: The function is entirely determined by the values of its derivates at one single point a. Since the values of derivatives at the point a of an unspecified function are known until the order n, we have to wonder if the sum defined above is an appropriate approximation of the function at the neighborhood of a.

Definition A.86 (Taylor polynomial). If $f : D_f \to \mathbb{R}$ is n-times differentiable at the point $a \in D_f$ $(n \in \mathbb{N})$, the expression:

$$p_{n,a}(x) = \sum_{v=0}^{n} \frac{f^{(v)}(a)}{v!} (x - a)^v, \tag{A.53}$$

is called nth *Taylor polynomial of* f *expanded at the point* a, *and the expression;*

$$r_{n,a}(x) = f(x) - p_{n,a}(x), \tag{A.54}$$

is called nth *Taylor remainder of* f *expanded at the point* a.

The approximation polynomial $f(a) + (x - a)f'(a)$ introduced previously is thus the *first Taylor polynomial* of f. The nth *Taylor remainder defines a function which is canceled at the point* a and continue at a.

Furthermore there exists a stronger property:

Theorem A.21 (Taylor polynomial). *If* $f : D_f \to \mathbb{R}$ *is n-times differentiable at the point* $a \in D_f$ $(n \in \mathbb{N}^*)$, *if* a *is an interior point of* D_f, *and if* f *is* $(n-1)$-*times differentiable on a neighborhood of* a, *then* $\lim_{x \to a} \dfrac{r_{n,a}(x)}{(x - a)^n} = 0$.

One says then also that (1) the nth Taylor remainder is an infinitely small order strictly higher than n at a, or also (2) $p_{n,a}(x)$ is an approximation of $f(x)$ of order strictly higher than n at a. Therefore the nth Taylor polynomial is in general an approximation which becomes better when n increases. To obtain approximation statements more precise, we have to study the nth Taylor remainder. In the case of $(n+1)$-times differentiable functions, this study is made easier by means of the following results (coming from the application of the mean-value theorem resulting from an extension of the Rolle theorem).

Theorem A.22. *If $f : D_f \to \mathbb{R}$ is $(n+1)$-times differentiable on an open interval containing a ($n \in \mathbb{N}$), then for any x of this interval:*

$$r_{n,a}(x) = \frac{f^{(n+1)}(a + \vartheta(x-a))}{n!}(1-\vartheta)^n(x-a)^{n+1} \tag{A.55}$$

(Cauchy writing of the Taylor remainder).

$$r_{n,a}(x) = \frac{f^{(n+1)}(a + \overline{\vartheta}(x-a))}{(n+1)!}(x-a)^{n+1}$$

(Lagrange writing of the Taylor remainder), where $\vartheta, \overline{\vartheta} \in]0,1[$.

When $a = 0$, we obtain the particular cases:

$$r_{n,0}(x) = \frac{f^{(n+1)}(\vartheta x)}{n!}(1-\vartheta)^n x^{n+1},$$

$$r_{n,0}(x) = \frac{f^{(n+1)}(\overline{\vartheta} x)}{(n+1)!}x^{n+1}.$$

The following theorem shows that the nth Taylor polynomial associated with a function (satisfying the conditions hereafter) can be characterized as the unique polynomial of degree n approaching this function at an order $>n$:

Theorem A.23. *If $f : D_f \to \mathbb{R}$ is n-times differentiable on a neighborhood of the point $a \in D_f$ ($n \in \mathbb{N}^*$) and if there exists a polynomial of degree n, $p(x)$, such that $\frac{f(x) - f(a)}{(x-a)^n}$ define a continuous function verifying $\lim\limits_{x \to a} \frac{f(x) - p(x)}{(x-a)^n} = 0$, then $p(x) = p_{n,a}(x)$.*

By contrast if we take other approximation criteria (e.g. uniform convergence of a sequence of polynomials towards the function on an interval), we obtain different polynomials.

A.7.2 Applications to Local Extrema

In order to have a local extremum at a point c, it is necessary to satisfy the following condition: $f'(c) = 0$ (refer to Rolle theorem). If we know likewise the derivatives of higher order at the point c, it is possible to formulate sufficient conditions which insure that $f'(x)$ changes its sign at the point c.

Theorem A.24. *Given $f : D_f \to \mathbb{R}$ n-times continuously differentiable on a neighborhood of the point c, such that $\forall k \in \{1, 2, \ldots, n-1\}$, $f^{(k)}(c) = 0$ and $f^{(n)}(c) \neq 0$. If n is even and $f^{(n)}(c) > 0$ then f admits at c a strict local minimum. If n is even and $f^{(n)}(c) < 0$ then f admits at c a strict local maximum. If n is odd, f does not admit a local extremum at c.*

In order to show this result, we consider the Taylor polynomial $p_{n-1, c}(x) = f(c)$ and the Taylor remainder $r_{n-1, c}(x) = f(x) - f(c) = \frac{f^{(n)}(c + \vartheta(x-c))}{n!}(x - c)^n$ and we note that on a neighborhood of c $f^{(n)}(c + \vartheta(x - c))$ has the same sign as $f^{(n)}(c)$ whereas $(x - c)^n$ changes its sign at c only if n is odd.

A.7.3 Taylor Series

Any function at least one time differentiable can be approached by one of its Taylor polynomials; We wonder if for an infinitely differentiable function, the approximation by $p_{n,a}(x)$ can be improved ad libitum (at will), i.e. if for any $\varepsilon \in \mathbb{R}_+^*$ and any interval I of D_f containing a, we can find $n \in \mathbb{N}$ such that $\forall x \in I$, $|r_{n,a}(x)| < \varepsilon$. (Generally when such a property is not satisfied, this characterizes a particularly important class of functions.)

Definition A.87 (Taylor series). If $f : D_f \to \mathbb{R}$ is indefinitely differentiable at the point $a \in D_f$, the following expression:

$$p_a(x) = \sum_{v=0}^{n} \frac{f^{(v)}(a)}{v!}(x - a)^v, \tag{A.56}$$

is called Taylor series of f expanded *at the point a*.

The Taylor polynomial $p_{n,a}(x)$ represents for any $n \in \mathbb{N}$ the partial sum of the Taylor series. The Taylor series converges towards $f(a)$ at $x = a$, but the series is not necessarily convergent elsewhere. Moreover, if the series converges at x, this is not always towards $f(x)$ as we can see hereafter: Let $f : \mathbb{R} \to \mathbb{R}$ defined by

$$x \mapsto f(x) = \begin{cases} e^{-1/x^2} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases} \tag{A.57}$$

f is indefinitely differentiable on $\mathbb{R}$ with $f^{(v)}(0) = 0$ for any $v \in \mathbb{N}$. The Taylor series expanded at the point $a = 0$ converges for any $x \in \mathbb{R}$ towards the value 0, therefore $p_0(x) \neq f(x)$ for any $x \neq 0$. $r_{n,0}(x)$ coincides here with $f(x)$ for any x, and does not converge towards 0 when n increases indefinitely.

A.7.4 Analytic Functions

In order to characterize the functions which do not have such a singular behavior, it is necessary first to note that Taylor series are entire series.

Definition A.88 (Analytic function). Let us denote $D^\circ f$ the interior[28] of D_f, $f : D_f \to \mathbb{R}$ is said analytic at $a \in D^\circ f$ if f is expandable in entire series around the point a, i.e. if there exists an entire series $\sum_{v=0}^\infty c_v(x-a)^v$ converging towards $f(x)$ on a neighborhood of a. f is said "analytic" on $D^\circ f$ if f is analytic at any point $a \in D^\circ f$.

If f is analytic at a, the entire series at $(x-a)$ which expands f around a in a sum which coincides with f on an open interval J containing a, with J included at the same time in $D^\circ f$ and in the convergence interval of the series. It is possible to show that at any point of J, f is analytic.

Theorem A.25 (Entire series with a nonnull radius of convergence defines a differentiable function). *An entire series $f(x) = \sum_{v=1}^\infty c_v(x-a)^v$ whose radius of convergence is different from zero defines a differentiable function f, such that: $f'(x) = \sum_{v=1}^\infty v c_v(x-a)^{v-1}$. Both series have the same radius of convergence.*

Therefore an analytic function is differentiable, its derivative is obtained by derivating term by term its entire series, and this is also an analytic function. Finally, any analytic function is indefinitely differentiable, and it is possible to associate with it its Taylor series at any point $a \in D^\circ f$. It comes the uniqueness theorem:

Theorem A.26 (Expansion of an entire series coincides with Taylor series). *The expansion in entire series of an analytic function f around a point $a \in D_f$ coincides with its expansion in Taylor series at the point a.*

For an analytic function a singular behavior analogous to the example hereafter (see figure) is excluded.

Example A.2 (Non-analytic function; Singular behavior of a particular function). Given $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto f(x) = e^{-1/x^2}$ for $x \neq 0$ and 0 for $x = 0$,

---

[28] *Interior:* The interior of a set is the union of all its open subsets. A set constituted only with interior points is an open set, e.g. a disk without its circumference in $\mathbb{R}^2$.
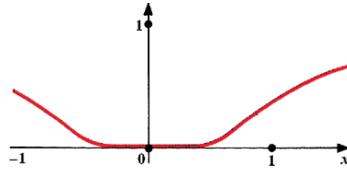
the differentiations give $f'(x) = 2/x^3 \cdot f(x)$, $f''(x) = ((4/x^6) - (6/x^4)) \cdot f(x)$, $f^{(v)}(x) = P_v(1/x) \cdot f(x)$ for r $x \neq 0$, where $P_v(1/x)$ is a polynomial of degree3 v at $1/x$. $f^{(v)}(0) = 0$ for any $v \in \mathbb{N}$. This function is indefinitely differentiable at the point 0, but does not admit expansion in entire series at this point. The Taylor series expanded at 0 is $p_0(x) = 0$. Except for the point 0, its value does not coincide with that of the function. By extending the function on the complexes, one obtains a non holomorphic function at $x = 0$.

As said previously, for an analytic function a singular behavior analogous to the example above is excluded. Such a function is not analytic at the point $x = 0$, although it is indefinitely differentiable. It is however analytic at all other points $a \in \mathbb{R}$, but the radius of convergence of the corresponding entire series is $|a|$, thus the point $x = 0$ is located at the frontier of the convergence interval. This behavior can be *explained* in a satisfactory way only by the study of *complex functions* (refer to theory of functions). Real analysis provides however sufficient conditions for the analysis of a function, for example:

Theorem A.27. *If successive derivatives of a function f indefinitely differentiable on an interval $[a,b]$ are uniformly "minorized", i.e. if there exists $m \in \mathbb{R}$ such that $f^{(n)}(x) > m$ for any $n \in \mathbb{N}$ and any $x \in [a,b]$, then f is analytic on $]a,b[$.*

The hypotheses are verified in particular if all derivatives are positives. An analogous theorem can be stated for an uniform majorization.

A.7.5 Binomial Series

The function $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto (1+x)^n$ is expanded easily in entire series around 0 for $n \in \mathbb{N}$. We have $(1+x)^n = \sum\limits_{v=0}^{\infty} \frac{f^{(v)}(0)}{v!} x^v = \sum\limits_{v=0}^{\infty} \binom{n}{v} x^v$. Binomial coefficients $\binom{n}{v}$ are defined by induction by $\binom{n}{0} := 1$, $\binom{n}{v+1} := \binom{n}{v} \cdot \frac{n-v}{v+1}$, $v \in \mathbb{N}$. Since $\binom{n}{v} = 0$ for $v > n$, only the $n+1$ first terms of the series are different from zero and the radius convergence is infinite. A simple calculation states the following property: $\binom{n}{v} + \binom{n}{v+1} = \binom{n+1}{v+1}$, illustrated by the Pascal's triangle. When we extend the definition of binomial coefficients to any real number r, we observe that for $r \in \mathbb{R} \backslash \mathbb{N}$ we have always $\binom{r}{v} \neq 0$. By using the definition of the power

function for an unspecified real exponent, it is possible to show that we have still $(1+x)^r = \sum_{v=0}^{\infty} \binom{r}{v} x^v$ for $|x| < 1$ (binomial series). The radius of convergence is moreover equal to 1 for $r \in \mathbb{R} \backslash \mathbb{N}$.

Example A.3. Example for $r \in \mathbb{R} \backslash \mathbb{N}$: $(1+x)^{-1} = \dfrac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 + \cdots$, $(1+x)^{1/2} = \sqrt{1+x} = 1 + (1/2)x - (1/8)x^2 + (1/16)x^3 - (5/128)x^4 + \cdots$.

A generalization is written:

$$(a+b)^r = \sum_{v=0}^{\infty} \binom{r}{v} a^{r-v} b^v, \quad \text{with } r \in \mathbb{R}, |b| < |a|. \tag{A.58}$$

A.8 Distribution Theory

Distribution theory, created in the 1950s by Laurent Schwartz, made possible to make rigorous certain heuristic process (i.e. *symbolic calculation of Heaviside*, *delta of Dirac*), to clarify the notion of weak solution of a *partial differential equation* (PDE), and lastly to provide *a general framework to Fourier transform*. Distribution theory is a vast *generalization of the function notion* (of several variables); The fundamental idea is that of *duality:* Distributions are, by definition, linear forms on a space of functions, named *test functions*,[29] which are indefinitely differentiable; this makes possible to define, by transposition, *partial derivatives of distributions* and their *product* with *indefinitely differentiable functions*, then consequently, to apply a *linear differential operator* (with coefficients C^∞) to a distribution, and therefore to search distribution-solutions for linear partial differential equations (with coefficients C^∞). Then the "elementary solution" notion of such a partial differential equation emerges, this notion plays a fundamental part in the resolution of the equation by means of the *convolution* operation of distributions. Let us note that the multiplication operation of functions does not extend to distributions.

Distributions are defined on open parts denoted $\Omega, \Omega', \dots$ spaces $\mathbb{R}^n$, the *test functions on Ω* are the functions with complex values, defined on Ω, of *class C^∞* and with *compact support*;[30] They form a *complex vector space* denoted $\mathfrak{D}(\Omega)$. (*Notation:* for any multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, we pose $|\alpha| = \alpha_1 + \cdots + \alpha_n$ and we denote D^α the differential operator which consists in derivating α_1 times with respect to the variable x_1, α_2 times with respect to the variable x_2, etc.)

Definition A.89 (Distribution). A distribution u on Ω is a linear form on the space $\mathfrak{D}(\Omega)$, denoted $\varphi \rightarrow (u, \varphi)$, satisfying the following continuity condition: If a sequence (φ_k) of "test functions" is such that:

---

[29] *Test function*: An infinitely differentiable function of several real variables used in studying solutions of partial differential equations.

[30] *Support*: A support "supp φ" of a function φ is the closure of the set of points where φ is nonzero.

- supp φ_k is included in a "compact", included in Ω and independent on k.
- For any α, the sequence $(D^\alpha \varphi_k)$ converges uniformly towards 0, then (u, φ_k) tends towards 0.

Distributions on Ω form a vector space denoted $\mathfrak{D}'(\Omega)$.

Example A.4. A function f on Ω, locally integrable (i.e. integrable on any "compact" included in Ω), defines a distribution u by the formula:

$$\langle f, \varphi \rangle = \int_\Omega \varphi(x) f(x) dx.$$

Then it is possible to say that u is the function f. This applies (in particular) in the case $n = 1$ to a Heaviside function H, to a characteristic function on the interval $[0, +\infty]$, and also to the function $\log[x]$; By contrast the function $\frac{1}{x}$ is not locally integrable; however we can associate with it a distribution denoted $pv \frac{1}{x}$ (for Cauchy "principal value": pv); See table hereafter:

(a) $\langle f, \varphi \rangle = \int_\Omega \varphi(x) f(x) dx,$

(b) $H(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$

(c) $\left\langle pv \dfrac{1}{x}, \varphi \right\rangle = \lim_{\varepsilon = 0} \int_{|x| \geq \varepsilon} \dfrac{\varphi(x)}{x} dx,$

(d) $\langle \delta_a, \varphi \rangle = \varphi(a), \ \langle \delta, \varphi \rangle = \varphi(0).$

When Ω contains the point 0, the distribution δ of Dirac, defined by $\langle \delta, \varphi \rangle = \varphi(0)$, plays a fundamental part; but this is not a function in the previous sense above. In an analogous way, we define the distributions δ_a, where $a \in \Omega$.

A next step will be the distributions $\varphi \to D^\alpha \varphi(0)$.

Remark A.6. A convolution is known as an integral that expresses the amount of overlap of one function g as it is shifted over another function f. It therefore *blends* one function with another. For example, it is possible to blend a *singularity* with a *stationary curve,* or *a waveform (e.g. a gauss-curve, or a wavelet,...) with an arbitrary signal while varying the size of the waveform.* A convolution is defined as a product of functions f and g and that are objects in the algebra of Schwartz functions in $\mathbb{R}^n$. The continuous writing of a convolution of two functions over an infinite range is written: $f * g \equiv \int_{-\infty}^{+\infty} f(\tau) g(t - \tau) d\tau = \int_{-\infty}^{+\infty} g(\tau) f(t - \tau) d\tau$. $f * g$ is also written $f \otimes g$. There is also a definition of the convolution which arises in *probability theory* given by $F(t) * G(t) = \int F(t - x) dG(x)$ where $\int F(t - x) dG(x)$ is a Stieltjes integral.

A.8.1 Derivation of Distributions

Definition A.90 (Derivation of distributions). For any multi-index such that $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ and any distribution $u \in \mathcal{D}'(\Omega)$, we define $D^\alpha u$ by the formula

$$\langle D^\alpha u, \varphi \rangle = (-1)^{|\alpha|} \langle u, D^\alpha \varphi \rangle. \tag{A.59}$$

We notice that, for the distributions, the derivation is an operation always possible. When $n = 1$, we write also u', u'', etc., like for the functions.

Example A.5. • If u is a function f of class $C^{|\alpha|}$, $D^\alpha u$ is the function $D^\alpha f$ (we carry out an integration by parts which explains the sign $(-1)^{|\alpha|}$).
• (Case $n = 1$). We have $H' = \delta$. More generally if a function is of class C^1 on $\mathbb{R}$ deprived of a sequence of "isolated points" a_i $(i = 1, 2, \ldots)$ where this function admits left-limits and right-limits $f(a_i-)$ and $f(a_i+)$, its derivative as distribution is equal to its derivative as function (defined only almost everywhere) augmented of the distribution $\sum_i (f(a_i+) - f(a_i-)) \delta_{a_i}$ (formula of jumps).
• (Case $n = 1$) $\log |x|$ has the following derivative $pv \dfrac{1}{x}$. See table hereafter:

$$\text{(a)} \quad \langle D^\alpha u, \varphi \rangle = (-1)^{|\alpha|} \langle u, D^\alpha \varphi \rangle,$$
$$\text{(b)} \quad (\log |x|)' = pv \dfrac{1}{x}.$$

A.8.2 Multiplication

The product of a distribution $u \in \mathcal{D}'(\Omega)$ by a function $f \in C^\infty(\Omega)$ is defined by the formula $\langle fu, \varphi \rangle = \langle u, f\varphi \rangle$. For example: $f\delta_a = f(a)\delta_a$, $f\delta' = f(0)\delta' - f(0)\delta$, $x \, pv \frac{1}{x} = 1$.

A.8.3 Support of Distributions

A distribution $u \in \mathcal{D}'(\Omega)$ is equal to zero on an open set $\Omega' \subset \Omega$ if we have $\langle u, \varphi \rangle = 0$ for any $\varphi \in \mathcal{D}(\Omega)$, with its support included in Ω'. Furthermore, if a distribution is equal to zero on each one of the open sets of an unspecified family, this distribution is equal to zero on their union. This makes it possible to define the *support of a distribution* as being the *complement* of the largest "open set" on which the distribution is equal to zero. See table hereafter:

$$\text{supp } D^\alpha \delta_a = \{a\}, \tag{A.60}$$
$$\text{supp } pv\frac{1}{x} = \text{supp } \log |x| = \mathbb{R}. \tag{A.61}$$

A.8.4 Convolution of Distributions

In such a framework, we pose $\Omega = \mathbb{R}^n$, and if we have to show the variable on which depend u and φ, we will write $\langle u(x), \varphi(x) \rangle$ instead of $\langle u, \varphi \rangle$. The product of convolution $u * v$ of two distributions $u, v \in \mathcal{D}'(\mathbb{R}^n)$ is defined formally by:

$$\langle u * v, \varphi \rangle = \langle u(x), \langle v(y), \varphi(x+y) \rangle \rangle, \tag{A.62}$$

such an expression has a meaning only if certain hypotheses are respected, e.g. (1) If for any compact $K \subset \mathbb{R}^n$, the set of couples $(x, y) \in$ supp $u \times$ supp v verifying $x + y \in K$ is compact, then it is said that supp u and supp v are convolutive; (2) If u and v are the integrable functions f and g, then in such a case $u * v$ is the function defined almost everywhere

$$(f * g)(x) = \int f(x - y)g(y)dy. \tag{A.63}$$

Therefore, the convolution is an operation which is not defined everywhere. (Distribution is always commutative, but associative only if certain conditions are verified.) See table hereafter concerning the convolution:

$$(f * g)(x) = \int f(x - y)g(y)dy, \tag{A.64}$$

$$\delta * u = u * \delta = u, \tag{A.65}$$

$$D^\alpha \delta * u = D^\alpha u. \tag{A.66}$$

A.8.5 Applications to Partial Differential Equations with Constant Coefficients

Let $A = \sum_{|\alpha| \leq m} a_\alpha D^\alpha$ be a differential operator of order m where the coefficients a_α are constants; The operator extends to a map from $\mathcal{D}'(\mathbb{R}^n)$ to itself:

$$A(u) = \sum_{|\alpha| \leq m} a_\alpha D^\alpha u. \tag{A.67}$$

It is possible to write $A(u) = A(\delta) * u$, so that a partial differential equation of the form $A(u) = v$ can be considered as an *algebraic equation*: $A(\delta) * u = v$ in the set $\mathcal{D}'(\mathbb{R}^n)$, *which is not however an algebra for the convolution.*

The *Malgrange–Ehrenpreis theorem* states that there exist always elementary solutions, i.e. distributions E verifying $A(E) = \delta$.

Example A.6 (Convolutions).
(a) (n = 1):

$$A = \frac{d^m}{dx^m} + a_{n-1}\frac{d^{m-1}}{dx^{m-1}} + \cdots + a_0,$$
$$E = zH,$$

where z is the unique solution of the differential equation $A(z) = 0$ with initial conditions

$$z(0) = \cdots = z^{(m-2)}(0) = 0, \quad z^{(m-1)}(0) = 1.$$

(b) ("Wave operator" in $\mathbb{R}^3$, also called "Alembert operator" in $\mathbb{R}^3$):

$$A = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} - \frac{\partial^2}{\partial x_3^2},$$
$$\langle E, \varphi \rangle = \int_{\mathbb{R}^3} \frac{\varphi(|x|, x)}{4\pi|x|} dx,$$

where $x = (x_1, x_2, x_3)$ and $|x| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$.
(c) ("Cauchy–Riemann operator", n = 2):

$$A = \bar{\partial} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right),$$
$$\langle E, \varphi \rangle = \int_{\mathbb{R}^2} \frac{\varphi(x, y)}{\pi(x + iy)} dxdy.$$

(d) ("Laplace operator" in $\mathbb{R}^3$):

$$A = \triangle = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2},$$
$$\langle E, \varphi \rangle = \int_{\mathbb{R}^3} \frac{-\varphi(x)}{4\pi|x|} dx.$$

A.8.6 Use of Elementary Solutions

If E is an elementary solution of a differential operator A, usually we have to look for a solution of an equation $A(u) = v$ under the form $u = E * v$, but this has a sense only if certain conditions are satisfied, for example if the supports of E and v are convolutive.

Example A.7. If we use the previous example (a) with $m = 2$, and we look for the resolution of the Cauchy problem

$$(1): A(f) = g, \quad f(0) = b_0, \quad f'(0) = b_1, \tag{A.68}$$

where g is a given function on $[0, +\infty[$, b_0 and b_1 are constants, f is an unknown function on $[0, +\infty[$. The previous system (1) is equivalent to $A(fH) = gH + (b_1 + a_1 b_0)\delta + b_0 \delta'$ and admits the unique solution $fH = (zH) * (gH + b_1 + a_1 b_0)\delta + b_0 \delta')$, which gives (for $x \geq 0$) the following formula:

$$f(x) = \int_0^x z(x-y)g(y)dy + (b_1 + a_1 b_0)z(x) + b_0 z'(x). \tag{A.69}$$

A.9 Approximation Theory

This section gives a brief overview of basic methods of approximation.[31] In short, the principle of an approximation can be described as follows. Let $f : [a,b] \to \mathbb{R}$ be a function and ε a positive real number, the purpose is to approach f by a function p of a particular set of functions, such a set is generally the set of polynomial functions, so that we have $|f(x) - p(x)| < \varepsilon$ for any $x \in [a,b]$ (see a symbolic example in Fig. A.20). The value at x of a function f, n-times continuously differentiable on an interval $I = [a - \alpha, a + \alpha]$ can be approached by $p_{n,a}(x)$. i.e. we have $\forall x \in I$,

$$|f(x) - p_{n,a}(x)| \leq 2\frac{\alpha^n}{n!}M_n = \mu_n, \tag{A.70}$$

with $M_n = \sup_I \left|f^{(n)}(x)\right|$. And the error made by replacing f by $p_{n,a}$ on I is uniformly "majorized" by μ_n.

Definition A.91 (Uniformly approachable map). $f : [a,b] \to \mathbb{R}$ is uniformly approachable by functions of a set F if for any $\varepsilon \in \mathbb{R}_+^*$ there is $p \in F$ such that:

$$|f(x) - p(x)| < \varepsilon \quad \text{for any } x \in [a,b]. \tag{A.71}$$

p is then called a uniform approximation of f on $[a,b]$. The polynomial functions can be used to calculate these approximations. Approximations by polynomial functions are particularly important.

Theorem A.28 (Weierstrass approximation). *Any continuous function $f : [a,b] \to \mathbb{R}$ is uniformly approachable by polynomial functions.*
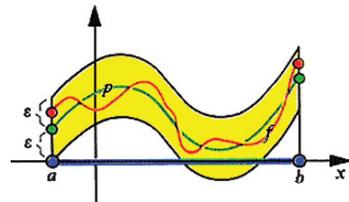
Fig. A.20 Approximation

---

[31] Approximation theory belongs to Differential calculus.

It is possible to apply this theorem to the functions f defined on $[0,1]$ by choosing particular approximation functions $b_n(f)$:

$$b_n(f)(x) = \sum_{v=0}^{n} f\left(\frac{v}{n}\right) \cdot \binom{n}{v} x^v(1-x)^{n-v}, \tag{A.72}$$

which corresponds to Bernstein polynomials, (such approximation functions are polynomial and converge uniformly towards f on $[0,1]$). We can use this approach for the functions f defined on $[a,b]$, $\varphi(t):[0,1] \to [a,b]$ by $t \to \varphi(t) = t(b-a)+q$ then $f \circ \varphi$ is approachable by the functions $b_n(f \circ \varphi)$ thus f is approachable by the functions $b_n(f \circ \varphi) \circ \varphi^{-1}$ which verify $b_n(f \circ \varphi) \circ \varphi^{-1}(x) = b_n(f \circ \varphi)\left(\frac{x-a}{b-a}\right)$ and are then polynomial.

A.9.1 Best Approximations

The convergence of the sequence $b_n(f)$ towards f can be relatively slow. We can wonder if there are better sequences of polynomial functions to build better approximations of f. In this aim, it is preferable to be located within the more general framework of *normed vector spaces*.

Definition A.92 (Best approximation). Let $(V, \| \ \|)$ be a *normed vector space* and $U \subseteq V$. $\bar{f} \in U$ is called best *approximation* of $f \in V$ relative to U and $\| \ \|$ if $\|f - \bar{f}\| \leq \|f - g\|$ for any $g \in U$.

The best approximation thus depends on the choice of U and the norm $\| \ \|$. The theorem which follows gives a condition of existence.

Theorem A.29 (Condition of existence). *If U is a vector subspace of finite dimension of $(V, \| \ \|)$, then for any $f \in V$ there exists a best approximation $\bar{f}$ of f relative to U and $\| \ \|$.*

A.9.1.1 Tchebychev Polynomials

Let $C^0[a,b]$ be a vector space of real continuous functions defined on $[0,1]$, on which one defines a norm $\| \ \|_\infty$ by

$$\|f\|_\infty := \sup_{x \in [a,b]} (|f(x)|) \tag{A.73}$$

which is the Tchebychev norm. The polynomial functions of degree at the most (n-2) constitute a vectorial subspace S_{n-1} of $C^0[a,b]$ of dimension n. It is possible to wonder what is the best approximation of the power function $f_n(x) = x^n$ relative to S_{n-1} and $\| \ \|_\infty$. If one denotes $\bar{f}_n$ this *best approximation* of f_n, then $t_n = f_n - \bar{f}_n$ is among all the polynomial functions g_n of degree n (whose coefficient of the

monomial of degree n is 1), the one for which the maximum of $|g_n(x)|$ is the smallest on $[a, b]$. The $t_n(x)$ are Tchebychev polynomials, and one can write:

$$t_n(x) = \frac{(b-a)^n}{2^{2n-1}} \cos\left(n \cdot \arccos\left(\frac{2x}{b-a} - \frac{b+a}{b-a}\right)\right). \tag{A.74}$$

With a maximum for their absolute value at $\frac{(b-a)^n}{2^{2n-1}}$. In particular for the interval $[-1, 1]$ we have $t_n(x) = 2^{1-n} \cos(n \cdot \arccos x)$.

A.9.1.2 Legendre Polynomials

If we use the set $C^0[a, b]$ with the Euclidean norm (and not the Tchebychev norm) which is written:

$$\|f\|_2 := \sqrt{\int_a^b [f(x)]^2 dx}, \tag{A.75}$$

the previous method leads to the definition of Legendre polynomials $l_n(x)$. l_n is among all the polynomial functions g_n of degree n (whose coefficient of the monomial of degree n is 1) the one for which $\int_a^b [g_n(x)]^2 dx$ is the smallest. In particular for the interval $[-1, 1]$ we have:

$$l_n(x) = \frac{n!}{(2n)!} \frac{d^n}{dx^n} (x^2 - 1)^n. \tag{A.76}$$

The norms $\|\ \|_\infty$ and $\|\ \|_2$ do not provide the same result.

A.9.1.3 Least-Square Methods

(Gauss) Least-square methods: The approximation $\bar{f}$ of f must be a linear combination of n linearly independent functions g_v:

$$\bar{f} = \sum_{v=1}^n \alpha_v g_v, \tag{A.77}$$

with $\alpha_v \in \mathbb{R}$. The coefficients α_v minimize the value of $(\sum_{\mu=1}^n (\sum_{v=1}^n \alpha_v g_v(x_\mu) - y_\mu)^2)^{1/2}$. To do this, we have to equalize to zero (for any $i \in \{1, \ldots, m\}$) the partial derivatives: $\frac{\partial}{\partial \alpha_i} \sum_{\mu=1}^m (\sum_{v=1}^n \alpha_v g_v(x_\mu) - y_\mu)^2$. Consequently, one obtains a system of n equations with n unknowns $\alpha_1, \ldots, \alpha_n$, which generally admits a solution. This technique presented quickly is that of least squares.

A.10 Interpolation Theory

Let $(n+1)$ be the points $P_\mu(x_\mu, y_\mu)$, $(\mu \in \{0,\ldots,n\}$, pairwise distinct abscissas $x_\mu)$, the purpose is to find a polynomial function p_n of degree n which takes the value y_μ at each point x_μ. Then we determine the coefficients α_v of the expression:

$$p_n(x) = \sum_{v=0}^{n} \alpha_v x^v, \qquad (A.78)$$

in order to have: $\sum_{v=0}^{n} \alpha_v x_\mu^v = y_\mu, \forall \mu \in \{0,\ldots,n\}$. It is a system of linear equations whose unknowns are $\alpha_0,\ldots,\alpha_n$ and which admits a single solution. The associated homogeneous system is written: $\sum_{v=0}^{n} \alpha_v x_\mu^v = 0$, which corresponds to $p_n(x_\mu) = 0$ with $\mu \in \{0,\ldots,n\}$. If this system had a solution $(a_0,\ldots,a_n) \neq (0,\ldots,0)$ the polynomial p_n would have $(n+1)$ roots (while not being identically null), what is impossible since it is of degree n at most. Therefore, the homogeneous system admits only the trivial solution $(0,\ldots,0)$, this proves that the determinant of the associated matrix is nonzero and demonstrates the (previous) announced result according to Weierstrass approximation theorem. There are several methods which allow to construct interpolation polynomials.

A.10.1 Lagrange Method

This method is also called *Lagrange interpolation* polynomial. The interpolation polynomial associated with the points $P_\mu(x_\mu, y_\mu)$, can be written in the form:

$$p_n(x) = \sum_{v=0}^{n} \prod_{\substack{\mu=0 \\ \mu \neq v}}^{n} \frac{x - x_\mu}{x_v - x_\mu} y_v. \qquad (A.79)$$

Calculations resulting from this method are often rather long. Furthermore, if one adds new points to the set to be interpolated, it is necessary to restart all calculations. Unlike Lagrange method, the Newton–Gregory method avoids such a problem.

A.10.2 Newton–Gregory method

By using the following writing:

$$p_n(x) = c_0 + c_1(x - x_0) + \cdots + c_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}), \qquad (A.80)$$

then the coefficient c_v are easily calculated starting from the equations $P_n(x_\mu) = y_u, \mu \in \{0,\ldots,n\}$. Indeed, $p_n(x_0) = y_0$ leads to $c_0 = y_0$ and by resolving successively the equations after having replaced the unknowns already determined

by their values, we have then: $c_1 = \frac{y_1 - y_0}{x_1 - x_0}$, $c_2 = \frac{\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}}{x_2 - x_0}$, etc. Formally, if we define by recurrence for all $\mu, \nu \in \mathbb{N}$, $\mu \geq \nu$: $[x_\mu] := y_\mu$ (Increase of order 0), $\frac{[x_\mu x_{\mu-1}, \ldots, x_{\mu-\nu+1}] - [x_{\mu-1}, \ldots, x_{\mu-\nu}]}{x_\mu - x_{\mu-\nu}}$ (increase of order ν with $\nu > 0$), then we get $c_\nu = [x_\nu x_{\nu-1}, \ldots, x_0]$ for any $\nu \in \{0, \ldots, n\}$. Therefore, it is possible to evaluate all coefficients starting from a simple calculation method. If we add new points to the set to be interpolated, it only suffices of calculating new coefficients. The order of the abscissas of points to be interpolated does not play any part. The method is particularly useful when these abscissas are equidistant. If x_0 is the smallest abscissa, and if $x_{\mu+1} - x_\mu = h$ for any $\mu \in \{0, \ldots, n-1\}$, we write, instead of the increases of order ν, the differences of order ν defined by:

$$\Delta^0 y_\mu := y_\mu,$$
$$\Delta^\nu y_\mu := \Delta^{\nu-1} y_{\mu+1} - \Delta^{\nu-1} y_\mu, \quad \nu > 0,$$
$$\text{Then } c_\nu = [x_\nu x_{\nu-1}, \ldots, x_0] = \frac{1}{\nu!} \cdot \frac{\Delta^\nu y_0}{h^\nu} \text{ for any } \nu \in \{0, \ldots, n\}.$$

Remark A.7. We obtain an analogous interpolation formula if (always in the case of equidistant abscissas) x_0 is not the smallest but the greatest abscissa or the median abscissa.

A.10.3 Approximation by Interpolation Polynomials

To approach a continuous function f, it is possible to choose on its graph $n+1$ pairwise distinct interpolation points $(x_0 < \cdots < x_\nu < \cdots x_n)$ and construct the corresponding interpolation polynomial p_n. At each interpolation abscissa we have then $f(x_\nu) = p_n(x_\nu)$. *The error* $(= \varepsilon)$ made during the interpolation is denoted:

$$\text{Error} = f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0), \ldots, (x - x_n), \qquad \text{(A.81)}$$

where ξ corresponds to a suitable value belonging to an interval containing $[x_0, x_n]$ and $\{x\}$. Taylor polynomials appear thus as borderline cases of interpolation polynomials, when all interpolation points are confused.

It is possible to generalize the approximation principle by interpolation (presented previously) to sets of points non necessarily pairwise distinct. If k points are confused $(k > 1)$, to define a single interpolation polynomial it is necessary to give itself k independent conditions which will be obtained by writing the equality of successive derivatives of the function to be approached and interpolation polynomial until the order $k - 1$. Lastly, we wonder how choose interpolation abscissas in an appropriate way, on a segment [a,b] so that the polynomial interpolation of a given (sufficiently differentiable) function is most precise possible. *The answer depends on the norm of the space of considered functions.* If we take Tchebychev norm,

the interpolation which leads to the best approximation is the one which uses the abscissas of zeros of the Tchebychev polynomial $t_{n+1}(x)$. In a similar way, for the Euclidean norm, we have to use the abscissas of zeros of the Legendre polynomial $l_{n+1}(x)$.

A.11 Numerical Resolution of Equations

A.11.1 Simple Iterative Methods

A main purpose of the numerical calculation is to resolve the equation $f(x) = 0$, $x \in [a,b]$. If we pose $g(x) = x + h(x) \cdot f(x)$, where $h(x)$ is a function can be selected arbitrarily (only condition: h different from zero on [a,b]); resolve the previous problem is equivalent to resolve: $g(x) = x$ with $x \in [a,b]$. To approach solutions of this type of equation we use a well-known iterative method, with an initial value x_0, and we can write $x_{v+1} = g(x_v), \forall v \in \mathbb{N}$. Under certain conditions the sequence represented by (x_v) converges and its limit denoted ξ is a solution of $f(x) = 0$.

Theorem A.30 (Convergence criterion). *If $g : [a,b] \to [a,b]$ defined by $x \to g(x)$ is continuously differentiable on $[a,b]$ and if $|g'(x)| < 1$ for any $x \in [a,b]$, then $g(x) = x$ admits exactly a solution ξ and $\xi = \lim_{v \to \infty} x_v$ where the sequence (x_v) is defined by $x_0 \in [a,b]$ and $x_{v+1} := g(x_v), \forall v \in \mathbb{N}$.*

Speed of convergence: In fact the convergence speed can be evaluated even starting from the first iteration by means of the equality:

$$|\xi - x_v| \leq \frac{L^v}{1-L} |x_1 - x_0|, \tag{A.82}$$

with $L = \sup_{x \in [a,b]} |g'(x)|$. If f is continuously differentiable on $[a,b]$, to apply the method to $f(x) = 0$ we choose $g = \mathbf{1}_\mathbb{R} + f \cdot h$ with h *satisfying the convergence criterion* (e.g. if $-2 < f'(x) < 0$, we will choose $h = 1$, and if $0 < f'(x) < 2$, $h = -1$ will be adapted).

A.11.2 Newton–Raphson Method

We assume that f' does not cancel itself, we want to choose h such that $g'(\xi) = 0$. From $g'(x) = 1 + h'(x) \cdot f(x) + h(x) \cdot f'(x)$, then it comes $h(\xi) = -1/f'(\xi)$, and it seems appropriate to choose $h(x) = -1/f'(x)$. In the case of a twice continuously differentiable function, the convergence criterion of a sequence which satisfies the recurrence relation: $x_{v+1} = x_v - \frac{f(x_v)}{f'(x_v)}$, is written $\forall x \in [a,b] : \left| \frac{f(x)f''(x)}{[f'(x)]^2} \right| < 1$. Such a method admits a simple geometrical interpretation: If we draw the tangent to the

graph of f at the point $(x_v, f(x_v))$, x_{v+1} is the abscissa of the intersection point of this tangent with the axis of abscissas.

A.11.3 Linear Interpolation Method (Regula Falsi)

If we replace, in Newton–Raphson method, the slope of the tangent $f'(x_v)$ by the slope of the secant passing by the points $(x_v, f(x_v))$ and $(x_{v-1}, f(x_{v-1}))$, we get the recurrence relation $x_{v+1} = x_v - \dfrac{x_v - x_{v-1}}{f(x_v) - f(x_{v-1})} f(x_v)$ for $v \geq 1$. x_{v+1} is the abscissa of the intersection point of the considered secant with the axis of abscissas. The coefficient before $f(x_v)$ not depending only on x_v, the convergence criteria of this method are more complicated.

A.11.4 Horner's Schema

To evaluate numerically the expressions $f(x_v)$ or $f'(x_v)$ which intervene for example in the previous methods, it is possible to use, in the case of a polynomial function f, the Horner schema. It is based on the relation: $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = ((\cdots (a_n x + a_{n-1}) x + \cdots + a_2) x + a_1) x + a_0$. The figure hereafter explains briefly the calculation method of $f(x_v)$, which uses additions and only n multiplications by x_v. Calculation method explained briefly is as follows: Given $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, we look for the value of $f(x_v)$,
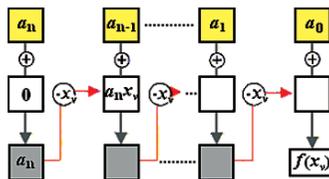
The iteration of this process makes possible to calculate the coefficients $\frac{1}{\mu!} f^{(\mu)}(x_v)$ of the Taylor expansion of f at the point x_v. The repetition of this method makes possible to calculate $f'(x_v)$. Indeed, the n first numbers of the sixth row of the table above are the coefficients of the polynomial defining the continuous extension of the increase rate function $m_{x_v}(x) = \frac{f(x) - f(x_v)}{x - x_v}$ that verifies $\overline{m}_{x_v}(x_v) = f'(x_v)$.

Remark A.8. More generally, the iteration of this method makes possible to calculate the coefficients $\frac{1}{\mu!} f^{(\mu)}(x_v)$ of Taylor expansion of f at the point x_v.

Example A.8. Given $f(x) = x^4 - 2x^3 + x^2 - 7x + 3$, we look for $f(3)$, $f'(3)$, $f''(3)$, $f'''(3)$:

1	−2	1	−7	3	
0	3	3	12	15	
1	1	4	5	18	$= f(3)$
0	3	12	48		
1	4	16	53		$= f'(3)$
0	3	21			
1	7	37			$= \frac{1}{2!} f''(3)$
0	3				
1	10				$= \frac{1}{3!} f'''(3)$

A.11.5 Graeffe Method

When we consider a polynomial function f_0 of degree n defined by $x \mapsto a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ as a function on $\mathbb{C}$, the fundamental theorem of the complex algebra insures that its expression is decomposed into the product of a constant with n factors of the form $a_n (x - \xi_1)(x - \xi_2) \cdots (x - \xi_n)$. By developing this product and by identifying the coefficients of the obtained polynomial with those of the initial polynomial, we show that (for $1 \leq m \leq n$):

$$\sum_{v_1 < v_2 < \cdots < v_m} (-\xi_{v_1})(-\xi_{v_2}) \cdots (-\xi_{v_m}) = \frac{a_{n-m}}{a_n}. \tag{A.83}$$

This property of roots (also called Vieta th.) makes possible to evaluate them when their absolute values are pairwise distinct. Indeed, if for example $|\xi_1| > |\xi_2| > \cdots > |\xi_n|$, the absolute values of roots being distant to each other, we have $\xi_v \approx -\dfrac{a_{n-v}}{a_{n-v+1}}$ for any $v \in \{1, \ldots, n\}$. This approximative calculation is possible only if all coefficients a_v are nonzero. Otherwise, or if we want to improve the precision of results, we build by recurrence $f_k(x^{2^k}) = (-1)^n \cdot f_{k-1}(x^{2^{k-1}}) \cdot f_{k-1}(-x^{2^{k-1}}) = a_{k,n} x^{2^k \cdot n} + a_{k,n-1} x^{2^k \cdot (n-1)} + \cdots + a_{k,0}$. For values of k sufficiently large, the coefficients $a_{k,v}$ are all nonzero and $\lim\limits_{k \to \infty} \sqrt[2^k]{-\dfrac{a_{k,n-v}}{a_{k,n-v+1}}} = \pm \xi_v$ for any $v \in \{1, \ldots, n\}$ with an appropriate choice of the sign.

A.12 Second-Order Differential Equations

Second-order differential equations (by using explicit representation) can be written $y'' = f(x, y, y')$, where f is a function defined on $G \subset \mathbb{R}^3$ having real values. A function $F : I_x \to \mathbb{R}$ defined by $x \mapsto y = F(x)$ is a solution of the differential

equation if it is twice differentiable and if for any $x \in I_x$, $(x, F(x), F'(x)) \in G$ and $F''(x) = f(x, F(x), F'(x))$. A solution denoted F_A is said solution of the problem of initial values $(x_0, y_0, y_0') \in G$, if $(x_0, F_A(x_0), F_A'(x_0)) = (x_0, y_0, y_0')$. Resolve a second-order differential equation is usually more difficult than a first-order equation. There are many methods to resolve particular types of equation. Generally we look for to reduce an equation of second-order to a system of first-order equations by a *change of variables*. Within this framework, differential equations of order 2 are very important.

A.12.1 General Resolution of Linear Differential Equations of Second-Order

A linear second-order differential equation can be written $y'' + a_1(x) \cdot y' + a_0(x) \cdot y = s(x)$, where a_0, a_1, s are continuous functions on I_x having real values. If $s = 0$ on any I_x, the equation becomes *homogeneous*; By contrast it is said to be non-homogeneous in the opposite case. For the general resolution we use the notion of linear independence of two functions $f_1 : I_x \to \mathbb{R}$ and $f_2 : I_x \to \mathbb{R}$: Both functions f_1 and f_2 are said linearly independent if the functional relation on I_x $c_1 \cdot f_1 + c_2 \cdot f_2 = 0$ where $(c_1, c_2) \in \mathbb{R}^2$ requires $c_1 = c_2 = 0$. We define the *Wronskian*[32] of two functions by $W(F_1, F_2)(x) = F_1(x) \cdot F_2'(x) - F_1'(x) \cdot F_2(x)$. We show that two solutions (of the homogeneous equation) defined on I_x, F_1 and F_2 are linearly independent if for any $x \in I_x$, $W(F_1, F_2)(x) \neq 0$ and

Proposition. (1) There exist two linearly independent solutions F_1 and F_2 defined on I_x for the homogeneous equation $y'' + a_1(x) \cdot y' + a_0(x) \cdot y = 0$. The set of solutions of the homogeneous equation is the set of $F_h(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x)$ with $(\alpha_1, \alpha_2) \in \mathbb{R}^2$. (2) There exists one solution F_p defined on I_x for the non-homogeneous equation. The set of solutions of the non-homogeneous equation is the set of $F_{nh}(x) = F_p(x) + \alpha_1 F_1(x) + \alpha_2 F_2(x)$ with $(\alpha_1, \alpha_2) \in \mathbb{R}^2$. (Furthermore we know that any problem of initial conditions admits a single solution.)

A.12.2 Resolution of Linear Homogeneous Equations

Then we are looking for two linearly independent solutions. If we do not know at the beginning the nonzero solution, the problem becomes very difficult; this is why we often use in practice *approached solutions* (when the tables or intuitions are not

---

[32] *Wronskian:* An $n \times n$ matrix whose ith row is a list of the $(i-1)$st derivatives of a set of functions $f_1, \ldots, f_n$; ordinarily used to determine linear independence of solutions of linear homogeneous differential equations.

sufficient). By contrast, if we know already a solution F_1 (which furthermore is not canceled on I_x) the following method makes it possible to calculate another solution F_2 independent of F_1. We pose $F_2(x) = v(x) \cdot F_1(x)$, and v is then solution of the differential equation: $v'' = - \left(2\frac{F_1'(x)}{F_1(x)} + a_1(x) \right) \cdot v'$. The integration is easy and we take for example $F_2(x) = F_1(x) \int_{x_0}^x \frac{1}{|F_1(t)|^2} \exp\left(- \int_{x_0}^t a_1(u)du \right) dt$ with $x_0 \in I_x$.

A.12.3 Particular Solution of a Non-Homogeneous Equation

If F_1 and F_2 are two linearly independent solutions of the corresponding homogeneous equation, it is possible to find a particular solution of the non-homogeneous equation by the *method of variation of constants* by posing $F_p(x) = v_1(x) \cdot F_1(x) + v_2(x) \cdot F_2(x)$. Then it suffices to resolve both differential equations: $v_2'(x) = -\frac{s(x)F_2(x)}{W(F_1,F_2)(x)}$ and $v_1'(x) = +\frac{s(x)F_1(x)}{W(F_1,F_2)(x)}$.

A.12.4 Linear Differential Equations of Second-Order with Constant Coefficients

We are interested here in the equations of the type: $y'' + p \cdot y' + q \cdot y = s(x)$ where p, q are two real numbers. First, we resolve the homogeneous equation $y'' + p \cdot y' + q \cdot y = 0$. So, we consider the *characteristic equation* $z^2 + pz + q = 0$: its resolution leads to distinguish three cases (see Fig. A.21):

(a) $p^2 - 4q > 0$: two distinct real solutions z_1 and z_2.
(b) $p^2 - 4q = 0$: a single real solutions z.
(c) $p^2 - 4q < 0$: two complex (conjugate) solutions z_1 and z_2.

The solutions of the homogeneous equation are then:

(a) $F_h(x) = \alpha_1 e^{z_1 x} + \alpha_2 e^{z_2 x}$.
(b) $F_h(x) = (\alpha_1 + \alpha_2 x)e^{zx}$.
(c) $F_h(x) = e^{\mathrm{Re}\, z_1 x}(\alpha_1 \cos(\mathrm{Im}\, z_1 x) + \alpha_2 \sin(\mathrm{Im}\, z_1 x))$.

In order to find a particular solution of the non-homogeneous equation, it is possible to use the methods of variation of constants. However there exist simpler methods if s is a *trigonometric function*, *exponential* or an *entire series*.

Let us consider a (standard) linear homogeneous equation of second order with constant coefficients $(p \geq 0, q > 0)$, $y'' + p \cdot y' + q \cdot y = 0$ where $p = \frac{r}{m}, q = \frac{D}{m}, x = t$ $(r \in \mathbb{R}^+, m \in \mathbb{R}_+^*, D \in \mathbb{R}_+^*)$.
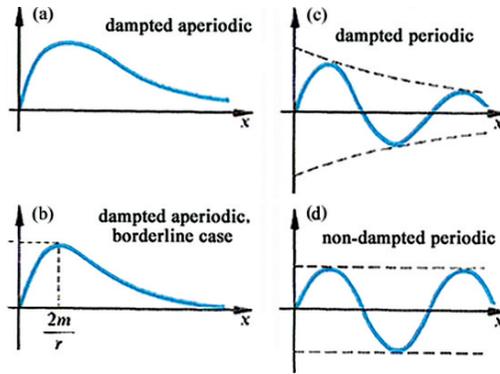
Fig. A.21 Qualitative graph of solutions

General solution	For initial conditions $(0,0,v_0)$
(a) $p^2 - 4q > 0$, i.e. $r^2 > 4mD$: $F(x) = \alpha_1 e^{z_1 x} + \alpha_2 e^{z_2 x}$, with $z_{1/2} = \dfrac{-r \pm \sqrt{r^2 - 4mD}}{2m}$, where $z_2 < z_1 < 0$	$\alpha_1 + \alpha_2 = 0 \wedge \alpha_1 z_1 + \alpha_2 z_2 = v_0$ $\Leftrightarrow \alpha_1 = \dfrac{v_0}{z_1 - z_2} \wedge \alpha_2 = \dfrac{v_0}{z_2 - z_1}$ $F_A(x) = \dfrac{v_0}{z_1 - z_2}\left(e^{z_1 x} - e^{z_2 x}\right)$
(b) $p^2 - 4q = 0$, i.e. $r^2 = 4mD$: $F(x) = (\alpha_1 + \alpha_2 x)e^{z x}$, with $z = -\dfrac{r}{2m}$	$\alpha_1 = 0 \wedge \alpha_1 z + \alpha_2 = v_0$ $\Leftrightarrow \alpha_1 = 0 \wedge \alpha_2 = v_0$ $F_A(x) = v_0\, x\, e^{-\frac{r}{2m}x}$
(c) $p^2 - 4q < 0$, i.e. $r^2 < 4mD$: $F(x) = e^{\operatorname{Re} z_1 x}(\alpha_1 \cos \operatorname{Im} z_1 x + \alpha_2 \sin \operatorname{Im} z_1 x)$, with $z_{1/2} = \dfrac{-r \pm i\sqrt{4mD - r^2}}{2m}$	$\alpha_1 = 0 \wedge \alpha_1 \operatorname{Re} z_1 + \alpha_2 \operatorname{Im} z_1 = v_0$ $\Leftrightarrow \alpha_1 = 0 \wedge \alpha_2 = \dfrac{v_0}{\operatorname{Im} z_1}$ $F_A(x) = \dfrac{2mv_0}{\sqrt{4mD - r^2}}\, e^{-\frac{r}{2m}x}$ $\times \sin \dfrac{\sqrt{4mD - r^2}}{2m} x$
Particular solution: $r = 0$ (without friction)	$F_A(x) = v_0 \sqrt{\dfrac{m}{D}} \sin \sqrt{\dfrac{m}{D}} x$

A.13 Other Reminders

A.13.1 Basic Reminders in Mathematics and Statistics

A.13.1.1 Matrix Diagonalization

The matrix diagonalization is the process of taking a square matrix and converting it into a diagonal matrix. Matrix diagonalization is equivalent to transforming the underlying system of equations into a (special) set of coordinate axes in which the matrix takes this *canonical form*. Matrix diagonalization is also equivalent to finding the matrix's eigenvalues, which turn out to be the "entries" of the diagonalized matrix. Similarly, the eigenvectors make up the new set of axis corresponding to the diagonal matrix. A "change of basis" in a *vector space E* on *K* (i.e. a linear map from *E* to *E*) of dimension *n*. The matrix which allows to pass from a basis $(V_1, \ldots, V_n)$ to a basis $(\widehat{V}_1, \ldots, \widehat{V}_n)$ is the square matrix *H* of order *n* whose terms of the *i*th column are the coordinates of $\widehat{V}_i$ in the basis $(V_1, \ldots, V_n)$, $1 \leqslant i \leqslant n$. Given $(x_1 \ldots x_n)^T, (\hat{x}_1 \ldots \hat{x}_n)^T$, the column vectors X and $\widehat{X}$, coordinates of a vector belonging to *E* in these two bases verify: $X = H\widehat{X}$ and $\widehat{X} = H^{-1}X$. Matrices *A* and $\widehat{A}$ associated with an endomorphism of *E* in these bases verify:

$$A = H\widehat{A}H^{-1} \quad \text{and} \quad \widehat{A} = H^{-1}AH. \tag{A.84}$$

Here is a diagram of the diagonalization:

$\widehat{A}$: Matrix associated with $\{f\}$
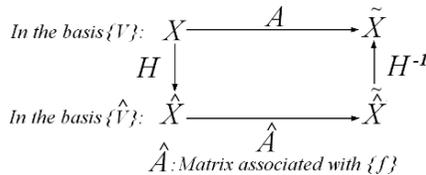
An *endomorphism f* of *E* is *diagonalizable* if there exists a basis $(\widehat{V}_1, \ldots, \widehat{V}_n)$ of *E* in which the matrix associated with *f* is the *diagonal matrix* $\widehat{A}$, i.e. such that.: $f(\widehat{V}_k) = \lambda_k \widehat{V}_k$, $1 \leqslant k \leqslant n$; $(\widehat{V}_1, \ldots, \widehat{V}_n)$ is a basis of *eigenvectors* of *f* and $\lambda_1, \ldots, \lambda_n$ are *eigenvalues* of *f*. The *subspace* $F(\lambda) = Ker(f - \lambda I) = \{V \in E; f(V) = \lambda V\}$ is of dimension higher or equal to 1, if and only if λ is an eigenvalue of *f*; it is the Eigen-subspace associated with the eigenvalue λ. The same notions are applied to a square matrix *A* of order *n* with elements in K associated with an endomorphism, in particular *associated canonically with the endomorphism A* (in the *canonical basis* of K^n). Vectorial equality: $(A - \lambda 1)(V) = 0$ is equivalent to the *matrix equality:*

$$(A - \lambda I)(V) = 0. \tag{A.85}$$

The *characteristic polynomial* of *f* and *A* (*A* matrix associated with *f*) is:

$$P_A(\lambda) = \det(A - \lambda I) \tag{A.86}$$

is a *polynomial of degree n* on λ which depends only on f. The *eigenvalues* of f and on A are the *roots* of P_A. If A has real (or complex) elements, the n roots (distinct or not) of P_A are the *eigenvalues* of A and f; *their sum is equal to the trace* of A (which is the sum of n elements of the *main diagonal*, also called *principal diagonal*) and their product is equal to $det(A)$. These quantities depend only on f and are known as *trace* of f and *determinant* of f. In the particular case where n roots of P_A are all distinct, A is *diagonalizable*, and by choosing for each eigenvalue λ_k a non-zero eigenvector which is solution of $(A - \lambda_k I) V = O$, we obtain a *basis of eigenvectors*. The *roots* of P_A are thus the *eigenvalues* of A and f. $Trace(A) = \sum_k \lambda_k$ or $Trace(f) = \sum_k \lambda_k$, $\det(A) = \prod_K \lambda_k$ or $\det(f) = \prod_K \lambda_k$ with $(A - \lambda_k I) V = O$ we obtain a basis of eigenvectors. $P_A(\lambda) = \det(A - \lambda I)$ and $f(\widehat{V}_k) = \lambda_k \widehat{V}_k \Leftrightarrow f\widehat{V}_k - \lambda_k \widehat{V}_k = 0$. Thus λ_k *are the eigenvalues* of f and $(\widehat{V}_1, \ldots, \widehat{V}_n)$ is a *basis* of *eigenvectors* of f.

A.13.1.2 Some Statistical Definitions

Definition A.93 (Process). A process is a sequence $(X_n, n \in I)$ of random vectors defined on a same space (Ω, a, P) with values in $\mathbb{R}^p$.[33] The set of indexes I is generally $\mathbb{N}, \mathbb{N}^*$ or $\mathbb{Z}$. Thus "for each state of the nature" ω corresponds an element $(X_n(\omega), n \in I)$ of $[\mathbb{R}^{pI}]$; such an element is called *trajectory* of the process. A process admits a law. One can show that this one is characterized by the knowledge of the laws of all finite subfamily $(X_{n1}, \ldots, X_{nk})$ extracted from the sequence $(X_n, n \in I)$. From this characterization it is possible to define particular classes of process.

Definition A.94 (Gaussian Vector). A random vector X with value in $\mathbb{R}^p$ is known as Gaussian, if any linear combination of its co-ordinates $\alpha'X = \sum_{i=1}^p \alpha_i X_i$, follows a normal law on $\mathbb{R}$.

Law of large numbers: In one of its versions (Gourieroux and Monfort 1989) the law of large numbers ensures that under the assumption of independence and equal-distribution the empirical averages converge towards the corresponding theoretical averages.

Definition A.95 (Strong law of large numbers). Let $(X_n, n \in \mathbb{N}^*)$ be a sequence of random vectors (with value in $\mathbb{R}^p$) which are integrable, independent and with a same law, then: $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \overset{as}{\rightarrow} E(X_j)$, when $n \rightarrow \infty$. Such a convergence is valid also when each variable X_j is replaced by $g(X_j)$ where g is a given function such that $E \left\| g(X_j) \right\| < +\infty$.

Definition A.96 (Central limit theorem). If $X_n, n \in \mathbb{N}^*$ are random independent vectors of the same law of average m, with variance-covariance matrix Σ, the law of large numbers shows that $\bar{X}_n - m$ converges strongly towards zero. (We can wonder

---

[33] (Ω, a, P), where Ω is the *set of the possible results, a* is a σ-algebra, i.e. *a set of parts* of Ω (called *events*), containing the parts ϕ and Ω, and P is a function of set, i.e. a map from a to $\mathbb{R}^+$, which satisfies $P(\Omega) = 1$.

what is the speed of convergence and seek a simple equivalent of the difference between empirical and theoretical average.) In the case of an independent sample, the sequence $\sqrt{n}(\bar{X}_n - m)$ "converges in law" towards a vector according to a normal law $N[0, \Sigma] : \sqrt{n}(\bar{X}_n - m) \rightarrow N[0, \Sigma]$.

Definition A.97 (Characteristic function). The law of X is characterized by the values taken by:[34] $E\left[exp(itX)\right] = E\left[cos(tX)\right] + iE\left[sin(tX)\right]$, where E is the expected value and i indicates the pure imaginary number and $t \in \mathbb{R}$. A *characteristic function* of (the law of) X is the map φ_x defines as follows: $\varphi_x(t) = E\left[exp(itX)\right]$.

Definition A.98 (White noise). A white noise is an infinite stationary random signal whose function of autocorrelation is proportional to a Dirac. i.e. a constant complex spectrum on all the frequency band.

This means that the values of a time series, at two different moments, are non-correlated, even if the moments are very close. (From this definition, it is possible to say that two centered Gaussian white-noises $w(t)$ and $v(t)$ *are entirely defined by means of the spectral densities* $\hat{w}(t)$ and $\hat{v}(t) : E\left[w(t)w(t+\tau)^T\right] = \hat{w}(t)\delta(\tau)$ and $E\left[v(t)v(t+\tau)^T\right] = \hat{v}(t)\delta(\tau)$. And if it is postulated (ref: Calman theory) that the Gaussian white noises $w(t)$ and $v(t)$ are stationary and independent, then it is possible to write $E\left[w(t)w(t+\tau)^T\right] = \hat{w}\delta(\tau)$, $E\left[v(t)v(t+\tau)^T\right] = \hat{v}\delta(\tau)$, $E[v(t)w(t+\tau)^T] = 0$.)

Definition A.99 (Markov chain). A sequence (x_n) of random variables is a *Markov chain,* if the conditional law of x_t knowing $x_{t-1}, x_{t-2}, \ldots$ is the same one as that of x_t knowing x_{t-1}.

Definition A.100 (Homogeneous Markov chain). The chain is *homogeneous* if the law of $(x_{t_1}, \ldots, x_{t_k})$ knowing x_{t_0} is the same one as that of $(x_{t_1-t_0}, x_{t_2-t_0}, \ldots, x_{t_k-t_0})$ knowing x_0 for any k and any $(k+1)$-tuple $t_0 \leq t_1 \leq \cdots \leq t_k$.

An interesting technique in the area of nonlinear state space is the Monte Carlo Markov Chain (MCMC), which is an important numerical technique for Bayesian statistics (Gelfand and Smith 1990). *Typically rooted in conditional probability computation, it is a very powerful numerical technique, which has completely transformed the Bayesian statistics in practice.*

Definition A.101 (Monte Carlo Markov Chain). One calls the MCMC algorithm any method producing an ergodic Markov chain $(x^{(t)})$ of stationary law, the distribution of interest. (The rediscovery of MCMC methods by the statisticians in the year 1990 allowed a very significant development of the inference by simulation.)

---

[34] The integrals.

A.13.1.3 Fourier Series and Fourier Transform

Fourier Series

A Fourier series can be defined in the following way: Let ε_n be a map such that:

$$\varepsilon_n : t \longrightarrow e^{i\omega t}, \quad (\mathbb{R} \to \mathbb{C}). \tag{A.87}$$

Let c_n be Fourier coefficients of order $n \in \mathbb{Z}$, such that:

$$c_n = \frac{1}{2\pi} \int_{\alpha}^{\alpha+2\pi} e^{i\omega t} f(t) dt, \tag{A.88}$$

then a *Fourier series* can be written as follows:

$$\sum c_n \varepsilon_n. \tag{A.89}$$

Any periodic function $f(x)$ can be decomposed into several intervals in the period-icity domain $-\pi \leq x \leq \pi$. If $f(x)$ can be replaced in this interval by a continuous curve, $f(x)$ can be represented by a convergent series of the form:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]. \tag{A.90}$$

With: $x = \omega t$.

$$f_p = \sum_{k=-\infty}^{+\infty} a_k e^{2\pi i k t / p}. \tag{A.91}$$

The coefficients for $k = 0, 1, 2, \ldots$ are calculated by means of formulas:

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx, \tag{A.92}$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx. \tag{A.93}$$

Fourier Transform

Continuous Writing

In the Fourier transform $F\{s(t)\}$, through the Fourier integral, the function $s(t)$ is replaced by a continuous spectral function $S(\omega)$ (*i.e. Spectral density*) where the frequency ω corresponds to the density of the spectrum. Inverse transform $F^{-1}\{S(\omega)\}$ is the function $s(t)$.

$$F\{s(t)\} = S(\omega) = \int_{-\infty}^{+\infty} s(t) e^{-i\omega t} dt; \quad \text{with} \quad (i = \sqrt{-1}). \tag{A.94}$$

$$F^{-1}\{S(\omega)\} = s(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega) e^{i\omega t} d\omega; \quad \text{with} \quad (i = \sqrt{-1}). \tag{A.95}$$

$$\text{Spectral energy:} \int_{-\infty}^{+\infty} |s(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |S(\omega)|^2 \, d\omega. \qquad (A.96)$$

Furthermore, we notice that $F\{s(t-\tau)\} = S(\omega) \cdot e^{-i\omega\tau}$ with $(i = \sqrt{-1})$. Another type of notations:

$$F(\omega) = \int_{-\infty}^{+\infty} e^{-2\pi i \omega t} f(t) dt, \qquad (A.97)$$

$$f(t) = \int_{-\infty}^{+\infty} e^{2\pi i \omega t} F(\omega) d\omega. \qquad (A.98)$$

Discrete Writing

If we write $f_p = f(t)$, $\frac{-P}{2} < t < \frac{P}{2}$.

$$f_p = \frac{A_0}{2} + \sum_{k=1}^{+\infty} \left(A_n \cos \frac{2\pi k t}{p} + B_n \sin \frac{2\pi k t}{p} \right). \qquad (A.99)$$

Taking into account the *Moivre formula* we write:

$$f_p = \sum_{k=-\infty}^{+\infty} a_k e^{2\pi i k t / p}, \qquad (A.100)$$

$$a_k = \sum_{k=-\infty}^{+\infty} f_p(t) e^{-2\pi i k t / p}. \qquad (A.101)$$

Three Continuous Writings of the Fourier transform

There are three different writings for the Fourier transform according to the location of the "2π". Each one of these writings comprises two alternatives according to the sign. Here are the various writings:

(1) ω is measured in *Hertz* and the 2π are in the exponent:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{2\pi i \omega t} dt \quad \text{and} \quad f(t) = \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{-2\pi i \omega t} d\omega,$$

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-2\pi i \omega t} dt \quad \text{and} \quad f(t) = \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{2\pi i \omega t} d\omega.$$

(2) ω are measured in *radian* (per second) and the 2π are in the inverse transform:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{i\omega t} dt \quad \text{and} \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{-\pi i \omega t} d\omega,$$

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \quad \text{and} \quad f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{\pi i \omega t} d\omega.$$

(3) ω are measured in *radian* (per second) and the 2π are distributed between the direct transform and the inverse transform:

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t)e^{i\omega t}\,dt \quad \text{and} \quad f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{f}(\omega)e^{-\pi i\omega t}\,d\omega,$$

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}\,dt \quad \text{and} \quad f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{f}(\omega)e^{\pi i\omega t}\,d\omega.$$

Commonly, the following writings are used:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} e^{-2\pi i\omega t} f(t)\,dt, \tag{A.102}$$

$$f(t) = \int_{-\infty}^{+\infty} e^{2\pi i\omega t} \hat{f}(\omega)\,d\omega, \tag{A.103}$$

$$f_p = \frac{A_0}{2} + \sum_{k=1}^{+\infty} \left(A_n \cos\frac{2\pi kt}{p} + B_n \sin\frac{2\pi kt}{p} \right), \tag{A.104}$$

$$f_p = \sum_{k=-\infty}^{+\infty} a_k e^{2\pi ikt/p}. \tag{A.105}$$

Bibliography

Abarbanel HDI (1996) Analysis of observed chaotic data. Springer, Berlin.

Abraham R, Marsden JE (1978) Foundations of classical mechanics. Addison-Wesley, Massachusetts.

Abraham R, Stewart HB (1986) A chaotic blue sky catastrophe in forced relaxation oscillations. Physica 21D:394–400.

Abraham RH, Shaw CD (1988) Dynamics, the geometry of behavior. Part I and II. Aerial, Santa Cruz.

Abraham-Frois G, Berrebi E (1995) Instabilité, cycles, Chaos. Economica.

Abraham-Frois G, Larbre F (1997) Prices, profits and rhythms of accumulation. Cambridge University Press, Cambridge.

Abraham-Frois G (1998) Nonlinear dynamics and endogenous cycles. Springer, Berlin.

Abry P, Veitch D (1998a) Wavelet analysis of long-range dependent traffic. IEEE Trans Info Theory 44(1):2–15.

Abry P, Veitch D (1998b) A wavelet-based joint estimator of the parameters of long-range dependence. IEEE Trans Info Theory 45(1):878–897.

Abry P, Flandrin P, Taqqu M, Veitch D (2000) Wavelets for the analysis, estimation and synthesis of scaling data, In: Park K, Willinger W (eds.), Self-similar network Traffic and performance evaluation. Wiley, New York.

Abry P, Flandrin P, Taqqu M, Veitch D (2002a) Self-similarity and long range dependence through the wavelets lens. In: Doukhan P, Oppenheim G, Taqqu M (eds.), Theory and applications of long-range dependence. Birhauser, Basel.

Abry P, Veitch D, Taqqu M (2002b) Meaningful MRA initialization for discrete time series. Signal Processing 80(9):1971–1983.

Aeyels D (1981) Generic observability of differentiable systems. SIAM J Control Optim 19:595–603.

Aghion P, Howitt P (1999) Endogenous growth theory. MIT, Cambridge, MA.

Amin R, Smith R, Fife-Schaw C, Wood P (1997) Evaluation of the use of signal processing, image processing and data compression for the analysis of statistical data. Final Report. Commission of the European Communities.

Andronov AA, Leontovich EA (1937) Some cases of dependence of limit cycles on a parameter. Uchenye zapiski Gorkovskogo Universiteta 6:3–24.

Andronov, Leontovich, Gordon, Maier (1971) The theory of bifurcations of dynamical systems on a plane. Israel program of scientific translation, Jerusalem.

Arneado A, Bacry E, Muzy JF (1999) The thermodynamics of fractals revisited with wavelets. Wavelets in Physics. Cambridge University Press, Cambridge, p. 352.

Arnold V (1989) Mathematical methods of classical mechanics. Springer, Berlin.

Assaf D IV, Gadbois S (1992) Definition of chaos. Am Math Monthly 99:865.

Bacry E, Delour J, Muzy JF (2001) Modelling financial time series using multifractal random walks. Physica A, 299(1–2):84–92.

Banks J, Brooks J, Cairns G, Davis G, Stacey P (1992) Amer. Math. Monthly 99(4):332–334.

Benhabib J, Nishimura K (1979) The Hopf bifurcation and the existence of closed orbits in multisectorors of optimal economic growth. Journal of Economic Theory 21.

Bergé P, Pomeau Y, Vidal Ch (1987) Order within chaos: towards a deterministic approach to turbulence. Wiley-Interscience, New York.

Bergé P, Pomeau Y, Vidal Ch (1988) L'ordre dans le chaos. Collection Enseignement des Sciences. Herman Editeur, Paris.

Bishop C, Nabney I (1995) Neural networks for pattern recognition. Oxford University Press, Oxford.

Blanchard O (1989) Lectures on Macroeconomics. MIT Press.

Blanchard O (2005) Macroeconomics. Prentice Hall, Upper Saddle River, NJ.

Block L, Coppel W (1992) Dynamics in one dimension. Lecture Notes in Mathematics, vol. 1513. Springer, New York.

Bourbonnais R, Terraza M (1998) Analyse des séries temporelles en économie. PUF, Paris, pp. 122–124.

Broomhead DS, King GP (1986). On the qualitative analysis of experimental dynamical systems. In: Sarkas S, (ed.) Nonlinear phenomena and chaos. Adam Hilger, Bristol, pp. 113–144.

Broomhead DS, Jones R, King GP (1986) Extracting qualitative dynamics from experimental date. Physica D 20:217–236.

Broomhead DS, Jones R, King GP (1988) Comment on "Singular-value decomposition and embedding dimension". Phys Rev A 37:5004.

Bucheit J, Chen S, Donoho D, Johnstone L (n.d.) Wavelab.701. Stanford University. J Scargle NASA-Ames.

Burmeister E (1980) Capital theory and dynamics. Cambridge University Press, Cambridge.

Campbell J, Lo A, MacKinlay C (1997) The econometrics of financial markets. Princeton University Press, Princeton, NJ.

Carmona R, Hwang WL, Torresani B (1998) Practical time-frequency analysis. Gabor and wavelet transforms with an implementation in S. Wavelet ananlysis and its applications, vol. 9. Academic, San Diego.

Chen S, Donoho D (1995) Atomic decomposition by pursuit. In: SPIE International Conference on Wavelets, San Diego, 1995.

Clegg R (2003) Analysing long-range dependence in teletraffic. Networks and Nonlinear Dynamics Group. Department of Mathematics, University of York UK.

Cohen L (1989) Time-frequency distributions a review. Proceedings of the IEEE 77:941–981.

Coifman RR, Wickerhauser MV (1992) Entropy-based algoritms for best basis selection. IEEE Transactions on information theory 38(2):713–718.

Croux C, Forni M, Reichlin L (2001) A measure of comovement for economic indicators: theory and empirics. Review of Economics and Statistics 83:232–241.

Dang-Vu H, Delcarte C (2000) Bifurcation et chaos. Une introduction à la dynamique contemporaine avec des programmes en Pascal, Fortran et Mathematica. Ellipse.

Devaney R (1989) An introduction to chaotic dynamical systems. Addison-Wesley, Reading, MA.

Eckmann JP, Ruelle D (1985) Ergodic theory of chaos and strange attractors. Rev Mod Phys 57(3):617–656.

Farge M (1992) The continuous wavelet transform of two-dimensional fluid flows. In: Ruskai MB (ed.) Wavelets and their applications. Jones and Bartlett, Boston, MA, p. 275.

Friedman J, Stuetzel W (1981) Projection pursuit regression. Journal of the American Statistical Association 76:817–823.

Gavrilov N, Shilnikov A (2000) Example of a Blue Sky Catastrophe. Amer. Math. Soc. Transl. Vol. 200, AMS, 99–105.

Glass L, Mackey MC (1988) From clocks to chaos: the rhythms of life. Princeton University Press, Princeton, NJ.

Gleick J (1987) Chaos: making a new science. Viking, New York.

Goergen A (1994a) Chaos Macro-économique. Revue d'Economie Politique 104.

Goergen A (1994b) Générations imbriquées, optimisation et chaos. Revue d'Economie Politique 104.

Goergen A (1999) Reswitching: a null measure phenomenon. The Current State of Economic Science 1:195–203.

Goergen A (2001) Analyse rétrospective du modèle de Ramsey. Contribution to the symposium of Pise, Oct. 2001.

Goergen A (2006) Dynamique Economique. Solutions de problèmes avec Maple et Matlab. Economica.

Gondolfo G (1983) Economic dynamics: methods and models, 2nd edition. North Holland, Amsterdam.

Goodwin R (1967) A growth cycle. In: C.H. Feinstein (ed), Socialism, Capitalism and Economic Growth. Cambridge University Press, Cambridge.

Gourieroux C, Monfort A (1989) Statistique et modèles économétriques. Economica.

Gourieroux C, Monfort A (1995) Séries temporelles et modèles dynamiques. Economica.

Granger CWJ (1957) A statistical model for sunspot activity. Astrophys J 126: 152–158.

Granger CWJ (1963) Economic processes involving feedback. Information Control 6:28–48.

Granger CWJ (1964) Spectral Analysis of Economic Time Series, in association with M. Hatanaka, Princeton University Press, Princeton, NJ. October 1964. (French translation: Analyze spectrale des series temporelles en economie, Dunod, Paris, 1969.)

Granger CWJ (1966) The typical spectral shape of an economic variable. Econometrica 34:150–161.

Granger CWJ, Rees H (1968) Spectral analysis of the term structure of interest rates. Rev Econ Stud 35:67–76.

Granger CWJ (1969) Investigation causal relations by econometrics models and cross spectral methods. Econometrica 37.

Granger CWJ (1977) Forecasting Economic Time Series, with Paul Newbold, Academic Press, March 1977. Second edition, October, 1986.

Granger CWJ (1978a) Nonlinear time series modelling, with A Andersen. In: DF Findley (ed.) Applied time series analysis. Academic, New York.

Granger CWJ (1978b) Introduction to bilinear time series models, with A Andersen. Vandenhoeck & Ruprect, Gottingen.

Granger CWJ (1980) An introduction to long-memory time series, with R Joyeux. J Time Series Anal 1:15–30.

Granger CWJ (1983) Forecasting white noise. In: A Zellner (ed.) Applied Time Series Analysis of Economic Data, Proceedings of the Conference on Applied Time Series Analysis of Economic Data (October 1981). US Government Printing Office.

Granger CWJ (1984) Applications of spectral analysis in econometrics, with R. Engle. Time series and frequency domain, Handbook of Statistics, vol. 3. North-Holland, Amsterdam.

Granger CWJ (1991a) Nonlinear transformations of integrated time series, with J Hallman. J Time Series Anal 12:207–224.

Granger CWJ (1991b) Developments in the nonlinear analysis of economic series. Scandinavian J Econom 93:263–276.

Granger CWJ (1991c) Long memory processes with attractors, with J Hallman. Oxford Bull Econ Stat 53:11–26.

Granger CWJ (1991d) Long run economic relationships: readings in cointegration. Edited with R. Engle. Oxford University Press, Oxford.

Granger CWJ (1993a) The power of the neural network linearity test, with T Teräsvirta and Chien-Fu Lin. J Time Series Anal 14:209–220.

Granger CWJ (1993b) Modelling nonlinear dynamic relationships, with T Teräsvirta. Oxford University Press, Oxford.

Granger CWJ (1994) Is chaotic economic theory relevant for economics? A review essay. J Int Comp Econ 3:139–145.

Granger CWJ (1995) Nonlinear relationships between extended memory series. Econometrica 63:265–279.

Granger CWJ (1996) Modeling volatility persistence of speculative returns, with Z Ding. J Econom 73:185–215.

Granger CWJ, Morgenstern O (1963) Spectral analysis of New York stock market prices. Kyklos 16:1–27. Reprinted in PH Cootner (ed.) (1964) Random Character of Stock Market Prices, MIT, Cambridge, MA.

Granger CWJ, Morgenstern O (1970) The predictability of stock market prices. Heath Lexington Books, Lexington, MA.

Grossman SJ, Stiglitz JE (1980) On the impossibility of informationally efficient markets. The American Economic Review 70(3):393–408.

Guckenheimer J, Holmes P (1983) Nonlinear oscillations, dynamical systems and bifurcations of vector fields. Springer, New York.

Guegan D (2003) Les chaos en finance. Economica.

Hale J, Koçak H (1991) Dynamics and bifurcations. Springer, Berlin.

Haller G, Iacono R (2003) Stretching, alignment, and shear in slowly varying velocity fields. Phys Rev E 68:056304.

Hamilton JD (1994) Times Series Analysis. Princeton University Press, Princeton.

Hramov A, Koronovskii A, Ponomarenko V, Prokhorov M (2006) Detecting synchronization of self-sustained oscillators by external driving with varying frequency. Phys Rev E 73:026208.

Hubbard BB (1998) The world according to wavelets. A.K. Peters, Wellesley, MA.

Hubbard BB (2000) Ondes et ondelettes. La Saga d'un outil mathématique. Pour la Science.

Hubbard JH, West BH (1995a) Differential equations: a dynamical systems approach. Ordinary differential equations. Text in Applied Mathematics, vol. 5. Springer, New York.

Hubbard JH, West BH (1995b) Differential equations: a dynamical systems approach. Higher-Dimensional Systems. Text in Applied Mathematics, vol. 18. Springer, New York.

Hudgins L, Friehe CA, Mayer ME (1993) Wavelet transforms and atmospheric turbulence. Phys Rev Lett 71(20):3279–3282.

Hidalgo C, Brañas B, Uckan T, Harris JH, Isler R, Ritz ChP, Wootton A (1994) On the role of neutral particles on edge turbulence and electric fields in the Advanced Toroidal Facility. Phys Plasmas 1:3–5.

Johnston J, Dinardo J (1999) Méthodes économétriques. Economica.

Jones LK (1987) On a conjoncture of Huber concerning the convergence of projection pursuit regression. Ann Stat 15(2):880–882.

Judd Kenneth L (1998) Numerical method in economics. MIT, Cambridge, MA.

Invernizzi S, Medio A (1991) On lags and chaos in economics dynamic models. J Math Econ 20:521–550.

Iooss G, Joseph DD (1980) Elementary stability and bifurcation theory. Springer, New York.

Kantz H, Schreiber T (1997) Nonlinear Time series analysis. Cambridge Nonlinear Science Series. Cambridge University Press, Cambridge.

Kim YC, Powers EJ (1978) Digital bispectral analysis of self-excited fluctuation spectra. Phys Fluids 21:1452–1453.

Körner T (1988) Fourier Analysis. Cambridge University Press, Cambridge.

Kravtsov YuA, Kadtke JB (1996) Predictability of complex dynamical systems. Springer, Berlin.

Kurz M (1968) The general instabilty of class of competitive growth processes. Review of Economic Studies XXV:102.

Laffont JJ (1991) Economie de l'incertain et de l'information. Cours de théorie microéconomique. Economica.

Lardic S, Mignon V (2002) Econométrie des séries temporelles macroéconomiques et financières (Econometrics of the macroeconomic and financial time series). Economica.

Li TY, Yorke JA (1975) Period three implies chaos. Am Math Monthly 82:481–485.

Lo A (1991) Long-term memory in stock market prices. Econometrica 59(5):1279–1313.

Lo A, MacKinlay C (1988) Stock market prices do not follow random walks: evidence from a simple specification test. Rev Finan Stud 1(1):41–66.

Lo A, MacKinlay AC (1989) The size and power of the variance ratio test in finite samples. J Econom 40:203–238.

Lo A, MacKinlay AC (1990) An econometric analysis of non-synchronous trading. J Econom 45:181–211.

Lo A, MacKinlay AC (1999) A non-random walk down Wall Street. Princeton University Press, Princeton, NJ.

Lorenz EN (1963) Deterministic non-periodic flow. J Atmos Sci 20:130–141.

Lucas R (1972) Expectations and the neutrality of money. Jounal of Economic Theory 4(2):103–124.

Lucas R (1976) Econometric policy evaluation: a critique. Journal of Monetary Economics 1(2):19–46.

Lucas R (1981) Studies in business: Cycle theory. MIT, Cambridge, MA.

Lucas R (1988) On the mechanics of economic development. Journal of Monetary Economics 22:3–42.

Lyon JF, Carreras BA, Chipley KK, Cole MJ, Harris JH, Jernigan TC, Johnson RL, Lynch VE, Nelson BE, Rome JA, Sheffield J, Thompson PB (1986) The advanced toroidal facility. Fusion Technol 10:179.

Mallat S (1998) A wavelet tour of signal processing. Academic press, New York.

Mallat S, Zhang Z (1993) Matching pursuit with time-fréquency dictionaries. IEEE Transaction on signal processing 41(12):3395–3415.

Mandelbrot B (1975) Les objets Fractals. Flammarion, Paris.

Mandelbrot B (1997) Fractals and scaling in finance. Discontinuity, concentration, risk. Springer, Berlin.

Mankiw N Gregory, Romer D, Shapiro MD (1991) Stock market forecastability and volatility: A statistical appraisal. The Review of Economic Studies 58:455–477.

Mankiw G, Romer D, Weil D (1992) A contribution to the empirics of economic growth. Quaterly Journal of Economics 107:407–438.

Marotto F (1978) Snap-back repellers imply chaos in R. J Math Anal Appl 63:199–223.

May R.M (1983) Nonlinear problems in ecology and resource management. In: G Ioos, RHG Helleman, R Stora (eds.) Chaotic behaviour of deterministic systems. North Holland, Amsterdam, pp. 514–563.

Medio A (1991) Discrete and continuous-time models of chaotic dynamics in economics. Structural Change and Economic Dynamic 2(1).

Medio A (1992) Chaotic dymamics. Theory and applications to economics. Cambridge University press, Cambridge.

Meinecke F, Ziehe A, Kurths J, Müeller K (2005) Measuring phase synchronization of superimposed signals. Phys Rev Lett 94:84102.

Mendel JM (1991) Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. Proc IEEE 79(3):278–305.

Meneveau C (1991) Analysis of turbulence in the orthonormal wavelet representation. J Fluid Mech 232:469.

Meyer Y (ed.) (1989) Wavelets and applications. Research notes in applied mathématics, RMA 20. Series editors: PG Ciarlet and JL Lions. Springer, Berlin.

Meyer Y (1992) Les ondelettes, algorithmes et applications. Armand Colin, Paris.

Milnor J (1968) Singular points of complex hypersurfaces. Princeton University Press, Princeton.

Milnor J (1999) Dynamics in one complex variable. Vieweg, Wiesbaden, Germany.

Minsky M, Papert S (1988) Perceptrons. MIT, Cambridge.

Noakes L (1991) The Takens embedding theorem. Int J Bifur Chaos 1(4):867–872.

Oseledec V (1968) A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. Trans Moskov Math 19:197–221.

Ott E (1993) Chaos in dynamical systems. Cambridge University Press, Cambridge, pp. 266–291.

Ott E, Gregori EC, Yorke JA (1990) Controlling chaos. Phys Rev Lett 64(11):1196–1199.

Ott E, Sauer T, Yorke JA (1994) Coping with chaos : Analysis of chaotic data and the exploitation of chaotic systems. Wiley Interscience, New York.

Pati Y, Rezaifar R, Krishnaprasad P (1993) Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. Asimolar conf. on Signal, System and Comp., November 1993.

Perli R, Sakellaris P (1998) Human capital formation and business cycle persistence. Journal of Monetary Economics 42:67–92.

Pikovsky A, Rosenblum M, Kurths J (2001) Synchronization: a universal concept in nonlinear sciences. Cambridge University Press, Cambridge.

Pilgram B, et al. (1997) Dynamic detection of rhythmic oscillations in heart-rate tracings: a state-space approach based on 4th-order cumulants. Biol Cybern 76(4):299–309.

Qian S, Chen D (1988) Signal representation via adaptative normalized gaussian functions. Signal Processing 36(1):1–11.

Ramsey JB, Zhang Z (1996) The apllication of wave form dictionnaries to stock market index data. YuA Kravtsov, JB Kadtke (eds.) Predictibility of complex dynamical systems. Springer, Berlin.

Romer P (1986) Inceasing returns and long-run growth. Journal of Political Economy 94:1002–1032.

Romer D (1996) Advanced macroeconomy. McGraw-Hill, New York.

Romer D (2000) Macroeconomie approfondie. Collect. sciences economiques. McGraw-Hill/Ediscience, New York/Paris.

Rosenblatt F (1962) Principles of Neurodynamics. Spartan Books, New York.

Roughan M, Veitch D, Abry P (2000) Real-time estimation of the parameters of long-range dependence. IEEE/ACM Transaction on networking 8(4):467–478.

Ruelle D (1979) Ergodic theory of differentiable dynamic systems. IHES Publ Math 50:27–58.

Ruelle D (1980) Measures describing a turbulence flow. Ann N Y Acad Sci 357:1–9.

Ruelle D (1981) Small random perturbations of dynamical systems and the definition of attractors. Commun Math Phys 82:137.

Ruelle (Ed.) (1995) Turbulence, strange attractors, and chaos. Series A, vol. 16. World Scientific series on nonlinear science. Series Editor: LO Chua.

Ruelle D, Takens F (1971) On the nature of turbulence. Commun Math Phys 20: 167–192.

Sapoval B (2001) Universalités et fractales. Champs. Flammarion.

Sauer T (1994) Reconstruction of dynamical systems from interspike intervals. Phys Rev Lett 72:3811–3814.

Sauer T, Yorke JA, Casdagli M (1991) Embedology. J Stat Phys 65:579–616.

Schäfer C, Rosenblum MG, Abel H-H, Kurths J (1999) Synchronization in human cardiorespiratory system. Phys Rev E 60:857.

Schuster H G (1989) Deterministic chaos: An introduction, 2nd edn. VCH, Veinheim.

Sharkovsky A (1964) Ukrain Math Zh 16(1); and 17(3):104–111, 1965.

Shilnikov LP, Turaev DV (1995a) On a blue sky catastrophe. Dokl Akad Nauk SSSR 342(5):596–599.

Shilnikov LP, Turaev DV (1995b) Simple bifurcations leading to hyperbolic attractors. Comput Math Appl 34(2–4):174–180.

Shilnikov A, Shilnikov L, Turaev D (2003) On some mathematical topics in classic synchronization. WIAS, Berlin.

Sibbertsen (2002) Long memory in volatilities of German stock returns. Dordmund University, Germany.

Smale S (1967) Differentiable Dynamical Systems. Bull Am Math Soc 73:747–817.

Smale S (1980) The mathematics of time. Essays on dynamical systems, Economic processes, and related topics. Springer, Berlin.

Solow R (1957) Technical change and the aggregate production function. Review of Economics and Statistics 39:312–320.

Sparrow C (1980) Bifurcation and chaotic behavior in simple feedback systems. J Theor Biol 83:93–105.

Takens F (1981) Detecting strange attractors in turbulence. In: DA Rand, LS Young (eds.) Dynamical systems and turbulence. Lecture Notes in Mathematics, vol. 898. Springer, Berlin, pp. 366–381.

Tin-Yau Tam (2001) Generalized Schur-concave functions and eaton triples. Department of mathematics, Auburn University, USA, April 11, 2001.

Tobin J (1965) Money and capacity growth. Econometrica, October 1965.

Torresani B (1991) Wavelets associated with representations of the affine. Weyl Heysenberg group. J Math Phys 32:1273–1279.

Torresani B (1995) Analyse continue par ondelettes. EDP Sciences. CNRS Editions, Paris.

Urbach RMA (2000) Footprints of chaos in the markets. Financial Times/Prentice Hall, London/New York.

Van den Berg JC (ed.) (1999) Wavelets in physics. Cambridge university press, Cambridge.

van Milligen BPh (1999) Long-range correlation analysis of plasma turbulence. In: 26th EPS conf. on Contr. Fusion and Plasma Physics, Maastricht, 14–18 June 1999. ECA 23J:49–52.

van Milligen BPh, Hidalgo C, Sanchez E (1995a) Nonlinear phenomena and intermittency in plasma turbulence. Phys Rev Lett 74(3):395–398.

van Milligen BPh, Sanchez E, Estrada, Hidalgo C, Brana B (1995b) Wavelet bicoherence: A new turbulence analysis tool. Phys Plasma 2(8):3017.

Varian HR (1979) Catastrophe theory and the business cycle. Economic Inquiry 17.

Veitch D, Taqqu M, Abry P (2003) Meaningful MRA initialization for discrete time series. Signal Processing 80(9):1971–1973.

Vellekoop M, Berglund R (1994) On intervals, transitivity = chaos. Am Math Monthly 101(4):353–355.

Verhulst F (1996) Nonlinear differential equations and dynamical systems. Universitext. Springer, Berlin.

von Bremen H, Udwadia FE, Proskurowski W (1997) An efficient QR based method for the computation of Lyapunov Exponents. Physica D 101(2):1–16.

Watson MW (1993) Measures of fit for calibrated models. Journal of Political Economy 101:1011–1041.

Wen Y (1998) Can a real business cycle model pass the Watson test? Journal of Monetary Economics 42:185–203.

Whitney H (1936) Differentiable manifolds. Ann Math 37:645–680.

Wiggins S (1988) Global bifurcations and chaos-analytical methods. Springer, New York.

Wiggins S (2003) Introduction to applied nonlinear dynamical systems and chaos. Text in applied mathematics. Springer, Berlin.

Wolfram S (2002) A new kind of science. Wolfram media, Champaign, IL, p. 972.

Index