Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge
A Series of Modern Surveys in Mathematics 60

Michel Talagrand

Upper and Lower Bounds for Stochastic Processes

Modern Methods and Classical Problems

Springer

Ergebnisse der Mathematik und ihrer Grenzgebiete

Volume 60

3. Folge

A Series of Modern Surveys in Mathematics

For further volumes:
www.springer.com/series/728

Michel Talagrand

Upper and Lower Bounds for Stochastic Processes

Modern Methods and Classical Problems

Springer

Michel Talagrand
Institut de Mathématiques
Université Paris VI
Paris Cedex 05, France

ISSN 0071-1136 ISSN 2197-5655 (electronic)
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics
ISBN 978-3-642-54074-5 ISBN 978-3-642-54075-2 (eBook)
DOI 10.1007/978-3-642-54075-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014932986

Mathematics Subject Classification: 60G17, 60G15, 60G52, 60B11, 46B99

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated to the memory of Marc Yor

Preface

What is the maximum level a certain river is likely to reach over the next 25 years? What is the likely magnitude of the strongest earthquake to occur during the life of a planned nuclear plant, or the speed of the strongest wind a suspension bridge will have to stand? The present book does not deal with such fundamental practical questions, but rather with some (arguably also fundamental) mathematics which have emerged from the consideration of these questions. All these situations can be modeled in the same manner. The value X_t of the quantity of interest (be it water level or speed of wind) at time t is a random variable. What can be said about the maximum value of X_t over a certain range of t? In particular, how can we guarantee that, with probability close to one, this maximum will not exceed a given threshold?

A collection of random variables $(X_t)_{t \in T}$, where t belongs to a certain index set T, is called a stochastic process, and the topic of this book is the study of the supremum of certain stochastic processes, and more precisely to find upper and lower bounds for these suprema. The key word of the book is

INEQUALITIES.

It is not required that T be a subset of the real line, and large parts of the book do not deal directly with the "classical theory of processes" which is mostly concerned with this situation. The book is by no means a complete treatment of the hugely important question of bounding stochastic processes, in particular because it does not really expand on the most basic and robust results which are the most important for the "classical theory of processes". Rather, its specific goal is to demonstrate the impact and the range of modern abstract methods, in particular through their treatment of several classical questions which are not accessible to "classical methods".

The most important idea about bounding stochastic processes is called "chaining", and was invented by A. Kolmogorov. This method is wonderfully efficient. With little effort it suffices to answer a number of natural questions. It is however not a panacea, and in a number of natural situations it fails to provide a complete understanding. This is best discussed in the case of Gaussian processes, that is processes for which the family $(X_t)_{t \in T}$ consists of jointly Gaussian random variables (r.v.s). These are arguably the most important of all. A Gaussian process defines in a canonical manner a distance d on its index set T by the formula

$$d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2} \ . \tag{0.1}$$

Probably the single most important conceptual progress about Gaussian processes was the gradual realization that the metric space (T,d) is the key object to understand them, irrelevant of the other possible structures of the index set. This led R. Dudley to develop in 1967 an abstract version of Kolmgorov's chaining argument adapted to this situation. This provides a very efficient bound for Gaussian processes. Unfortunately, there are natural situations where this bound is not tight. Roughly speaking, one might say that "there sometimes remains a parasitic logarithmic factor in the estimates".

The discovery around 1985 (by X. Fernique and the author) of a precise (and in a sense, *exact*) relationship between the "size" of a Gaussian process and the "size" of this metric space provided the missing understanding in the case of these processes. Attempts to extend this result to other processes spanned a body of work which forms the core of this book.

A significant part of this book is devoted to situations where one has to use some skills to "remove the last parasitic logarithm in the estimates." These situations occur with unexpected frequency in all kinds of problems. A particularly striking example is as follows. Consider n^2 independent uniform random points $(X_i)_{i \le n^2}$ which are uniformly distributed in the unit square $[0,1]^2$. We want to understand how far a typical sample is from being very uniformly spread on the unit square. To measure this we construct a one to one map π from $\{1,\ldots,n^2\}$ to the vertices $v_1,\ldots,v_{n^2}$ of a uniform $n \times n$ grid in the unit square. If we try to minimize the *average* distance between X_i and $v_{\pi(i)}$ we can do as well as about $\sqrt{\log n}/n$ but no better. If we try to minimize the *maximum* distance between X_i and $v_{\pi(i)}$, we can do as well as about $(\log n)^{3/4}/n$ but no better. The factor $1/n$ is just due to scaling. It is the fractional powers of $\log n$ that require work.

Even though the book is largely self-contained, it mostly deals with rather subtle questions such as the previous one. It also devotes considerable energy to the problem of finding *lower* bounds for certain processes, a topic considerably more difficult and less developed than the search for upper bounds. Therefore it should probably be considered as an advanced text, even though I hope that eventually the main ideas of at least Chapter 2 will become part of every probabilist's tool kit. In a sense this book is a second edition (or, rather, a continuation) of the monograph [1], or at least of the part of that work which was devoted to the present topic. I made no attempt to cover again all the relevant material of [1]. Familiarity with [1] is certainly not a prerequisite, and maybe not even helpful, because the way certain results are presented there is arguably obsolete. The present book incorporates (with much detail added) the material of a previous (and, in retrospect, far too timid) attempt [2] in the same direction, but its goal is much broader. I am really trying here to communicate as much as possible of my experience working in the area of boundedness of stochastic processes, and consequently I have in particular covered most of the subjects related to this area on which

I ever worked, and I have included all my pet results, whether or not they have yet generated activity. I have also included a number of recent results by others in the same general direction. I find that these results are deep and very beautiful. They are also sometimes rather difficult to access for the non-specialist (or even for the specialists themselves). I hope that explaining them here in a unified (and often simplified) presentation will serve a useful purpose. Bitter experience has taught me that I should not attempt to write about anything on which I have not meditated enough to make it part of my flesh and blood (and that even this is very risky). Consequently this book covers only topics and examples about which I have at least the illusion that I might write as well as anybody else, a severe limitation. I can only hope that it still covers the state-of-art knowledge about sufficiently many fundamental questions to be useful, and that it contains sufficiently many deep results to be of lasting interest.

A number of seemingly important questions remain open, and one of my main goals is to popularize these. Of course opinions differ as to what constitutes a really important problem, but I like those I explain in the present book. Several of them were raised a generation ago in [1], but have seen little progress since. One deals with the geometry of Hilbert space, a topic that can hardly be dismissed as being exotic. These problems might be challenging. At least, I made every effort to make some progress on them. The great news is that when this book was nearly complete, Witold Bednorz and Rafał Latała solved the Bernoulli Conjecture on which I worked for years in the early nineties (Theorem 5.1.5). In my opinion this is the most important result in abstract probability for at least a generation. I offered a prize of $ 5000 for the solution to this problem, and any reader understanding this amazing solution will agree that after all this was not such a generous award (specially since she did not have to sign this check). But solving the Bernoulli Conjecture is only the first step of a vast (and potentially very difficult) research program, which is the object of Chapter 12. I now offer a prize of $ 1000 for a positive solution of the possibly even more important problem raised at the end of Chapter 12 (see also [3]). The smaller amount reflects both the fact that I am getting wiser and my belief that a positive solution to this question would revolutionize our understanding of fundamentally important structures (so that anybody making this advance will not care about money anyway). I of course advise to claim this prize before I am too senile to understand the solution, for there can be no guarantee of payment afterwards.

I am very much indebted to Jian Ding and James Lee, who motivated me to start this project (by kindly but firmly pointing out that they found [2] far too difficult to read), and to Joseph Yukich, whose unflinching advice helped me to make this text more of a book and less of a gigantic research paper.

I must apologize for the countless inaccuracies and mistakes, small or big, that this book is bound to contain despite all the efforts made to remove them. I was very much helped in this endeavor by a number of colleagues, and in particular by Albert Hanen who read the entire book. Very special thanks are also due to Tim Austin, Witold Bednorz, Jian Ding, Rafał Latała, Nicholas Harvey, Joseph Lehec, Shahar Mendelson and Marc Yor (to whom I owe in particular the idea of adding Appendix A). Of course, all the remaining mistakes are my sole responsibility.

I am happy to acknowledge here the extraordinary help that I have received over the last 10 years from Albert Hanen. During that period I wrote over 2600 pages of book material. Albert Hanen has read every single of them in complete detail, often in several versions, attempting with infinite patience to check every single statement. He has corrected thousands of typos, hundreds of mistakes and helped me clarify countless obscurities. Without his endless labor of love, my efforts to communicate would have been significantly less productive during this entire period. I am very grateful to him.

The untimely death of Marc Yor while this book was in production is an irretrievable loss for Probability Theory. Marc had donated much time to improve this work (as well as the author's previous two books), and it is only befitting that it be dedicated to his memory.

References

1. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
2. Talagrand, M.: The Generic Chaining. Springer Monographs in Mathematics. Springer, Berlin (2005). viii+222 pp. ISBN: 3-540-24518-9
3. Talagrand, M.: Are many small sets explicitly small? In: STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, pp. 13–35. ACM, New York (2010). Available at http://research.microsoft.com/apps/video/dl.aspx?id=137091

Contents

1. Philosophy and Overview of the Book

1.1 Underlying Philosophy

This short chapter will describe the philosophy underlying this book, and some of its highlights. This will be done using words rather than formulas, so that the description is necessarily imprecise, and is only intended to provide some insight into our point of view.

The practitioner of stochastic processes is likely to be struggling at any given time with his favorite model of the moment, a model that will typically involve a rather rich and complicated structure. There is a near infinite supply of such models. Fashions come and go, and the importance with which we view any specific model is likely to strongly vary over time.

The first advice the author received from his advisor Gustave Choquet was as follows: Always consider a problem under the minimum structure in which it makes sense. This advice will probably be as fruitful in the future as it has been in the past, and it has strongly influenced this work. By following it, one is naturally led to the study of problems with a kind of minimal and intrinsic structure. Besides the fact that it is much easier to find the crux of the matter in a simple structure than in a complicated one, there are not so many really basic structures, so one can hope that they will remain of interest for a very long time. This book is devoted to the study of a few of these structures.

It is of course very nice to enjoy the feeling, real or imaginary, that one is studying structures that might be of intrinsic importance, but the success of the approach of studying "minimal structures" has ultimately to be judged by its results. More often than not general principles are insufficient to answer specific questions. Yet, as we shall demonstrate, the tools arising from this approach have provided the final words to a number of classical problems.

1.2 Peculiarities of Style

The author has tried to make this book as self contained as he could, but some readers may be disturbed to see that certain standard considerations are given little or no attention. You will find rather little about "convergence"

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*,
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of
Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_1,

here, at least explicitly. There are no apparent σ-algebras, and measurability is hardly mentioned at all. Essentially we prove inequalities, and for this one can basically pretend that every index set is finite. This is why we shall shamelessly consider suprema of families of r.v.s without ever defining "essential supremum" or "separable processes", and why (for example) when studying convergence of random Fourier series, we put much more emphasis on the estimates of finite partial sums than on convergence itself. All these missing "details" belong to pre-1950 mathematics. While these are fundamentally important, there is already a plethora of material available about them, which this author has no special competences to rewrite.

1.3 What This Book Is *Really* About

Readers (should there be any) who are trained in the classical theory of processes (semi-martingales, etc.) may find it difficult to understand what this book is all about.

For us a stochastic process is a collection of random variables (r.v.s) $(X_t)_{t\in T}$, where T is an index set. This index set may be, or not, a subset of $\mathbb{R}^m$. Most importantly, it may be a subset of $\mathbb{R}^m$, but its structure as a subset of $\mathbb{R}^m$ is of no help. A fundamental example of stochastic processes is a random series

$$X_t = \sum_{k\geq 1} \xi_k f_k(t) \,, \tag{1.1}$$

where f_k are functions and ξ_k are independent random variables. The study of such series will occupy a large part of this book.

Our main objective is the study of stochastic processes, and to find conditions under which their trajectories are bounded or continuous.

Let us first discuss a very simple situation, processes $(X_t)_{t\in T}$ where $T = [0,1]^m$, that satisfy the Kolmogorov conditions, that is

$$\forall\, s,t \in [0,1]^m \ , \ \mathsf{E}|X_s - X_t|^p \leq d(s,t)^\alpha \ , \tag{1.2}$$

where $d(s,t)$ denotes the Euclidean distance and $p > 0, \alpha > m$.

Kolmogorov's chaining idea is to use successive approximations $\pi_n(t)$ of a point t of T. Here it is natural to assume for example that $\pi_n(t) \in G_n$ where G_n is the set of points x in $[0,1[$ (the set which you might denote $[0,1)$ if you are Anglo-Saxon) such that the coordinates of $2^n x$ are integers. Thus $\operatorname{card} G_n = 2^{nm}$.

For $n \geq 0$, let us define

$$U_n = \{s \in G_n \ , \ t \in G_{n+1} \ ; \ d(s,t) \leq 3\sqrt{m}2^{-n}\} \ .$$

The somewhat arbitrary choice of the constant $3\sqrt{m}$ is related to the fact that each point of T is within distance $\sqrt{m}2^{-n}$ of a point of G_n. A very important feature here is that even though there are 2^{nm} choices for s and $2^{(n+1)m}$ possible choices for t we have

$$\operatorname{card} U_n \le K(m) 2^{nm} \;, \tag{1.3}$$

where $K(m)$ denotes a number depending only on m. This is due to the fact that (ignoring "edge effects") the space T is really "m-dimensional" and "the same around each point". *Much of the work done in this book is to handle situations where such a homogeneity does not occur,* and these situations do occur in classical problems. In Appendix A we give a self-contained proof that processes satisfying (1.2) have a continuous version, and we explain several more basic chaining results which might help to provide perspective. We also discuss some classical tools such as the Garsia-Rodemich-Rumsey lemma [1]. This appendix is designed to be read independently of the rest of the book, and now is probably the best time to study it.

1.4 Gaussian Processes and the Generic Chaining

This section gives an overview of Chapter 2. More generally, Section n gives the overview for Chapter $(n-2)$.

The most important question considered in this book is the boundedness of Gaussian processes. As we already noticed, the intrinsic distance (0.1) points to the relevance of the metric space (T,d) where T is the index set. This metric space is far from being arbitrary, since it is isometric to a subset of a Hilbert space. (By its very nature, this introduction is going to contain many statements, like the previous one, which, depending on the reader's background, may or may not sound obvious. The best way to obtain complete clarification about these statements is to start reading from the next chapter on.) Something properly extraordinary happens here. It is a deadly trap to try to use the specific properties of the metric space (T,d). The proper approach is to just think of it as a general metric space. Since there is only so much one can do with a bare metric space structure, nothing can get really complicated then.

One of the most important properties of Gaussian processes is the following "increment condition"

$$\forall u > 0 \;,\; \mathsf{P}(|X_s - X_t| \ge u) \le 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right) \;, \tag{1.4}$$

which simply expresses a (somewhat suboptimal) bound on the tail of the Gaussian r.v. $X_s - X_t$. When proving regularity conditions on Gaussian processes, this is actually the only property we shall use. There is an obvious similarity between (1.4) and (1.2), and pursuing Kolmogorov's ideas in the present abstract setting led in 1967 to the celebrated Dudley's bound. This is arguably the most important result about regularity of Gaussian processes, so it deserves to be stated here. For any $\delta > 0$,

$$\mathsf{E} \sup_{d(s,t)\le\delta} |X_s - X_t| \le L \int_0^\delta \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon \; . \tag{1.5}$$

Here L is a number and $N(T,d,\epsilon)$ denotes the smallest number of balls for the distance d, of radius ϵ, which is needed to cover T. A quick proof of this bound in this exact form may be found in the self-contained Appendix A on page 601.

If one thinks of chaining as a succession of steps that provide successive approximations of the space (T,d), in the Kolmogorov chaining for each n the "variation of the process during the n-th step is controlled uniformly over all possible chains".

In Section 2.2 we explain the basic idea of the generic chaining. The twist over the classical method is simply that the "variation of the process during the n-th step may depend on which chain we follow". Once the argument is properly organized, it is not any more complicated than the classical argument. It is in fact exactly the same, and requires no more energy than most books spend e.g. to prove the continuity of Brownian motion by weaker methods. Yet, while Dudley's classical bound is not always sharp, the bound obtained through the generic chaining is optimal, as will be explained later.

Many processes of importance do not have tails as simple to describe as in (1.4). For example, when one controls these tails through Bernstein's inequality, two distances rather than one get involved. To deal with that situation it is convenient to formulate the generic chaining bound using special sequences of partitions of the metric space (T,d), that we shall call *admissible sequences* throughout the book.

To make the generic chaining bound useful, the basic issue is then to be able to construct admissible sequences. These admissible sequences measure an aspect of the "size" of the metric space. In Section 2.3 we introduce another method to measure the size of the metric space, through the behavior of certain "functionals", that are simply numbers attached to each subset of the entire space. The fundamental fact is that the measure of the size of the metric space one obtains through admissible sequences and through functionals are equivalent in full generality. This is proved in Section 2.3 for the easy part (that the admissible sequence approach provides a larger measure of size than the functional approach) and in Section 2.6 for the converse. This converse is, in effect, a method to construct sequences of partitions in a metric space. The point of this equivalence is that in practice, as will be demonstrated throughout the book, it is much easier in concrete situations to guess the size of a given space through the functional approach than the admissible sequences approach.

In Section 2.4 we prove that the generic bound can be reversed for Gaussian processes, therefore providing a characterization of their sample-boundedness. Gaussian processes are deeply related to the geometry of Hilbert space. In some sense, a Gaussian process *is* nothing but a subset of Hilbert space. A number of basic questions related to the geometry of Hilbert space remain unanswered, such as how to relate certain measures of size of a subset of Hilbert space with the corresponding measures of size of its convex hull.

The conventional wisdom among mainstream probabilists has long been that Dudley's bound "suffices for all practical purposes" and that the cases where it is not sharp are "exotic". To dispel this belief, in Section 2.5 we investigate in detail the case of ellipsoids of a Hilbert space. Dudley's bound fails to explain the size of the Gaussian processes indexed by ellipsoids. Ellipsoids will play a basic role in Chapter 4.

In Sections 2.6 we detail in the simplest case (which is also the most important) the fundamental "partitioning scheme", a method to construct partitions in a general metric space. Interestingly the construction is quite immediate, in that it is performed following a simple "greedy algorithm". It does, however, require some care to prove that the partition the algorithm constructs possesses the properties we wish. This method, and its later generalizations, are of fundamental importance. The good news is that the *statements* of these later generalizations will require more abstraction, but that their *proofs* will be absolutely identical to that of the fundamental case of this section. A first generalization is given in Section 2.7.

1.5 Random Fourier Series and Trigonometric Sums, I

In Section 3.2 we investigate Gaussian processes in "the stationary case," where e.g. the underlying space is a compact group and the distance is translation invariant. This is relevant to the study of random Fourier series, which are simply series of the type (1.1) where f_k are characters of the group. The basic example is that of series of the type

$$X_t = \sum_{k\geq 1} \xi_k \exp(2\pi i k t) , \tag{1.6}$$

where $t \in [0,1]$ and the r.v.s ξ_k are independent. The fundamental case where $\xi_k = a_k g_k$ for numbers a_k and independent Gaussian r.v.s (g_k) is of great historical importance. We prove some of the classical results of M. Marcus and G. Pisier, which provide a complete solution in this case, and are also quite satisfactory in the more general case (1.6) when the random coefficients (ξ_k) are square-integrable. In Section 3.3 we explain a result of X. Fernique on vector-valued random Fourier series, which had an important part in the genesis of this book.

1.6 Matching Theorems, I

Despite the fact demonstrated in Section 2.5 that the generic chaining is already required to understand ellipsoids in Hilbert space, the misconception that such oddities will not occur in "real" situations might persist. Chapter 4 makes the point that the generic chaining (or of course some equivalent form of it) is already required to really understand the irregularities occurring in the distribution of N points $(X_i)_{i\le N}$ independently and uniformly distributed in the unit square. These irregularities are measured by the "cost" of pairing (=matching) these points with N fixed points that are very uniformly spread, for various notions of cost.

In Section 4.3 we investigate the situation where the cost of a matching is measured by the average distance between paired points. We prove the result of Ajtai, Komlós, Tusnády, that the expected cost of an optimal matching is at most $L\sqrt{\log N}/\sqrt{N}$ where L is a number. In Section 4.4 we investigate the situation where the cost of a matching is measured instead by the maximal distance between paired points. We prove the theorem of Leighton and Shor that the expected cost of a matching is at most $L(\log N)^{3/4}/\sqrt{N}$. The factor $1/\sqrt{N}$ is simply a scaling factor, but the fractional powers of log are indeed fascinating, and all the more since as we shall prove later in Chapter 6 that they are optimal.

In Section 4.2 we provide a general background on matchings, and we show that one can often reduce the proof of a matching theorem to the proof of a suitable bound for a quantity of the type

$$\sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}(f(X_i)-\int f\mathrm{d}\lambda)\Big|$$

where $\mathcal{F}$ is a class of functions on the unit square and λ is Lebesgue's measure. That is, we have to bound a complicated random process. The main issue is to control in the appropriate sense the size of the class $\mathcal{F}$. For this we parametrize this class of functions by a suitable ellipsoid of Hilbert space using Fourier transforms.

This approach illustrates particularly well the benefits of an abstract point of view: we are able to trace the mysterious fractional powers of log back to the geometry of ellipsoids in Hilbert space. This is why we start the chapter with an investigation of these ellipsoids in Section 4.1. The philosophy of the main result, the Ellipsoid Theorem, is that an ellipsoid is in some sense somewhat smaller than what one might think at first. This is due to the fact that an ellipsoid is sufficiently convex, and that, somehow, it gets "thinner" when one gets away from its center. The Ellipsoid Theorem is a special case of a more general result (with the same proof) about the structure of sufficiently convex bodies, one that will have important applications in Chapter 16.

With the exception of Section 4.1, the results of Chapter 4 are not connected to any subsequent material before Chapter 14.

1.7 Bernoulli Processes

In Chapter 5 we investigate Bernoulli processes, where the individual random variables X_t are linear combinations of independent random signs, a special case of the general setting (1.1). Random signs are obviously important r.v.s, and occur frequently in connection with "symmetrization procedures", a very useful tool. Each Bernoulli process is associated with a Gaussian process in a canonical manner, when one replaces the random signs by independent standard Gaussian r.v.s. The Bernoulli process has better tails than the corresponding Gaussian process (it is "subgaussian") and is bounded whenever the Gaussian process is bounded. There is however a completely different reason for which a Bernoulli process might be bounded, namely that the sum of the absolute values of the coefficients of the random signs remains bounded independently of the index t. A natural question is then to decide whether these two extreme situations are the only fundamental reasons why a Bernoulli process can be bounded, in the sense that a suitable "mixture" of them occurs in every bounded Bernoulli process. This was the "Bernoulli Conjecture" (to be stated formally on page 130), which has been so brilliantly solved by W. Bednorz and R. Latała. The proof of their fundamental result occupies much of this chapter. Many of the previous ideas it builds upon will be further developed in subsequent chapters.

In Section 5.2 we describe an efficient method to organize chaining arguments for Bernoulli processes. This method is an essential step of the proof of the Bednorz-Latała theorem, but it also turns out to be very useful in practical situations, and in particular in the study of random Fourier series.

In Section 5.3 we present some fundamental facts about Bernoulli processes, which are the building blocks of the proof of the Bednorz-Latała theorem.

A linear combination of independent random signs looks like a Gaussian r.v. when the coefficients of the random signs are small. We can expect that a Bernoulli process will look like a Gaussian process when these coefficients are suitably small. The purpose of Section 5.4 is to make this idea precise. The following sections then complete the proof of the Bednorz-Latała theorem.

1.8 Trees and the Art of Lower Bounds

We describe different notions of trees, and show how one can measure the "size" of a metric space by the size of the largest trees it contains, in a way which is equivalent to the measures of size introduced in Chapter 2. This idea played an important part in the history of Gaussian processes. Its appeal is mostly that trees are easy to visualize. Building a large tree in a metric space is an efficient method to bound its size from below. We perform such an explicit construction in the toy case of certain ellipsoids as a warm-up. We then use similar ideas to prove (using also one of the main results of Chapter 5) that the upper bounds obtained in the matching problems of Chapter 4 are sharp: we prove lower bounds of the same order as these upper bounds.

1.9 Random Fourier Series and Trigonometric Sums, II

In order to demonstrate the efficiency of the chaining method of Section 5.2, we return in this chapter to the study of random Fourier series, but now without making any assumption of integrability on the random coefficients, which we simply assumed to be independent symmetric r.v.s. This chapter also develops one of the fundamental ideas of this work: many processes can be exactly controlled, not by using one or two distances, but by using an entire family of distances. With these tools, we are able to give in full generality necessary and sufficient conditions for convergence of random Fourier series. These conditions can be formulated in words by saying that convergence is equivalent to the finiteness of (a proper generalization of) a certain "entropy integral". We then give examples of application of the abstract theorems to the case of ordinary random Fourier series.

1.10 Processes Related to Gaussian Processes

It is natural to expect that our increased understanding of the properties of Gaussian processes will also bring information about processes that are, in various senses, related to Gaussian processes. Such was the case with the Bernoulli processes in Chapter 5.

It turns out that p-stable processes, an important class of processes, are conditionally Gaussian, and in Section 8.1 we use this property to provide lower bounds for such processes. Although these bounds are in general very far from being upper bounds, they are in a sense extremely accurate (in certain situations, where there is "stationarity", these lower bounds can be reversed). We are able to obtain these bounds rather easily, even in the most difficult case, the case $p = 1$. Essentially more general results are proved later in Chapter 11 for infinitely divisible processes, but the proofs are considerably simpler in the case of p-stable processes.

Another natural class of processes that are conditionally Gaussian are order 2 Gaussian chaos (which are essentially second degree polynomials of Gaussian random variables). It seems at present a hopelessly difficult task to give lower and upper bounds of the same order for these processes, but in Section 8.2 we obtain a number of results in the right direction. Chaos processes are also very instructive because there exists other methods than chaining to control them (a situation which we do not expect to occur for processes defined as sums of a random series).

In Section 8.3 we study the tails of a single multiple order Gaussian chaos, and present a deep result of R. Latała which provides a rather complete description of these tails.

1.11 Theory and Practice of Empirical Processes

Let us first hurry to insist that despite the title this chapter covers only a very special (yet of fundamental importance) topic about empirical processes: how to control the supremum of the empirical process over a class of functions.

The fundamental theoretical question in this direction is whether there exists a "best possible" method to control this supremum at a given size of the random sample. In Section 9.1 we offer a natural candidate for such a "best possible" method, in the spirit of the Bednorz-Latała result of Chapter 5. Whether this natural method is actually optimal is a major open problem (Problem 9.1.3), which could well be difficult. We then demonstrate again the power of the chaining scheme of Section 5.2 by providing a sharper version of Ossiander's bracketing theorem with a very simple proof.

Does meditating on the "theoretically best possible" way to control the empirical process provide help for the "practical" matter of controlling empirical processes under actual sets of conditions that occur naturally in applications of this theory? In order to convince the reader that this might well be the case, we selected (in a very arbitrary manner) two deep recent results, of which we present somewhat streamlined proofs: in Section 9.3, a recent theorem of G. Paouris and S. Mendelson, and in Section 9.4 a recent theorem of R. Adamczak, A.E. Litvak, A. Pajor and N. Tomczak-Jaegermann.

1.12 Partition Scheme for Families of Distances

As we already pointed out in Chapter 7 the description of the size of the tail of a r.v. often requires an entire sequence of parameters, and for processes that consist of such r.v.s, the natural underlying structure is not a metric space, but a space equipped with a suitable family of distances. In Section 10.1 we extend the tools of Section 2.7 to this setting.

In Section 10.2 we apply these tools to the situation of "canonical processes" where the r.v.s X_t are linear combinations of independent copies of symmetric r.v.s with density proportional to $\exp(-|x|^\alpha)$ where $\alpha \geq 1$ (and to considerably more general situations as discovered by R. Latała). In these situations, the size of the process can be completely described as a function of the geometry of the index space, a far reaching extension of the Gaussian case.

1.13 Infinitely Divisible Processes

We study these processes in a much more general setting than what mainstream probability theory has yet investigated: we make no assumption of stationarity of increments of any kind and our processes are actually indexed by an abstract set. These processes are to Lévy processes what a general Gaussian process is to Brownian motion.

Our main tool to study infinitely divisible processes is a representation theorem due to J. Rosinski, which makes them appear as conditionally Bernoulli processes. Unfortunately they do not seem to be conditionally Gaussian. Since we do not understand Bernoulli processes as well as Gaussian processes, it is technically challenging to use this fact. Bringing in the tools of Chapter 5, for a large class of these processes, we are able to prove lower bounds that extend those given in Section 8.1 for p-stable process. These lower bounds are not upper bounds in general, but we succeed in showing in a precise sense that they are upper bounds for "the part of boundness of the process which is due to cancellation". Thus, whatever bound might be true for the "remainder of the process" owes nothing to cancellation. The results are described in complete detail with all definitions in Section 11.2.

1.14 The Fundamental Conjectures

In Chapter 12 we outlay a long range research program. We believe that it might well be true in considerable generality that for processes of the type (1.1) (as was proved in special cases in Chapter 11) "chaining explains all the part of the boundedness which is due to cancellation", and in Section 12.3 we state a precise conjecture to that effect. Even if this conjecture is true, there would remain to describe the "part of the boundedness which owes nothing to cancellation", and for this part also we propose sweeping conjectures, which, if true, would revolutionize our understanding. At the heuristic level, the underlying idea of these conjectures is that ultimately, a bound for a stochastic process always arises from the use of the 'union bound' $\mathsf{P}(\cup_n A_n) \leq \sum_n \mathsf{P}(A_n)$ in a simple situation, the use of basic principles such as linearity and positivity, or combinations of these.

1.15 Convergence of Orthogonal Series; Majorizing Measures

The old problem of characterizing the sequences (a_m) such that for each orthonormal sequence (φ_m) the series $\sum_{m \geq 1} a_m \varphi_m$ converges a.s. has recently been solved by A. Paszkiewicz. Using a more abstract point of view, we present a very much simplified proof of his results (due essentially to

W. Bednorz). This leads us to the question of discussing when a certain condition on the "increments" of a process implies its boundedness. When the increment condition is of "polynomial type", this is more difficult than in the case of Gaussian processes, and requires the notion of "majorizing measure". We present several elegant results of this theory, in their seemingly final form recently obtained by W. Bednorz.

1.16 Matching Theorems II: Shor's Matching Theorem

This chapter continues Chapter 4. We prove a deep improvement of the Ajtai, Komlós, Tusnády theorem due to P. Shor. Unfortunately, due mostly to our lack of geometrical understanding, the best conceivable matching theorem, which would encompass this result as well as those of Chapter 4, and much more, remains as a challenging problem, "the ultimate matching conjecture" (a conjecture which is solved in the next chapter in dimension ≥ 3).

1.17 The Ultimate Matching Theorem in Dimension ≥ 3

In this case, which is easier than the case of dimension 2 (but still apparently rather non-trivial), we are able to obtain the seemingly final result about matchings, a strong version of "the ultimate matching conjecture". There are no more fractional powers of $\log N$ here, but in a random sample of N points uniformly distributed in $[0,1]^3$, local irregularities occur at all scales between $N^{-1/3}$ and $(\log N)^{1/3} N^{-1/3}$, and our result can be seen as a precise global description of these irregularities. Strictly speaking the proof does not use chaining, although it is in the same spirit, and it remains to crystallize the abstract principle that might lay behind it.

1.18 Applications to Banach Space Theory

Chapter 16 gives applications to Banach space theory. The sections of this Chapter are largely independent of each other, and the link between them is mostly that they all reflect past interests of the author. The results of this chapter do not use those of Chapter 10. In Section 16.1, we study the cotype of operators from ℓ_N^∞ into a Banach space. In Section 16.2, we prove a comparison principle between Rademacher (=Bernoulli) and Gaussian averages of vectors in a finite dimensional Banach space, and we use it to compute the Rademacher cotype-2 of a finite dimensional space using only a few vectors. In Section 16.3 we discover how to classify the elements of the unit ball of L^1 "according to the size of the level sets". In Section 16.4 we explain, given a Banach space E with an 1-unconditional basis (e_i), how to "compute" the quantity $\mathsf{E}\|\sum_i g_i e_i\|$ when g_i are independent Gaussian r.v.s, a further variation on the fundamental theme of the interplay between the L^1, L^2 and L^∞ norms. In Section 16.5 we study the norm of the restriction of an operator from ℓ^q_N to the subspace generated by a randomly chosen small proportion of the coordinate vectors, and in Section 16.6 we use these results to obtain a sharpened version of the celebrated results of J. Bourgain on the Λ_p problem. In Section 16.7, given a uniformly bounded orthonormal system, we study how large a subset we can find on the span of which the L^2 and L^1 norms are close to each other. In Section 16.8, given a k-dimensional subspace of L^p for $1 < p < 2$ we investigate for which values of N we can embed nearly isometrically this subspace as a subspace of ℓ^p_N. We prove that we may choose N as small as about $k \log k (\log \log k)^2$. A recent proof by G. Schechtman of a theorem of Y. Gordon concludes this chapter in Section 16.9.

1.19 Appendix B: Continuity

Most of the book is devoted to the task of bounding stochastic processes. The connoisseur knows this is the hard work, and that once it is understood it is a simple matter to study continuity. This appendix samples some results in the direction of moduli of continuity, in particular for Gaussian processes.

Reference

1. Garsia, A.M., Rodemich, E., Rumsey, H.: A real variable lemma and the continuity of path of some Gaussian processes. Indiana Univ. Math. J. **20**, 565–578 (1970/1971)

2. Gaussian Processes and the Generic Chaining

2.1 Overview

The overview of this chapter is given in Chapter 1, Section 1.4. More generally, Section 1.n is the overview of Chapter $n-2$.

2.2 The Generic Chaining

In this section we consider a metric space (T,d) and a process $(X_t)_{t\in T}$ that satisfies the increment condition:

$$\forall u>0\ ,\ \mathsf{P}(|X_s-X_t|\geq u)\leq 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right)\ . \qquad (1.4)$$

In particular this is the case when $(X_t)_{t\in T}$ is a Gaussian process and $d(s,t)^2 = \mathsf{E}(X_s-X_t)^2$. Unless explicitly specified otherwise (and even when we forget to repeat it) we will *always* assume that the process is centered, i.e.

$$\forall t\in T\ ,\ \ \mathsf{E}X_t=0\ . \qquad (2.1)$$

We will measure the "size of the process $(X_t)_{t\in T}$" by the quantity $\mathsf{E}\sup_{t\in T} X_t$. (The reader who is impatient to understand why this quantity is a good measure of the "size of the process" can peek ahead to Lemma 2.2.1 below.)

A side issue (in particular when T is uncountable) is that what is meant by the quantity $\mathsf{E}\sup_{t\in T} X_t$ is not obvious. An efficient method is to *define* this quantity by the following formula:

$$\mathsf{E}\sup_{t\in T} X_t=\sup\Big\{\mathsf{E}\sup_{t\in F} X_t\ ;\ F\subset T\ ,\ F \text{ finite}\Big\}\ , \qquad (2.2)$$

where the right-hand side makes sense as soon as each r.v. X_t is integrable. This will be the case in almost all the situations considered in this book. For the next few dozens of pages, we make the effort to explain in every case how to reduce the study of the supremum of the r.v.s under consideration to the supremum of a finite family, until the energy available for this sterile exercise runs out, see Section 1.2.

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*,
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_2,

Let us say that a process $(X_t)_{t \in T}$ is *symmetric* if it has the same law as the process $(-X_t)_{t \in T}$. Almost all the processes we shall consider are symmetric (although for some of our results this hypothesis is not necessary). The following justifies using the quantity $\mathsf{E} \sup_t X_t$ to measure "the size of a symmetric process".

Lemma 2.2.1. *If the process $(X_t)_{t \in T}$ is symmetric then*

$$\mathsf{E} \sup_{s,t \in T} |X_s - X_t| = 2\mathsf{E} \sup_{t \in T} X_t \ .$$

Proof. We note that

$$\sup_{s,t \in T} |X_s - X_t| = \sup_{s,t \in T} (X_s - X_t) = \sup_{s \in T} X_s + \sup_{t \in T} (-X_t) \,,$$

and we take expectations. □

Exercise 2.2.2. Consider a symmetric process $(X_t)_{t \in T}$. Given any t_0 in T prove that

$$\mathsf{E} \sup_{t \in T} |X_t| \le 2\mathsf{E} \sup_{t \in T} X_t + \mathsf{E}|X_{t_0}| \le 3\mathsf{E} \sup_{t \in T} |X_t| \ . \tag{2.3}$$

Generally speaking, and unless mentioned otherwise, the exercises have been designed to be easy. The author however never taught this material in a classroom, so it might happen that some exercises are not that easy after all for the beginner. Please do not be discouraged if this should be the case. (In fact, as it would have taken supra-human dedication for the author to write in detail all the solutions, there is no real warranty that each of the exercise is really feasible or even correct.) The exercises have been designed to shed some light on the material at hand, and to shake the reader out of her natural laziness by inviting her to manipulate some simple objects. (Please note that it is probably futile to sue me over the previous statement, since the reader is referred as "she" through the entire book and not only in connection with the word "laziness".)

In this book, we often state inequalities about the supremum of a symmetric process using the quantity $\mathsf{E} \sup_{t \in T} X_t$ simply because this quantity looks typographically more elegant than the equivalent quantity $\mathsf{E} \sup_{s,t \in T} |X_s - X_t|$. Of course, it is not always enough to control the first moment of $\sup_{s,t \in T} |X_s - X_t|$. We also need to control the tails of this r.v. Emphasis is given to the first moment simply because, as the reader will eventually realize, this is the difficult part, and once this is achieved, control of higher moments is often provided by the same arguments.

Our goal is to find bounds for $\mathsf{E} \sup_{t \in T} X_t$ depending on the structure of the metric space (T, d). We will assume that T is finite, which, as shown by (2.2), does not decrease generality.

Given any t_0 in T, the centering hypothesis (2.1) implies

$$\mathsf{E}\sup_{t\in T} X_t = \mathsf{E}\sup_{t\in T}(X_t - X_{t_0}) \; . \tag{2.4}$$

The latter form has the advantage that we now seek estimates for the expectation of the non-negative r.v. $Y = \sup_{t\in T}(X_t - X_{t_0})$. Then,

$$\mathsf{E}Y = \int_0^\infty \mathsf{P}(Y > u)\,\mathrm{d}u \; . \tag{2.5}$$

Thus it is natural to look for bounds of

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \geq u\Big) \; . \tag{2.6}$$

The first bound that comes to mind is the "union bound"

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \geq u\Big) \leq \sum_{t\in T} \mathsf{P}(X_t - X_{t_0} \geq u) \; . \tag{2.7}$$

It seems worthwhile to draw right away some consequences from this bound, and to discuss at leisure a number of other simple, yet fundamental facts. This will take a bit over three pages, after which we will come back to the main story of bounding Y. Throughout this work, $\Delta(T)$ denotes the diameter of T,

$$\Delta(T) = \sup_{t_1,t_2\in T} d(t_1,t_2) \; . \tag{2.8}$$

When we need to make clear which distance we use in the definition of the diameter, we will write $\Delta(T,d)$ rather than $\Delta(T)$. Consequently (1.4) and (2.7) imply

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \geq u\Big) \leq 2\operatorname{card} T \exp\Big(-\frac{u^2}{2\Delta(T)^2}\Big) \; . \tag{2.9}$$

Let us now record a simple yet important computation, that will allow us to use the information (2.9).

Lemma 2.2.3. *Consider a r.v. $Y \geq 0$ which satisfies*

$$\forall\, u > 0 \; , \; \mathsf{P}(Y \geq u) \leq A \exp\Big(-\frac{u^2}{B^2}\Big) \tag{2.10}$$

for certain numbers $A \geq 2$ and $B > 0$. Then

$$\mathsf{E}Y \leq LB\sqrt{\log A} \; . \tag{2.11}$$

Here, as in the entire book, L denotes a universal constant. We make the convention that this constant **is not necessarily** the same on each occurrence. This convention is very convenient, but it certainly needs to get used to, as e.g. in the formula $\sup_x xy - Lx^2 = y^2/L$. This convention should be remembered at all times.

When meeting an unknown notation such as this previous L, the reader might try to look at the **index**, where some of the most common notation is recorded.

Proof of Lemma 2.2.3. We use (2.5) and we observe that since $\mathsf{P}(Y \geq u) \leq 1$, then for any number u_0 we have

$$\begin{aligned} \mathsf{E}Y = \int_0^\infty \mathsf{P}(Y \geq u)\mathrm{d}u &= \int_0^{u_0} \mathsf{P}(Y \geq u)\mathrm{d}u + \int_{u_0}^\infty \mathsf{P}(Y \geq u)\mathrm{d}u \\ &\leq u_0 + \int_{u_0}^\infty A\exp\Big(-\frac{u^2}{B^2}\Big)\mathrm{d}u \\ &\leq u_0 + \frac{1}{u_0}\int_{u_0}^\infty uA\exp\Big(-\frac{u^2}{B^2}\Big)\mathrm{d}u \\ &= u_0 + \frac{AB^2}{2u_0}\exp\Big(-\frac{u_0^2}{B^2}\Big)\,, \end{aligned} \tag{2.12}$$

and the choice of $u_0 = B\sqrt{\log A}$ completes the proof. □

Combining (2.11) and (2.9) we obtain that (considering separately the case where $\operatorname{card} T = 1$)

$$\mathsf{E}\sup_{t\in T} X_t \leq L\Delta(T)\sqrt{\log \operatorname{card} T}\,. \tag{2.13}$$

The following special case is fundamental.

Lemma 2.2.4. *If $(g_k)_{k\geq 1}$ are standard Gaussian r.v.s then*

$$\mathsf{E}\sup_{k\leq N} g_k \leq L\sqrt{\log N}\,. \tag{2.14}$$

Exercise 2.2.5. (a) Prove that (2.14) holds as soon as the r.v.s g_k are centered and satisfy

$$\mathsf{P}(g_k \geq t) \leq 2\exp\Big(-\frac{t^2}{2}\Big) \tag{2.15}$$

for $t > 0$.
(b) For $N \geq 2$ construct N centered r.v.s $(g_k)_{k\leq N}$ satisfying (2.15), and taking only the values $0, \pm\sqrt{\log N}$ and for which $\mathsf{E}\sup_{k\leq N} g_k \geq \sqrt{\log N}/L$. (You are not yet asked to make these r.v.s independent.)
(c) After learning (2.16) below, solve (b) with the further requirement that the r.v.s g_k are independent. If this is too hard, look at Exercise 2.2.7, (b) below.

This is taking us a bit ahead, but an equally fundamental fact is that when the r.v.s (g_k) are jointly Gaussian, and "significantly different from each other" i.e. $\mathsf{E}(g_k - g_\ell)^2 \geq a^2 > 0$ for $k \neq \ell$, the bound (2.14) can be reversed, i.e. $\mathsf{E}\sup_{k\leq N} g_k \geq a\sqrt{\log N}/L$, a fact known as Sudakov's minoration. Sudakov's minoration is a non-trivial fact, but it should be really helpful to solve Exercise 2.2.7 below. Before that let us point out a simple fact, that will be used many times.

Exercise 2.2.6. Consider independent events $(A_k)_{k\ge 1}$. Prove that

$$\mathsf{P}\Big(\bigcup_{k\le N} A_k\Big) \ge 1 - \exp\Big(-\sum_{k\le N}\mathsf{P}(A_k)\Big)\ . \tag{2.16}$$

(Hint: $\mathsf{P}(\cup_{k\le N} A_k) = 1 - \prod_{k\le N}(1-\mathsf{P}(A_k))$.)

In words: independent events such that the sum of their probabilities is small are basically disjoint.

Exercise 2.2.7. (a) Consider independent r.v.s $Y_k \ge 0$ and $u > 0$ with

$$\sum_{k\le N}\mathsf{P}(Y_k \ge u) \ge 1\ . \tag{2.17}$$

Prove that

$$\mathsf{E}\sup_{k\le N} Y_k \ge \frac{u}{L}\ .$$

(Hint: use (2.16) to prove that $\mathsf{P}(\sup_{k\le N} Y_k \ge u) \ge 1/L$.)
(b) We assume (2.17), but now Y_k need not be ≥ 0. Prove that

$$\mathsf{E}\sup_{k\le N} Y_k \ge \frac{u}{L} - \mathsf{E}|Y_1|\ .$$

(Hint: observe that for each event Ω we have $\mathsf{E}\mathbf{1}_\Omega \sup_k Y_k \ge -\mathsf{E}|Y_1|$.)
(c) Prove that if $(g_k)_{k\ge 1}$ are independent standard Gaussian r.v.s then $\mathsf{E}\sup_{k\le N} g_k \ge \sqrt{\log N}/L$.

Before we go back to our main story, it might be worth for the reader to consider in detail consequences of an "exponential decay of tails" such as in (2.10). This is the point of the next exercise.

Exercise 2.2.8. (a) Assume that for a certain $B > 0$ the r.v. $Y \ge 0$ satisfies

$$\forall\, u > 0\ ,\ \mathsf{P}(Y \ge u) \le 2\exp\Big(-\frac{u}{B}\Big)\ . \tag{2.18}$$

Prove that

$$\mathsf{E}\exp\Big(\frac{Y}{2B}\Big) \le L\ . \tag{2.19}$$

Prove that for $a > 0$ one has $(x/a)^a \le \exp x$. Use this for $a = p$ and $x = Y/2B$ to deduce from (2.19) that for $p \ge 1$ one has

$$(\mathsf{E}Y^p)^{1/p} \le LpB\ . \tag{2.20}$$

(b) Assuming now that for a certain $B > 0$ one has

$$\forall\, u > 0\ ,\ \mathsf{P}(Y \ge u) \le 2\exp\Big(-\frac{u^2}{B^2}\Big)\ , \tag{2.21}$$

prove similarly (or deduce from (a)) that $\mathsf{E}\exp(Y^2/2B^2) \le L$ and that for $p \ge 1$ one has

$$(\mathsf{E}Y^p)^{1/p} \le LB\sqrt{p}\ . \tag{2.22}$$

In words, (2.20) states that "as p increases, the L^p norm of an exponentially integrable r.v. does not grow faster than p," and (2.22) asserts that if the square of the r.v. is exponentially integrable, then its L^p norm does not grow faster than $\sqrt{p}$. (These two statements are closely related.) More generally it is very classical to relate the size of the tails of a r.v. with the rate of growth of its L^p norm. This is not explicitly used in the sequel, but is good to know as background information. As the following shows, (2.22) provides the correct rate of growth in the case of Gaussian r.v.s.

Exercise 2.2.9. If g is a standard Gaussian r.v. it follows from (2.22) that for $p \geq 1$ one has $(\mathsf{E}|g|^p)^{1/p} \leq L\sqrt{p}$. Prove one has also

$$(\mathsf{E}|g|^p)^{1/p} \geq \frac{\sqrt{p}}{L} \,. \tag{2.23}$$

One knows how to compute exactly $\mathsf{E}|g|^p$, from which one can deduce (2.23). You are however asked to provide a proof in the spirit of this work by deducing (2.23) solely from the information that, say, for $u > 0$ we have (choosing on purpose crude constants) $\mathsf{P}(|g| \geq u) \geq \exp(-u^2/3)/L$.

You will find basically no exact computations in this book. The aim is different. We study quantities which are far too complicated to be computed exactly, and we try to bound them from above, and sometimes from below by simpler quantities with as little a gap as possible between the upper and the lower bounds, the gap being ideally only a multiplicative constant.

We go back to our main story. The bound (2.13) will be effective if the variables $X_t - X_{t_0}$ are rather uncorrelated (and if there are not too many of them). But it will be a disaster if many of the variables $(X_t)_{t \in T}$ are nearly identical. Thus it seems a good idea to gather those variables X_t which are nearly identical. To do this, we consider a subset T_1 of T, and for t in T we consider a point $\pi_1(t)$ in T_1, which we think of as a (first) approximation of t. The elements of T to which corresponds the same point $\pi_1(t)$ are, at this level of approximation, considered as identical. We then write

$$X_t - X_{t_0} = X_t - X_{\pi_1(t)} + X_{\pi_1(t)} - X_{t_0} \,. \tag{2.24}$$

The idea is that it will be effective to use (2.7) for the variables $X_{\pi_1(t)} - X_{t_0}$, because there are not too many of them, and they are rather different (at least in some global sense and if we have done a good job at finding $\pi_1(t)$). On the other hand, since $\pi_1(t)$ is an approximation of t, the variables $X_t - X_{\pi_1(t)}$ are "smaller" than the original variables $X_t - X_{t_0}$, so that their supremum should be easier to handle. The procedure will then be iterated.

Let us set up the general procedure. For $n \geq 0$, we consider a subset T_n of T, and for $t \in T$ we consider $\pi_n(t)$ in T_n. (The idea is of course that the points $\pi_n(t)$ are successive approximations of t.) We assume that T_0 consists of a single element t_0, so that $\pi_0(t) = t_0$ for each t in T. The fundamental relation is

$$X_t - X_{t_0} = \sum_{n\geq 1}\bigl(X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\bigr)\,, \tag{2.25}$$

which holds provided we arrange that $\pi_n(t) = t$ for n large enough, in which case the series is actually a finite sum. Relation (2.25) decomposes the increments of the process $X_t - X_{t_0}$ along the "chain" $(\pi_n(t))_{n\geq 0}$ (and this is why this method is called "chaining").

It will be convenient to control the set T_n through its cardinality, with the condition

$$\operatorname{card} T_n \leq N_n \tag{2.26}$$

where

$$N_0 = 1\,;\; N_n = 2^{2^n} \quad \text{if} \quad n \geq 1\,. \tag{2.27}$$

The notation (2.27) will be used throughout the book. The reader who has studied Appendix A will observe that the procedure to control T_n is rather different here. This is a crucial point of the generic chaining method.

It is good to notice right away that $\sqrt{\log N_n}$ is about $2^{n/2}$, which explains the ubiquity of this latter quantity. The occurrence of the function $\sqrt{\log x}$ itself is related to the fact that in some sense this is the inverse of the function $\exp(-x^2)$ that governs the size of the tails of a Gaussian r.v. Let us also observe the fundamental inequality

$$N_n^2 \leq N_{n+1}\,,$$

which makes it very convenient to work with this sequence.

Since $\pi_n(t)$ approximates t, it is natural to assume that

$$d(t,\pi_n(t)) = d(t,T_n) = \inf_{s\in T_n} d(t,s)\,. \tag{2.28}$$

For $u > 0$, (1.4) implies

$$\mathsf{P}\bigl(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \geq u2^{n/2} d(\pi_n(t),\pi_{n-1}(t))\bigr) \leq 2\exp(-u^2 2^{n-1})\,.$$

The number of possible pairs $(\pi_n(t),\pi_{n-1}(t))$ is bounded by

$$\operatorname{card} T_n \cdot \operatorname{card} T_{n-1} \leq N_n N_{n-1} \leq N_{n+1} = 2^{2^{n+1}}\,.$$

Thus, if we denote by Ω_u the event defined by

$$\forall n \geq 1\,,\, \forall t\,,\, |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \leq u2^{n/2} d(\pi_n(t),\pi_{n-1}(t))\,, \tag{2.29}$$

we obtain

$$\mathsf{P}(\Omega_u^c) \leq p(u) := \sum_{n\geq 1} 2\cdot 2^{2^{n+1}} \exp(-u^2 2^{n-1})\,. \tag{2.30}$$

Here again, at the crucial step, we have used the "union bound": indeed we bound the probability that one of the events (2.29) fails by the sum of the probabilities that the individual events fail. When Ω_u occurs, (2.25) yields

$$|X_t - X_{t_0}| \le u \sum_{n\ge1} 2^{n/2} d(\pi_n(t), \pi_{n-1}(t)) \,,$$

so that

$$\sup_{t\in T} |X_t - X_{t_0}| \le uS$$

where

$$S := \sup_{t\in T} \sum_{n\ge1} 2^{n/2} d(\pi_n(t), \pi_{n-1}(t)) \,.$$

Thus

$$\mathsf{P}\Big(\sup_{t\in T} |X_t - X_{t_0}| > uS\Big) \le p(u) \,.$$

For $n \ge 1$ and $u \ge 3$ we have

$$u^2 2^{n-1} \ge \frac{u^2}{2} + u^2 2^{n-2} \ge \frac{u^2}{2} + 2^{n+1} \;,$$

from which it follows that

$$p(u) \le L \exp\Big(-\frac{u^2}{2}\Big) \;.$$

We observe here that since $p(u) \le 1$ the previous inequality holds not only for $u \ge 3$ but also for $u > 0$. (This type or argument will be used repeatedly.) Therefore

$$\mathsf{P}\Big(\sup_{t\in T} |X_t - X_{t_0}| > uS\Big) \le L \exp\Big(-\frac{u^2}{2}\Big) \,. \tag{2.31}$$

In particular (2.31) implies

$$\mathsf{E} \sup_{t\in T} X_t \le LS \,.$$

The triangle inequality and (2.5) yield

$$\begin{aligned} d(\pi_n(t), \pi_{n-1}(t)) &\le d(t, \pi_n(t)) + d(t, \pi_{n-1}(t)) \\ &\le d(t, T_n) + d(t, T_{n-1}) \,, \end{aligned}$$

so that $S \le L \sup_{t\in T} \sum_{n\ge0} 2^{n/2} d(t, T_n)$, and we have proved the fundamental bound

$$\mathsf{E} \sup_{t\in T} X_t \le L \sup_{t\in T} \sum_{n\ge0} 2^{n/2} d(t, T_n) \,. \tag{2.32}$$

Now, how do we construct the sets T_n? It is obvious that we should try to make the right-hand side of (2.32) small, but this is obvious only because we have used an approach which naturally leads to this bound. The "traditional chaining method" (as used e.g. in Appendix A) chooses them so that

$$\sup_{t\in T} d(t, T_n)$$

is as small as possible for $\operatorname{card} T_n \le N_n$, where of course

$$d(t, T_n) = \inf_{s \in T_n} d(t, s) . \tag{2.33}$$

Thus we define

$$e_n(T) = e_n(T, d) = \inf \sup_t d(t, T_n) , \tag{2.34}$$

where the infimum is taken over all subsets T_n of T with $\operatorname{card} T_n \le N_n$. (Since here T is finite, the infimum is actually a minimum.) We call the numbers $e_n(T)$ the **entropy numbers**. This definition is convenient for our purposes. It is unfortunate that it is not consistent with the conventions of Operator Theory, which uses e_{2^n} to denote what we call e_n, but we can't help it if Operator Theory gets it wrong. When T is infinite, the numbers $e_n(T)$ are also defined by (2.34) but are not always finite (e.g. when $T = \mathbb{R}$).

It is good to observe that (since $N_0 = 1$) ,

$$\frac{\Delta(T)}{2} \le e_0(T) \le \Delta(T) . \tag{2.35}$$

Recalling that T is finite, let us then choose for each n a subset T_n of T with $\operatorname{card} T_n \le N_n$ and $e_n(T) = \sup_{t \in T} d(t, T_n)$. Since $d(t, T_n) \le e_n(T)$ for each t, (2.32) implies the following.

Proposition 2.2.10 (Dudley's entropy bound [2]). *Under the increment condition (1.4), we have*

$$\mathsf{E} \sup_{t \in T} X_t \le L \sum_{n \ge 0} 2^{n/2} e_n(T) . \tag{2.36}$$

We proved this bound only when T is finite, but using (2.2) it also extends to the case where T is infinite, as is shown by the following easy fact.

Lemma 2.2.11. *If U is a subset of T, we have $e_n(U) \le 2e_n(T)$.*

Proof. Indeed, if $a > e_n(T)$, by definition one can cover T by N_n balls (for the distance d) with radius a, and the intersections of these balls with U are of diameter $\le 2a$, so U can be covered by N_n balls in U with radius $2a$. □

The reader already familiar with Dudley's entropy bound might not recognize it. Usually this bound is formulated as in (1.5) using covering numbers. The covering number $N(T, d, \epsilon)$ is defined to be the smallest integer N such that there is a subset F of T, with $\operatorname{card} F \le N$ and

$$\forall t \in T \ , \ d(t, F) \le \epsilon .$$

The covering numbers relate to the entropy numbers by the formula

$$e_n(T) = \inf\{\epsilon \ ; \ N(T, d, \epsilon) \le N_n\} .$$

Indeed, it is obvious by definition of $e_n(T)$ that for $\epsilon > e_n(T)$, we have $N(T,d,\epsilon) \le N_n$, and that if $N(T,d,\epsilon) \le N_n$ we have $e_n(T) \le \epsilon$. Consequently,

$$\begin{aligned}\epsilon < e_n(T) &\Rightarrow N(T,d,\epsilon) > N_n \\ &\Rightarrow N(T,d,\epsilon) \ge 1 + N_n\ .\end{aligned}$$

Therefore

$$\sqrt{\log(1+N_n)}(e_n(T) - e_{n+1}(T)) \le \int_{e_{n+1}(T)}^{e_n(T)} \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\ .$$

Since $\log(1+N_n) \ge 2^n \log 2$ for $n \ge 0$, summation over $n \ge 0$ yields

$$\sqrt{\log 2}\sum_{n\ge 0} 2^{n/2}(e_n(T) - e_{n+1}(T)) \le \int_0^{e_0(T)} \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\ . \tag{2.37}$$

Now,

$$\begin{aligned}\sum_{n\ge 0} 2^{n/2}(e_n(T) - e_{n+1}(T)) &= \sum_{n\ge 0} 2^{n/2} e_n(T) - \sum_{n\ge 1} 2^{(n-1)/2} e_n(T) \\ &\ge \left(1 - \frac{1}{\sqrt{2}}\right)\sum_{n\ge 0} 2^{n/2} e_n(T)\,,\end{aligned}$$

so (2.37) yields

$$\sum_{n\ge 0} 2^{n/2} e_n(T) \le L\int_0^\infty \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\,.$$

Hence Dudley's bound now appears in the familiar form

$$\mathsf{E}\sup_{t\in T} X_t \le L\int_0^\infty \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon\ . \tag{2.38}$$

Of course, since $\log 1 = 0$, the integral takes place in fact over $0 \le \epsilon \le \Delta(T)$. The right-hand side is often called Dudley's entropy integral.

Exercise 2.2.12. Prove that

$$\int_0^\infty \sqrt{\log N(T,d,\epsilon)}\,\mathrm{d}\epsilon \le L\sum_{n\ge 0} 2^{n/2} e_n(T)\,,$$

showing that (2.36) is not an improvement over (2.38).

Exercise 2.2.13. Assume that for each $\epsilon > 0$ we have $\log N(t,d,\epsilon) \le (A/\epsilon)^\alpha$. Prove that $e_n(T) \le K(\alpha)A2^{-n/\alpha}$.

Here of course $K(\alpha)$ is a number depending only on α. This, and similar notation are used throughout the book. It is understood that such numbers *need not be the same on every occurrence* and it would help to **remember this at all times**. The difference between the notations K and L is that L is a universal constant, i.e. a number that does not depend on anything, while K might depend on some parameters, such as α here.

How does one estimate covering numbers (or, equivalently, entropy numbers)? The next exercise introduces the reader to "volume estimates", a simple yet fundamental method for this purpose. It deserves to be fully understood. If this exercise is too hard, you can find all the details below in the proof of Lemma 2.5.5.

Exercise 2.2.14. (a) If (T, d) is a metric space, define the packing number $N^*(T, d, \epsilon)$ as the largest integer N such that T contains N points with mutual distances $\geq \epsilon$. Prove that $N(T, d, \epsilon) \leq N^*(T, d, \epsilon)$. Prove that if $\epsilon' > 2\epsilon$ then $N^*(T, d, \epsilon') \leq N(T, d, \epsilon)$.
(b) Let us denote by d the Euclidean distance in $\mathbb{R}^m$, and by B the unit Euclidean ball of center 0. Let us denote by $\mathrm{Vol}(A)$ the m-dimensional volume of a subset A of $\mathbb{R}^m$. By comparing volumes, prove that for any subset A of $\mathbb{R}^m$,

$$N(A, d, \epsilon) \geq \frac{\mathrm{Vol}(A)}{\mathrm{Vol}(\epsilon B)} \tag{2.39}$$

and

$$N(A, d, 2\epsilon) \leq N^*(A, d, 2\epsilon) \leq \frac{\mathrm{Vol}(A + \epsilon B)}{\mathrm{Vol}(\epsilon B)} \,. \tag{2.40}$$

(c)Conclude that

$$\left(\frac{1}{\epsilon}\right)^m \leq N(B, d, \epsilon) \leq \left(\frac{2+\epsilon}{\epsilon}\right)^m \,. \tag{2.41}$$

(d) Use (c) to find estimates of $e_n(B)$ for the correct order for each value of n. (Hint: $e_n(B)$ is about $\min(1, 2^{-2^n/m})$. This decreases very fast as n increases.) Estimate Dudley's bound for B provided with the Euclidean distance.
(e) Use (c) to prove that if T is a subset of $\mathbb{R}^m$ and if n_0 is any integer such that $m2^{-n_0} \leq 1$ then for $n > n_0$ one has $e_n(T) \leq L2^{-2^n/2m} e_{n_0}(T)$. (Hint: cover T by N_{n_0} balls of radius $2e_{n_0}(T)$ and cover each of these by balls of smaller radius using (c).)
(f) This part provides a generalization of (2.39) and (2.40) to a more abstract setting, but with the same proofs. Consider a metric space (T, d) and a positive measure μ on T such all balls of a given radius have the same measure, $\mu(B(t, \epsilon)) = \varphi(\epsilon)$ for each $\epsilon > 0$ and $t \in T$. For a subset A of T and $\epsilon > 0$ let $A_\epsilon = \{t \in T ; d(t, A) \leq \epsilon\}$, where $d(t, A) = \inf_{s \in A} d(t, s)$. Prove that

$$\frac{\mu(A)}{\varphi(2\epsilon)} \leq N(A, d, 2\epsilon) \leq \frac{\mu(A_\epsilon)}{\varphi(\epsilon)} \,.$$

There are many simple situations where Dudley's bound is not of the correct order. Although this takes us a bit ahead, we give such an example in the next exercise. There the set T is particularly appealing: it is a simplex in $\mathbb{R}^m$. Another classical example which is in a sense canonical occurs on page 44. Yet other examples based on fundamental geometry (ellipsoids in $\mathbb{R}^m$) are explained in Section 2.5.

Exercise 2.2.15. Consider an integer m and an i.i.d. standard Gaussian sequence $(g_i)_{i\le m}$. For $t = (t_i)_{i\le m}$, let $X_t = \sum_{i\le m} t_i g_i$. This is called the canonical Gaussian process on $\mathbb{R}^m$. Its associated distance is the Euclidean distance on $\mathbb{R}^m$. It will be much used later. Consider the set

$$T = \Big\{ (t_i)_{i\le m}\ ;\ t_i \ge 0\ ,\ \sum_{i\le m} t_i = 1 \Big\}\ , \tag{2.42}$$

the convex hull of the canonical basis. By (2.14) we have $\mathsf{E}\sup_{t\in T} X_t = \mathsf{E}\sup_{i\le m} g_i \le L\sqrt{\log m}$. Prove that however the right-hand side of (2.36) is $\ge (\log m)^{3/2}/L$. (Hint: For an integer $k \le m$ consider the subset T_k of T consisting of sequences $t = (t_i)_{i\le m} \in T$ for which $t_i \in \{0, 1/k\}$. Using part (f) of Exercise 2.2.14 with $T = A = T_k$ and μ the counting measure prove that $\log N(T_k, d, 1/(L\sqrt{k})) \ge k\log(em/k)/L$ and conclude. You need to be fluent with Stirling's formula to succeed.) Thus in this case Dudley's bound is off by a factor about $\log m$. Exercise 2.3.4 below will show that in $\mathbb{R}^m$ the situation cannot be worse than this.

The bound (2.32) seems to be genuinely better than the bound (2.36) because when going from (2.32) to (2.36) we have used the somewhat brutal inequality

$$\sup_{t\in T} \sum_{n\ge 0} 2^{n/2} d(t, T_n) \le \sum_{n\ge 0} 2^{n/2} \sup_{t\in T} d(t, T_n)\,.$$

The method leading to the bound (2.32) is probably the most important idea of this work. Of course the fact that it appears now so naturally does not reflect the history of the subject, but rather that the proper approach is being used. When using this bound, we will choose the sets T_n in order to minimize the right-hand side of (2.32) instead of choosing them as in (2.34). The true importance of this procedure is that as will be demonstrated later, this provides essentially the best possible bound for $\mathsf{E}\sup_{t\in T} X_t$. To understand that matters are not trivial, the reader should try, in the situation of Exercise 2.2.15, to find sets T_n such that the right-hand side of (2.32) is of the correct order $\sqrt{\log m}$. It would probably be quite an athletic feat to succeed at this stage, but the reader is encouraged to keep this question in mind as her understanding deepens.

The next exercise provides a simple (and somewhat "extremal") situation showing that (2.32) is an actual improvement over (2.36).

Exercise 2.2.16. (a) Consider a finite metric space (T,d). Assume that it contains a point t_0 with the property that for $n \geq 0$ we have $\operatorname{card}(T \setminus B(t_0, 2^{-n/2})) \leq N_n - 1$. Prove that T contains sets T_n with $\operatorname{card} T_n \leq N_n$ and $\sup_{t\in T}\sum_{n\geq 0} 2^{n/2} d(t,T_n) \leq L$. (Hint: $T_n = \{t_0\} \cup \{t \in T; d(t,t_0) > 2^{-n/2}\}$.)
(b) Given an integer $s \geq 10$, construct a finite metric space (T,d) with the above property, such that $\operatorname{card} T \leq N_s$ and that $e_n(T) \geq 2^{-n/2}/L$ for $1 \leq n \leq s-1$, so that Dudley's integral is of order s. (Hint: this might be hard if you really never though about metric spaces. Try then a set of the type $T = \{a_\ell f_\ell ; \ell \leq M\}$ where $a_\ell > 0$ is a number and $(f_\ell)_{\ell \leq M}$ is the canonical basis of $\mathbb{R}^M$.)

It turns out that the idea behind the bound (2.32) admits a technically more convenient formulation.

Definition 2.2.17. *Given a set T an* admissible sequence *is an increasing sequence $(\mathcal{A}_n)$ of partitions of T such that $\operatorname{card}\mathcal{A}_n \leq N_n$, i.e. $\operatorname{card}\mathcal{A}_0 = 1$ and $\operatorname{card}\mathcal{A}_n \leq 2^{2^n}$ for $n \geq 1$.*

By an increasing sequence of partitions we mean that every set of $\mathcal{A}_{n+1}$ is contained in a set of $\mathcal{A}_n$. Throughout the book we denote by $A_n(t)$ the unique element of $\mathcal{A}_n$ which contains t. The double exponential in the definition of (2.27) of N_n occurs simply since for our purposes the proper measure of the "size" of a partition $\mathcal{A}$ is $\log\operatorname{card}\mathcal{A}$. This double exponential ensures that "the size of the partition $\mathcal{A}_n$ doubles at every step". This offers a number of technical advantages which will become clear gradually.

Theorem 2.2.18 (The generic chaining bound). *Under the increment condition (1.4) (and if $\mathsf{E}X_t = 0$ for each t) then for each admissible sequence $(\mathcal{A}_n)$ we have*

$$\mathsf{E}\sup_{t\in T} X_t \leq L \sup_{t\in T}\sum_{n\geq 0} 2^{n/2}\Delta(A_n(t)) \; . \tag{2.43}$$

Here of course, as always, $\Delta(A_n(t))$ denotes the diameter of $A_n(t)$ for d. One could think that (2.43) could be much worse than (2.32), but is will turn out that this is not the case when the sequence $(\mathcal{A}_n)$ is appropriately chosen.

Proof. We may assume T to be finite. We construct a subset T_n of T by taking exactly one point in each set A of $\mathcal{A}_n$. Then for $t \in T$ and $n \geq 0$, we have $d(t,T_n) \leq \Delta(A_n(t))$ and the result follows from (2.32). □

Definition 2.2.19. *Given $\alpha > 0$, and a metric space (T,d) (that need not be finite) we define*

$$\gamma_\alpha(T,d) = \inf \sup_{t\in T}\sum_{n\geq 0} 2^{n/\alpha}\Delta(A_n(t)) \, ,$$

where the infimum is taken over all admissible sequences.

It is useful to observe that since $A_0(t) = T$ we have $\gamma_\alpha(T,d) \geq \Delta(T)$.

Exercise 2.2.20. (a) Prove that if $d \le d'$ then $\gamma_2(T,d) \le \gamma_2(T,d')$.
(b) More generally prove that if $d \le Bd'$ then $\gamma_2(T,d) \le B\gamma_2(T,d')$.

Exercise 2.2.21. (a) If T is finite, prove that $\gamma_2(T,d) \le L\Delta(T)\sqrt{\log \operatorname{card} T}$. (Hint: Ensure that $\Delta(A_n(t)) = 0$ if $N_n \ge \operatorname{card} T$.)
(b) Prove that for $n \ge 0$ we have

$$2^{n/2} e_n(T) \le L\gamma_2(T,d) \ . \tag{2.44}$$

(Hint: observe that $2^{n/2} \max\{\Delta(A); A \in \mathcal{A}_n\} \le \sup_{t \in T} \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t))$.)
(c) Prove that, equivalently, for $\epsilon > 0$ we have

$$\epsilon\sqrt{\log N(T,d,\epsilon)} \le L\gamma_2(T,d) \ .$$

The reader should compare (2.44) with Corollary 2.3.2 below.

Combining Theorem 2.2.18 with Definition 2.2.19 yields

Theorem 2.2.22. *Under (1.4) and (2.1) we have*

$$\mathsf{E} \sup_{t \in T} X_t \le L\gamma_2(T,d) \ . \tag{2.45}$$

Of course to make (2.45) of interest we must be able to control $\gamma_2(T,d)$, i.e. we must learn how to construct admissible sequences, a topic we shall first address in Section 2.3.

Let us also point out, recalling (2.31), and observing that

$$|X_s - X_t| \le |X_s - X_{t_0}| + |X_t - X_{t_0}| \ , \tag{2.46}$$

we have actually proved

$$\mathsf{P}\Big(\sup_{s,t \in T} |X_s - X_t| \ge Lu\gamma_2(T,d)\Big) \le 2\exp(-u^2) \ . \tag{2.47}$$

There is no reason other than the author's fancy to feature the phantom coefficient 1 in the exponent of the right-hand side, but it might be good at this stage for the reader to write every detail on how this is deduced from (2.31). The different exponents in (2.31) and (2.47) are of course made possible by the fact that the constant L is not the same in these inequalities.

We note that (2.47) implies a lot more than (2.45). Indeed, for each $p \ge 1$, using (2.22)

$$\mathsf{E}\big(\sup_{s,t} |X_s - X_t|\big)^p \le K(p)\gamma_2(T,d)^p \ , \tag{2.48}$$

and in particular

$$\mathsf{E}\big(\sup_{s,t} |X_s - X_t|\big)^2 \le L\gamma_2(T,d)^2 \ . \tag{2.49}$$

One of the ideas underlying Definition 2.2.19 is that partitions of T are really handy. For example, given a partition $\mathcal{B}$ of T whose elements are "small" for a certain distance d_1 and a partition $\mathcal{C}$ whose elements are "small" for another distance d_2, then the elements of the partition generated by $\mathcal{B}$ and $\mathcal{C}$, i.e. the partition which consists of the sets $B \cap C$ for $B \in \mathcal{B}$ and $C \in \mathcal{C}$, are "small" for both d_1 and d_2. This is illustrated in the proof of the following theorem, which applies to processes with a weaker tail condition than (1.4). This theorem will be used many times (the reason being that a classical inequality of Bernstein naturally produces tail conditions such as in (2.50)).

Theorem 2.2.23. *Consider a set T provided with two distances d_1 and d_2. Consider a centered process $(X_t)_{t \in T}$ which satisfies*

$$\forall s, t \in T \,,\ \forall u > 0 \,,$$

$$\mathsf{P}(|X_s - X_t| \geq u) \leq 2 \exp\left(-\min\left(\frac{u^2}{d_2(s,t)^2}, \frac{u}{d_1(s,t)}\right)\right) . \tag{2.50}$$

Then

$$\mathsf{E} \sup_{s,t \in T} |X_s - X_t| \leq L(\gamma_1(T, d_1) + \gamma_2(T, d_2)) . \tag{2.51}$$

This theorem will be applied when d_2 is the ℓ_2 distance and d_1 is the ℓ_∞ distance (but it sounds funny, when considering two distances, to call them d_2 and d_∞).

Proof. We denote by $\Delta_j(A)$ the diameter of the set A for d_j. We consider an admissible sequence $(\mathcal{B}_n)_{n \geq 0}$ such that

$$\forall t \in T \,,\ \sum_{n \geq 0} 2^n \Delta_1(B_n(t)) \leq 2\gamma_1(T, d_1) \tag{2.52}$$

and an admissible sequence $(\mathcal{C}_n)_{n \geq 0}$ such that

$$\forall t \in T \,,\ \sum_{n \geq 0} 2^{n/2} \Delta_2(C_n(t)) \leq 2\gamma_2(T, d_2) . \tag{2.53}$$

Of course here $B_n(t)$ is the unique element of $\mathcal{B}_n$ that contains t (etc.). We define partitions $\mathcal{A}_n$ of T as follows. We set $\mathcal{A}_0 = \{T\}$, and, for $n \geq 1$, we define $\mathcal{A}_n$ as the partition generated by $\mathcal{B}_{n-1}$ and $\mathcal{C}_{n-1}$, i.e. the partition that consists of the sets $B \cap C$ for $B \in \mathcal{B}_{n-1}$ and $C \in \mathcal{C}_{n-1}$. Thus

$$\operatorname{card} \mathcal{A}_n \leq N_{n-1}^2 \leq N_n \,,$$

and the sequence $(\mathcal{A}_n)$ is admissible. (Let us repeat here that the fundamental inequality $N_n^2 \leq N_{n+1}$ is the reason why it is so convenient to work with the sequence N_n.) For each $n \geq 0$ let us consider a set T_n that intersects each element of $\mathcal{A}_n$ in exactly one point, and for $t \in T$ let us denote by $\pi_n(t)$ the element of T_n that belongs to $A_n(t)$. To use (2.50) we observe that for $v > 0$ it implies

$$\mathsf{P}(|X_s - X_t| \geq v d_1(s,t) + \sqrt{v} d_2(s,t)) \leq 2\exp(-v) ,$$

and thus, given $u \geq 1$, we have, since $u \geq \sqrt{u}$,

$$\mathsf{P}\Big(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \geq u\big(2^n d_1(\pi_n(t), \pi_{n-1}(t)) + 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t))\big)\Big) \\ \leq 2\exp(-u2^n) , \tag{2.54}$$

so that, proceeding as in (2.30), with probability $\geq 1 - L\exp(-u)$ we have

$$\forall n \ , \ \forall t \ , \ |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \leq u\big(2^n d_1(\pi_n(t), \pi_{n-1}(t)) \\ + 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t))\big) . \tag{2.55}$$

Now, under (2.55) we get

$$\sup_{t\in T} |X_t - X_{t_0}| \leq u \sup_{t\in T} \sum_{n\geq 1} \big(2^n d_1(\pi_n(t), \pi_{n-1}(t)) + 2^{n/2} d_2(\pi_n(t), \pi_{n-1}(t))\big) .$$

When $n \geq 2$ we have $\pi_n(t), \pi_{n-1}(t) \in A_{n-1}(t) \subset B_{n-2}(t)$, so that

$$d_1(\pi_n(t), \pi_{n-1}(t)) \leq \Delta_1(B_{n-2}(t)) .$$

Hence, since $d_1(\pi_1(t), \pi_0(t)) \leq \Delta_1(B_0(t)) = \Delta_1(T)$, using (2.52) in the last inequality, (and remembering that the value of L need not be the same on each occurrence)

$$\sum_{n\geq 1} 2^n d_1(\pi_n(t), \pi_{n-1}(t)) \leq L \sum_{n\geq 0} 2^n \Delta_1(B_n(t)) \leq 2L\gamma_1(T,d) = L\gamma_1(T,d) .$$

Proceeding similarly for d_2 shows that under (2.55) we obtain

$$\sup_{s,t\in T} |X_t - X_{t_0}| \leq Lu(\gamma_1(T,d_1) + \gamma_2(T,d_2)) ,$$

and therefore using (2.46),

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \geq Lu(\gamma_1(T,d_1) + \gamma_2(T,d_2))\Big) \leq L\exp(-u) , \tag{2.56}$$

which using (2.5) implies the result. □

Exercise 2.2.24. Consider a space T equipped with two different distances d_1 and d_2. Prove that

$$\gamma_2(T, d_1 + d_2) \leq L(\gamma_2(T,d_1) + \gamma_2(T,d_2)) . \tag{2.57}$$

(Hint: given an admissible sequence of partitions $\mathcal{A}_n$ (resp. $\mathcal{B}_n$) which behaves well for d_1 (resp. d_2) consider as in the beginning of the proof of Theorem 2.2.23 the partitions generated by $\mathcal{A}_n$ and $\mathcal{B}_n$.)

Exercise 2.2.25 (R. Latała, S. Mendelson). Consider a process $(X_t)_{t\in T}$ and for a subset A of T and $n \ge 0$ let

$$\Delta_n(A) = \sup_{s,t\in A} (\mathsf{E}|X_s - X_t|^{2^n})^{2^{-n}} .$$

Consider an admissible sequence of partitions $(\mathcal{A}_n)_{n\ge 0}$.
(a) Prove that

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \le \sup_{t\in T} \sum_{n\ge 0} \Delta_n(A_n(t)) .$$

(Hint: Use chaining and (A.11) for $\varphi(x) = x^{2^n}$.)
(b) Explain why this result implies Theorem 2.2.23. (Hint: Use Exercise 2.2.8.)

The following exercise assumes that you are familiar with the contents of Appendix B. It develops the theme of "chaining with varying distances" of Exercise 2.2.25 in a different direction. Variations on this idea will turn out later to be fundamental.

Exercise 2.2.26. Assume that for $n \ge 0$ we are given a distance d_n on T and a convex function φ_n with $\varphi_n(0) = 0$, $\varphi_n(x) = \varphi_n(-x) \ge 0$. Assume that

$$\forall\, s,t \in T\ ,\ \mathsf{E}\varphi_n\Big(\frac{X_s - X_t}{d_n(s,t)}\Big) \le 1 .$$

Consider a sequence $\epsilon_n > 0$ and assume that $N(T, d_0, \epsilon_0) = 1$. Prove that

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \le \sum_{n\ge 0} \epsilon_n \varphi_n^{-1}(N(T, d_n, \epsilon_n)) .$$

Prove that this implies Theorem B.2.3. (Hint: simple modification of the argument of Theorem B.2.3.)

We now prove some more specialized results, which may be skipped at first reading. This is all the more the case since for many processes of importance the machinery of "concentration of measure" allows one to find very competent bounds for the quantity $\mathsf{P}(|\sup_{t\in T} X_t - \mathsf{E}\sup_{t\in T} X_t| \ge u)$. For example in the case of Gaussian processes, (2.58) below is a consequence of (2.96) and (2.45). The point of (2.58) is that it improves on (2.47) using only the increment condition (1.4).

Theorem 2.2.27. *If the process (X_t) satisfies (1.4) then for $u > 0$ one has*

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge L(\gamma_2(T,d) + u\Delta(T))\Big) \le L\exp(-u^2) . \tag{2.58}$$

Proof. This is one of a very few instances where one must use some care when using the generic chaining. Consider an admissible sequence $(\mathcal{A}_n)$ of partitions with $\sup_{t\in T}\sum_{n\geq 0} 2^{n/2}\Delta(A_n(t)) \leq 2\gamma_2(T,d)$ and for each n consider a set T_n with card $T_n \leq N_n$ and such that every element A of $\mathcal{A}_n$ meets T_n. Let $U_n = \cup_{q\leq n} T_q$, so that $U_0 = T_0$ and $\operatorname{card} U_n \leq 2N_n$, and the sequence (U_n) increases. For $u > 0$ consider the event $\Omega(u)$ given by

$$\forall\, n \geq 1\ ,\ \forall\, s,t \in U_n\ ,\ |X_s - X_t| \leq 2(2^{n/2}+u)d(s,t)\ , \tag{2.59}$$

so that (somewhat crudely)

$$\mathsf{P}(\Omega^c(u)) \leq 2\sum_{n\geq 1}(\operatorname{card} U_n)^2 \exp(-2(2^n+u^2)) \leq L\exp(-2u^2)\ . \tag{2.60}$$

Consider now $t \in T$. We define by induction over $q \geq 0$ integers $n(t,q)$ as follows. We start with $n(t,0)=0$, and for $q \geq 1$ we define

$$n(t,q) = \inf\Big\{n\ ;\ n \geq n(t,q-1)\ ;\ d(t,U_n) \leq \frac{1}{2} d(t,U_{n(t,q-1)})\Big\}\ . \tag{2.61}$$

We then consider $\pi_q(t) \in U_{n(t,q)}$ with $d(t,\pi_q(t)) = d(t,U_{n(t,q)})$. Thus, by induction, and denoting by t_0 the unique element of $T_0 = U_0$, for $q \geq 0$, it holds

$$d(t,\pi_q(t)) \leq 2^{-q} d(t,t_0) \leq 2^{-q}\Delta(T)\ . \tag{2.62}$$

Also, when $\Omega(u)$ occurs, using (2.59) for $n = n(t,q)$, and since $\pi_q(t) \in U_n$ and $\pi_{q-1}(t) \in U_{n(t,q-1)} \subset U_n$,

$$|X_{\pi_q(t)} - X_{\pi_{q-1}(t)}| \leq 2(2^{n(t,q)/2}+u)d(\pi_q(t),\pi_{q-1}(t))\ .$$

Assuming that $\Omega(u)$ occurs, we thus obtain

$$\begin{aligned}
|X_t - X_{t_0}| &\leq \sum_{q\geq 1} |X_{\pi_q(t)} - X_{\pi_{q-1}(t)}| \\
&\leq \sum_{q\geq 1} 2(2^{n(t,q)/2}+u)d(\pi_q(t),\pi_{q-1}(t)) \\
&\leq \sum_{q\geq 1} 2(2^{n(t,q)/2}+u)d(t,\pi_q(t)) \\
&\quad + \sum_{q\geq 1} 2(2^{n(t,q)/2}+u)d(t,\pi_{q-1}(t))\ .
\end{aligned} \tag{2.63}$$

We now control the four summations on the right-hand side. First,

$$\begin{aligned}
\sum_{q\geq 1} 2^{n(t,q)/2} d(t,\pi_q(t)) &\leq \sum_{q\geq 1} 2^{n(t,q)/2} d(t,T_{n(t,q)}) \\
&\leq \sum_{n\geq 0} 2^{n/2} d(t,T_n) \leq 2\gamma_2(T,d)\ ,
\end{aligned}$$

since by definition of $n(t,q)$ we have $n(t,q) > n(t,q-1)$ unless $d(t,T_{n(t,q-1)}) = 0$. Now, the definition of $n(t,q)$ implies

$$d(t,\pi_{q-1}(t)) = d(t,U_{n(t,q-1)}) \le 2d(t,U_{n(t,q)-1}) ,$$

so that

$$\sum_{q\ge1} 2^{n(t,q)/2} d(t,\pi_{q-1}(t)) \le 2\sum_{q\ge1} 2^{n(t,q)/2} d(t,T_{n(t,q)-1}) ,$$

and as above this is $\le L\gamma_2(T,d)$. Next, (2.62) implies $\sum_{q\ge1} d(\pi_q(t),t) \le 2\Delta(T)$ and $\sum_{q\ge1} d(\pi_{q-1}(t),t) \le 2\Delta(T)$. In summary, when $\Omega(u)$ occurs, we have $|X_t - X_{t_0}| \le L(\gamma_2(T,d) + u\Delta(T))$. □

One idea underlying the proof (and in particular the definition (2.61) of $n(t,q)$) is that for an efficient chaining the distance $d(\pi_n(t),\pi_{n+1}(t))$ decreases geometrically. In Chapter 15 we shall later see situations where this is not the case.

We will at times need the following more precise version of (2.56), in the spirit of Theorem 2.2.27.

Theorem 2.2.28. *Under the conditions of Theorem 2.2.23, for all values* $u_1, u_2 > 0$ *we have*

$$\mathsf{P}\Big(\sup_{s,t\in T} |X_s - X_t| \ge L\big(\gamma_1(T,d_1) + \gamma_2(T,d_2) + u_1 D_1 + u_2 D_2\big)\Big) \le L\exp(-\min(u_2^2, u_1)) , \tag{2.64}$$

where for $j = 1,2$ *we set* $D_j = \sum_{n\ge0} e_n(T,d_j)$.

We observe from (2.44) that $e_n(T,d_2) \le L2^{-n/2}\gamma_2(T,d_2)$ so that by summation $D_2 \le L\gamma_2(T,d_2)$ and similarly for D_1. Thus (2.64) recovers (2.56). Moreover, in many practical situations, one has $D_j \le Le_0(T,d_j) \le L\Delta_j(T) = L\Delta(T,d_j)$. Still, the occurrence of the unwieldy quantities D_j makes the statement of Theorem 2.2.28 a bit awkward. It would be pleasing if in the statement of this theorem one could replace D_j by the smaller quantity $\Delta_j(T)$. Unfortunately this does not seem to be true. The reader might like to consider the case where card $T = N_n$ and $d_1(s,t) = 1$ for $s \ne t$ to understand where the difficulty lies.

Proof. There exists a partition $\mathcal{U}_n$ of T into N_n sets, each of which have a diameter $\le 2e_n(T,d_1)$ for d_1. Consider the partition $\mathcal{B}'_n$ generated by $\mathcal{U}_0,\dots,\mathcal{U}_{n-1}$. These partitions form an admissible sequence such that

$$\forall B \in \mathcal{B}'_n ,\ \Delta_1(B) \le 2e_{n-1}(T,d_1) . \tag{2.65}$$

Let us also consider an admissible sequence $(\mathcal{C}'_n)$ which has the same property for d_2,

$$\forall C \in \mathcal{C}'_n \, , \; \Delta_2(C) \le 2e_{n-1}(T, d_2) \; .$$

We define $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$, and for $n \ge 2$ we define $\mathcal{A}_n$ as being the partition generated by $\mathcal{B}_{n-2}$, $\mathcal{B}'_{n-2}$, $\mathcal{C}_{n-2}$ and $\mathcal{C}'_{n-2}$, where $\mathcal{B}_n$ and $\mathcal{C}_n$ are as in (2.52) and (2.53) respectively. Let us define a chaining $\pi_n(t)$ associated as usual to the sequence $(\mathcal{A}_n)$ of partitions. (That is we select a set T_n which meets every element of $\mathcal{A}_n$ in exactly one point, and $\pi_n(t)$ denote the element of T_n which belongs to $A_n(t)$.)

$$U = (2^n + u_1) d_1(\pi_n(t), \pi_{n-1}(t)) + (2^{n/2} + u_2) d_2(\pi_n(t), \pi_{n-1}(t)) \; ,$$

so that (2.50) implies somewhat crudely that

$$\mathsf{P}(|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| \ge U) \le 2\exp(-2^n - \min(u_2^2, u_1)) \; .$$

For $n \ge 3$ we have $\pi_n(t), \pi_{n-1}(t) \in B_{n-3}(t)$, so that $d_1(\pi_n(t), \pi_{n-1}(t)) \le \Delta_1(B_{n-3}(t))$, and $\pi_n(t), \pi_{n-1}(t) \in B'_{n-3}(t)$ so that, using (2.65) in the last inequality,

$$d_1(\pi_n(t), \pi_{n-1}(t)) \le \Delta_1(B'_{n-3}(t)) \le 2e_{n-3}(T, d_1) \; .$$

Proceeding in the same fashion for d_2 it follows that with probability at least $1 - L\exp(-\min(u_2^2, u_1))$ we have

$$\begin{aligned} \forall n \ge 3 \, , \; \forall t \in T \, , \; |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| &\le 2^n \Delta_1(B_{n-3}(t)) + 2^{n/2} \Delta_2(C_{n-3}(t)) \\ &\quad + 2u_1 e_{n-3}(T, d_1) + 2u_2 e_{n-3}(T, d_2) \; . \end{aligned}$$

This inequality remains true for $n = 1, 2$ if in the right-hand side one replaces $n - 3$ by 0, and chaining (i.e. use of (2.25)) completes the proof. □

2.3 Functionals

To make Theorem 2.2.18 useful, we must be able to construct good admissible sequences. In this section we explain our basic method. This method, and its variations, are at the core of the book.

Let us recall that we have defined $\gamma_a(T, d)$ as

$$\gamma_\alpha(T, d) = \inf \sup_{t \in T} \sum_{n \ge 0} 2^{n/\alpha} \Delta(A_n(t))$$

where the infimum is taken over all admissible sequences $(\mathcal{A}_n)$ of partitions of T. Let us now define the quantity

$$\gamma^*_\alpha(T, d) = \inf \sup_{t \in T} \sum_{n \ge 0} 2^{n/\alpha} d(t, T_n) \, ,$$

where the infimum is over all choices of the sets T_n with $\operatorname{card} T_n \le N_n$.

It is rather obvious that $\gamma^*_\alpha(T,d) \le \gamma_\alpha(T,d)$. To prove this, consider an admissible sequence $(\mathcal{A}_n)$ of partitions of T. Choose T_n such that each set of $\mathcal{A}_n$ contains one element of T_n. Then for each $t \in T$

$$\sum_{n\ge0} 2^{n/\alpha} d(t,T_n) \le \sum_{n\ge0} 2^{n/\alpha} \Delta(A_n(t)) \ ,$$

and this proves the claim. This is simply the argument of Theorem 2.2.18. Now, we would like to go the other way, that is to prove

$$\gamma_\alpha(T,d) \le K(\alpha)\gamma^*_\alpha(T,d) \ . \tag{2.66}$$

This is achieved by the following result.

Theorem 2.3.1. *Consider a metric space (T,d), an integer $\tau' \ge 0$ and for $n \ge 0$, consider subsets T_n of T with* $\operatorname{card} T_0 = 1$ *and* $\operatorname{card} T_n \le N_{n+\tau'} = 2^{2^{n+\tau'}}$ *for $n \ge 1$. Consider numbers $\alpha > 0$, $S > 0$, and let*

$$U = \Big\{ t \in T \, ; \sum_{n\ge0} 2^{n/\alpha} d(t,T_n) \le S \Big\} \ .$$

Then $\gamma_\alpha(U,d) \le K(\alpha,\tau')S$.

Of course here $K(\alpha,\tau')$ denotes a number depending on α and τ' only. When $U = T$ and $\tau' = 0$, this proves (2.66), and shows that the bound (2.43) is as good as the bound (2.32), if one does not mind the possible loss of a constant factor. The superiority of the bound (2.43) is that it uses admissible sequences, and as explained before Theorem 2.2.23 these are very convenient.

It is also good to observe that Theorem 2.3.1 allows us to control $\gamma_\alpha(U,d)$ using sets T_n that need not be subsets of U.

It seems appropriate to state the following obvious consequence of (2.66).

Corollary 2.3.2. *For any metric space (T,d) we have*

$$\gamma_\alpha(T,d) \le K(\alpha) \sum_{n\ge0} 2^{n/\alpha} e_n(T) \ .$$

Exercise 2.3.3. Find a simple direct proof of Corollary 2.3.2. (Hint. You do have to construct the partitions. If this is too difficult, try first to read the proof of Theorem 2.3.1, and simplify it suitably.)

Exercise 2.3.4. Use (2.44) and Exercise 2.2.14 (d) to prove that if $T \subset \mathbb{R}^m$ then

$$\sum_{n\ge0} 2^{n/2} e_n(T) \le L \log(m+1)\gamma_2(T,d) \ . \tag{2.67}$$

In words, Dudley's bound is never off by more than a factor about $\log(m+1)$ in $\mathbb{R}^m$.

The following simple observation allows one to construct a sequence which is admissible from one which is slightly too large. It will be used a great many times.

Lemma 2.3.5. *Consider* $\alpha > 0$, *an integer* $\tau \geq 0$ *and an increasing sequence of partitions* $(\mathcal{B}_n)_{n\geq 0}$ *with* $\operatorname{card}\mathcal{B}_n \leq N_{n+\tau}$. *Let*

$$S := \sup_{t\in T}\sum_{n\geq 0} 2^{n/\alpha}\Delta(B_n(t)) .$$

Then we can find an admissible sequence $(\mathcal{A}_n)_{n\geq 0}$ *such that*

$$\sup_{t\in T}\sum_{n\geq 0} 2^{n/\alpha}\Delta(A_n(t)) \leq 2^{\tau/\alpha}(S + K(\alpha)\Delta(T)) . \tag{2.68}$$

Of course (for the last time) here $K(\alpha)$ denotes a number depending on α only (that need not be the same at each occurrence).

Proof. We set $\mathcal{A}_n = \{T\}$ if $n < \tau$ and $\mathcal{A}_n = \mathcal{B}_{n-\tau}$ if $n \geq \tau$ so that $\operatorname{card}\mathcal{A}_n \leq N_n$ and

$$\sum_{n\geq\tau} 2^{n/\alpha}\Delta(A_n(t)) = 2^{\tau/\alpha}\sum_{n\geq 0} 2^{n/\alpha}\Delta(B_n(t)) .$$

Using the bound $\Delta(A_n(t)) \leq \Delta(T)$, we obtain

$$\sum_{n\leq\tau} 2^{n/\alpha}\Delta(A_n(t)) \leq K(\alpha)2^{\tau/\alpha}\Delta(T) .$$

□

Exercise 2.3.6. Prove that (2.68) might fail if one replaces the right-hand side by $K(\alpha,\tau)S$. (Hint: S does not control $\Delta(T)$.)

Proof of Theorem 2.3.1. There is no other way than to roll up our sleeves and actually construct a partition. For $u \in T_n$, let

$$V(u) = \{t \in U\,;\, d(t,T_n) = d(t,u)\} .$$

(This is a well known construction, the sets $V(u)$ are simply the closures of the Voronoï cells associated to the points of T_n.) The sets $V(u)$ cover U i.e. $U = \bigcup_{u\in T_n} V(u)$, but they are not disjoint. First find a partition $\mathcal{C}_n$ of U, with $\operatorname{card}\mathcal{C}_n \leq N_{n+\tau'}$, and the property that

$$\forall C \in \mathcal{C}_n\,,\ \exists u \in T_n\,,\ C \subset V(u) .$$

This cannot be the partitions we are looking for since the sequence $(\mathcal{C}_n)$ need not be increasing. A more serious problem is that for $t \in V(u)$ it might happen that $d(t,T_n) \ll \Delta(V(u))$, and hence that $\Delta(C_n(t)) \gg d(t,T_n)$, in which case we have no control over $\Delta(C_n(t))$. To alleviate this problem, we will suitably break the sets of $\mathcal{C}_n$ into smaller pieces. Consider C as above, let b be the smallest integer $b > 1/\alpha + 1$, and consider the set

$$C_{bn} = \{t \in C \, ; \, d(t,u) \le 2^{-bn} \Delta(U)\} \, ,$$

so that $\Delta(C_{bn}) \le 2^{-bn+1}\Delta(U)$. Similarly, consider, for $0 \le k < bn$, the set

$$C_k = \{t \in C \, ; \, 2^{-k-1}\Delta(U) < d(t,u) \le 2^{-k}\Delta(U)\} \, .$$

Thus $\Delta(C_k) \le 2^{-k+1}\Delta(U)$, and, when $k < bn$, and since $C \subset V(u)$,

$$\forall t \in C_k \, , \; \Delta(C_k) \le 2^{-k+1}\Delta(U) \le 4d(t,u) \le 4d(t,T_n) \, .$$

Therefore,

$$\forall k \le bn \, , \, \forall t \in C_k \, , \; \Delta(C_k) \le 4d(t,T_n) + 2^{-bn+1}\Delta(U) \, . \tag{2.69}$$

Consider the partition $\mathcal{B}_n$ consisting of the sets C_k for $C \in \mathcal{C}_n$, $0 \le k \le bn$, so that $\operatorname{card} \mathcal{B}_n \le (bn+1)N_{n+\tau'}$. Consider the partition $\mathcal{A}_n$ generated by $\mathcal{B}_0$, … , $\mathcal{B}_n$, so that the sequence $(\mathcal{A}_n)$ increases, and $\operatorname{card} \mathcal{A}_n \le N_{n+\tau}$, where τ depends on α and τ' only. (The reader is advised to work out this fact in complete detail.) From (2.69) we get

$$\forall A \in \mathcal{A}_n \, , \, \forall t \in A \, , \; \Delta(A) \le 4d(t,T_n) + 2^{-bn+1}\Delta(U) \, ,$$

and thus

$$\begin{aligned}\sum_{n\ge 0} 2^{n/\alpha}\Delta(A_n(t)) &\le 4\sum_{n\ge 0} 2^{n/\alpha} d(t,T_n) + \Delta(U)\sum_{n\ge 0} 2^{n/\alpha - bn+1} \\ &\le 4(S + \Delta(U)) \, .\end{aligned}$$

Since for t in U we have $d(t,T_0) \le S$ where T_0 contains a unique point, we have $\Delta(U) \le 2S$, and the conclusion follows from Lemma 2.3.5. □

Exercise 2.3.7. In Theorem 2.3.1, carry out the correct dependence of $K(\alpha,\tau')$ upon τ'.

Let us now explain the crucial idea of functionals (and the reason behind the name). We will say that a map F is a *functional* on a set T if, to each subset H of T it associates a number $F(H) \ge 0$, and if it is increasing, i.e.

$$H \subset H' \subset T \Rightarrow F(H) \le F(H') \, . \tag{2.70}$$

Intuitively a functional is a measure of "size" for the subsets of T. It allows to identify which subsets of T are "large" for our purposes. Suitable partitions of T will then be constructed through an exhaustion procedure that selects first the large subsets of T.

When reading the words "measure of the size of a subset of T" the reader might form the picture of the functional $F(H) = \mu(H)$ where μ is a measure on T. For our purposes, this picture is incorrect, because our goal is to understand in a sense what are the smallest functionals which satisfy a certain property to be explained below, and these do not look at all like the measure μ example. A first fundamental example of a functional is

$$F(H) = \gamma_2(H, d) \,. \tag{2.71}$$

A second, equally important, is the quantity

$$F(H) = \mathsf{E} \sup_{t \in H} X_t$$

where $(X_t)_{t \in T}$ is a process indexed by T.

Now we wish to explain the basic property needed for a functional. That this property is relevant is by no means intuitively obvious yet (but we shall soon see that the functional (2.71) does enjoy this property). Let us first try it in words: if we consider a set that is the union of many small pieces far enough from each other, then this set is significantly larger (as measured by the functional) than the *smallest* of its pieces. “Significantly larger” depends on the scale of the pieces, and on their number. This is a kind of “growth condition”.

First, let us explain what we mean by “small pieces far from each other”. There is a scale, say $a > 0$ at which this happens. The pieces are small at that scale: they are contained in balls with radius $a/100$. The balls are far from each other: any two centers of such balls are at mutual distance $\geq a$. Wouldn't you say that such pieces are “well separated”? Of course there is nothing specific about the choice of the radius $a/100$, and sometimes the radius has to be smaller, so we introduce a parameter $r \geq 4$, and we ask that the “small pieces” be contained in balls with radius a/r rather than $a/100$. The reason why we require $r \geq 4$ is that we want the following: two points taken in different balls with radius a/r whose centers are at distance $\geq a$ cannot be too close to each other. This would not be true for $r = 2$, so we give ourselves some room, and take $r \geq 4$. Here is the formal definition.

Definition 2.3.8. *Given $a > 0$ and an integer $r \geq 4$ we say that subsets $H_1, \ldots, H_m$ of T are (a, r)-separated if*

$$\forall \ell \leq m \,,\ H_\ell \subset B(t_\ell, a/r) \,, \tag{2.72}$$

where the points $t_1, t_2, \ldots, t_m$ in T satisfy

$$\forall \ell \leq m \,,\ t_\ell \in B(s, ar) \ ;\ \forall \ell \,,\ \ell' \leq m \,,\ \ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a \tag{2.73}$$

for a certain point $s \in T$.

Of course here $B(s, a)$ denotes the closed ball with center s and radius a in the metric space (T, d). A secondary feature of this definition is that the small pieces H_ℓ are not only well separated (on a scale a), but they are in the “same region of T” (on the larger scale ra). This is the content of the first part of condition (2.73):

$$\forall \ell \leq m \,,\ t_\ell \in B(s, ar) \,.$$

Exercise 2.3.9. Find interesting examples of metric spaces for which there are no points $t_1, \ldots, t_m$ as in (2.73), for all values of n (respectively all large enough values of n).

Now, what do we mean by "the union of the pieces is significantly larger than the *smallest* of these pieces"? In this first version of the growth condition, this means that the size of this union is larger than the size of the smallest piece by a quantity $a\sqrt{\log N}$ where N is the number of pieces. (We remind the reader that the function $\sqrt{\log y}$ arises from the fact that this it in a sense the inverse of the function $\exp(-x^2)$.) Well, sometimes it will only be larger by a quantity of say $a\sqrt{\log N}/100$. This is how the parameter c^* below comes into the picture. Of course, one could also multiply the functionals by a suitable constant (i.e. $1/c^*$) to always reduce to the case $c^* = 1$ but this is a matter of taste.

Another feature is that we do not need to consider the case with N pieces for a general value of N, but only for the case where $N = N_n$ for some n. This is simply because we care about the value of $\log N$ only within, say, a factor of 2, and this is precisely what motived the definition of N_n. In order to understand the definition below one should also recall that $\sqrt{\log N_n}$ is about $2^{n/2}$.

It will be rather convenient to consider not only a single functional but a whole sequence (F_n) of functionals, but at first reading one might assume that F_n does not depend on n. So, consider a metric space (T, d) (that need *not* be finite), and a decreasing sequence $(F_n)_{n \geq 0}$ of functionals on T, that is

$$\forall H \subset T\,,\; F_{n+1}(H) \leq F_n(H)\;. \tag{2.74}$$

Definition 2.3.10. *We say that the functionals F_n satisfy the* growth condition *with parameters $r \geq 4$ and $c^* > 0$ if for any integer $n \geq 0$ and any $a > 0$ the following holds true, where $m = N_{n+1}$. For each collection of subsets $H_1, \ldots, H_m$ of T that are (a, r)-separated we have*

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq c^* a 2^{n/2} + \min_{\ell \leq m} F_{n+1}(H_\ell)\;. \tag{2.75}$$

We observe that the functional F_n occurs on the left-hand side of (2.75), while the smaller functional F_{n+1} occurs on the right-hand side (which gives us a little extra room to check this condition).

Exercise 2.3.11. Find example of spaces (T, d) where the growth condition holds while $F_n(H) = 0$ for each n and each $H \subset T$. (Hint: use Exercise 2.3.9.)

We now note the non-obvious fact that condition (2.75) imposes strong restrictions on the metric space (T, d), and we explain this now. We prove that (2.75) implies that if $a > 2^{-n/2} F_0(T)/c^*$, each ball $B(s, ar)$ can be covered by N_{n+1} balls $B(t, a)$. Consider points $t_1, \ldots, t_k$ in $B(s, ar)$ such that $d(t_\ell, t_{\ell'}) \geq a$ whenever $\ell \neq \ell'$. Assume that $k \geq m = N_{n+1}$ for a certain $n \geq 0$.

Taking $H_\ell = \{t_\ell\}$, and since $F_{n+1} \geq 0$, the separation condition implies $F_0(T) \geq F_n(T) \geq c^* a 2^{n/2}$. Consequently, if we assume that $c^* a 2^{n/2} > F_0(T)$, we must have $k < N_{n+1}$. If k is as large as possible, the ball $B(s, ar)$ is covered by the balls $B(t_\ell, a)$ for $\ell \leq k$, proving the claim.

Exercise 2.3.12. If $r \geq 4$, prove that the preceding property yields the inequality $2^{n/2} e_n(T) \leq K(r) F_0(T)/c^* + L\Delta(T)$. (Hint: Iterate the process of covering a ball with radius ar by balls with radius a to bound the minimum number $N(T, d, \epsilon)$ of balls with radius ϵ needed to cover T and use Exercise 2.2.13.) Explain why the term $L\Delta(T)$ is necessary. (Hint: use Exercise 2.3.11.)

The following illustrates how we shall use the first part of (2.73).

Exercise 2.3.13. Let (T, d) be isometric to a subset of $\mathbb{R}^k$ provided with the distance induced by a norm. Prove that in order to check that a sequence of functionals satisfies the growth condition of Definition 2.3.10, it suffices to consider the values of n for which $N_{n+1} \leq (1 + 2/r)^k$. (Hint: it follows from (2.41) that for larger values of n there are no points $t_1, \ldots, t_m$ as in (2.73).)

As we shall soon see, the existence of a sequence of functionals satisfying the separation property will give us a lot more information than the crude result of Exercise 2.3.12.

Before we come to this, what is the point of considering such sequences of functionals? As the following result shows, decreasing sequences of functionals satisfying the growth condition of Definition 2.3.10 are “built into” the definition of $\gamma_2(T, d)$.

Theorem 2.3.14. *For any metric space (T, d) there exists a decreasing sequence of functionals $(F_n)_{n \geq 0}$ with $F_0(T) = \gamma_2(T, d)$ which satisfies the growth condition of Definition 2.3.10 for $r = 4$ and $c^* = 1/2$.*

In words, one can find a decreasing sequence of functionals satisfying the growth condition with $F_0(T)$ as large as $\gamma_2(T, d)$.

Proof. This proof provides a good opportunity to understand the typical way a sequence (F_n) of functionals might depend on n. For a subset H of T we define

$$F_n(H) = \inf \sup_{t \in H} \sum_{k \geq n} 2^{k/2} \Delta(A_k(t)) \ ,$$

where the infimum is taken over all admissible sequences $(\mathcal{A}_n)$ of partitions of H. (The dependence on n is that the summation starts at $k = n$. This feature will often occur.) Thus $F_0(T) = \gamma_2(T, d)$. To prove the growth condition of Definition 2.3.10, consider $m = N_{n+1}$ and consider points $(t_\ell)_{\ell \leq m}$ of T, with $d(t_\ell, t_{\ell'}) \geq a$ if $\ell \neq \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/4)$, and the set $H = \bigcup_{\ell \leq m} H_\ell$. Consider an admissible sequence $(\mathcal{A}_n)$ of H, and

$$I = \{\ell \leq m \ ; \ \exists A \in \mathcal{A}_n \ , \ A \subset H_\ell\} \ .$$

Since the sets $(H_\ell)_{\ell\le m}$ are disjoint, we have $\operatorname{card} I \le N_n$, and thus there exists $\ell \le m$ with $\ell \notin I$. Then for $t \in H_\ell$, we have $A_n(t) \not\subset H_\ell$, so that since $A_n(t) \subset H$, the set $A_n(t)$ must meet a set $H_{\ell'}$ for a certain $\ell' \ne \ell$, and consequently it meets the ball $B(t_{\ell'}, a/4)$. Since $d(t, B(t_{\ell'}, a/4)) \ge a/2$, this implies that $\Delta(A_n(t)) \ge a/2$. Therefore

$$\sum_{k\ge n} 2^{k/2}\Delta(A_k(t)) \ge \frac{1}{2}a2^{n/2} + \sum_{k\ge n+1} 2^{k/2}\Delta(A_k(t)\cap H_\ell)\ . \tag{2.76}$$

Since $\mathcal{A}'_n = \{A \cap H_\ell; A \in \mathcal{A}_n\}$ is an admissible sequence of H_ℓ, we have by definition

$$\sup_{t\in H_\ell} \sum_{k\ge n+1} 2^{k/2}\Delta(A_k(t)\cap H_\ell) \ge F_{n+1}(H_\ell)\ .$$

Hence, taking the supremum over t in H_ℓ in (2.76) we get

$$\sup_{t\in H_\ell} \sum_{k\ge n} 2^{k/2}\Delta(A_k(t)) \ge \frac{1}{2}a2^{n/2} + F_{n+1}(H_\ell)\ .$$

Since the admissible sequence $(\mathcal{A}_n)$ is arbitrary, we have shown that

$$F_n(H) \ge \frac{1}{2}a2^{n/2} + \min_\ell F_{n+1}(H_\ell)\ ,$$

which is (2.75) for $c^* = 1/2$. □

The previous proof demonstrates how to use functionals F_n which actually depend on n. This will be a very useful technical device. However it is not really needed here, since we also have the following.

Theorem 2.3.15. *The functionals $F_n(H) = \gamma_2(H,d)$ satisfy the growth condition of Definition 2.3.10 for $r = 8$ and $c^* = 1/4$.*

Proof. The proof is almost the same as that of Theorem 2.3.14. Consider points $(t_\ell)_{\ell\le m}$ of T, with $d(t_\ell, t_{\ell'}) \ge a$ if $\ell \ne \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/8)$, and the set $H = \bigcup_{\ell\le m} H_\ell$. Consider an admissible sequence $(\mathcal{A}_n)$ of H, and

$$I = \{\ell \le m\ ;\ \exists A \in \mathcal{A}_n\ ,\ A \subset H_\ell\}\ .$$

Since the sets $(H_\ell)_{\ell\le m}$ are disjoint, we have $\operatorname{card} I \le N_n$, and thus there exists $\ell \le m$ with $\ell \notin I$. Then for $t \in H_\ell$, we have $A_n(t) \not\subset H_\ell$, so that since $A_n(t) \subset H$, the set $A_n(t)$ must meet a set $H_{\ell'}$ for a certain $\ell' \ne \ell$, and consequently it meets the ball $B(t_{\ell'}, a/8)$. Since $d(t, B(t_{\ell'}, a/8)) \ge a/2$, this implies that $\Delta(A_n(t)) \ge a/2$. Therefore, since $\Delta(A_n(t)\cap H_\ell) \le \Delta(H_\ell) \le a/4$,

$$\sum_{k\ge 0} 2^{k/2}\Delta(A_k(t)) \ge \frac{1}{4}a2^{n/2} + \sum_{k\ge 0} 2^{k/2}\Delta(A_k(t)\cap H_\ell)\ . \tag{2.77}$$

From this point on the proof is identical to that of Theorem 2.3.14. □

Our next result is fundamental. It is a kind of converse to Theorem 2.3.14. The size of $\gamma_2(T,d)$ cannot be really larger than $F_0(T)$ when the sequence (F_n) of functionals satisfies the growth condition of Definition 2.3.10. Put it another way, it says that in a sense $F(H) = \gamma_2(H,d)$ is the smallest functional which satisfies the growth condition of Definition 2.3.10. This also explains why we did not give any very simple example of functional satisfying the growth condition. This seems to be the simplest example.

Theorem 2.3.16. *Let (T,d) be a metric space. Assume that there exists a decreasing sequence of functionals $(F_n)_{n\geq 0}$ which satisfies the growth condition of Definition 2.3.10. Then*

$$\gamma_2(T,d) \leq \frac{Lr}{c^*} F_0(T) + Lr\Delta(T) . \tag{2.78}$$

This theorem and its generalizations form the backbone of this book. The essence of this theorem is that it produces (by actually constructing them) a sequence of partitions that witnesses the inequality (2.78). For this reason, it could be called "the fundamental partitioning theorem." The proof of Theorem 2.3.16 is not really difficult, but since one has to construct the partitions, it does require again to roll up our sleeves and even get a bit of grease on our hands. Thus this proof will be better presented (in Section 2.6) after the power of this principle has been demonstrated in Section 2.4 and the usefulness of its consequences illustrated again in Section 2.5.

Exercise 2.3.17. Consider a metric space T consisting of exactly two points. Prove that the sequence of functionals given by $F_n(H) = 0$ for each $H \subset T$ satisfies the growth condition of Definition 2.3.10 for $r = 4$ and any $c^* > 0$. Explain why we cannot replace (2.78) by the inequality $\gamma_2(T,d) \leq LrF_0(T)/c^*$.

Given the functionals F_n, Theorem 2.3.16 yields partitions, but it does not say how to find these functionals. One must understand that there is no magic. Admissible sequences are not going to come out of thin air, but rather they will reflect the geometry of the space (T,d). Once this geometry is understood, it is usually possible to guess a good choice for the functionals F_n. Many examples will be given in subsequent chapters. It seems, at least to the author, that it is much easier to guess the functionals F_n rather than the partitions that witness the inequality (2.78). Besides, as Theorem 2.3.14 shows, we really have no choice. Functionals with the growth property are intimately connected with admissible sequences of partitions.

2.4 Gaussian Processes and the Mysteries of Hilbert Space

Consider a Gaussian process $(X_t)_{t\in T}$, that is, a jointly Gaussian family of centered r.v.s indexed by T. We provide T with the canonical distance

$$d(s,t) = \big(\mathsf{E}(X_s - X_t)^2\big)^{1/2} \,. \tag{2.79}$$

Recall the functional γ_2 of Definition 2.2.19.

Theorem 2.4.1 (The Majorizing Measure Theorem). *For some universal constant L we have*

$$\frac{1}{L}\gamma_2(T,d) \le \mathsf{E}\sup_{t\in T} X_t \le L\gamma_2(T,d) \,. \tag{2.80}$$

The reason for the name is explained in Section 6.2. We can reformulate this theorem by the statement

$$\textbf{Chaining suffices to explain the size of a Gaussian process}. \tag{2.81}$$

By this statement we simply means that (as witnessed by the left-hand side inequality in (2.80)) the "natural" chaining bound for the size of a Gaussian process (as witnessed by the right-hand side inequality in (2.80)) is of correct order, provided of course one uses the best possible chaining. The author believes that this is an occurrence of a much more general phenomenon, several aspects of which will be investigated in later chapters.

The right-hand side inequality in (2.80) follows from Theorem 2.2.22. To prove the lower bound we will use Theorem 2.3.16 and the functionals

$$F_n(H) = F(H) = \sup_{H^*\subset H, H^*\,\mathrm{finite}} \mathsf{E}\sup_{t\in H^*} X_t \,,$$

so that F_n does not depend on n. To apply (2.78) we need to prove that the functionals F_n satisfy the growth condition with c^* a universal constant and to bound $\Delta(T)$ (which is easy). We strive to give a proof that relies on general principles, and lends itself to generalizations.

Lemma 2.4.2 (Sudakov minoration). *Assume that*

$$\forall p\,,\, q \le m\,,\quad p \ne q \quad \Rightarrow d(t_p, t_q) \ge a \,.$$

Then we have

$$\mathsf{E}\sup_{p\le m} X_{t_p} \ge \frac{a}{L_1}\sqrt{\log m} \,. \tag{2.82}$$

Here and below $L_1, L_2, \ldots$ are specific universal constants. Their values remain the same (at least within the same section).

Exercise 2.4.3. Prove that Lemma 2.4.2 is equivalent to the following statement. If $(X_t)_{t\in T}$ is a Gaussian process, and d is the canonical distance, then

$$e_n(T,d) \le 2^{-n/2}\mathsf{E}\sup_{t\in T} X_t \,. \tag{2.83}$$

Compare with Exercise 2.3.12.

A proof of Sudakov minoration can be found in [5], p. 83. The same proof is actually given further in the present book, and the ambitious reader may like to try to understand this now, using the following steps.

Exercise 2.4.4. Use Lemma 8.3.6 and Lemma 16.8.10 to prove that for a Gaussian process $(X_t)_{t\in T}$ we have $e_n(T,d) \le 2^{-n/2}\mathsf{E}\sup_{t\in T}|X_t|$. Then use Exercise 2.2.2 to deduce (2.83).

To understand the relevance of Sudakov minoration, let us consider the case where $\mathsf{E}X_{t_p}^2 \le 100a^2$ (say) for each p. Then (2.82) means that the bound (2.13) is of the correct order in this situation.

Exercise 2.4.5. Prove (2.82) when the r.v.s X_{t_p} are independent. (Hint: use Exercise 2.2.7 (b).)

Exercise 2.4.6. A natural approach ("the second moment method") to prove that $\mathsf{P}(\sup_{p\le m} X_{t_p} \ge u)$ is at least $1/L$ for a certain value of u is as follows. Consider the r.v. $Y = \sum_p \mathbf{1}_{\{X_{t_p}\ge u\}}$, prove that $\mathsf{E}Y^2 \le L(\mathsf{E}Y)^2$, and then use the Paley-Zygmund inequality (7.30) below to prove that $\sup_{p\le m} X_{t_p} \ge a\sqrt{\log m}/L_1$ with probability $\ge 1/L$. Prove that this approach works when the r.v.s X_{t_ℓ} are independent, but find examples showing that this naive approach does not work in general to prove (2.82).

The following is a very important property of Gaussian processes, and one of the keys to Theorem 2.4.1. It is a facet of the theory of concentration of measure, a leading idea of modern probability theory. The reader is referred to the (very nice) book of M. Ledoux [4] to learn about this.

Lemma 2.4.7. *Consider a Gaussian process $(X_t)_{t\in U}$, where U is finite and let $\sigma = \sup_{t\in U}(\mathsf{E}X_t^2)^{1/2}$. Then for $u > 0$ we have*

$$\mathsf{P}\Big(\Big|\sup_{t\in U} X_t - \mathsf{E}\sup_{t\in U} X_t\Big| \ge u\Big) \le 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big)\ . \tag{2.84}$$

Let us stress in words what this means. The size of the fluctuations of $\mathsf{E}\sup_{t\in U} X_t$ is governed by the size of the individual r.v.s X_t, rather than by the (typically much larger) quantity $\mathsf{E}\sup_{t\in U} X_t$.

Exercise 2.4.8. Find an example of a Gaussian process for which

$$\mathsf{E}\sup_{t\in T} X_t \gg \sigma = \sup_{t\in T}(\mathsf{E}X_t^2)^{1/2}\ ,$$

whereas the fluctuations of $\sup_{t\in T} X_t$ are of order σ, e.g. the variance of $\sup_t X_t$ is about σ^2. (Hint: $T = \{(t_i)_{i\le n}; \sum_{i\le n} t_i^2 \le 1\}$ and $X_t = \sum_{i\le n} t_i g_i$ where g_i are independent standard Gaussian. Observe first that $(\sup_t X_t)^2 = \sum_{i\le n} g_i^2$ is of order n and has fluctuations of order $\sqrt{n}$ by the central limit theorem. Conclude that $\sup_t X_t$ has fluctuations of order 1 whatever the value of n.)

Proposition 2.4.9. *Consider points $(t_\ell)_{\ell\le m}$ of T. Assume that $d(t_\ell,t_{\ell'})\ge a$ if $\ell\ne\ell'$. Consider $\sigma>0$, and for $\ell\le m$ a finite set $H_\ell\subset B(t_\ell,\sigma)$. Then if $H=\bigcup_{\ell\le m}H_\ell$ we have*

$$\mathsf{E}\sup_{t\in H}X_t\ge\frac{a}{L_1}\sqrt{\log m}-L_2\sigma\sqrt{\log m}+\min_{\ell\le m}\mathsf{E}\sup_{t\in H_\ell}X_t\ . \tag{2.85}$$

When $\sigma\le a/(2L_1L_2)$, (2.85) implies

$$\mathsf{E}\sup_{t\in H}X_t\ge\frac{a}{2L_1}\sqrt{\log m}+\min_{\ell\le m}\mathsf{E}\sup_{t\in H_\ell}X_t\ , \tag{2.86}$$

which can be seen as a generalization of (2.82).

Proof. We can and do assume $m\ge 2$. For $\ell\le m$, we consider the r.v.

$$Y_\ell=\big(\sup_{t\in H_\ell}X_t\big)-X_{t_\ell}=\sup_{t\in H_\ell}(X_t-X_{t_\ell})\ .$$

We set $U=H_\ell$ and for $t\in U$ we set $Z_t=X_t-X_{t_\ell}$. Since $H_\ell\subset B(t_\ell,\sigma)$ we have $\mathsf{E}Z_t^2=d(t,t_\ell)^2\le\sigma^2$ and, for $u\ge 0$ equation (2.84) used for the process $(Z_t)_{t\in U}$ implies

$$\mathsf{P}(|Y_\ell-\mathsf{E}Y_\ell|\ge u)\le 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big)\ .$$

Thus if $V=\max_{\ell\le m}|Y_\ell-\mathsf{E}Y_\ell|$ then

$$\mathsf{P}(V\ge u)\le 2m\exp\Big(-\frac{u^2}{2\sigma^2}\Big)\,, \tag{2.87}$$

and (2.11) implies $\mathsf{E}V\le L_2\sigma\sqrt{\log m}$. Now, for each $\ell\le m$,

$$Y_\ell\ge\mathsf{E}Y_\ell-V\ge\min_{\ell\le m}\mathsf{E}Y_\ell-V\,,$$

and thus

$$\sup_{t\in H_\ell}X_t=Y_\ell+X_{t_\ell}\ge X_{t_\ell}+\min_{\ell\le m}\mathsf{E}Y_\ell-V$$

so that

$$\sup_{t\in H}X_t\ge\max_{\ell\le m}X_{t_\ell}+\min_{\ell\le m}\mathsf{E}Y_\ell-V\ .$$

We then take expectations and use (2.82). □

Exercise 2.4.10. Prove that (2.86) might fail if one allows $\sigma=a$. (Hint: the intersection of the balls $B(t_\ell,a)$ might contain a ball with positive radius.)

Exercise 2.4.11. Prove that

$$\mathsf{E}\sup_{t\in H}X_t\le La\sqrt{\log m}+\max_{\ell\le m}\mathsf{E}\sup_{t\in H_\ell}X_t\ . \tag{2.88}$$

Try to find improvements on this bound. (Hint: peek at (16.81) below.)

Proof of Theorem 2.4.1. We fix $r \geq 2L_1L_2$. To prove the growth condition for the functionals F_n we simply observe that (2.86) implies that (2.75) holds for $c^* = 1/L$. Using Theorem 2.3.16, it remains only to control the term $\Delta(T)$. But

$$\mathsf{E}\max(X_{t_1}, X_{t_2}) = \mathsf{E}\max(X_{t_1} - X_{t_2}, 0) = \frac{1}{\sqrt{2\pi}} d(t_1, t_2)\,,$$

so that $\Delta(T) \leq \sqrt{2\pi}\mathsf{E}\sup_{t\in T} X_t$. □

The proof of Theorem 2.4.1 displays an interesting feature. This theorem aims at understanding $\mathsf{E}\sup_{t\in T} X_t$, and for this we use functionals that are based on precisely this quantity. This is not a circular argument. The content of Theorem 2.4.1 is that there is simply no other way to bound a Gaussian process than to control the quantity $\gamma_2(T,d)$. Of course, to control this quantity in a specific situation, we must in some way gain understanding of the underlying geometry of this situation.

The following is a noteworthy consequence of Theorem 2.4.1.

Theorem 2.4.12. *Consider two processes $(Y_t)_{t\in T}$ and $(X_t)_{t\in T}$ indexed by the same set. Assume that the process $(X_t)_{t\in T}$ is Gaussian and that the process $(Y_t)_{t\in T}$ satisfies the condition*

$$\forall u > 0\ ,\ \forall s\ ,\ t \in T\ ,\ \mathsf{P}(|Y_s - Y_t| \geq u) \leq 2\exp\left(-\frac{u^2}{d(s,t)^2}\right)\ ,$$

where d is the distance (2.79) associated to the process X_t. Then we have

$$\mathsf{E}\sup_{s,t\in T} |Y_s - Y_t| \leq L\mathsf{E}\sup_{t\in T} X_t\ .$$

Proof. We combine (2.49) with the left-hand side of (2.80). □

Let us now turn to a simple (and classical) example that illustrates well the difference between Dudley's bound (2.38) and the bound (2.32). Basically this example reproduces, for a metric space associated to an actual Gaussian process, the metric structure that was described in an abstract setting in Exercise 2.2.16. Consider an independent sequence $(g_i)_{i\geq 1}$ of standard Gaussian r.v.s and for $i \geq 2$ set

$$X_i = \frac{g_i}{\sqrt{\log i}}\ . \tag{2.89}$$

Consider an integer $s \geq 3$ and the process $(X_i)_{2\leq i\leq N_s}$ so the index set is $T = \{2, 3, \ldots, N_s\}$. The distance d associated to the process satisfies for $p \neq q$

$$\frac{1}{\sqrt{\log(\min(p,q))}} \leq d(p,q) \leq \frac{2}{\sqrt{\log(\min(p,q))}}\ . \tag{2.90}$$

Consider $1 \leq n \leq s-2$ and $T_n \subset T$ with $\operatorname{card} T_n = N_n$. There exists $p \leq N_n + 1$ with $p \notin T_n$, so that (2.90) implies $d(p, T_n) \geq 2^{-n/2}/L$ (where the distance from a point to a set is defined in (2.33)). This proves that $e_n(T) \geq 2^{-n/2}/L$. Therefore

$$\sum_n 2^{n/2} e_n(T) \geq \frac{s-2}{L} \,. \tag{2.91}$$

On the other hand, for $n \leq s$ let us now consider $T_n = \{2, 3, \ldots, N_n, N_s\}$, integers $p \in T$ and $m \leq s-1$ such that $N_m < p \leq N_{m+1}$. Then $d(p, T_n) = 0$ if $n \geq m+1$, while, if $n \leq m$,

$$d(p, T_n) \leq d(p, N_s) \leq L2^{-m/2}$$

by (2.90) and since $p \geq N_m$ and $N_s \geq N_m$. Hence we have

$$\sum_n 2^{n/2} d(p, T_n) \leq \sum_{n \leq m} L 2^{n/2} 2^{-m/2} \leq L \,. \tag{2.92}$$

Comparing (2.91) and (2.92) proves that the bound (2.38) is worse than the bound (2.32) by a factor about s.

Exercise 2.4.13. Prove that when T is finite, the bound (2.38) cannot be worse than (2.32) by a factor greater than about $\log\log \operatorname{card} T$. This shows that the previous example is in a sense extremal. (Hint: use $2^{n/2} e_n(T) \leq L\gamma_2(T, d)$ and $e_n(T) = 0$ if $N_n \geq \operatorname{card} T$.)

Exercise 2.4.14. Prove that the estimate (2.67) is essentially optimal. (Hint: if $m \geq \exp(10s)$, one can produce the situation of Example 2.2.16 (b) inside $\mathbb{R}^m$.)

It follows from (2.92) and (2.32) that $\mathsf{E} \sup_{i \geq 1} X_i < \infty$. A simpler proof of this fact is given in Proposition 2.4.16 below.

Now we generalize the process of Exercise 2.2.15 to Hilbert space. We consider the Hilbert space $\ell^2 = \ell^2(\mathbb{N}^*)$ of sequences $(t_i)_{i \geq 1}$ such that $\sum_{i \geq 1} t_i^2 < \infty$, provided with the norm

$$\|t\| = \|t\|_2 = \Big(\sum_{i \geq 1} t_i^2\Big)^{1/2} \,. \tag{2.93}$$

To each t in ℓ^2 we associate a Gaussian r.v.

$$X_t = \sum_{i \geq 1} t_i g_i \tag{2.94}$$

(the series converges in $L^2(\Omega)$). In this manner, for each subset T of ℓ^2 we can consider the Gaussian process $(X_t)_{t \in T}$. The distance induced on T by the process coincides with the distance of ℓ^2 since from (2.94) we have $\mathsf{E} X_t^2 = \sum_{i \geq 1} t_i^2$.

The importance of this construction is that it is generic. *All* Gaussian processes can be obtained in this way, at least when there is a countable subset T' of T that is dense in the space (T,d), which is the only case of importance for us. Indeed, it suffices to think of the r.v. Y_t of a Gaussian process as a point in $L^2(\Omega)$, where Ω is the underlying probability space, and to identify $L^2(\Omega)$, which is then separable, and ℓ^2 by choosing an orthonormal basis of $L^2(\Omega)$.

Here is the place to make a general observation. It is not true that all processes of interest can be represented as the sum of a random series as in (2.94). Suppose, however, that one is interested in the boundedness of a random series of functions, $X_u = \sum_{i\ge1} \xi_i f_i(u)$ for $u \in U$. Then all that matters is the set T of coefficients $T = \{t = (f_i(u))_i; u \in U\}$. For a sequence $t = (t_i)$ we then define

$$X_t = \sum_i t_i \xi_i \tag{2.95}$$

and the fundamental issue becomes to understand the boundedness of the process $(X_t)_{t\in T}$. This is why processes of the type (2.95) play such an important role in this book.

A subset T of ℓ^2 will always be provided with the distance induced by ℓ^2, so we may also write $\gamma_2(T)$ rather than $\gamma_2(T,d)$. We denote by $\operatorname{conv} T$ the convex hull of T, and we write

$$T_1 + T_2 = \{t_1 + t_2 \; ; \; t_1 \in T_1 \; , \; t_2 \in T_2\} \; .$$

Theorem 2.4.15. *For a subset T of ℓ^2, we have*

$$\gamma_2(\operatorname{conv} T) \le L\gamma_2(T) \; . \tag{2.96}$$

For two subsets T_1 and T_2 of ℓ^2, we have

$$\gamma_2(T_1 + T_2) \le L\big(\gamma_2(T_1) + \gamma_2(T_2)\big) \; . \tag{2.97}$$

Proof. To prove (2.96) we observe that since $X_{a_1t_1+a_2t_2} = a_1X_{t_1} + a_2X_{t_2}$ we have

$$\sup_{t\in\operatorname{conv} T} X_t = \sup_{t\in T} X_t \; . \tag{2.98}$$

We then use (2.80) to write

$$\frac{1}{L}\gamma_2(\operatorname{conv} T) \le \mathsf{E} \sup_{\operatorname{conv} T} X_t \le \mathsf{E} \sup_T X_t \le L\gamma_2(T) \; .$$

The proof of (2.97) is similar. □

We recall the ℓ^2 norm $\|\cdot\|$ of (2.93). Here is a simple fact.

Proposition 2.4.16. *Consider a set $T = \{t_k \; ; \; k \ge 1\}$ where*

$$\forall\, k \ge 1 \; , \; \|t_k\| \le 1/\sqrt{\log(k+1)} \; .$$

Then $\mathsf{E}\sup_{t\in T} X_t \le L$.

Proof. We have

$$\mathsf{P}\Big(\sup_{k\geq 1}|X_{t_k}|\geq u\Big)\leq\sum_{k\geq 1}\mathsf{P}(|X_{t_k}|\geq u)\leq\sum_{k\geq 1}2\exp\Big(-\frac{u^2}{2}\log(k+1)\Big) \tag{2.99}$$

since X_{t_k} is Gaussian with $\mathsf{E}X_{t_k}^2\leq 1/\log(k+1)$. Now for $u\geq 2$, the right-hand side of (2.99) is at most $L\exp(-u^2/L)$. □

Exercise 2.4.17. Deduce Proposition 2.4.16 from (2.32). (Hint: see Exercise 2.2.16 (a).)

Combining with (2.98), Proposition 2.4.16 proves that $\mathsf{E}\sup_{t\in T}X_t\leq L$, where $T=\operatorname{conv}\{t_k;k\geq 1\}$. The following shows that this situation is in a sense generic.

Theorem 2.4.18. *Consider a countable set $T\subset\ell^2$, with $0\in T$. Then we can find a sequence (t_k), such that each element t_k is a multiple of the difference of two elements of T, with*

$$\forall\, k\geq 1\;,\;\|t_k\|\sqrt{\log(k+1)}\leq L\mathsf{E}\sup_{t\in T}X_t$$

and

$$T\subset\operatorname{conv}(\{t_k\;;\;k\geq 1\})\;.$$

Proof. By Theorem 2.4.1 we can find an admissible sequence $(\mathcal{A}_n)$ of T with

$$\forall t\in T\,,\;\sum_{n\geq 0}2^{n/2}\Delta(A_n(t))\leq L\mathsf{E}\sup_{t\in T}X_t:=S\;. \tag{2.100}$$

We construct sets $T_n\subset T$, such that each $A\in\mathcal{A}_n$ contains exactly one element of T_n. We ensure in the construction that $T=\bigcup_{n\geq 0}T_n$ and that $T_0=\{0\}$. (To do this, we simply enumerate the elements of T as $(v_n)_{n\geq 1}$ with $v_0=0$ and we ensure that v_n is in T_n.) For $n\geq 1$ consider the set U_n that consists of all the points

$$2^{-n/2}\frac{t-v}{\|t-v\|}$$

where $t\in T_n, v\in T_{n-1}$ and $t\neq v$. Thus each element of U_n has norm $2^{-n/2}$, and U_n has at most $N_nN_{n-1}\leq N_{n+1}$ elements. Let $U=\bigcup_{k\geq 1}U_k$. We observe that U contains at most N_{n+2} elements of norm $\geq 2^{-n/2}$. If we enumerate $U=\{t_k;k=1,\ldots\}$ where the sequence $(\|t_k\|)$ is non-increasing, then if $\|t_k\|\geq 2^{-n/2}$ we have $k\leq N_{n+2}$ and this implies that $\|t_k\|\leq L/\sqrt{\log(k+1)}$.

Consider $t\in T$, so that $t\in T_m$ for some $m\geq 0$. Writing $\pi_n(t)$ for the unique element of $T_n\cap A_n(t)$, since $\pi_0(t)=0$ we have

$$t=\sum_{1\leq n\leq m}\pi_n(t)-\pi_{n-1}(t)=\sum_{1\leq n\leq m}a_n(t)u_n(t)\;, \tag{2.101}$$

where

$$u_n(t) = 2^{-n/2} \frac{\pi_n(t) - \pi_{n-1}(t)}{\|\pi_n(t) - \pi_{n-1}(t)\|} \in U \,;\; a_n(t) = 2^{n/2} \|\pi_n(t) - \pi_{n-1}(t)\| \,.$$

Since

$$\sum_{1 \le n \le m} a_n(t) \le \sum_{n \ge 1} 2^{n/2} \Delta(A_{n-1}(t)) \le 2S$$

and since $u_n(t) \in U_n \subset U$ we see from (2.101) that

$$t \in 2S \operatorname{conv}(U \cup \{0\}) \,.$$

This concludes the proof. □

It is good to meditate a little about the significance of Theorem 2.4.18. First, we reformulate this theorem in a way which is suitable for generalizations. Consider the class $\mathcal{C}$ of sets of the type $C = \operatorname{conv}\{t_k ; k \ge 1\}$ and for $C \in \mathcal{C}$ define the size $s(C)$ as $\inf \sup_k \|t_k\| \sqrt{\log(k+1)}$, where we assume without loss of generality that the sequence $(\|t_k\|)_{k \ge 1}$ decreases, and where the infimum is over all possible choices of the sequence (t_k) for which $C = \operatorname{conv}\{t_k ; k \ge 1\}$. Proposition 2.4.16 implies $\mathsf{E} \sup_{t \in C} X_t \le L s(C)$. Theorem 2.4.18 implies that given a countable set T with $0 \in T$ we can find $T \subset C \in \mathcal{C}$ with $s(C) \le L \mathsf{E} \sup_{t \in T} X_t$. In words, the size of T for the Gaussian process is witnessed by the smallest size (as measured by s) of an element of $\mathcal{C}$ containing T.

Also worthy of detailing is a remarkable geometric consequence of Theorem 2.4.18. Consider an integer N and let us provide ℓ^2_N ($= \mathbb{R}^N$ provided with the Euclidean distance) with the canonical Gaussian measure μ, i.e. the law of the i.i.d. Gaussian sequence $(g_i)_{i \le N}$. Let us view an element t of ℓ^2_N as a function on ℓ^2_N by the canonical duality, so t is a r.v. Y_t on the probability space (ℓ^2_N, μ). The processes (X_t) and (Y_t) have the same law, hence they are really the same object viewed in two different ways. Consider a subset T of ℓ^2_N, and assume that $T \subset \operatorname{conv}\{t_k ; k \ge 1\}$. Then for any $v > 0$ we have

$$\Big\{ \sup_{t \in T} t \ge v \Big\} \subset \bigcup_{k \ge 1} \{t_k \ge v\} \,. \tag{2.102}$$

The somewhat complicated set on the left-hand side is covered by a countable union of much simpler sets: the sets $\{t_k \ge v\}$ are *half-spaces*. Assume now that for $k \ge 1$ and a certain S we have $\|t_k\| \sqrt{\log(k+1)} \le S$. Then (2.99) implies that for $u \ge 2$

$$\sum_{k \ge 1} \mu(\{t_k \ge Su\}) \le L \exp(-u^2/L) \,.$$

Theorem 2.4.18 implies that may take $S \le L \mathsf{E} \sup_t X_t$. Therefore for $v \ge L \mathsf{E} \sup_t X_t$, the fact that the set in the left-hand side of (2.102) is small (in the sense of probability) can be witnessed by the fact that this set can be covered by a countable union of simple sets (half-spaces) the *sum* of the probabilities of which is small.

Of course, one may hope that the two remarkable phenomena described above occur (at least in some form) in many other settings, a topic to which we shall come back many times.

Exercise 2.4.19. Prove that if $T \subset \ell^2$ and $0 \in T$, then (even when T is not countable) we can find a sequence (t_k) in ℓ^2, with $\|t_k\|\sqrt{\log(k+1)} \le L\mathsf{E}\sup_{t\in T} X_t$ for all k and

$$T \subset \overline{\mathrm{conv}}\{t_k \; ; \; k \ge 1\} \; ,$$

where $\overline{\mathrm{conv}}$ denotes the closed convex hull. (Hint: do the obvious thing, apply Theorem 2.4.18 to a dense countable subset of T.) Denoting now $\mathrm{conv}^*(A)$ the set of infinite sums $\sum_i \alpha_i a_i$ where $\sum_i |\alpha_i| = 1$ and $a_i \in A$, prove that one can also achieve

$$T \subset \mathrm{conv}^*\{t_k \; ; \; k \ge 1\} \; .$$

Exercise 2.4.20. Consider a set $T \subset \ell^2$ with $0 \in T \subset B(0,\delta)$. Prove that we can find a sequence (t_k) in ℓ^2, with the following properties:

$$\forall\, k \ge 1 \; , \; \|t_k\|\sqrt{\log(k+1)} \le L\mathsf{E}\sup_{t\in T} X_t \; , \tag{2.103}$$

$$\|t_k\| \le L\delta \; , \tag{2.104}$$

$$T \subset \overline{\mathrm{conv}}\{t_k \; ; \; k \ge 1\} \; , \tag{2.105}$$

where $\overline{\mathrm{conv}}$ denotes the closed convex hull. (Hint: copy the proof of Theorem 2.4.18, observing that since $T \subset B(0,\delta)$ one may chose $\mathcal{A}_n = \{T\}$ and $T_n = \{0\}$ for $n \le n_0$, where n_0 is the smallest integer for which $2^{n_0/2} \ge \delta^{-1}\mathsf{E}\sup_{t\in T} X_t$, and thus $U_n = \emptyset$ for $n \le n_0$.)

The purpose of the next exercise is to derive from Exercise 2.4.20 some results of Banach space theory due to S. Artstein [1]. This exercise is more elaborate, and may be omitted at first reading. A Bernoulli r.v. ε is such that $\mathsf{P}(\varepsilon = \pm 1) = 1/2$. (The reader will not confuse Bernoulli r.v.s ε_i with positive numbers ϵ_k!)

Exercise 2.4.21. In this exercise we consider a subset $T \subset \mathbb{R}^N$, where $\mathbb{R}^N$ is provided with the Euclidean distance. We assume that for some $\delta > 0$, we have

$$0 \in T \subset B(0,\delta) \; .$$

We consider independent Bernoulli r.v.s $(\varepsilon_{i,p})_{i,p\ge 1}$ and for $q \le N$ we consider the random operator $U_q : \mathbb{R}^N \to \mathbb{R}^q$ given by

$$U_q(x) = \Big(\sum_{i\le N} \varepsilon_{i,p} x_i\Big)_{p\le q} \; .$$

The purpose of the exercise is to show that there exists a number L such that if

$$q \geq \delta^{-1} \mathsf{E} \sup_{t \in T} \sum_{i \leq N} g_i t_i \ , \tag{2.106}$$

then with high probability

$$U_q(T) \subset B(0, L\delta\sqrt{q}) \ . \tag{2.107}$$

(a) Use the subgaussian inequality (3.2.2) to prove that if $\|x\| = 1$, then

$$\mathsf{E} \exp\Big(\frac{1}{4}\Big(\sum_{i \leq N} \varepsilon_{i,p} x_i\Big)^2\Big) \leq L \ . \tag{2.108}$$

(b) Use (2.108) and independence to prove that for $x \in \mathbb{R}^n$ and $v \geq 1$,

$$\mathsf{P}(\|U_q(x)\| \geq Lv\sqrt{q}\|x\|) \leq \exp(-v^2 q) \ . \tag{2.109}$$

(c) Use (2.109) to prove that with probability close to 1, for each of the vectors t_k of Exercise 2.4.20 one has $\|U_q(t_k)\| \leq L\delta\sqrt{q}$ and conclude.

The simple proof of Theorem 2.4.15 hides the fact that (2.96) is a near miraculous result. It does not provide any real understanding of what is going on. Here is a simple question.

Research problem 2.4.22. Given a subset T of the unit ball of ℓ^2, give a geometrical proof that $\gamma_2(\operatorname{conv} T) \leq L\sqrt{\log \operatorname{card} T}$.

The issue is that, while this result is true whatever the choice of T, the structure of an admissible sequence which witnesses that $\gamma_2(\operatorname{conv} T) \leq L\sqrt{\log \operatorname{card} T}$ must depend on the "geometry" of the set T.

A geometrical proof should of course not use Gaussian processes but only the geometry of Hilbert space. A really satisfactory argument would give a proof that holds in Banach spaces more general than Hilbert space, for example by providing a positive answer to the following, where the concept of q-smooth Banach space is explained in [6].

Research problem 2.4.23. Given a 2-smooth Banach space, is it true that for each subset T of its unit ball $\gamma_2(\operatorname{conv} T) \leq K\sqrt{\log \operatorname{card} T}$? More generally, is it true that for each finite subset T one has $\gamma_2(\operatorname{conv} T) \leq K\gamma_2(T)$? (Here K may depend on the Banach space, but not on T.)

Here of course we use the distance induced by the norm to compute the γ_2 functional.

Research problem 2.4.24. Still more generally, is it true that for a finite subset T of a q-smooth Banach space, one has $\gamma_q(\operatorname{conv} T) \leq K\gamma_\alpha(T)$?

Even when the Banach space is ℓ^p, I do not know the answer to these problems (unless $p = 2$!). (The Banach space ℓ^p is 2-smooth for $p \geq 2$ and q-smooth for $p < 2$, where $1/p + 1/q = 1$.) One concrete case is when the set T consists of the first N vectors of the unit basis of ℓ^p. It is possible to show in this case that $\gamma_q(\operatorname{conv} T) \leq K(p)(\log N)^{1/q}$, where $1/p + 1/q = 1$. We leave this as a challenge to the reader. The proof for the general case is pretty much the same as for the case $p = q = 2$ which was already proposed as a challenge after Exercise 2.2.15.

2.5 A First Look at Ellipsoids

We have illustrated the gap between Dudley's bound (2.38) and the sharper bound (2.32), using the examples (2.42) and (2.89). Perhaps the reader deems these examples artificial, and believes that "in all practical situations" Dudley's bound suffices. Before we prove Theorem 2.3.16 (thus completing the proof of the Majorizing Measure Theorem 2.4.1) in the next section, we feel that it may be useful to provide some more motivation by demonstrating that the gap between Dudley's bound (2.38) and the generic chaining bound (2.32) already exists for *ellipsoids* in Hilbert space. It is hard to argue that ellipsoids are artificial, unnatural or unimportant. Moreover, understanding ellipsoids will be fundamental in several subsequent questions, such as the matching theorems of Chapter 4.

Given a sequence $(a_i)_{i\geq 1}, a_i > 0$, we consider the ellipsoid

$$\mathcal{E} = \left\{ t \in \ell^2 \; ; \; \sum_{i \geq 1} \frac{t_i^2}{a_i^2} \leq 1 \right\} . \tag{2.110}$$

Proposition 2.5.1. *We have*

$$\frac{1}{L} \left(\sum_{i \geq 1} a_i^2 \right)^{1/2} \leq \mathsf{E} \sup_{t \in \mathcal{E}} X_t \leq \left(\sum_{i \geq 1} a_i^2 \right)^{1/2} . \tag{2.111}$$

Proof. The Cauchy-Schwarz inequality implies

$$Y := \sup_{t \in \mathcal{E}} X_t = \sup_{t \in \mathcal{E}} \sum_{i \geq 1} t_i g_i \leq \left(\sum_{i \geq 1} a_i^2 g_i^2 \right)^{1/2} . \tag{2.112}$$

Taking $t_i = a_i^2 g_i / (\sum_{j \geq 1} a_j^2 g_j^2)^{1/2}$ yields that actually $Y = (\sum_{i \geq 1} a_i^2 g_i^2)^{1/2}$ and thus $\mathsf{E} Y^2 = \sum_{i \geq 1} a_i^2$. The right-hand side of (2.111) follows from the Cauchy-Schwarz inequality:

$$\mathsf{E} Y \leq (\mathsf{E} Y^2)^{1/2} = \left(\sum_{i \geq 1} a_i^2 \right)^{1/2} . \tag{2.113}$$

For the left-hand side, let $\sigma = \max_{i\geq 1} |a_i|$. Since $Y = \sup_{t\in\mathcal{E}} X_t \geq |a_i||g_i|$ for any i, we have $\sigma \leq L\mathsf{E}Y$. Also,

$$\mathsf{E}X_t^2 = \sum_i t_i^2 \leq \max_i a_i^2 \sum_j \frac{t_j^2}{a_j^2} \leq \sigma^2 \ . \tag{2.114}$$

Then (2.84) implies

$$\mathsf{E}(Y - \mathsf{E}Y)^2 \leq L\sigma^2 \leq L(\mathsf{E}Y)^2 \ ,$$

so that $\sum_{i\geq 1} a_i^2 = \mathsf{E}Y^2 = \mathsf{E}(Y-\mathsf{E}Y)^2 + (\mathsf{E}Y)^2 \leq L(\mathsf{E}Y)^2$. □

As a consequence of Theorem 2.4.1,

$$\gamma_2(\mathcal{E}) \leq L\Big(\sum_{i\geq 1} a_i^2\Big)^{1/2} \ . \tag{2.115}$$

This statement is purely about the geometry of ellipsoids. The proof we gave was rather indirect, since it involved Gaussian processes. Later on, in Theorem 4.1.11, we shall give a "purely geometric" proof of this result that will have many consequences.

Let us now assume that the sequence $(a_i)_{i\geq 1}$ is non-increasing. Since

$$2^n \leq i \leq 2^{n+1} \Rightarrow a_{2^n} \geq a_i \geq a_{2^{n+1}}$$

we get

$$\sum_{i\geq 1} a_i^2 = \sum_{n\geq 0} \sum_{2^n\leq i<2^{n+1}} a_i^2 \leq \sum_{n\geq 0} 2^n a_{2^n}^2$$

and

$$\sum_{i\geq 1} a_i^2 \geq \sum_{n\geq 0} 2^n a_{2^{n+1}}^2 = \frac{1}{2}\sum_{n\geq 1} 2^n a_{2^n}^2 \ ,$$

and thus $\sum_{n\geq 0} 2^n a_{2^n}^2 \leq 3\sum_{i\geq 1} a_i^2$. So we may rewrite (2.111) as

$$\frac{1}{L}\Big(\sum_{n\geq 0} 2^n a_{2^n}^2\Big)^{1/2} \leq \mathsf{E}\sup_{t\in\mathcal{E}} X_t \leq \Big(\sum_{n\geq 0} 2^n a_{2^n}^2\Big)^{1/2} \ . \tag{2.116}$$

Proposition 2.5.1 describes the size of ellipsoids with respect to Gaussian processes. Our next result describes their size with respect to Dudley's entropy bound (2.36).

Proposition 2.5.2. *We have*

$$\frac{1}{L}\sum_{n\geq 0} 2^{n/2} a_{2^n} \leq \sum_{n\geq 0} 2^{n/2} e_n(\mathcal{E}) \leq L\sum_{n\geq 0} 2^{n/2} a_{2^n} \ . \tag{2.117}$$

The right-hand sides in (2.116) and (2.117) are distinctively different. Dudley's bound fails to describe the behavior of Gaussian processes on ellipsoids. This is a simple occurrence of a general phenomenon. In some sense an ellipsoid is smaller than what one would predict just by looking at its entropy numbers $e_n(\mathcal{E})$. This idea will be investigated further in Section 4.1.

Exercise 2.5.3. Prove that for an ellipsoid $\mathcal{E}$ of $\mathbb{R}^m$ one has

$$\sum_{n\geq 0} 2^{n/2} e_n(\mathcal{E}) \leq L\sqrt{\log(m+1)}\gamma_2(T,d) ,$$

and that this estimate is essentially optimal. Compare with (2.67).

The proof of (2.117) hinges on ideas which are at least 50 years old, and which relate to the methods of Exercise 2.2.14. The left-hand side is the easier part (it is also the most important for us). It follows from the next lemma, the proof of which is basically a special case of (2.39).

Lemma 2.5.4. *We have* $e_n(\mathcal{E}) \geq \frac{1}{2} a_{2^n}$*.*

Proof. Consider the following ellipsoid in $\mathbb{R}^{2^n}$:

$$\mathcal{E}_n = \left\{ (t_i)_{i\leq 2^n} \; ; \; \sum_{i\leq 2^n} \frac{t_i^2}{a_i^2} \leq 1 \right\} .$$

Since $\mathcal{E}_n$ is the image of $\mathcal{E}$ by a contraction (namely the "projection on the first 2^n coordinates") it holds that $e_n(\mathcal{E}_n) \leq e_n(\mathcal{E})$.

Let us denote by B the centered unit Euclidean ball of $\mathbb{R}^{2^n}$ and by Vol the volume in this space. Let us consider a subset T of $\mathcal{E}_n$, with $\operatorname{card} T \leq 2^{2^n}$, and $\epsilon > 0$; then

$$\mathrm{Vol}\Big(\bigcup_{t\in T} (\epsilon B + t)\Big) \leq \sum_{t\in T} \mathrm{Vol}(\epsilon B + t) \leq 2^{2^n} \epsilon^{2^n} \mathrm{Vol} B = (2\epsilon)^{2^n} \mathrm{Vol} B \,.$$

On the other hand, since $a_i \geq a_{2^n}$ for $i \leq 2^n$, we have $a_{2^n} B \subset \mathcal{E}_n$, so that $\mathrm{Vol}\mathcal{E}_n \geq a_{2^n}^{2^n} \mathrm{Vol} B$. Thus when $2\epsilon < a_{2^n}$, we cannot have $\mathcal{E}_n \subset \bigcup_{t\in T}(\epsilon B + t)$. Therefore $e_n(\mathcal{E}_n) \geq \epsilon$. □

We now turn to the upper bound, which relies on a special case of (2.40).

Lemma 2.5.5. *We have*

$$e_{n+3}(\mathcal{E}) \leq 3 \max_{k\leq n} (a_{2^k} 2^{k-n}) \,. \tag{2.118}$$

Proof. We keep the notation of the proof of Lemma 2.5.4. First we show that

$$e_{n+3}(\mathcal{E}) \leq e_{n+3}(\mathcal{E}_n) + a_{2^n} \,. \tag{2.119}$$

To see this, we observe that when $t \in \mathcal{E}$, then

$$1 \geq \sum_{i \geq 1} \frac{t_i^2}{a_i^2} \geq \sum_{i > 2^n} \frac{t_i^2}{a_i^2} \geq \frac{1}{a_{2^n}^2} \sum_{i > 2^n} t_i^2$$

so that $(\sum_{i>2^n} t_i^2)^{1/2} \leq a_{2^n}$ and, viewing $\mathcal{E}_n$ as a subset of $\mathcal{E}$, we have $d(t, \mathcal{E}_n) \leq a_{2^n}$. Thus if we cover $\mathcal{E}_n$ by certain balls with radius ϵ, the balls with the same centers but radius $\epsilon + a_{2^n}$ cover $\mathcal{E}$. This proves (2.119).

Consider now $\epsilon > 0$, and a subset Z of $\mathcal{E}_n$ with the following properties:

$$\text{any two points of } Z \text{ are at mutual distance} \geq 2\epsilon \tag{2.120}$$

$$\operatorname{card} Z \text{ is as large as possible under (2.120).} \tag{2.121}$$

Then by (2.121) the balls centered at points of Z and with radius $\leq 2\epsilon$ cover $\mathcal{E}_n$. Thus

$$\operatorname{card} Z \leq N_{n+3} \Rightarrow e_{n+3}(\mathcal{E}_n) \leq 2\epsilon \,. \tag{2.122}$$

The balls centered at the points of Z, with radius ϵ, have disjoint interiors, so that

$$\operatorname{card} Z \ \operatorname{Vol}(\epsilon B) \leq \operatorname{Vol}(\mathcal{E}_n + \epsilon B) \,. \tag{2.123}$$

Now for $t = (t_i)_{i \leq 2^n} \in \mathcal{E}_n$, we have $\sum_{i \leq 2^n} t_i^2 / a_i^2 \leq 1$, and for t' in ϵB, we have $\sum_{i \leq 2^n} t_i'^2 / \epsilon^2 \leq 1$. Let $c_i = 2 \max(\epsilon, a_i)$. Since

$$\frac{(t_i + t_i')^2}{c_i^2} \leq \frac{2t_i^2 + 2t_i'^2}{c_i^2} \leq \frac{1}{2}\left(\frac{t_i^2}{a_i^2} + \frac{t_i'^2}{\epsilon^2}\right) ,$$

we have

$$\mathcal{E}_n + \epsilon B \subset \mathcal{E}^1 := \left\{ t \, ; \, \sum_{i \leq 2^n} \frac{t_i^2}{c_i^2} \leq 1 \right\} .$$

Therefore

$$\operatorname{Vol}(\mathcal{E}_n + \epsilon B) \leq \operatorname{Vol} \mathcal{E}^1 = \operatorname{Vol} B \prod_{i \leq 2^n} c_i$$

and comparing with (2.123) yields

$$\operatorname{card} Z \leq \prod_{i \leq 2^n} \frac{c_i}{\epsilon} = 2^{2^n} \prod_{i \leq 2^n} \max\left(1, \frac{a_i}{\epsilon}\right) .$$

Assume now that for any $k \leq n$ we have $a_{2^k} 2^{k-n} \leq \epsilon$. Then $a_i \leq a_{2^k} \leq \epsilon 2^{n-k}$ for $2^k < i \leq 2^{k+1}$, so that

$$\begin{aligned}\prod_{i \leq 2^n} \max\left(1, \frac{a_i}{\epsilon}\right) &= \prod_{k \leq n-1} \prod_{2^k < i \leq 2^{k+1}} \max\left(1, \frac{a_i}{\epsilon}\right) \\ &\leq \prod_{k \leq n-1} \left(2^{n-k}\right)^{2^k} = 2^{\sum_{k \leq n} (n-k) 2^k} \\ &\leq 2^{2^{n+2}}\end{aligned}$$

since $\sum_{i \geq 0} i 2^{-i} = 4$.

To sum up, if $\epsilon = \max_{k \le n} a_{2^k} 2^{k-n}$, we have shown that

$$\operatorname{card} Z \le 2^{2^n} \cdot 2^{2^{n+2}} \le N_{n+3} ,$$

so that $e_{n+3}(\mathcal{E}_n) \le 2\epsilon$. The conclusion follows from (2.119). □

Proof of Proposition 2.5.2. We have, using (2.118)

$$\begin{aligned}\sum_{n \ge 3} 2^{n/2} e_n(\mathcal{E}) &= \sum_{n \ge 0} 2^{(n+3)/2} e_{n+3}(\mathcal{E}) \\ &\le L \sum_{n \ge 0} 2^{n/2} \Big(\sum_{k \le n} 2^{k-n} a_{2^k} \Big) \\ &\le L \sum_{k \ge 0} 2^k a_{2^k} \sum_{n \ge k} 2^{-n/2} \\ &\le L \sum_{k \ge 0} 2^{k/2} a_{2^k} .\end{aligned}$$

Since $\mathcal{E}$ is contained in the ball centered at the origin with radius a_1, we have $e_n(\mathcal{E}) \le a_1$ for each n. The result follows. □

2.6 Proof of the Fundamental Partitioning Theorem

In this section we prove Theorem 2.3.16.

Theorem 2.6.1. *Assume that on the metric space (T, d) there exists a decreasing sequence of functionals $(F_n)_{n \ge 0}$ that satisfies the growth condition of Definition 2.3.10. Then we can find an increasing sequence of partitions $(\mathcal{A}_n)$ with* $\operatorname{card} \mathcal{A}_n \le N_{n+1}$ *and*

$$\sup_{t \in T} \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \le \frac{Lr}{c^*} F_0(T) + Lr\Delta(T) . \tag{2.124}$$

This is not exactly Theorem 2.3.16 because here we have $\operatorname{card} \mathcal{A}_n \le N_{n+1}$ rather than $\operatorname{card} \mathcal{A}_n \le N_n$, but Theorem 2.3.16 follows by combining Theorem 2.6.1 with Lemma 2.3.5.

Replacing F_n by F_n/c^* it suffices to consider the case $c^* = 1$, so we assume this condition throughout this section.

Before going into the details let us first explain the principle of the construction. We construct the increasing sequence $(\mathcal{A}_n)$ of partitions by induction, starting of course with $\mathcal{A}_0 = \{T\}$. Together with $C \in \mathcal{A}_n$, we will construct a point $t_{n,C}$ of T, and an integer $j_n(C)$ in $\mathbb{Z}$. We assume

$$C \subset B(t_{n,C}, r^{-j_n(C)}) , \tag{2.125}$$

so that in particular

$$\Delta(C) \le 2r^{-j_n(C)} . \tag{2.126}$$

Thus, we may think of $j_n(C)$ as keeping track of the diameter of C. More accurately, $j_n(C)$ keeps track of a convenient upper bound for the diameter of C, as it may well happen that $\Delta(C)$ is much smaller than $2r^{-j_n(C)}$. We do *not* require that $t_{n,C}$ belongs to C.

To start the construction, we set $\mathcal{A}_0 = \{T\}$, and we choose any point $t_{0,T} \in T$. We then take for $j_0(T)$ the largest possible integer such that $T \subset B(t_{0,T}, r^{-j_0(T)})$, so that

$$r^{-j_0(T)} \le r\Delta(T) . \tag{2.127}$$

Let us now assume that for a certain $n \ge 0$ we have already constructed the partition $\mathcal{A}_n$ with $\operatorname{card} \mathcal{A}_n \le N_{n+1}$. To construct $\mathcal{A}_{n+1}$ we will split each set of $\mathcal{A}_n$ in at most N_{n+1} pieces according to Lemma 2.6.2 below. Since $N_{n+1}^2 \le N_{n+2}$ we will have $\operatorname{card} \mathcal{A}_{n+1} \le N_{n+2}$, and in this manner we will construct the corresponding increasing sequence of partitions $\mathcal{A}_n$.

All the magic of course is in the procedure by which we will split a given element of $\mathcal{A}_n$ into pieces and in the information that we gather while doing so. To describe this procedure, let us fix $C \in \mathcal{A}_n$, and let $j = j_n(C)$.

Lemma 2.6.2 (The Decomposition Lemma). *Consider a subset C of T, an integer $n \ge 0$ and $j \in \mathbb{Z}$. Let $m = N_{n+1}$. Assume that for a certain $t_C \in T$ we have $C \subset B(t_C, r^{-j})$. Then we can find $m' \le m$ and a partition $(A_\ell)_{\ell \le m'}$ such that for each $\ell \le m'$ we have* **either**

$$\exists t_\ell \in C \ , \ A_\ell \subset B(t_\ell, r^{-j-1}) \ , \tag{2.128}$$

or else

$$r^{-j-1} 2^{n/2-1} + \sup_{t \in A_\ell} F_{n+1}(A_\ell \cap B(t, r^{-j-2})) \le F_n(C) . \tag{2.129}$$

Thus we split C into two kinds of pieces. Those that satisfy (2.128) are of "smaller diameter" than C itself. For those that satisfy (2.129), we gain some (still mysterious) control on the behavior of the functionals F_n. Two noticeable features of this proof are that it is "algorithmic" (the construction is obtained by repeating a basic simple step until the entire set C has been used up) and "greedy" in that the basic simple step maximizes some simple measure of "gain".

Proof. The proof will show in fact that for $\ell < m$ the set A_ℓ satisfies (2.128) and that if $\ell = m = m'$ the set $A_\ell = A_m$ satisfies (2.129). (The present formulation is motivated by pedagogical reason, as it makes the exposition easier in more complicated cases.) To avoid being distracted by secondary issues, let us first assume that T is finite. By induction over $1 \le \ell \le m = N_{n+1}$ we construct points $t_\ell \in C$ and sets $A_\ell \subset C$ as follows.

First, we set $D_0 = C$ and we choose t_1 in C such that

$$F_{n+1}(C \cap B(t_1, r^{-j-2})) = \sup_{t \in C} F_{n+1}(C \cap B(t, r^{-j-2})) \,. \tag{2.130}$$

We then set $A_1 = C \cap B(t_1, r^{-j-1})$. The idea is simply that "we take the largest possible piece of C" (it is in this sense that the method is "greedy"). The reader notices that the radius of the balls in (2.130) is r^{-j-2} while it is r^{-j-1} in the definition of A_1. This is the main idea of the proof. A "large piece" of C is a piece of the type $A_1 = C \cap B(t_1, r^{-j-1})$ for which $F_{n+1}(C \cap B(t_1, r^{-j-2}))$ (rather than $F_{n+1}(A_1)$) is large. This construction is perfectly appropriate in order to use the growth condition of Definition 2.3.10, as it naturally creates well separated "large" pieces (of which $C \cap B(t_1, r^{-j-2})$ is the first one). The drawback of the construction is that the information we produce "skips a level" since it pertains to smaller balls than those we would like (with radius r^{-j-2} rather than r^{-j-1}), and the key point of the proof will be to show that we can at some stage recover the information about the "skipped level".

To continue the construction, assume now that $t_1, \ldots, t_\ell$ and $A_1, \ldots, A_\ell$ have already been constructed, and set $D_\ell = C \setminus \bigcup_{1 \le p \le \ell} A_p$. If $D_\ell = \emptyset$, we set $m' = \ell$ and the construction stops. Otherwise, we choose $t_{\ell+1}$ in D_ℓ such that

$$F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) = \sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, r^{-j-2})) \,. \tag{2.131}$$

We set $A_{\ell+1} = D_\ell \cap B(t_{\ell+1}, r^{-j-1})$ and we continue in this manner until either we stop or we construct

$$D_{m-1} = C \setminus \bigcup_{\ell < m} A_\ell \,.$$

If D_{m-1} is empty, the construction is finished. Otherwise we set $A_m = D_{m-1}$, so that $A_1, \ldots, A_m$ form a partition of C. In this manner we have partitioned C in at most m pieces.

If $\ell < m$ it is obvious by construction that (2.128) holds, so that to finish the proof it suffices to show that (2.129) holds for $\ell = m$. The proof relies on the growth condition. (Let us observe for future use that it actually suffices for the proof that the growth condition holds whenever a is of the type $a = r^{-j'-1}$ for a certain $j' \in \mathbb{Z}$, and that other values of a are not needed.) Then (2.73) rewrites as

$$\forall \ell \le m \,,\, t_\ell \in B(s, r^{-j}) \,;\, \forall \ell \,,\, \ell' \le m \,,\, \ell \ne \ell' \Rightarrow d(t_\ell, t_{\ell'}) \ge r^{-j-1} \,, \tag{2.132}$$

and the content of the growth condition is that this implies (since $c^* = 1$)

$$\begin{aligned} &\forall \;\ell \le m \,,\, H_\ell \subset B(t_\ell, r^{-j-2}) \\ &\Rightarrow F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge r^{-j-1} 2^{n/2} + \min_{\ell \le m} F_{n+1}(H_\ell) \,. \end{aligned} \tag{2.133}$$

Let us construct a point $t_m \in A_m = D_{m-1}$ as in (2.131) for $\ell = m-1$. All the points $(t_\ell)_{\ell \le m}$ belong to $C \subset B(t_C, r^{-j})$. For $\ell < m$ we have by construction

$$t_{\ell+1} \in D_\ell = C \setminus \bigcup_{1 \le p \le \ell} A_p = C \setminus \bigcup_{1 \le p \le \ell} B(t_p, r^{-j-1}) ,$$

and therefore $d(t_{\ell+1}, t_p) \ge r^{-j-1}$ for $p \le \ell$. Consequently these points satisfy (2.132) for $s = t_C$, and therefore (2.133) holds for $H_\ell = D_{\ell-1} \cap B(t_\ell, r^{-j-2})$, where we recall that $D_0 = C$. Since $H_\ell \subset C$, we obtain

$$F_n(C) \ge F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge r^{-j-1} 2^{n/2} + \min_{\ell \le m} F_{n+1}(H_\ell) . \tag{2.134}$$

Now, it follows from (2.131) that for $1 \le \ell \le m-1$

$$\begin{aligned}\sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, r^{-j-2})) &\le F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) \\ &= F_{n+1}(H_{\ell+1}) ,\end{aligned}$$

and (2.130) implies that this is also true when $\ell = 0$. Since the sequence (D_ℓ) decreases, this implies that for $0 \le \ell < m$ we have

$$\sup_{t \in D_{m-1}} F_{n+1}(D_{m-1} \cap B(t, r^{-j-2})) \le F_{n+1}(H_{\ell+1})$$

and therefore

$$\sup_{t \in D_{m-1}} F_{n+1}(D_{m-1} \cap B(t, r^{-j-2})) \le \min_{1 \le \ell \le m} F_{n+1}(H_\ell) .$$

Combining with (2.134) we finally obtain (since $A_m = D_{m-1}$)

$$r^{-j-1} 2^{n/2} + \sup_{t \in A_m} F_{n+1}(A_m \cap B(t, r^{-j-2})) \le F_n(C) , \tag{2.135}$$

and this finishes the proof when T is finite. When T need not be finite, we set $\epsilon = r^{-j-1} 2^{n/2-1}$ and we replace (2.131) by

$$F_{n+1}(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) \ge \sup_{t \in D_\ell} F_{n+1}(D_\ell \cap B(t, r^{-j-2})) - \epsilon , \tag{2.136}$$

and rather than (2.135) we reach

$$r^{-j-1} 2^{n/2} + \sup_{t \in A_m} F_{n+1}(A_m \cap B(t, r^{-j-2})) \le F_n(C) + \epsilon .$$

Recalling the value of ϵ finishes the proof. □

We now continue the construction proving Theorem 2.6.1. We split the set $C \in \mathcal{A}_n$ into at most m pieces using the Decomposition Lemma (Lemma 2.6.2), and we consider one of these pieces A.

If $A = A_\ell$ satisfies (2.128), we define $j_{n+1}(A) = j + 1 = j_n(C) + 1$ and $t_{n+1,A} = t_\ell$, so that

$$A = A_\ell \subset B(t_\ell, r^{-j-1}) = B(t_{n+1,A}, r^{-j_{n+1}(A)}) \ .$$

Let us stress for further use that in that case $t_{n+1,A} \in C$.

If $A = A_\ell$ satisfies (2.129), we define instead $j_{n+1}(A) = j (= j_n(C))$ and $t_{n+1,A} = t_{n,C}$, so that

$$A \subset C \subset B(t_{n,C}, r^{-j_n(C)}) = B(t_{n+1,A}, r^{-j_{n+1}(A)}) \ .$$

This completes the basic procedure and the construction, and we turn to the proof of (2.124). First we observe that for any $t \in T$, (2.126) implies

$$\sum_{n \geq 0} 2^{n/2} \Delta(A_n(t)) \leq 2 \sum_{n \geq 0} r^{-j_n(A_n(t))} 2^{n/2} \ , \tag{2.137}$$

and our objective is to bound the right-hand side. We fix t in T once and for all. It turns out that in the right-hand side of (2.137) only certain terms really contribute. We develop this idea in the next lemma. The basic observation is simply that the sum of a geometric series can be basically bounded by either the first or the last term of the series.

Lemma 2.6.3. *Consider numbers $(a_n)_{n \geq 0}$, $a_n \geq 0$, and assume $\sup_n a_n < \infty$. Consider $\alpha > 1$ and define*

$$I = \left\{k \geq 0 \ ; \ \forall n \geq 0 \ , n \neq k \ , \ a_n < a_k \alpha^{|k-n|}\right\} . \tag{2.138}$$

Then

$$\sum_{n \geq 0} a_n \leq \frac{2\alpha}{\alpha - 1} \sum_{k \in I} a_k \ . \tag{2.139}$$

Proof. Let us write $n \prec k$ when $a_k \geq a_n \alpha^{|n-k|}$. This relation is a partial order: if $n \prec k$ and $k \prec p$ then $a_p \geq a_n \alpha^{|p-k|+|k-n|} \geq a_n \alpha^{|p-n|}$, so that $n \prec p$. Let us observe that the set I defined above is the set of elements k of $\mathbb{N}$ that are maximal, i.e. $k \prec k' \Rightarrow k = k'$. Since we assume that the sequence (a_n) is bounded, there cannot exist an increasing sequence for the order $\prec$. Consequently, for each n in $\mathbb{N}$ there exists $k \in I$ with $n \prec k$. Then $a_n \leq a_k \alpha^{-|n-k|}$, and therefore

$$\sum_{n \geq 0} a_n \leq \sum_{k \in I} \sum_{n \geq 0} a_k \alpha^{-|k-n|} \leq \frac{2}{1 - \alpha^{-1}} \sum_{k \in I} a_k \ . \qquad \square$$

We go back to the control of the right-hand side of (2.137). We recall that $r \ge 4$. To lighten notation we set $j(n) = j_n(A_n(t))$, and we set $a_n = r^{-j(n)}2^{n/2}$. This sequence is bounded because either $j(n) > j(n-1)$ and then $a(n) \le a(n-1)$, or else $a_n \le F_0(T)$ by (2.129). Consider the set I provided by Lemma 2.6.3 for $\alpha = \sqrt{2}$. We observe the following fundamental relation:

$$k \in I\ ,\ k \ge 1 \Rightarrow j(k-1) = j(k)\ ,\ j(k+1) = j(k) + 1\ . \tag{2.140}$$

Indeed, if $j(k+1) = j(k)$, then $a_{k+1} = \sqrt{2}a_k$, so that $k \notin I$ by the definition of I, and if $j(k-1) = j(k) - 1$ then $a_{k-1} = (r/\sqrt{2})a_k \ge 2a_k$, and again $k \notin I$ by definition of I.

Lemma 2.6.4. *Consider elements* $1 \le k < k'$ *of* I. *Then*

$$\frac{1}{4r}a_k \le F_{k-1}(A_{k-1}(t)) - F_{k'+1}(A_{k'+1}(t))\ . \tag{2.141}$$

Proof. It follows from (2.125) that if we define $A^* := A_{k'+1}(t)$ and $t^* := t_{k'+1,A^*}$ then

$$A^* \subset B(t^*, r^{-j(k'+1)})\ .$$

Moreover, since $k' \in I$ we have $j(k'+1) = j(k') + 1$, and as noted we have $t^* \in A_{k'}(t) \subset A_k(t)$. Also $j(k') \ge j(k+1)$, and $j(k+1) = j(k) + 1$ since $k \in I$ and $k \ge 1$. Consequently, $j(k'+1) \ge j(k) + 2$ and therefore

$$A^* \subset A_k(t) \cap B(t^*, r^{-j(k)-2})\ .$$

Moreover, since $k \in I$ and $k \ge 1$, we have $j(k-1) = j(k)$. By construction, (2.129) used for $n = k-1$ and $C = A_n(t) = A_{k-1}(t)$ implies

$$r^{-j(k)-1}2^{(k-1)/2-1} + \sup_{u \in A_k(t)} F_k(A_k(t) \cap B(u, r^{-j(k)-2})) \le F_{k-1}(A_{k-1}(t))\ , \tag{2.142}$$

so that since $r^{-j(k)-1}2^{(k-1)/2-1} \ge a_k/4r$, (2.142) implies

$$\frac{1}{4r}a_k + F_k(A^*) \le F_{k-1}(A_{k-1}(t))\ . \tag{2.143}$$

Since $k \le k'$ and since the sequence (F_n) decreases, we have $F_k(A^*) \ge F_{k'+1}(A_{k'+1}(t))$ and (2.143) proves (2.141). □

Proof of Theorem 2.6.1. Let

$$x(n) = F_n(A_n(t))\ ,$$

so that (2.141) implies

$$\frac{1}{4r}a_k \le x(k-1) - x(k'+1)\ .$$

Moreover, since the sequence (F_n) of functionals decreases, and since the sequence of sets $(A_n(t))$ decreases, the sequence $(x(n))$ decreases.

Let us assume first that I is infinite and let us enumerate I as an increasing sequence $(k_i)_{i\geq 1}$. For $i \geq 1$ let us define $y(i) = x(k_i)$, so that the sequence $(y(i))$ decreases since the sequence $(x(n))$ decreases. For $i \geq 2$ we have $k_i - 1 \geq k_{i-1}$ so that $x(k_i - 1) \leq y(i-1)$. Similarly, $x(k_{i+1}+1) \geq x(k_{i+2}) = y(i+2)$. Since $k_i \geq 1$ (2.141) implies

$$\frac{1}{4r} a_{k_i} \leq y(i-1) - y(i+2) \,. \tag{2.144}$$

Since $y(i) \leq x(0) = F_0(A_0(t)) = F_0(T)$, summation of the inequalities (2.144) yields

$$\sum_{i\geq 2} a_{k_i} \leq LrF_0(T) \,. \tag{2.145}$$

It only remains to control a_{k_1}. When $k_1 = 0$, then $a_0 = r^{-j_0(T)} \leq r\Delta(T)$. Otherwise $k_1 \geq 1$, and then (2.143) implies $a_{k_1} \leq 4rF_0(T)$. This completes the proof when I is infinite. Only small changes are required when I is finite, and this is left to the reader. □

2.7 A General Partitioning Scheme

Theorem 2.6.1 admits considerable generalizations, which turn out to be very useful. These generalizations admit basically the same proof as Theorem 2.6.1. They require an extension of the "growth condition" of Definition 2.3.10. We consider a function

$$\theta : \mathbb{N} \cup \{0\} \to \mathbb{R}^+ \,.$$

Definition 2.3.10 corresponds to the case $\theta(n) = 2^{(n-1)/2}$.

The condition we are about to state involves two new parameters β and τ. Definition 2.3.10 corresponds to the case $\beta = 1$ and $\tau = 1$. The parameter $\tau \in \mathbb{N}$ is of secondary importance. The larger τ, the more "room there is".

Let us recall that since Definition 2.3.8, we say that sets $(H_\ell)_{\ell\leq m}$ are (a, r) separated if there exist $s, t_1, \ldots, t_m$ for which

$$\forall \ell \leq m \,,\, t_\ell \in B(s, ar) \,;\, \forall \ell \,,\, \ell' \leq m \,,\, \ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a \,, \tag{2.146}$$

and

$$\forall \ell \leq m \,,\, H_\ell \subset B(t_\ell, a/r) \,.$$

Definition 2.7.1. *We say that the functionals F_n satisfy the* growth condition *if for a certain integer $\tau \geq 1$, and for certain numbers $r \geq 4$ and $\beta > 0$, the following holds true. Consider $a > 0$, any integer $n \geq 0$, and set*

$m = N_{n+\tau}$. *Then, whenever the subsets* $(H_\ell)_{\ell \le m}$ *of* T *are* (a, r) *separated in the sense of Definition 2.3.8, then*

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge a^\beta \theta(n+1) + \min_{\ell \le m} F_{n+1}(H_\ell) . \tag{2.147}$$

In the right-hand side of (2.147), the term $a^\beta\theta(n+1)$ is the product of a^β, which accounts for the scale at which the sets H_ℓ are separated, and of the term $\theta(n+1)$, which accounts for the number of these sets. The "linear case" $\beta = 1$ is by far the most important. The role of the parameter τ is to give some room. When τ is large, there are more sets and it should be easier to prove (2.147).

The reader noticed that we call "growth condition" both the condition of Definition 2.3.10 and the more general condition of Definition 2.7.1. It is not practical to give different names to these conditions because we shall eventually consider several more conditions in the same spirit. We shall always make precise to which condition we refer.

We will assume the following regularity condition for θ. For some $1 < \xi \le 2$, and all $n \ge 0$, we have

$$\xi\theta(n) \le \theta(n+1) \le \frac{r^\beta}{2}\theta(n) . \tag{2.148}$$

When $\theta(n) = 2^{(n-1)/2}$, (2.148) holds for $\xi = \sqrt{2}$. The main result of this section is as follows.

Theorem 2.7.2. *Under the preceding conditions we can find an increasing sequence* $(\mathcal{A}_n)$ *of partitions of* T *with* $\operatorname{card} \mathcal{A}_n \le N_{n+\tau}$ *such that*

$$\sup_{t \in T} \sum_{n \ge 0} \theta(n) \Delta(A_n(t))^\beta \le L(2r)^\beta \Big(\frac{F_0(T)}{\xi - 1} + \theta(0)\Delta(T)^\beta\Big) . \tag{2.149}$$

In all the situations we shall consider, it will be true that $F_0(\{t_1, t_2\}) \ge \theta(0) d(t_1, t_2)^\beta$ for any points t_1 and t_2 of T. (Since $F_1(H) \ge 0$ for any set H, this condition is essentially weaker in spirit than (2.147) for $n = 0$.) Then $\theta(0)\Delta(T)^\beta \le F_0(T)$.

The sequence $(\mathcal{A}_n)$ of Theorem 2.7.2 need not be admissible because $\operatorname{card} \mathcal{A}_n$ is too large. To construct good admissible sequences we will combine Theorem 2.7.2 with Lemma 2.3.5.

Not surprisingly, the key to the proof of Theorem 2.7.2 is the following, which is simply an adaptation of Lemma 2.6.2 to the present setting.

Lemma 2.7.3. *If the functionals* F_n *satisfy the growth condition, then, given integers* $n \ge 0$ *and* $j \in \mathbb{Z}$*, for any subset* C *of* T *such that*

$$\exists s \in T \; ; \; C \subset B(s, r^{-j}) ,$$

we can find a partition $(A_\ell)_{\ell \le m'}$ *of* C*, where* $m' \le m = N_{n+\tau}$*, such that for each* $\ell \le m'$ *we have* **either**

$$\exists t_\ell \in C \; ; \; A_\ell \subset B(t_\ell, r^{-j-1}) \; , \tag{2.150}$$

or else

$$\frac{1}{2} r^{-\beta(j+1)} \theta(n+1) + \sup_{t \in A_\ell} F_{n+1}(A_\ell \cap B(t, r^{-j-2})) \le F_n(C) \; . \tag{2.151}$$

The proof is nearly identical to the proof of Lemma 2.6.2 so it is left to the reader.

Proof of Theorem 2.7.2. We construct the sequence of partitions $(\mathcal{A}_n)$ and $t_{n,A}, j_n(A)$ for $A \in \mathcal{A}_n$ as in Theorem 2.6.1, using the Decomposition Lemma at each step. Since, however, there is no point in repeating the same proof, we will organize the argument differently.

The basic idea is that (since we have "skipped levels") we must keep track not only of what we do in the current step of the construction but also of what we do in the previous step. This is implemented by keeping track for each set $C \in \mathcal{A}_n$ of three different "measures of its size", namely

$$a_i(C) = \sup_{t \in C} F_n(C \cap B(t, r^{-j_n(C)-i})) \; ,$$

for $i = 0, 1, 2$. This quantity depends also on n, in the sense that if $C \in \mathcal{A}_n$ and $C \in \mathcal{A}_{n+1}$ than $a_i(C)$ need not be the same whether we see C as an element of $\mathcal{A}_n$ or of $\mathcal{A}_{n+1}$. To lighten notation we shall not indicate this dependence. For technical reasons keeping track of the values $a_j(C)$ is not very convenient, and instead we will keep track of three quantities $b_j(C)$ for $j = 0, 1, 2$, where $b_j(C) (\ge a_j(C))$ is a kind of "regularized version" of $a_j(C)$. (These quantities also depend on n.) We rewrite the conditions $a_j(C) \le b_j(C)$:

$$F_n(C) \le b_0(C) \tag{2.152}$$

$$\forall t \in C \, , \; F_n(C \cap B(t, r^{-j_n(C)-1})) \le b_1(C) \tag{2.153}$$

$$\forall t \in C \; , \; F_n(C \cap B(t, r^{-j_n(C)-2})) \le b_2(C) \; . \tag{2.154}$$

We will also require the following two technical conditions:

$$b_1(C) \le b_0(C) \tag{2.155}$$

and

$$b_0(C) - \frac{1}{2} r^{-\beta(j_n(C)+1)} \theta(n) \le b_2(C) \le b_0(C) \; . \tag{2.156}$$

Moreover, the quantities b_i will satisfy the following fundamental relation: if $n \ge 0$, $A \in \mathcal{A}_{n+1}$, $C \in \mathcal{A}_n$, $A \subset C$, then

$$\sum_{0\le i\le 2} b_i(A) + \frac{1}{2}(1-\frac{1}{\xi})r^{-\beta(j_{n+1}(A)+1)}\theta(n+1)$$
$$\le \sum_{0\le i\le 2} b_i(C) + \frac{1}{4}(1-\frac{1}{\xi})r^{-\beta(j_n(C)+1)}\theta(n)\ . \tag{2.157}$$

As we shall show below in the last step of the proof, summation of these relations over $n \ge 0$ implies (2.149). Let us make a first comment about (2.157). When $j_{n+1}(A) > j_n(C)$, since $r^{-\beta}\theta(n+1) \le \theta(n)/2$ by (2.148) we have

$$r^{-\beta(j_{n+1}(A)+1)}\theta(n+1) \le \frac{1}{2} r^{-\beta(j_n(C)+1)}\theta(n)\ , \tag{2.158}$$

and in that case (2.157) is satisfied as soon as $\sum_{0\le i\le 2} b_i(A) \le \sum_{0\le i\le 2} b_i(C)$. This is related to the idea, already made explicit in the proof of Theorem 2.6.1, that this case "does not matter". It will be harder to satisfy (2.157) when $j_{n+1}(A) = j_n(C)$.

Before we go into the details of the construction, and of the recursive definition of the numbers $b_j(C)$, we explain how this proof was found. It is difficult here to give a "big picture" why the approach works. We simply gather in each case the available information to make sensible definitions. Analysis of these definitions in the two main cases below will convince the reader that this is exactly how we have proceeded. Of course, when starting such an approach, it is difficult to know whether it will succeed, so we simply crossed our fingers and tried. The overall method seems powerful.

We now define the numbers $b_j(C)$ by induction over n. We start with

$$b_0(T) = b_1(T) = b_2(T) = F_0(T)\,.$$

For the induction step from n to $n+1$, let us first consider the case where, when applying the Decomposition Lemma, the set $A = A_\ell$ satisfies (2.150). We then define

$$b_0(A) = b_2(A) = b_1(C)\,,\ b_1(A) = \min(b_1(C), b_2(C))\ .$$

Relations (2.155) and (2.156) for A are obvious. To prove (2.152) for A, we write

$$\begin{aligned} F_{n+1}(A) &\le F_{n+1}(C \cap B(t_\ell, r^{-j-1})) \\ &\le F_n(C \cap B(t_\ell, r^{-j-1})) \le b_1(C) = b_0(A)\,, \end{aligned}$$

using (2.153) for C. In a similar manner, we have, if $t \in A$, and since $j_{n+1}(A) = j+1$,

$$\begin{aligned} F_{n+1}(A \cap B(t, r^{-j_{n+1}(A)-1})) &\le F_{n+1}(C \cap B(t, r^{-j-2})) \\ &\le F_n(C \cap B(t, r^{-j-2})) \\ &\le \min(b_1(C), b_2(C)) = b_1(A)\ , \end{aligned}$$

and this proves (2.153) for A. Also, (2.154) for A follows from (2.152) for A since $b_2(A) = b_0(A)$.

To prove (2.157), we observe that

$$\sum_{0 \le i \le 2} b_i(A) \le 2b_1(C) + b_2(C) \le \sum_{0 \le i \le 2} b_i(C) \ , \tag{2.159}$$

since $b_1(C) \le b_0(C)$ by (2.155). We observe that, since $j_{n+1}(A) = j_n(C) + 1$, (2.158) holds and combining with (2.159) this proves (2.157).

Next, we consider the case where $A = A_\ell$ satisfies (2.151). We define

$$b_0(A) = b_0(C) \, ; \, b_1(A) = b_1(C) \ ; \ b_2(A) = b_0(C) - \frac{1}{2} r^{-\beta(j+1)} \theta(n+1) \ .$$

It is obvious that A and $n+1$ in place of C and n satisfy the relations (2.125), (2.155) and (2.156). The relations (2.152) and (2.153) for A follow from the fact that similar relations hold for C rather than A, that $F_{n+1} \le F_n$, and that the functional F_{n+1} is increasing. Moreover (2.154) follows from (2.151) and (2.152).

To prove (2.157), we observe that by definition

$$\begin{aligned}
&\sum_{0 \le i \le 2} b_i(A) + \frac{1}{2}(1 - \frac{1}{\xi}) r^{-\beta(j+1)} \theta(n+1) \\
&= 2b_0(C) + b_1(C) - \frac{1}{2\xi} r^{-\beta(j+1)} \theta(n+1) \\
&\le 2b_0(C) + b_1(C) - \frac{1}{2} r^{-\beta(j+1)} \theta(n) \ ,
\end{aligned} \tag{2.160}$$

using the regularity condition (2.148) on $\theta(n)$ in the last inequality. But (2.156) implies

$$b_0(C) \le b_2(C) + \frac{1}{2} r^{-\beta(j+1)} \theta(n) \ ,$$

so that (2.160) implies (2.157).

We have completed the construction, and we turn to the proof of (2.149). By (2.157), for any t in T, any $n \ge 0$, we have, setting $j_n(t) = j_n(A_n(t))$

$$\begin{aligned}
&\sum_{0 \le i \le 2} b_i(A_{n+1}(t)) + \frac{1}{2}(1 - \frac{1}{\xi}) r^{-\beta(j_{n+1}(t)+1)} \theta(n+1) \\
&\le \sum_{0 \le i \le 2} b_i(A_n(t)) + \frac{1}{4}(1 - \frac{1}{\xi}) r^{-\beta(j_n(t)+1)} \theta(n) \ .
\end{aligned}$$

Since $b_i(T) = F_0(T)$ and since $b_i(A) \ge 0$ by (2.152) to (2.154), summation of these relations for $0 \le n \le q$ implies

$$\frac{1}{2}(1 - \frac{1}{\xi}) \sum_{0 \le n \le q} r^{-\beta(j_{n+1}(t)+1)} \theta(n+1) \tag{2.161}$$

$$\le 3F_0(T) + \frac{1}{4}(1 - \frac{1}{\xi}) \sum_{0 \le n \le q} r^{-\beta(j_n(t)+1)} \theta(n) \tag{2.162}$$

and thus

$$\frac{1}{4}(1-\frac{1}{\xi})\sum_{0\le n\le q} r^{-\beta(j_n(t)+1)}\theta(n) \le 3F_0(T) + \frac{1}{4}(1-\frac{1}{\xi})r^{-\beta(j_0(T)+1)}\theta(0) \; .$$

By (2.125), we have $\Delta(A_n(t)) \le 2r^{-j_n(t)}$, and the choice of $j_0(T)$ implies $r^{-j_0(T)-1} \le \Delta(T)$ so that, since $\xi \le 2$,

$$\sum_{n\ge 0}\theta(n)\Delta^\beta(A_n(t)) \le \frac{L(2r)^\beta}{\xi-1}(F_0(T)+\Delta^\beta(T)\theta(0)) \; . \qquad \square$$

Exercise 2.7.4. Write in complete detail the proof of Theorem 2.7.2 along the lines of the proof of Theorem 2.6.1.

To illustrate how the parameter τ in Theorem 2.7.2 may be used we give another proof of Theorem 2.3.1. Recall the definition 2.2.19 of $\gamma_\alpha(T,d)$.

Second proof of Theorem 2.3.1. We will use Theorem 2.7.2 with $r = 4$, $\beta = 1$ and $\tau = \tau' + 1$. For $n \ge 0$ and a subset A of U we define

$$F_n(A) = \sup_{t\in A}\sum_{k\ge n} 2^{k/\alpha} d(t, T_k) \; .$$

In order to check (2.147), consider $m = N_{n+\tau'+1}$, and assume that there exist points $t_1, \dots, t_m$ of U such that

$$1 \le \ell < \ell' \le m \Rightarrow d(t_\ell, t_{\ell'}) \ge a \, .$$

Consider then subsets $H_1, \dots, H_m$ of U with $H_\ell \subset B(t_\ell, a/4)$. By definition of F_{n+1}, given any $\epsilon > 0$, we can find $u_\ell \in H_\ell$ such that

$$\sum_{k\ge n+1} 2^{k/\alpha} d(u_\ell, T_k) \ge F_{n+1}(H_\ell) - \epsilon \; .$$

Since $d(t_\ell, t_{\ell'}) \ge a$ for $\ell \ne \ell'$, the open balls $B(t_\ell, a/2)$ are disjoint. Since there are $N_{n+\tau'+1}$ of them, whereas $\operatorname{card} T_n \le N_{n+\tau'}$, one of these balls cannot meet T_n. Thus there is $\ell \le m$ with $d(t_\ell, T_n) \ge a/2$. Since $u_\ell \in H_\ell \subset B(t_\ell, a/4)$, the inequality $d(u_\ell, T_n) \ge a/4$ holds, and

$$\begin{aligned}\sum_{k\ge n} 2^{k/\alpha} d(u_\ell, T_k) &\ge 2^{n/\alpha}\frac{a}{4} + \sum_{k\ge n+1} 2^{k/\alpha} d(u_\ell, T_k) \\ &\ge 2^{n/\alpha-2}a + F_{n+1}(H_\ell) - \epsilon \; .\end{aligned}$$

Since $u_\ell \in H_\ell$ this shows that

$$F_n\Big(\bigcup_{p\le m} H_p\Big) \ge 2^{n/\alpha-2}a + F_{n+1}(H_\ell) - \epsilon \, ,$$

and since ϵ is arbitrary, this proves that (2.147) holds with $\theta(n+1) = 2^{n/\alpha-2}$. (Condition (2.148) holds only when $\alpha \geq 1$, which is the most interesting case. We leave to the reader to complete the case $\alpha < 1$ by using a different value of r.) We have $F_0(U) \leq S$, and since $d(t, T_0) \leq S$ for $t \in U$, and $\operatorname{card} T_0 = 1$, we have $\Delta(U) \leq 2S$. To finish the proof one simply applies Theorem 2.7.2 and Lemma 2.3.5. □

We now collect some simple facts, the proof of which will also serve as another (easy) application of Theorem 2.7.2.

Theorem 2.7.5. *(a) If U is a subset of T, then*

$$\gamma_\alpha(U,d) \leq \gamma_\alpha(T,d) \,.$$

(b) If $f : (T,d) \to (U,d')$ is onto and satisfies

$$\forall x, y \in T \,,\; d'(f(x), f(y)) \leq A d(x,y) \,,$$

for some constant A, then

$$\gamma_\alpha(U,d') \leq K(\alpha) A \gamma_\alpha(T,d) \,.$$

(c) We have

$$\gamma_\alpha(T,d) \leq K(\alpha) \sup \gamma_\alpha(F,d) \,, \tag{2.163}$$

where the supremum is taken over $F \subset T$ and F finite.

It seems plausible that with different methods than those used below one should be able to obtain (b) and (c) with $K(\alpha) = 1$, although there is little motivation to do this.

Proof. Part (a) is obvious. To prove (b) we consider an admissible sequence of partitions $\mathcal{A}_n$ with $\sup_t \sum_{n\geq 0} 2^{n/\alpha} \Delta(A_n(t), d) \leq 2\gamma_a(T,d)$. Consider then sets $T_n \subset T$ with $\operatorname{card} T_n \leq N_n$ and $\operatorname{card}(T_n \cap A) = 1$ for each $A \in \mathcal{A}_n$ so that $\sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} d(t, T_n) \leq 2\gamma_\alpha(T,d)$. We observe that $\sup_{s\in U} \sum_{n\geq 0} 2^{n/\alpha} d'(s, f(T_n)) \leq 2A\gamma_\alpha(T,d)$, and we apply Theorem 2.3.1.

To prove (c) we essentially repeat the argument in the proof of Theorem 2.3.14. We define

$$\gamma_{\alpha,n}(T,d) = \inf \sup_{t\in T} \sum_{k\geq n} 2^{k/\alpha} \Delta(A_k(t))$$

where the infimum is over all admissible sequences $(\mathcal{A}_k)$. We consider the functionals

$$F_n(A) = \sup \gamma_{\alpha,n}(G,d)$$

where the supremum is over all finite subsets G of A. We will use Theorem 2.7.2 with $\beta = 1$, $\theta(n+1) = 2^{n/\alpha-1}$, $\tau = 1$, and $r = 4$. (As in the proof of Theorem 2.3.1 this works only for $\alpha \geq 1$, and the case $\alpha < 1$ requires a different choice of r.) To prove (2.147), consider $m = N_{n+1}$ and consider points

$(t_\ell)_{\ell \le m}$ of T, with $d(t_\ell, t_{\ell'}) \ge a$ if $\ell \ne \ell'$. Consider sets $H_\ell \subset B(t_\ell, a/4)$ and $c < \min_{\ell \le m} F_{n+1}(H_\ell)$. For $\ell \le m$, consider finite sets $G_\ell \subset H_\ell$ with $\gamma_{\alpha,n+1}(G_\ell, d) > c$, and $G = \bigcup_{\ell \le m} G_\ell$. Consider an admissible sequence $(\mathcal{A}_n)$ of G, and

$$I = \{\ell \le m \; ; \; \exists A \in \mathcal{A}_n \, , \, A \subset G_\ell\}$$

so that, since the sets G_ℓ for $\ell \le m$ are disjoint, we have $\operatorname{card} I \le N_n$, and thus there exists $\ell \le m$ with $\ell \notin I$. Then for $t \in G_\ell$, we have $A_n(t) \not\subset G_\ell$, so $A_n(t)$ meets a ball $B(t_{\ell'}, a/4)$ for $\ell \ne \ell'$, and hence $\Delta(A_n(t)) \ge a/2$; so that

$$\sum_{k \ge n} 2^{k/\alpha} \Delta(A_k(t)) \ge \frac{a}{2} 2^{n/\alpha} + \sum_{k \ge n+1} 2^{k/\alpha} \Delta(A_k(t) \cap G_\ell)$$

and hence

$$\sup_{t \in G_\ell} \sum_{k \ge n} 2^{k/\alpha} \Delta(A_k(t)) \ge a 2^{n/\alpha - 1} + \gamma_{\alpha,n+1}(G_\ell, d) \, .$$

Since the admissible sequence $(\mathcal{A}_n)$ is arbitrary, we have shown that

$$\gamma_{\alpha,n}(G, d) \ge a 2^{n/\alpha - 1} + c$$

and thus

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge a 2^{n/\alpha - 1} + \min_{\ell \le m} F_{n+1}(H_\ell) \, ,$$

which is (2.147). Finally, we have $F_0(T) = \sup \gamma_\alpha(G, d)$, where the supremum is over all finite subsets G of T, and since $\Delta(G) \le \gamma_\alpha(G, d)$, we have that $\Delta(T) \le F_0(T)$ and we conclude by Lemma 2.3.5 and Theorem 2.7.2. □

There are many possible variations about the scheme of proof of Theorem 2.7.2. We end this section with such a version. This specialized result will be used only in Section 16.8, and its proof could be omitted at first reading.

There are natural situations, where, in order to be able to prove (2.147), we need to know that $H_\ell \subset B(t_\ell, \eta a)$ where η is very small. In order to apply Theorem 2.7.2, we have to take $r \ge 1/\eta$, which (when $\beta = 1$) produces a loss of a factor $1/\eta$. We will give a simple modification of Theorem 2.7.2 that produces only the loss of a factor $\log(1/\eta)$.

For simplicity, we assume $r = 4, \beta = 1, \theta(n) = 2^{n/2}$ and $\tau = 1$. We consider an integer $s \ge 2$.

Theorem 2.7.6. *Assume that the hypotheses of Theorem 2.7.2 are modified as follows. Whenever $t_1, \ldots, t_m$ are as in (2.146), and whenever $H_\ell \subset B(t_\ell, a4^{-s})$, we have*

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge a 2^{(n+1)/2} + \min_{\ell \le m} F_{n+s}(H_\ell) \, . \tag{2.164}$$

Then there exists an increasing sequence of partitions $(\mathcal{A}_n)$ *in* T *such that* $\operatorname{card}\mathcal{A}_n \le N_{n+1}$ *and*

$$\sup_{t\in T}\sum_{n\ge 0} 2^{n/2}\Delta(A_n(t)) \le Ls(F_0(T)+\Delta(T)) .$$

Now the reader observes that in the last term of (2.164) we have $F_{n+s}(H_\ell)$ rather than the larger quantity $F_{n+1}(H_\ell)$. This will be essential in Section 16.8. It is of course unimportant in the first term of the right-hand side to have the exponent $n+1$ rather than n. We use $n+1$ to mirror (2.73).

Proof. We closely follow the proof of Theorem 2.7.2. For clarity we assume that T is finite. First, we copy the proof of the Decomposition Lemma (Lemma 2.6.2), and rather than obtaining (2.129) (which occurs exactly for $\ell = m$) we now have

$$\frac{1}{2}4^{-j-1}2^{(n+1)/2} + \sup_{t\in A_m} F_{n+s}(A_m \cap B(t, 4^{-j-1-s})) \le F_n(C) . \tag{2.165}$$

Together with each set C in $\mathcal{A}_n$, we construct numbers $b_i(C) \ge 0$ for $0 \le i \le s+1$, such that

$$\forall i\,,\ 1\le i\le s+1\,,\ b_i(C)\le b_0(C)$$

$$b_0(C) \ge b_{s+1}(C) \ge b_0(C) - \frac{1}{2}4^{-j_n(C)-1}2^{n/2}$$

$$F_n(C)\le b_0(C)$$

$$\forall i\,,\ 1\le i\le s+1\,,\ \forall t\in C\,,\ F_{n+i-1}(C\cap B(t,4^{-j_n(C)-i}))\le b_i(C) . \tag{2.166}$$

The reader observes that (2.166) is not a straightforward extension of (2.154), since it involves F_{n+i-1} rather than the larger quantity F_n. We set $b_i(T) = F_0(T)$ for $0\le i\le s+1$. For the induction from n to $n+1$ we consider one of the pieces A of the partition of $C \in \mathcal{A}_n$ and $j = j_n(C)$. If $A = A_m$, we set

$$\forall i\,,\ 0\le i\le s\,,\ b_i(A)=b_i(C)$$

$$b_{s+1}(A) = b_0(A) - \frac{1}{2}4^{-j-1}2^{(n+1)/2} . \tag{2.167}$$

Since $j_{n+1}(A) = j_n(C)$, for $i = s+1$ condition (2.166) for A follows from (2.165) and (2.167) since $F_n(C)\le b_0(C) = b_0(A)$. For $i\le s$, condition (2.166) for A follows from the same condition for C since $F_{n+i}\le F_{n+i-1}$.

If $A = A_\ell$ with $\ell < m$ we then set

$$b_{s+1}(A) = b_1(C)\ ;\ \forall i\le s\,,\ b_i(A)=\min(b_{i+1}(C), b_1(C)) .$$

Since $j_{n+1}(A) = j_n(A)+1$, condition (2.166) for $i = s+1$ and A follows from the same condition for C and $i = 1$, while condition (2.166) for $i\le s$ and A

follows form the same condition for C and $i+1$. Exactly as previously we show in both cases that

$$\sum_{0\le i\le s+1} b_i(A) + \frac{1}{2}(1-\frac{1}{\sqrt{2}})4^{-j_{n+1}(A)-1}2^{(n+1)/2} \le \sum_{0\le i\le s+1} b_i(C) + \frac{1}{4}(1-\frac{1}{\sqrt{2}})4^{-j_n(C)-1}2^{n/2} + \epsilon_{n+1}$$

and we finish the proof in the same manner. □

2.8 Notes and Comments

It seems necessary to say a few words about the history of Gaussian processes. I have heard people saying that the problem of characterizing continuity and boundedness of Gaussian processes goes back (at least implicitly) to Kolmogorov.

The understanding of Gaussian processes was long delayed by the fact that in the most immediate examples the index set is a subset of $\mathbb{R}$ or $\mathbb{R}^n$ and that the temptation to use the special structure of this index set is nearly irresistible. Probably the single most important conceptual progress about Gaussian processes is the realization, in the late sixties, that the boundedness of a (centered) Gaussian process is determined by the structure of the metric space (T,d), where d is the usual distance $d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2}$. It is of course difficult now to realize what a tremendous jump in understanding this was, since this seems so obvious *a posteriori*.

In 1967, R. Dudley obtained the inequality (2.36), which however cannot be reversed in general. (Actually, as R. Dudley pointed out repeatedly, he did not state (2.36). Nonetheless since he performed all the essential steps it seems appropriate to call (2.36) Dudley's bound. It simply does not seem worth the effort to find who deserves the very marginal credit of having stated (2.36) first.) A few years later, X. Fernique proved that in the "stationary case" Dudley's inequality can be reversed [3], i.e. he proved in that case the lower bound of Theorem 2.4.1. This result is historically important, because it was central to the work of Marcus and Pisier [7], [8] who build on it to solve all the classical problems on random Fourier series. A part of their results was presented in Section 3.2. Interestingly, now that the right approach has been found, the proof of Fernique's result is not really easier than that of Theorem 2.4.1.

Another major contribution of Fernique (building on earlier ideas of C. Preston) was an improvement of Dudley's bound based on a new tool called majorizing measures. Fernique conjectured that his inequality was essentially optimal. Gilles Pisier suggested in 1983 that I should work on this conjecture. In my first attempt I proved quite fast that Fernique's conjecture held in the case where the metric space (T, d) is ultrametric. I was quite disappointed to learn that Fernique had already done this, so I was discouraged for a while. In the second attempt, I tried to decide whether a majorizing measure existed on ellipsoids. I had the hope that some simple density with respect to the volume measure would work. It was difficult to form any intuition, and I really struggled in the dark for months. At some point I decided not to use the volume measure, but rather a combination of suitable point masses, and easily found a direct construction of the majorizing measure on ellipsoids. This of course made it quite believable that Fernique's conjecture was true, but I still tried to disprove it. At some point I realized that I did not understand why a direct approach to prove Fernique's conjecture using a partition scheme should fail, while this understanding should be useful to construct a counter example. Once I tried this direct approach, it was only a matter of a few days to prove Fernique's conjecture. Gilles Pisier made two comments about this discovery. The first one was "you are lucky", by which he of course meant that I was lucky that Fernique's conjecture was true, since a counter example would have been of limited interest. I am grateful to this day for his second comment: "I wish I had proved this myself, but I am very glad you did it."

Fernique's concept of majorizing measures is very difficult to grasp at the beginning, and was consequently dismissed by the main body of probabilists as a mere curiosity. (I must admit that I myself did find it very difficult to understand.) However, in 2000, while discussing one of the open problems of this book with K. Ball (be he blessed for his interest in it!) I discovered that one could replace majorizing measures by the totally natural variation on the usual chaining arguments that was presented here. That this was not discovered much earlier is a striking illustration of the inefficiency of the human brain (and of mine in particular).

Some readers wondered why I do not mention Slepian's lemma. Of course this omission is done on purpose and must be explained. Slepian's lemma is very specific to Gaussian processes, and focusing on it seems a good way to guarantee that one will never move beyond these. One notable progress made by the author was to discover the scheme of proof of Proposition 2.4.9 that dispenses with Slepian's lemma, and that we shall use in many situations. Comparison results such as Slepian's lemma are not at the root of results such as the majorizing measure theorem, but rather are (at least qualitatively) a consequence of them. Indeed, if two centered Gaussian processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ satisfy $\mathsf{E}(X_s - X_t)^2 \le \mathsf{E}(Y_s - Y_t)^2$ whenever $s, t \in T$, then (2.80) implies $\mathsf{E}\sup_{t\in T} X_t \le L\mathsf{E}\sup_{t\in T} Y_t$. (Slepian's lemma asserts that this inequality holds with constant $L = 1$.)

It may happen in the construction of Lemma 2.6.2 that $C = A_1$. Thus it may happen in the construction of Theorem 2.6.1 that a same set A belongs both to $\mathcal{A}_n$ and $\mathcal{A}_{n+1}$. When this is the case, the construction shows that one has $j_{n+1}(A) = j_n(A) + 1$. It is therefore incorrect, as was done in the first edition, to use in the construction a number $j(A)$ depending only on A.

I am grateful to J. Lehec for having pointed out this mistake. Fortunately, the only change required in the proofs is to add the proper indexes to the quantities of the type $j(C)$.

In [9] the author presented a particularly simple proof of (an equivalent form of) Theorem 2.4.1. It is also based on a partitioning scheme. For the readers who are familiar with that proof, it might be useful to compare the partitioning scheme of [9] with the partitioning scheme presented here. We shall show that these schemes "produce the same pieces of T", the difference being that these are not gathered to form partitions in the same manner. Consider a metric space (T,d) and a functional $F(H)$ on T. Assume that for a certain number r it satisfies the following growth condition. Given $m \geq 2$, $k \in \mathbb{Z}$ and points $t_1, \ldots, t_m$ of T, with $d(t_\ell, t_{\ell'}) \geq r^{-k}$, and subsets H_ℓ of $B(t_\ell, r^{-k-1})$ then

$$F(\cup_{j\leq m} H_\ell) \geq r^{-k}\sqrt{\log m} + \min_{\ell \leq m} F(H_\ell) \ . \tag{2.168}$$

Let us then perform the construction of Theorem 2.4.1 for the functionals $F_n = F$. Let us define $j_0(T)$ as in (2.127) and we partition T using the Decomposition Lemma 2.6.2. That is, for $j = j_{0,T}$ we inductively construct sets D_ℓ, and we pick t_ℓ in D_ℓ such that

$$F(D_\ell \cap B(t_{\ell+1}, r^{-j-2})) = \sup_{t\in D_\ell} F(D_\ell \cap B(t, r^{-j-2})) \ ,$$

we set $A_{\ell+1} = D_\ell \cap B(t_{\ell+1}, r^{-j-1})$ and $D_{\ell+1} = D_\ell \setminus A_{\ell+1}$. Assume that the construction continues until we construct a non-empty last piece $C = A_m = D_{m-1}$, where $m = N_1$. Let us get investigate what happens to this set C at the next stage of the construction. Recall that we have defined $j_1(A) = j$. First, we find u_1 in C with

$$F(C \cap B(u_1, r^{-j-2})) = \sup_{t\in C} F(C \cap B(t, r^{-j-2})) \ ,$$

we set $A_1^* = C \cap B(t, r^{-j-1})$ and $D_1^* = C \setminus A_1^*$, and we continue in this manner. The point is that this construction is the exact continuation of the construction by which we obtained A_1, A_2, etc. In consequence, if we consider the sets $A_1, \ldots, A_{m-1}$ together with the sets $A_1^*, \ldots, A_{m^*-1}^*$, where $m^* = N_2$, these pieces are simply obtained by continuing the exhaustion procedure by which we constructed $A_1, \ldots, A_{m-1}$ until $m + m^* - 2$ (etc.). Therefore, as in [9] we construct all the pieces that are obtained by pursuing this exhaustion procedure until the entire space is exhausted, and the same is true at every level of the construction.

The generic chaining as presented here (and the use of a scheme where the functional F might depend on the stage n of the construction) offers at times considerable clarification over the previous approaches. This justifies presenting a proof of Theorem 2.4.1 which is not the simplest we know.

References

1. Artstein, S.: The change in the diameter of a convex body under a random sign-projection. In: Geometric Aspects of Functional Analysis. Springer Lecture Notes in Math., vol. 1850, pp. 31–39 (2004)
2. Dudley, R.M.: The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J. Funct. Anal. **1**, 290–330 (1967)
3. Fernique, X.: Regularité des trajectoires des fonctions aléatoires gaussiennes. In: Ecole d'été de Probabilités de Saint-Flour, IV-1974. Lecture Notes in Math., vol. 480, pp. 1–96. Springer, Berlin (1975)
4. Ledoux, M.: The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001). x+181 pp. ISBN 0-8218-2864-9
5. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
6. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces, II. Function Spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 97. Springer, Berlin (1979). x+243 pp. ISBN: 3-540-08888-1
7. Marcus, M.B., Pisier, G.: Random Fourier Series with Applications to Harmonic Analysis. Annals of Mathematics Studies, vol. 101. Princeton University Press, Princeton; University of Tokyo Press, Tokyo (1981). v+151 pp. ISBN: 0-691-08289-8; 0-691-08292-8
8. Marcus, M.B., Pisier, G.: Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. Acta Math. **152**(3–4), 245–301 (1984)
9. Talagrand, M.: A simple proof of the majorizing measure theorem. Geom. Funct. Anal. **2**, 119–125 (1992)

3. Random Fourier Series and Trigonometric Sums, I

3.1 Translation Invariant Distances

The superiority of the generic chaining bound (2.45) over Dudley's entropy bound (2.36) is its ability to take advantage of the lack of homogeneity of the underlying space metric space (T, d). When, however, there is homogeneity, the situation should be simpler and Dudley's bound should be optimal. A typical such case is when T is a compact metrizable Abelian group and d is a translation invariant distance, $d(s+v, t+v) = d(s,t)$. At the expense of minor complications, one may also consider the case where T is a subset with non-empty interior in a locally compact group, but to demonstrate how simple things are we treat only the compact case. We denote by μ the normalized Haar measure of T, that is $\mu(T) = 1$ and μ is translation invariant. Thus, all balls with a given radius have the same Haar measure.

It is very convenient in this setting to use as a "main parameter" the function $\epsilon \mapsto \mu(B_d(0, \epsilon))$. We recall that we defined $N_0 = 1$ and $N_n = 2^{2^n}$ for $n \geq 1$.

Theorem 3.1.1. *Consider a continuous translation invariant distance d on T. For $n \geq 0$ define*

$$\epsilon_n = \inf\left\{\epsilon > 0 \; ; \; \mu(B_d(0,\epsilon)) \geq 2^{-2^n} = N_n^{-1}\right\} . \tag{3.1}$$

Then

$$\frac{1}{L} \sum_{n \geq 0} \epsilon_n 2^{n/2} \leq \gamma_2(T, d) \leq L \sum_{n \geq 0} \epsilon_n 2^{n/2} . \tag{3.2}$$

Our first lemma shows that the numbers ϵ_n are basically the entropy numbers, so that (3.2) simply states (as expected in this homogeneous case) that $\gamma_2(T, d)$ is equivalent to Dudley's integral.

Lemma 3.1.2. *The entropy numbers $e_n(T) = e_n(T, d)$ satisfy*

$$\epsilon_n \leq e_n(T) \leq 2\epsilon_n . \tag{3.3}$$

Proof. Since μ is translation invariant, all the balls of T with the same radius have the same measure. Consequently if one can cover T by N_n balls with radius ϵ then $\epsilon_n \leq \epsilon$, and this proves the left-hand side inequality.

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_3,

To prove the converse we follow the "volumic" method of Exercise 2.2.14. Consider $n \geq 1$ and a subset S of T which is as large as possible, and such that the balls $B(t, \epsilon_n)$ for $t \in S$ are disjoint. Each of these balls has a measure $\geq N_n^{-1}$, so that $N_n^{-1} \cdot \operatorname{card} S \leq 1$ and therefore $\operatorname{card} S \leq N_n$. But since we assumed S to be as large as possible, the balls centered at S with radius $2\epsilon_n$ cover T, so that the entropy number $e_n(T,d)$ satisfies $e_n(T,d) \leq 2\epsilon_n$. □

Proof of Theorem 3.1.1. The right-hand side inequality follows from (3.3) and Corollary 2.3.2.

To prove the left-hand side inequality we consider an admissible sequence $(\mathcal{A}_n)$ of partitions of T with $\sup_{t\in T} \sum_{n\geq 0} 2^{n/2} \Delta(A_n(t), d) \leq 2\gamma_2(T,d)$. We construct by induction a decreasing sequence $A_n \in \mathcal{A}_n$ as follows. First we choose $A_0 = T$ (there is no other choice). Having constructed A_{n-1} we choose $A_n \subset A_{n-1}$, $A_n \in \mathcal{A}_n$ with the largest possible measure for μ, so that, since $\operatorname{card} \mathcal{A}_n \leq N_n$ we have $\mu(A_n) \geq N_n^{-1}\mu(A_{n-1})$, and since $N_n^2 = N_{n+1}$ we obtain by induction that $\mu(A_n) \geq N_{n+1}^{-1}$. Since d is translation invariant, it follows from (3.1) that A_n cannot be contained in a ball with radius $< \epsilon_{n+1}$, and thus that $\Delta(A_n, d) \geq \epsilon_{n+1}$.

Consider now $t \in A_k$, so that $A_p(t) = A_p$ for each $0 \leq p \leq k$ and thus

$$\sum_{0\leq n\leq k} \epsilon_{n+1} 2^{n/2} \leq \sum_{0\leq n\leq k} 2^{n/2}\Delta(A_n, d) = \sum_{0\leq n\leq k} 2^{n/2}\Delta(A_n(t), d) \leq 2\gamma_2(T,d) \ .$$

Since $\epsilon_0 \leq \Delta(A_0, d)$ this completes the proof of the left-hand side inequality of (3.2). The reader will of course object that there is no reason for which the sets of $\mathcal{A}_n$ should be measurable for μ, but our argument works anyway replacing "measure" by "outer measure". □

Exercise 3.1.3. With the notation of Theorem 3.1.1 prove that for a constant K depending only on α, for $\alpha \geq 1$ we have

$$\frac{1}{K} \sum_{n\geq 0} \epsilon_n 2^{n/\alpha} \leq \gamma_\alpha(T,d) \leq K \sum_{n\geq 0} \epsilon_n 2^{n/\alpha} \ . \tag{3.4}$$

The following theorem might look deceptively simple. It expresses however a deep fact, and is at the root of the main result of this chapter, the Marcus-Pisier theorem, Theorem 3.2.12.

Theorem 3.1.4. *Consider a translation-invariant distance d_ω on T, that depends on a random parameter ω. Assuming enough measurability and integrability consider the distance d given by*

$$d(s,t) = (\mathsf{E} d_\omega(s,t)^2)^{1/2} \ . \tag{3.5}$$

Then

$$(\mathsf{E}\gamma_2(T, d_\omega)^2)^{1/2} \leq L\gamma_2(T,d) + L(\mathsf{E}\Delta(T,d_\omega)^2)^{1/2} \ . \tag{3.6}$$

Proof. First we observe that d is indeed a distance, using the triangle inequality in L^2 to obtain

$$(\mathsf{E}(d_\omega(s,t)+d_\omega(t,u))^2)^{1/2} \le (\mathsf{E}d_\omega(s,t)^2)^{1/2} + (\mathsf{E}d_\omega(t,u)^2)^{1/2} .$$

Consider the sequence (ϵ_n) as in Theorem 3.1.1. For each $n \ge 1$ let us set $B_n = B_d(0,\epsilon_n)$, so that $\mu(B_n) \ge N_n^{-1}$ by definition of ϵ_n. Let us define

$$b_n(\omega) = \frac{1}{\mu(B_n)} \int_{B_n} d_\omega(0,t) \mathrm{d}\mu(t) .$$

Markov inequality implies

$$\mu(\{t \in B_n \; ; \; d_\omega(0,t) \le 2b_n(\omega)\}) \ge \frac{1}{2}\mu(B_n) \ge \frac{1}{2}N_n^{-1} \ge N_{n+1}^{-1} ,$$

so that the number $\epsilon_n(\omega)$ corresponding to the distance d_ω satisfies $\epsilon_{n+1}(\omega) \le 2b_n(\omega)$. Also, $\epsilon_0(\omega) \le \Delta(T,d_\omega)$, so that (3.2) implies

$$\begin{aligned} \gamma_2(T,d_\omega) &\le L\sum_{n\ge 0} \epsilon_n(\omega)2^{n/2} \\ &\le L\Delta(T,d_\omega) + L\sum_{n\ge 0}\epsilon_{n+1}(\omega)2^{(n+1)/2} \\ &\le L\Delta(T,d_\omega) + L\sum_{n\ge 0} b_n(\omega)2^{(n+1)/2} . \end{aligned} \tag{3.7}$$

Now, the Cauchy-Schwarz inequality implies

$$b_n(\omega)^2 \le \frac{1}{\mu(B_n)} \int_{B_n} d_\omega(0,t)^2 \mathrm{d}\mu(t) ,$$

and since $\mathsf{E}d_\omega(0,t)^2 \le \epsilon_n^2$ for $t \in B_n$ we have $\mathsf{E}b_n(\omega)^2 \le \epsilon_n^2$, so that (3.6) follows from (3.7), the triangle inequality in L^2 and (3.2). □

Exercise 3.1.5. (a) Show that if T is an arbitrary metric space and d_ω an arbitrary random metric, then (3.6) need not hold.
(b) Give examples showing that the last term is necessary in (3.6).

3.2 The Marcus-Pisier Theorem

A *character* χ on T is a continuous map from T to $\mathbb{C}$ such that $|\chi(t)| = 1$ for each t and $\chi(s+t) = \chi(s)\chi(t)$ for any $s,t \in T$. In particular $\chi(0) = 1$. A random Fourier series is a series

$$\sum_{i\ge i} \xi_i \chi_i$$

where ξ_i is a complex-valued r.v. and χ_i is a (non-random) character. In this section we consider only random Fourier series whose terms are symmetric independent r.v.s. We will study the convergence of such series when we return to this topic in Chapter 7, and for now we concentrate on the central part of this study, i.e. the study of *finite* sums $\sum_i \xi_i \chi_i$, which we call *trigonometric sums*. It is implicit when we write such a formula that i ranges over a finite set. We denote by $\|\cdot\|$ the supremum norm of such a sum, so that

$$\Big\|\sum_i \xi_i \chi_i\Big\| = \sup_{t\in T}\Big|\sum_i \xi_i \chi_i(t)\Big| \, .$$

The ultimate goal is to find upper and lower bounds for the quantity $\mathsf{E}\|\sum_i \xi_i\chi_i\|$ that are of the same order in full generality. In the present section we shall basically achieve this under the extra condition that the r.v.s "ξ_i have L^1 and L^2 norms of the same order."

Theorem 3.2.1. *Assume that the r.v.s ξ_i are symmetric, independent, have a second moment, and consider on T the distance d given by*

$$d(s,t)^2 = \sum_i \mathsf{E}|\xi_i|^2 |\chi_i(s) - \chi_i(t)|^2 \, . \tag{3.8}$$

Then

$$\Big(\mathsf{E}\Big\|\sum_i \xi_i \chi_i\Big\|^2\Big)^{1/2} \le L\bigg(\gamma_2(T,d) + \Big(\sum_i \mathsf{E}|\xi_i|^2\Big)^{1/2}\bigg) \, . \tag{3.9}$$

Here we control the size of $\sum_i \xi_i\chi_i$ via $(\mathsf{E}\|\sum_i \xi_i\chi_i\|^2)^{1/2}$ rather than via $\mathsf{E}\|\sum_i \xi_i\chi_i\|$ simply because this is the way this is done in [3]. It is known that this makes no difference, as these two quantities are always "of the same order" from general principles.

If $X_t = \sum_i \xi_i\chi_i(t)$, (3.8) implies $\mathsf{E}|X_s - X_t|^2 \le d(s,t)^2$, but it does not seem possible to say much more unless one assumes more on the r.v.s ξ_i, e.g. that they are Gaussian. Therefore it is at first surprising to obtain a conclusion as strong as (3.9). Theorem 3.2.1 is another deceptively simple result on which the reader should meditate.

Throughout the book we denote by ε_i independent Bernoulli (=coin flipping) r.v.s, that is

$$\mathsf{P}(\varepsilon_i = \pm 1) = \frac{1}{2} \, .$$

(Thus ε_i is a r.v. while ϵ_i is a small positive number.)

Let us now explain the **fundamental** idea to obtain upper bounds for $\mathsf{E}\|\sum_i \xi_i\chi_i\|$ when r.v.s ξ_i are symmetric: the sum $\sum_i \xi_i\chi_i$ has the same distribution as the sum $\sum_i \varepsilon_i\xi_i\chi_i$, where the Bernoulli r.v.s ε_i are independent and independent of the r.v.s ξ_i. Therefore, if we work given the randomness of the ξ_i we are dealing with a process of the type

$$X_t = \sum_i a_i(t)\varepsilon_i \ . \tag{3.10}$$

Such a process is called a Bernoulli process (the individual random variables are linear combinations of independent Bernoulli r.v.s). The fundamental class of Bernoulli processes will receive much attention in Chapter 5, and we will learn how to produce very efficient bounds using chaining arguments. For the time being we need only the simple and essential fact that Bernoulli processes "have better tails than the corresponding Gaussian process." This is a consequence of the following simple (yet fundamental) fact, for which we refer to [3] page 90, or to Exercise 3.2.3 below.

Lemma 3.2.2 (The Subgaussian Inequality). *Consider independent Bernoulli r.v.s ε_i (i.e. with $\mathsf{P}(\varepsilon_i = 1) = \mathsf{P}(\varepsilon_i = -1) = 1/2$)) and real numbers a_i. Then for each $u > 0$ we have*

$$\mathsf{P}\Big(\Big|\sum_i \varepsilon_i a_i\Big| \geq u\Big) \leq 2\exp\Big(-\frac{u^2}{2\sum_i a_i^2}\Big) \ . \tag{3.11}$$

Exercise 3.2.3. (a) For $\lambda \in \mathbb{R}$ prove that

$$\mathsf{E}\exp\lambda\varepsilon_i = \cosh\lambda \leq \exp\frac{\lambda^2}{2} \ .$$

(b) Prove that

$$\mathsf{E}\exp\Big(\lambda\sum_i \varepsilon_i a_i\Big) \leq \exp\Big(\frac{\lambda^2}{2}\sum_i a_i^2\Big) \ ,$$

and prove (3.11) using the formula $\mathsf{P}(X \geq u) \leq \exp(-\lambda u)\mathsf{E}\exp\lambda X$ for $u > 0$ and $\lambda > 0$.

Exercise 3.2.4. (a) Prove that

$$\Big\|\sum_i \varepsilon_i a_i\Big\|_p \leq L\sqrt{p}\Big(\sum_i a_i^2\Big)^{1/2} \ .$$

(Hint: use (2.22).)
(b) For a r.v. $X \geq 0$ prove that $(\mathsf{E}X^2)^2 \leq \mathsf{E}X\,\mathsf{E}X^3$.
(c) Use (a) and (b) to prove that

$$\mathsf{E}\Big|\sum_i \varepsilon_i a_i\Big| \geq \frac{1}{L}\Big(\sum_i a_i^2\Big)^{1/2} \ . \tag{3.12}$$

As a consequence of the subgaussian inequality (3.11), a process such as (3.10) satisfies the increment condition (1.4) with respect to the distance d^* given by

$$d^*(s,t)^2 := \sum_i (a_i(s) - a_i(t))^2 \ ,$$

and therefore from (2.49),

$$\Big(\mathsf{E}\sup_{s,t\in T}|X_s-X_t|^2\Big)^{1/2}\le L\gamma_2(T,d^*)\ . \tag{3.13}$$

Let us now observe that (3.13) holds also for complex-valued processes (i.e. the quantities $a_i(t)$ might be complex). This is seen simply by considering separately the real and imaginary parts.

Proof of Theorem 3.2.1. Since for each character χ we have $\chi(0)=1$, it holds that $\mathsf{E}|\sum_i \xi_i\chi_i(0)|^2=\sum_i\mathsf{E}|\xi_i|^2$, so it suffices to prove that

$$\Big(\mathsf{E}\sup_{t,s}\Big|\sum_i(\xi_i\chi_i(t)-\xi_i\chi_i(s))\Big|^2\Big)^{1/2}\le L\gamma_2(T,d)+L\Big(\sum_i\mathsf{E}|\xi_i|^2\Big)^{1/2}\ . \tag{3.14}$$

Since the r.v.s ξ_i are symmetric the sum $\sum_i\xi_i\chi_i$ has the same distribution as the sum $\sum_i\varepsilon_i\xi_i\chi_i$, where the Bernoulli r.v.s ε_i are independent and independent of the r.v.s ξ_i. For clarity let us assume that the underlying probability space is a product $\Omega\times\Omega'$, with a product probability, and that if (ω,ω') is the generic point of this product, then ξ_i depends on ω only and ε_i depends on ω' only. For each ω define the distance d_ω on T by

$$d_\omega(s,t)^2=\sum_i|\xi_i(\omega)|^2|\chi_i(s)-\chi_i(t)|^2\ ,$$

and observe that

$$\Delta(T,d_\omega)^2\le 4\sum_i|\xi_i(\omega)|^2 \tag{3.15}$$

and

$$\mathsf{E}\,d_\omega(s,t)^2=\sum_i\mathsf{E}|\xi_i|^2|\chi_i(s)-\chi_i(t)|^2=d(s,t)^2\ . \tag{3.16}$$

Next we observe using (3.13) that for each ω we have

$$\mathsf{E}'\sup_{t,s}\Big|\sum_i\big(\varepsilon_i(\omega')\xi_i(\omega)\chi_i(t)-\varepsilon_i(\omega')\xi_i(\omega)\chi_i(s)\big)\Big|^2\le L\gamma_2(T,d_\omega)^2\ , \tag{3.17}$$

where E' denotes expectation in ω' only.

The distances d_ω are translation-invariant, as follows from the facts that $\chi(s+u)=\chi(s)\chi(u)$ and $|\chi(u)|=1$. The result then follows combining (3.6), (3.16) and (3.17). □

Exercise 3.2.5. The present exercise deduces classical bounds for trigonometric sums from (3.9). It is part of the exercise to recognize that it deals with trigonometric sums. We consider the case where T is the unit circle in $\mathbb{C}$, and where $\chi_i(t)=t^i$, the i-th power of t. We observe the bound

$$|s^i - t^i| \le \min(2, |i||s-t|) . \tag{3.18}$$

Let $c_i = \mathsf{E}|\xi_i|^2$, and consider the distance d of (3.8), $d(s,t)^2 = \sum_i c_i |s^i - t^i|^2$. Let $b_0 = \sum_{|i|\le 3} c_i$ and for $n \ge 1$, let $b_n = \sum_{N_n \le |i| \le N_{n+1}} c_i$. Prove that

$$\gamma_2(T,d) \le L \sum_{n\ge 0} 2^{n/2}\sqrt{b_n} , \tag{3.19}$$

and consequently from (3.9)

$$\mathsf{E}\Big\|\sum_i \xi_i \chi_i\Big\| \le L \sum_{n\ge 0} 2^{n/2}\sqrt{b_n} . \tag{3.20}$$

(Hint: Here since the group is in multiplicative form the unit is 1 rather than 0. Observe that $d(t,1)^2 \le \sum_i c_i \min(4, |i|^2|t-1|^2)$. Use this bound to prove that the quantity ϵ_n of Theorem 3.1.1 satisfies $\epsilon_n^2 \le L \sum_i c_i \min(1, |i|^2 2^{-2^{n+1}})$ and conclude using (3.2). If you find this exercise too hard, you will find its solution in Section 7.7.)

We now turn to the proof of lower bounds for trigonometric sums. We start by a general principle. We denote by $\Re x$ and $\Im x$ the real part and the imaginary part of a complex number x.

Lemma 3.2.6. *Consider a complex-valued process $(X_t)_{t\in T}$ and assume that both $(\Re X_t)_{t\in T}$ and $(\Im X_t)_{t\in T}$ are Gaussian processes. Consider the distance $d(s,t) = (\mathsf{E}|X_s - X_t|^2)^{1/2}$ on T. Then*

$$\gamma_2(T,d) \le L\mathsf{E} \sup_{s,t\in T} |X_s - X_t| . \tag{3.21}$$

Proof. Consider the distances d_1 and d_2 on T given respectively by

$$d_1(s,t)^2 = \mathsf{E}\big(\Re(X_s - X_t)\big)^2$$

and

$$d_2(s,t)^2 = \mathsf{E}\big(\Im(X_s - X_t)\big)^2 .$$

Combining the left-hand side of (2.80) with Lemma 2.2.1 implies

$$\gamma_2(T,d_1) \le L\mathsf{E} \sup_{s,t\in T} |\Re X_s - \Re X_t| \le L\mathsf{E} \sup_{s,t\in T} |X_s - X_t|$$

and similarly $\gamma_2(T,d_2) \le L\mathsf{E}\sup_{s,t\in T}|X_s - X_t|$. Since $d \le d_1 + d_2$, (2.57) implies that $\gamma_2(T,d) \le L\mathsf{E}\sup_{s,t\in T}|X_s - X_t|$. □

Exercise 3.2.7. Extend the Majorizing Measure Theorem 2.4.1 to the case of complex-valued processes.

Lemma 3.2.8. *Consider a finite number of independent standard normal r.v.s g_i and complex numbers a_i. Then*

$$\mathsf{E}\Big\|\sum_i a_i g_i \chi_i\Big\| \geq \frac{1}{L}\gamma_2(T,d) + \frac{1}{L}\Big(\sum_i |a_i|^2\Big)^{1/2} , \tag{3.22}$$

where d is the distance on T given by

$$d(s,t)^2 = \sum_i |a_i|^2 |\chi_i(s) - \chi_i(t)|^2 . \tag{3.23}$$

Proof. First we observe that, since $\chi(0) = 1$ for each character χ,

$$\mathsf{E}\Big\|\sum_i a_i g_i \chi_i\Big\| \geq \mathsf{E}\Big|\sum_i a_i g_i\Big| \geq \frac{1}{L}\Big(\sum_i |a_i|^2\Big)^{1/2} , \tag{3.24}$$

where the last inequality is obtained by considering separately the real and imaginary parts. Combining with (3.21) completes the proof. □

The following classical simple facts are also very useful.

Lemma 3.2.9 (The Contraction Principle). *Consider independent and symmetric r.v.s η_i valued in a Banach space, and numbers α_i with $|\alpha_i| \leq 1$. Then*

$$\mathsf{E}\Big\|\sum_i \alpha_i \eta_i\Big\| \leq \mathsf{E}\Big\|\sum_i \eta_i\Big\| . \tag{3.25}$$

Proof. We consider the quantity $\mathsf{E}\|\sum_i \alpha_i \eta_i\|$ as a function of the numbers α_i. It is convex, therefore it attains its maximum at an extreme point of its domain. For such an extreme point $\alpha_i = \pm 1$ for each i, and in that case the left and right-hand sides of (3.25) coincide. □

Lemma 3.2.10. *Consider complex vectors x_i in a complex Banach space and independent symmetric real-valued r.v.s ξ_i. Then, if ε_i denote independent Bernoulli r.v.s we have*

$$\mathsf{E}\Big\|\sum_i \xi_i x_i\Big\| \geq \mathsf{E}\Big\|\sum_i \mathsf{E}|\xi_i| \varepsilon_i x_i\Big\| . \tag{3.26}$$

Proof. Assuming without loss of generality that the r.v.s ξ_i and ε_i are independent we use the symmetry of the r.v.s ξ_i to write

$$\mathsf{E}\Big\|\sum_i \xi_i x_i\Big\| = \mathsf{E}\Big\|\sum_i \varepsilon_i |\xi_i| x_i\Big\| .$$

In the right-hand side, taking the expectation in the randomness of the variables ξ_i inside the norm rather than outside can only decrease this quantity. □

In particular, since $\mathsf{E}|g| = \sqrt{2/\pi}$ when g is a standard Gaussian r.v.,

$$\mathsf{E}\Big\|\sum_i \varepsilon_i x_i\Big\| \leq \sqrt{\frac{\pi}{2}}\mathsf{E}\Big\|\sum_i g_i x_i\Big\| . \tag{3.27}$$

Exercise 3.2.11. Prove that the inequality (3.27) cannot be reversed in general. More precisely find a situation where the sum is of length n and the right-hand side is about $\sqrt{\log n}$ times larger than the left hand side.

Theorem 3.2.12 (The Marcus-Pisier theorem [4]). *Consider complex numbers a_i, independent Bernoulli r.v.s ε_i, and independent standard Gaussian r.v.s g_i. Then*

$$\mathsf{E}\Big\|\sum_i a_i g_i \chi_i\Big\| \leq L\mathsf{E}\Big\|\sum_i a_i \varepsilon_i \chi_i\Big\| \,. \tag{3.28}$$

That is, in the setting of these random Fourier series, when $x_i = a_i\chi_i$, we can reverse the general inequality (3.27).

Proof. Consider a number $c > 0$. Then

$$\mathsf{E}\Big\|\sum_i a_i g_i \chi_i\Big\| \leq \mathrm{I} + \mathrm{II}\,, \tag{3.29}$$

where

$$\mathrm{I} = \mathsf{E}\Big\|\sum_i a_i g_i \mathbf{1}_{\{|g_i|\leq c\}} \chi_i\Big\|$$

and

$$\mathrm{II} = \mathsf{E}\Big\|\sum_i a_i g_i \mathbf{1}_{\{|g_i|> c\}} \chi_i\Big\| \,.$$

Let us define $u(c) = (\mathsf{E}(g\mathbf{1}_{\{|g|\geq c\}})^2)^{1/2}$. Consider the distance d given by (3.23). When $\xi_i = a_i g_i \mathbf{1}_{\{|g_i|\geq c\}}$, the distance d' given by (3.8) satisfies $d' = u(c)d$, so that $\gamma_2(T, d') = u(c)\gamma_2(T, d)$ and (3.9) implies

$$\mathrm{II} \leq Lu(c)\bigg(\gamma_2(T,d) + \Big(\sum_i |a_i|^2\Big)^{1/2}\bigg) \,.$$

Recalling the lower bound (3.22), it follows that we can choose c large enough that $\mathrm{II} \leq (1/2)\mathsf{E}\|\sum_i a_i g_i \chi_i\|$. We fix such a value of c. Then (3.29) entails

$$\mathsf{E}\Big\|\sum_i a_i g_i \chi_i\Big\| \leq 2\cdot \mathrm{I}\,.$$

Consider independent Bernoulli r.v.s ε_i, that are independent of the r.v.s g_i, so that by symmetry

$$\mathrm{I} = \mathsf{E}\Big\|\sum_i a_i \varepsilon_i g_i \mathbf{1}_{\{|g_i|< c\}} \chi_i\Big\| \,.$$

The contraction principle (Lemma 3.2.9) used given the randomness of the variables g_i yields $\mathrm{I} \leq c\mathsf{E}\|\sum_i a_i \varepsilon_i \chi_i\|$, which completes the proof. □

Exercise 3.2.13. Show that (3.28) does not hold when χ_i are general maps from T to $\mathbb{C}$ with $|\chi_i(t)| = 1$, even if $\chi_i(0) = 1$.

Combining (3.28) with (3.22) and recalling the distance d of (3.23) we obtain the following fundamental result.

Corollary 3.2.14. *We have*

$$\mathsf{E}\Big\|\sum_i a_i \varepsilon_i \chi_i\Big\| \geq \frac{1}{L}\gamma_2(T,d) + \frac{1}{L}\Big(\sum_i |a_i|^2\Big)^{1/2} . \tag{3.30}$$

Proposition 3.2.15. *Consider complex numbers a_i, independent symmetric real valued random variables ξ_i and characters χ_i. Consider on T the two distances given by*

$$d_1(s,t)^2 = \sum_{i\geq 1} |a_i|^2 (\mathsf{E}|\xi_i|)^2 |\chi_i(s) - \chi_i(t)|^2$$

and

$$d_2(s,t)^2 = \sum_{i\geq 1} |a_i|^2 \mathsf{E}\,\xi_i^2 |\chi_i(s) - \chi_i(t)|^2 .$$

Then

$$\frac{1}{L}\Big(\gamma_2(T,d_1) + \Big(\sum_i |a_i|^2 (\mathsf{E}|\xi_i|)^2\Big)^{1/2}\Big) \leq \mathsf{E}\Big\|\sum_i a_i \xi_i \chi_i\Big\| \tag{3.31}$$
$$\leq L\Big(\gamma_2(T,d_2) + \Big(\sum_i |a_i|^2 \mathsf{E}\xi_i^2\Big)^{1/2}\Big) .$$

Proof. The right-hand side of (3.31) simply reproduces (3.9). The left-hand side follows by combining (3.26) and (3.30). □

Exercise 3.2.16. Prove that the right-hand side of (3.31) is at most LA the left-hand side where

$$A = \sup_i \frac{(\mathsf{E}\xi^2)^{1/2}}{\mathsf{E}|\xi_i|} . \tag{3.32}$$

As a consequence, we can claim that "when the r.v.s ξ_i behave well, we know how to estimate the quantity $\mathsf{E}\|\sum_i a_i \xi_i \chi_i\|$".

This concludes for now our study of trigonometric sums. We shall return to this topic in Section 7.2, where we shall be able to estimate the quantity $\mathsf{E}\|\sum_i a_i \xi_i \chi_i\|$ under the only assumption that the r.v.s ξ_i are independent and symmetric. In Chapter 7 we shall also investigate the convergence of random Fourier series. Let us simply mention here that for such a series where the quantity A of (3.32) is bounded, Proposition 3.2.15 allows to show that the necessary and sufficient condition for convergence is $\gamma_2(T,d_2) < \infty$.

It is not always easy to estimate the quantity $\gamma_2(T,d)$ in concrete situations. The book of Marcus and Pisier [4] contains a thorough account (which we will not reproduce) of the link between the present results and the "classical ones". To illustrate the problems that arise, consider for example the case where $T = \{-1,1\}^N$ and for $i \le N$ and $t = (t_i)_{i\le N} \in T$, let $\chi_i(t) = t_i$. For real numbers a_i, it should be obvious that $\|\sum_{i\le N} a_i\varepsilon_i t_i\| = \sum_{i\le N}|a_i|$. Combining with (3.9) and (3.30) we get

$$\frac{1}{L}\sum_{i\le N}|a_i| \le \gamma_2(T,d) \le L\sum_{i\le N}|a_i| \;, \tag{3.33}$$

where of course $d(s,t)^2 = \sum_{i\le N} a_i^2|\chi_i(s)-\chi_i(t)|^2 = 4\sum_{i\le N} a_i^2 \mathbf{1}_{\{t_i\ne s_i\}}$. The following exercise is in fact quite challenging.

Exercise 3.2.17. Find a direct proof of (3.33).

Some basic questions however remain unanswered, such as the following. Consider independent r.v.s $\delta_i = \delta_i(\omega)$ with $\mathsf{P}(\delta_i = 1) = 1/2$ and $\mathsf{P}(\delta_i = 0) = 1/2$ and complex numbers a_i. Consider on T the distances given by

$$d(s,t)^2 = \sum_i |a_i|^2|\chi_i(s)-\chi_i(t)|^2$$

and

$$d_\omega(s,t)^2 = \sum_i \delta_i(\omega)|a_i|^2|\chi_i(s)-\chi_i(t)|^2 \;.$$

It follows from (3.26) that

$$\frac{1}{2}\mathsf{E}\Big\|\sum_i a_i\varepsilon_i\chi_i\Big\| \le \mathsf{E}\Big\|\sum_i a_i\varepsilon_i\delta_i\chi_i\Big\| \;. \tag{3.34}$$

Denoting by E_ε expectation in the r.v.s ε_i only, (3.9) implies

$$\mathsf{E}_\epsilon\Big\|\sum_i a_i\varepsilon_i\delta_i\chi_i\Big\| \le L\gamma_2(T,d_\omega) + L(\sum_i |a_i|^2)^{1/2} \;.$$

Taking expectations and using (3.30) to bound the left-hand side of (3.34) from below yields

$$\gamma_2(T,d) \le L\Big(\mathsf{E}\gamma_2(T,d_\omega) + \Big(\sum_i |a_i|^2\Big)^{1/2}\Big) \;. \tag{3.35}$$

Research problem 3.2.18. Find a direct proof of (3.35).

In other words we would like to prove (3.35) using (3.2), i.e. we would like to be able to reverse inequality (3.6), at least in the present case.

3.3 A Theorem of Fernique

The present section explains a result of Fernique on vector-valued random Fourier series and some of its consequences. In a sense it is an appendix of Section 3.2. It is presented in a separate section as an homage to X. Fernique's decisive contributions to the ideas presented in this volume.

We consider a compact Abelian group T and a complex Banach space E (nothing is lost by assuming that E is finite-dimensional). We denote by $\|\cdot\|$ the norm of E. Consider (finitely many) vectors $a_i \in E$ and characters χ_i on T. Consider independent standard Gaussian r.v.s g_i. We are interested in the sum $\sum_i a_i g_i \chi_i(t)$, and more specifically in estimating the quantity

$$\mathsf{E} \sup_{t\in T} \Big\| \sum_i a_i g_i \chi_i(t) \Big\| \,. \tag{3.36}$$

We denote by x^* the generic element of the dual E^* of E.

Theorem 3.3.1 ([2]). *We have*

$$\mathsf{E} \sup_{t\in T} \Big\| \sum_i a_i g_i \chi_i(t) \Big\| \le L\Big(\mathsf{E}\Big\| \sum_i a_i g_i \Big\| + \sup_{\|x^*\|\le 1} \mathsf{E} \sup_{t\in T} \Big| \sum_i x^*(a_i) g_i \chi_i(t) \Big| \Big) \,. \tag{3.37}$$

Here $\|x^*\|$ denote the (dual) norm of x^*. The reader should observe that both terms on the right-hand side are obviously lower bounds for the left-hand side. The point of (3.37) is that it reduces the estimation of the left-hand side to that of two simpler quantities.

Let us denote by E_1^* the unit ball of E^*. For $(x^*, t) \in E_1^* \times T$ we set $X_{x^*,t} = \sum_i x^*(a_i) g_i \chi_i(t)$, so that the quantity (3.36) is

$$\mathsf{E} \sup_{(x^*,t)\in E_1^*\times T} |X_{x^*,t}| \,. \tag{3.38}$$

The canonical distance on $E_1^* \times T$ associated to the process $(X_{x^*,t})$ is given by

$$d((x^*,s),(y^*,t))^2 = \sum_i |x^*(a_i)\chi_i(s) - y^*(a_i)\chi_i(t)|^2 \,. \tag{3.39}$$

On E_1^* we consider the distance δ given by

$$\delta(x^*,y^*)^2 = \sum_i |x^*(a_i) - y^*(a_i)|^2 \,. \tag{3.40}$$

Since $|\chi_i(t)| = 1$, we have

$$d((x^*,t),(y^*,t)) = \delta(x^*,y^*) \,. \tag{3.41}$$

Given $z^* \in E_1^*$ we consider the following distance on T:

$$d_{z^*}(s,t)^2 = \sum_i |z^*(a_i)\chi_i(s) - z^*(a_i)\chi_i(t)|^2 \ , \tag{3.42}$$

so that

$$d((x^*,t),(x^*,s)) = d_{x^*}(s,t) \ . \tag{3.43}$$

Given $x^*, y^*, z^* \in E_1^*$ and $s, t \in T$, we have

$$\begin{aligned} d((x^*,s)(y^*,t)) &\le d((x^*,s),(z^*,s)) + d((z^*,s),(z^*,t)) + d((z^*,t),(y^*,t)) \\ &= \delta(x^*,z^*) + d_{z^*}(s,t) + \delta(y^*,z^*) \ . \end{aligned} \tag{3.44}$$

Proof of Theorem 3.3.1. We use Lemma 3.2.6 to obtain

$$\gamma_2(E_1^*,\delta) \le L\mathsf{E}\Big\|\sum_i a_i g_i\Big\| \ . \tag{3.45}$$

Also, given $z^* \in E_1^*$, denoting by LS the right-hand side of (3.37) we have

$$\mathsf{E}\sup_{t\in T}|X_{z^*,t}| = \mathsf{E}\sup_{t\in T}\Big|\sum_i x^*(a_i)g_i\chi_i(t)\Big| \le LS \tag{3.46}$$

and Lemma 3.2.6 again implies

$$\gamma_2(T,d_{z^*}) \le LS \ . \tag{3.47}$$

Since the distance d_{z^*} is translation invariant, combining (3.2) and (3.3) yields

$$\sum_{n\ge 0} 2^{n/2} e_n(T,d_{z^*}) \le LS \ . \tag{3.48}$$

In the remainder of the proof we deduce from (3.44), (3.45) and (3.48) that

$$\gamma_2(E_1^* \times T, d) \le LS \ , \tag{3.49}$$

which finishes the proof using Theorem 2.2.22. Let us consider an admissible sequence $(\mathcal{A}_n)$ of partitions of E_1^* such that

$$\sup_{x^*\in E_1^*}\sum_{n\ge 0} 2^{n/2}\Delta(A_n(x^*),\delta) \le LS \ . \tag{3.50}$$

Given $A \in \mathcal{A}_n$ let us select a point $z^*(n,A) \in A$ for which

$$e_n(T,d_{z^*(n,A)}) \le 2\inf\{e_n(T,d_{z^*}) \ ; \ z^* \in A\} \ . \tag{3.51}$$

We then construct a partition $\mathcal{C}_{A,n}$ of T in N_n sets, each of which are of diameter $\le 4e_n(T,d_{z^*(n,A)})$ for the distance $d_{z^*(n,A)}$. We consider the partition $\mathcal{B}'_n$ of $E_1^* \times T$ in sets of the type $A \times C$ where $A \in \mathcal{A}_n$ and $C \in \mathcal{C}_{A,n}$. Its cardinality is $\le N_n^2 = N_{n+1}$. Let us define $\mathcal{B}_n$ as the partition of $E_1^* \times T$ generated by $\mathcal{B}'_1, \ldots, \mathcal{B}'_n$ so that as usual the sequence $(\mathcal{B}_n)$ increases and $\operatorname{card}\mathcal{B}_n \le N_{n+2}$.

Consider a point $(x^*, t) \in E_1^* \times T$. Then, denoting by $B_n((x^*, t))$ the set of $\mathcal{B}_n$ which contains the point (x^*, t), we have

$$B_n((x^*, t)) \subset A \times C ,$$

where $A = A_n(x^*)$ and C is the element of the partition $\mathcal{C}_{A,n}$ that contains t. Using (3.44) for $z^* \in A$ we obtain

$$\Delta(B_n(x^*, t), d) \le L(\Delta(A_n(x^*), \delta) + \Delta(C, d_{z^*})) . \tag{3.52}$$

Now, using successively the definition of the partition $\mathcal{C}_{A,n}$ and the choice of $z^*(n, A)$,

$$\Delta(C, d_{z^*(n,A)}) \le 4e_n(T, d_{z^*(n,A)}) \le 8e_n(T, d_{x^*}) ,$$

and therefore using (3.52) for $z^* = z^*(n, A)$ we get

$$\Delta(B_n(x^*, t), d) \le L(\Delta(A_n(x^*), \delta) + e_n(T, d_{x^*})) .$$

It then follows from (3.47) and (3.48) that

$$\sum_{n\ge 0} 2^{n/2} \Delta(B_n(x^*, t), \delta) \le LS ,$$

so that combining with Lemma 2.3.5 yields (3.49) and finishes the proof. □

The following question was open for a long time. It was instrumental in formulating the Bernoulli conjecture of Chapter 5. With the notation of Theorem 3.3.1, if ε_i are independent Bernoulli r.v.s, is it true that

$$\mathsf{E} \sup_t \Big\| \sum_i \varepsilon_i a_i \chi_i(t) \Big\| \le L\mathsf{E} \Big\| \sum_i \varepsilon_i a_i \Big\| + L \sup_{x^* \in E^*} \mathsf{E} \sup_{t\in T} \Big| \sum_i \varepsilon_i x^*(a_i) \chi_i(t) \Big| \ ? \tag{3.53}$$

Exercise 3.3.2. After you have learned the statement of Theorem 5.1.5, prove (3.53).

If you find this exercise too difficult, its solution can be found in [3].

3.4 Notes and Comments

The discovery by X. Fernique that Dudley's bound could be reversed for stationary Gaussian processes [1] was a major progress, with considerable influence. In particular it opened the way to the work of M. Marcus and G. Pisier on random Fourier series.

References

1. Fernique, X.: Regularité des trajectoires des fonctions aléatoires gaussiennes. In: Ecole d'été de Probabilités de Saint-Flour, IV-1974. Lecture Notes in Math., vol. 480, pp. 1–96. Springer, Berlin (1975)
2. Fernique, X.: Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens. Université de Montréal, Centre de Recherches Mathématiques, Montreal (1977). iv+217 pp
3. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
4. Marcus, M.B., Pisier, G.: Random Fourier Series with Applications to Harmonic Analysis. Annals of Mathematics Studies, vol. 101. Princeton University Press, Princeton; University of Tokyo Press, Tokyo (1981). v+151 pp. ISBN: 0-691-08289-8; 0-691-08292-8

4. Matching Theorems, I

We remind the reader that, before attacking any chapter, she should find useful to read the overview of this chapter, which is provided in the appropriate section of Chapter 1, in Section 1.6 in the present case. For the present chapter this overview should help to understand the overall approach and especially the ultimate goal of the first section.

4.1 The Ellipsoid Theorem

As pointed out after Proposition 2.5.2, an ellipsoid $\mathcal{E}$ is in some sense quite smaller than what one would predict by looking only at the numbers $e_n(\mathcal{E})$. We will trace the roots of this phenomenon to a simple geometric property, namely that an ellipsoid is "sufficiently convex", and we will formulate a general version of this principle for sufficiently convex bodies. The case of ellipsoids already suffices to provide tight upper bounds on certain matchings, which is the main goal of the present chapter. The general case is at the root of certain very deep facts of Banach space theory, such as Bourgain's celebrated solution of the Λ_p-problem in Sections 16.5 and 16.6.

The ellipsoid $\mathcal{E}$ of (2.110):

$$\mathcal{E} = \left\{ t \in \ell^2 \; ; \; \sum_{i \ge 1} \frac{t_i^2}{a_i^2} \le 1 \right\} \tag{2.110}$$

is the unit ball of the norm

$$\|x\|_{\mathcal{E}} = \left(\sum_{i \ge 1} \frac{x_i^2}{a_i^2} \right)^{1/2} . \tag{4.1}$$

Lemma 4.1.1. *We have*

$$\|x\|_{\mathcal{E}} \, , \, \|y\|_{\mathcal{E}} \le 1 \Rightarrow \left\| \frac{x+y}{2} \right\|_{\mathcal{E}} \le 1 - \frac{\|x-y\|_{\mathcal{E}}^2}{8} \, . \tag{4.2}$$

Proof. The parallelogram identity implies

$$\|x-y\|_{\mathcal{E}}^2 + \|x+y\|_{\mathcal{E}}^2 = 2\|x\|_{\mathcal{E}}^2 + 2\|y\|_{\mathcal{E}}^2 \le 4$$

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_4,

so that

$$\|x+y\|_{\mathcal{E}}^2 \le 4 - \|x-y\|_{\mathcal{E}}^2$$

and

$$\Big\|\frac{x+y}{2}\Big\|_{\mathcal{E}} \le \Big(1-\frac{1}{4}\|x-y\|_{\mathcal{E}}^2\Big)^{1/2} \le 1 - \frac{1}{8}\|x-y\|_{\mathcal{E}}^2 \ . \qquad \square$$

Since (4.2) is the only property of ellipsoids we will use, it clarifies matters to state the following definition.

Definition 4.1.2. *Consider a number $p \ge 2$. A norm $\|\cdot\|$ in a Banach space is called p-convex if for a certain number $\eta > 0$ we have*

$$\|x\|\,,\ \|y\| \le 1 \Rightarrow \Big\|\frac{x+y}{2}\Big\| \le 1 - \eta\|x-y\|^p \ . \tag{4.3}$$

Thus (4.2) implies that the Banach space ℓ^2 provided with the norm $\|\cdot\|_{\mathcal{E}}$ is 2-convex. For $q < \infty$ the classical Banach space L^q is p-convex where $p = \min(2,q)$. The reader is referred to [5] for this result and any other classical facts about Banach spaces. Let us observe that, taking $y = -x$ we must have

$$2^p \eta \le 1 \ . \tag{4.4}$$

In this section we shall study the metric space (T,d) where T is the unit ball of a p-convex Banach space B, and where d is the distance induced on B by another norm $\|\cdot\|_\sim$. This concerns in particular the case where T is the ellipsoid (2.110) and $\|\cdot\|_\sim$ is the ℓ^2 norm.

Given a metric space (T,d), we consider the functionals

$$\gamma_{\alpha,\beta}(T,d) = \Big(\inf \sup_{t\in T} \sum_{n\ge 0} \big(2^{n/\alpha} \Delta(A_n(t),d)\big)^\beta\Big)^{1/\beta} \ , \tag{4.5}$$

where α and β are positive numbers, and where the infimum is over all admissible sequences $(\mathcal{A}_n)$. Thus, with the notation of Definition 2.2.19, we have $\gamma_{\alpha,1}(T,d) = \gamma_\alpha(T,d)$. For matchings, the important functionals are $\gamma_{2,2}(T,d)$ and $\gamma_{1,2}(T,d)$ (but it requires no extra effort to consider the general case). The importance of these functionals is that in certain conditions they nicely relate to $\gamma_2(T,d)$ through Hölder's inequality. For motivation we explain right now how this is done, even though this may spoil for the acute reader the surprise of how the terms $\sqrt{\log N}$ occur in Section 4.3.

Lemma 4.1.3. *Consider a finite metric space T, and assume that* $\operatorname{card} T \le N_m$. *Then*

$$\gamma_2(T,d) \le \sqrt{m}\,\gamma_{2,2}(T,d) \ . \tag{4.6}$$

Proof. Since T is finite there exists an admissible sequence $(\mathcal{A}_n)$ of T for which

$$\forall t \in T \,,\ \sum_{n \geq 0} 2^n \Delta(A_n(t), d)^2 \leq \gamma_{2,2}(T, d)^2 \,. \tag{4.7}$$

Since $\operatorname{card} T \leq N_m$, we may assume that $A_m(t) = \{t\}$ for each t, so that in (4.7) the sum is really over $n \leq m-1$. Since $\sum_{0 \leq n \leq m-1} a_n \leq \sqrt{m}(\sum_{0 \leq n \leq m} a_n^2)^{1/2}$ by the Cauchy-Schwarz inequality, it follows that

$$\forall t \in T \,,\ \sum_{n \geq 0} 2^{n/2} \Delta(A_n(t), d) \leq \sqrt{m}\gamma_{2,2}(T, d) \,. \qquad \square$$

How to relate the functionals $\gamma_{1,2}$ and γ_2 by a similar argument is shown in Lemma 4.4.6 below.

Of course one may wonder how it is possible, using something as simple as the Cauchy-Schwarz inequality in Lemma 4.1.3 that one can ever get essentially exact results. At a general level the answer is obvious: it is because we use this inequality in the case of near equality. That this is indeed the case for the ellipsoids of Corollary 4.1.7 below is of course a non-trivial fact about the geometry of these ellipsoids.

Theorem 4.1.4. *If T is the unit ball of a p-convex Banach space, if η is as in (4.3) and if the distance d on T is induced by another norm, then*

$$\gamma_{\alpha,p}(T, d) \leq K(\alpha, p, \eta) \sup_{n \geq 0} 2^{n/\alpha} e_n(T, d) \,. \tag{4.8}$$

The following exercise stresses the point of this theorem.

Exercise 4.1.5. (a) Prove that for a general metric space (T, d), it is true that

$$\gamma_{\alpha,p}(T, d) \leq K(\alpha) \Big(\sum_{n \geq 0} \big(2^{n/\alpha} e_n(T, d)\big)^p \Big)^{1/p} \,, \tag{4.9}$$

and that

$$\sup_n 2^{n/\alpha} e_n(T, d) \leq K(\alpha) \gamma_{\alpha,p}(T, d) \,. \tag{4.10}$$

(b) Prove that it is essentially impossible in general to improve on (4.9).

In words, the content of Theorem 4.1.4 is that the size of T, as measured by the functional $\gamma_{\alpha,p}$ is smaller than what one would expect when knowing only the numbers $e_n(T, d)$.

Corollary 4.1.6 (The Ellipsoid Theorem). *Consider the ellipsoid $\mathcal{E}$ of (2.110) and $\alpha \geq 1$. Then*

$$\gamma_{\alpha,2}(\mathcal{E}) \leq K(\alpha) \sup_{\epsilon > 0} \epsilon (\operatorname{card}\{i \,;\, a_i \geq \epsilon\})^{1/\alpha}. \tag{4.11}$$

Proof. Without loss of generality we may assume that the sequence (a_i) is non-increasing. We apply Theorem 4.1.4 to the case $\|\cdot\| = \|\cdot\|_{\mathcal{E}}$, and where d is the distance of ℓ^2, and we get, using (2.118) in the last inequality,

$$\gamma_{\alpha,2}(\mathcal{E}) \le K(\alpha) \sup_n 2^{n/\alpha} e_n(\mathcal{E}) \le K(\alpha) \sup_n 2^{n/\alpha} a_{2^n} \ .$$

Now, the choice $\epsilon = a_{2^n}$ implies

$$2^{n/\alpha} a_{2^n} \le \sup_{\epsilon > 0} \epsilon (\mathrm{card}\{i \ ; \ a_i \ge \epsilon\})^{1/\alpha}. \qquad \square$$

The restriction $\alpha \ge 1$ is inessential and can be removed by a suitable modification of (2.118). The important cases are $\alpha = 1$ and $\alpha = 2$.

The following immediate reformulation is useful when the ellipsoid is described by a condition of the type $\sum_j b_j^2 x_j^2 \le 1$ rather than by a condition of the type $\sum_i (x_i/a_i)^2 \le 1$.

Corollary 4.1.7. *Consider a countable set J, numbers $(b_i)_{i \in J}$ and the ellipsoid*

$$\mathcal{E} = \Big\{ x \in \ell^2(J) \ ; \ \sum_{j \in J} b_j^2 x_j^2 \le 1 \Big\} \ .$$

Then

$$\gamma_{\alpha,2}(\mathcal{E}) \le K(\alpha) \sup_{u>0} \frac{1}{u} (\mathrm{card}\{j \in J \ ; \ |b_j| \le u\})^{1/\alpha} \ .$$

Proof. Without loss of generality we can assume that $J = \mathbb{N}$. We then set $a_i = 1/b_i$, we apply Corollary 4.1.6, and we set $\epsilon = 1/u$. $\square$

We give right away a striking application of this result. This application is at the root of the results of Section 4.4.

Proposition 4.1.8. *Consider the set $\mathcal{L}$ of functions $f : [0,1] \to \mathbb{R}$ such that $f(0) = f(1) = 0$, f is continuous on $[0,1]$, f is differentiable outside a finite set and $\sup |f'| \le 1$. Then $\gamma_{1,2}(\mathcal{L}, d_2) \le L$, where $d_2(f,g) = \|f - g\|_2 = \big(\int_{[0,1]} (f-g)^2 \mathrm{d}\lambda\big)^{1/2}$.*

Proof. The very beautiful idea (due to Coffman and Shor [3]) is to use the Fourier transform to represent $\mathcal{L}$ as a subset of an ellipsoid. The Fourier coefficients are defined for $p \in \mathbb{Z}$ by

$$c_p(f) = \int_0^1 \exp(2\pi i p x) f(x) \mathrm{d}x \ .$$

The key fact is the Plancherel formula,

$$\|f\|_2 = \Big(\sum_{p \in \mathbb{Z}} |c_p(f)|^2\Big)^{1/2} \ , \qquad (4.12)$$

which states that the Fourier transform is an isometry from $L^2([0,1])$ into $\ell^2_{\mathbb{C}}(\mathbb{Z})$. Thus, if

$$\mathcal{D} = \{(c_p(f))_{p\in\mathbb{Z}} \; ; \; f \in \mathcal{L}\} ,$$

it suffices to prove that $\gamma_{1,2}(\mathcal{D}, d) < \infty$ where d is the distance induced by $\ell^2_{\mathbb{C}}(\mathbb{Z})$. By integration by parts, and since $f(0) = f(1) = 0$, $c_p(f') = -2\pi i p c_p(f)$, so that, using (4.12) for f', we get

$$\sum_{p\in\mathbb{Z}} p^2 |c_p(f)|^2 \le \sum_{p\in\mathbb{Z}} |c_p(f')|^2 \le \|f'\|_2 ,$$

and since $|c_0(f)| \le \|f\|_2 \le 1$, for $f \in \mathcal{L}$ we have

$$|c_0(f)|^2 + \sum_{p\in\mathbb{Z}} p^2 |c_p(f)|^2 \le 2 ,$$

so that $\mathcal{D}$ is a subset of the complex ellipsoid $\mathcal{E}$ in $\ell^2_{\mathbb{C}}(\mathbb{Z})$ defined by

$$\sum_{p\in\mathbb{Z}} \max(1, p^2) |c_p|^2 \le 2 .$$

Viewing each complex number c_p as a pair (x_p, y_p) of real numbers with $|c_p|^2 = x_p^2 + y_p^2$ yields that $\mathcal{E}$ is (isometric to) the real ellipsoid defined by

$$\sum_{p\in\mathbb{Z}} \max(1, p^2)(x_p^2 + y_p^2) \le 2 ,$$

and the result follows from Corollary 4.1.7. □

Exercise 4.1.9. (a) For $k \ge 1$ consider the space $T = \{0,1\}^{2^k}$. Writing $t = (t_i)_{i\le 2^k}$ a point of T, consider on T the distance $d(t,t') = 2^{-j}$, where $j = \min\{i \le 2^k ; t_i \ne t'_j\}$. Consider the set $\mathcal{L}$ of 1-Lipschitz functions on (T,d) which are zero at $t = (0,\dots,0)$. Prove that $\gamma_{1,2}(\mathcal{L}, d_\infty) \le L\sqrt{k}$, where of course d_∞ denotes the distance induced by the uniform norm (Hint: use (4.9) and Lemma 4.3.9 below.)
(b) Let μ denote the uniform probability μ on T and d_2 the distance induced by $L^2(\mu)$. It can be shown that $\gamma_{1,2}(\mathcal{L}, d_2) \ge \sqrt{k}/L$. (This is not very difficult but requires a thorough understanding of Section 6.3.) Meditate upon the difference with Proposition 4.1.8.

As pointed out, the Ellipsoid theorem, and the principle behind it, have sweeping consequences. There might be more applications of this principle, and this motivates us to give a general theorem, from which Theorem 4.1.4 will immediately follow. The proof of this more general result is identical to the proof of Theorem 4.1.4 itself. Its statement is more complicated, but at first reading one should assume below that $\theta(n) = B2^{np/\alpha}$ for some number $B > 0$, the only case which is relevant for the main results of this book.

Theorem 4.1.10. *Under the hypotheses of Theorem 4.1.4, consider a sequence* $(\theta(n))_{n\geq 0}$*, such that*

$$\forall n \geq 0\,,\, \theta(n) \leq \eta\Big(\frac{1}{4e_n(T,d)}\Big)^p \tag{4.13}$$

and that, for certain numbers $1 < \xi \leq 2\,,\, r \geq 4$ *we have*

$$\forall n \geq 0\,,\, \xi\theta(n) \leq \theta(n+1) \leq \frac{r^p}{2}\theta(n)\,. \tag{4.14}$$

Then there exists an increasing sequence $(\mathcal{A}_n)$ *of partitions of* T *satisfying* $\operatorname{card}\mathcal{A}_n \leq N_{n+1}$ *and*

$$\sup_{t\in T}\sum_{n\geq 0}\theta(n)\Delta(A_n(t),d)^p \leq L\frac{(2r)^p}{\xi-1}\,. \tag{4.15}$$

Proof of Theorem 4.1.4. When $\theta(n) = B2^{np/\alpha}$, condition (4.14) is automatically satisfied with

$$\xi = \min(2, 2^{p/\alpha}) \text{ and } r = \max(4, 2^{1/p+1/\alpha})\,,$$

and condition (4.13) becomes

$$\sup_{n\geq 0} 2^{n/\alpha}e_n(T,d) \leq A\,, \tag{4.16}$$

where $(4A)^p = \eta/B$. Then (4.15) entails

$$\sup_{t\in T}\sum_{n\geq 0}2^{np/\alpha}\Delta(A_n(t),d)^p \leq K(\alpha,p,\eta)A^p\,. \tag{4.17}$$

The sequence $(\mathcal{A}_n)$ need not be admissible, but as usual we define an admissible sequence $(\mathcal{B}_n)$ by $\mathcal{B}_n = \{T\}$ for $n = 0, 1$ and $\mathcal{B}_n = \mathcal{A}_{n-1}$ for $n \geq 1$ to obtain an admissible sequence which (when there is equality in (4.16)) witnesses (4.8), completing the proof of Theorem 4.1.4. □

In the case of general functions $\theta(n)$, the important condition remains (4.13), the technical condition (4.14) is a version of the technical condition (2.148), and one should simply think of $\theta(n)$ as a regularized version of the right-hand side of (4.13).

Proof of Theorem 4.1.10. We denote by $\|\cdot\|$ the norm of the p-convex Banach space of which T is the unit ball. We shall use Theorem 2.7.2 for $\tau = 1\,,\, \beta = p$ and the functionals $F_n = F$ given by

$$F(A) = 1 - \inf\{\|v\|\,;\; v \in \operatorname{conv}A\}\,. \tag{4.18}$$

It should be obvious that F is a functional. In particular $F(A) \geq 0$ since A is a subset of the unit ball T of the Banach space. To prove that these functionals satisfy the growth condition (2.147) of Definition 2.7.1 we consider $n \geq 0$, $m = N_{n+1}$, and points $(t_\ell)_{\ell \leq m}$ in T, such that $d(t_\ell, t_{\ell'}) \geq a$ whenever $\ell \neq \ell'$. Consider also sets $H_\ell \subset T \cap B_d(t_\ell, a/r)$, where the index d emphasizes that the ball is for the distance d rather than for the norm $\|\cdot\|$. Set

$$u = \inf\Big\{\|v\| \; ; \; v \in \operatorname{conv} \bigcup_{\ell \leq m} H_\ell\Big\} = 1 - F\Big(\bigcup_{\ell \leq m} H_\ell\Big) , \tag{4.19}$$

and consider u' such that

$$u' > \max_{\ell \leq m} \inf\{\|v\| \; ; \; v \in \operatorname{conv} H_\ell\} = 1 - \min_{\ell \leq m} F(H_\ell) . \tag{4.20}$$

Let us define $u'' := \min(u', 1)$. (Observe that unless we are in the very special situation where $F(H_\ell) = 0$ for some ℓ, i.e. one of the sets H_ℓ consists of a singleton of norm 1, we can already assume that $u' \leq 1$.) For $\ell \leq m$ consider $v_\ell \in \operatorname{conv} H_\ell$ with $\|v_\ell\| \leq u''$. It follows from (4.3) that for $\ell, \ell' \leq m$,

$$\Big\|\frac{v_\ell + v_{\ell'}}{2u''}\Big\| \leq 1 - \eta\Big\|\frac{v_\ell - v_{\ell'}}{u''}\Big\|^p . \tag{4.21}$$

Moreover, since $(v_\ell + v_{\ell'})/2 \in \operatorname{conv} \bigcup_{\ell \leq m} H_\ell$, we have $u \leq \|v_\ell + v'_\ell\|/2$, and (4.21) implies

$$\frac{u}{u''} \leq 1 - \eta\Big\|\frac{v_\ell - v_{\ell'}}{u''}\Big\|^p ,$$

so that, using that $u'' \leq 1$ in the second inequality,

$$\|v_\ell - v_{\ell'}\| \leq u''\Big(\frac{u'' - u}{\eta u''}\Big)^{1/p} \leq R := \Big(\frac{u'' - u}{\eta}\Big)^{1/p} ,$$

and hence the points $w_\ell := R^{-1}(v_\ell - v_1)$ belong to T. Now, since $H_\ell \subset B_d(t_\ell, a/r)$ we have $v_\ell \in B_d(t_\ell, a/r)$, because the ball $B_d(t_\ell, a/r)$ is convex since the distance d arises from a norm. Since $r \geq 4$, we have $d(v_\ell, v_{\ell'}) \geq a/2$ for $\ell \neq \ell'$, and, since the distance d arises from a norm, we have $d(w_\ell, w_{\ell'}) \geq R^{-1}a/2$ for $\ell \neq \ell'$. Therefore $e_{n+1}(T, d) \geq R^{-1}a/4$.

Since $u' - u \geq u'' - u = \eta R^p$ it follows that

$$u' \geq u + \eta\Big(\frac{a}{4e_{n+1}(T, d)}\Big)^p .$$

Since u' is arbitrary in (4.20) we deduce using (4.19)

$$F\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq \min_{\ell \leq m} F(H_\ell) + \eta\Big(\frac{a}{4e_{n+1}(T, d)}\Big)^p ,$$

and from (4.13) that

$$F\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq \min_{\ell \leq m} F(H_\ell) + a^p \theta(n+1) .$$

This completes the proof of the growth condition (2.147). It follows from Theorem 2.7.2 that we can find an increasing sequence $(\mathcal{A}_n)$ of partitions of T with $\operatorname{card}\mathcal{A}_n \le N_{n+1}$ such that

$$\sup_{t\in T}\sum_{n\ge 0}\theta(n)\Delta(A_n(t))^\beta \le L(2r)^\beta\Big(\frac{F_0(T)}{\xi-1}+\theta(0)\Delta(T)^\beta\Big)\,.$$

To complete the proof of (4.15) one observes that $F_0(T)=F(T)=1$ and that $\theta(0)\Delta(T)^p \le \theta(0)2^p e_0^p(T) \le \eta 2^{-p} \le 1$, using (4.13) for $n=0$ and (4.4) in the last inequality. □

The following generalization of Theorem 4.1.4 is a consequence of Theorem 4.1.10. When applied to ellipsoids, it yields very precise results. It will not be used in the sequel, and could be omitted at first reading.

Theorem 4.1.11. *Consider β , β' , $p>0$ with*

$$\frac{1}{\beta}=\frac{1}{\beta'}+\frac{1}{p}\,. \tag{4.22}$$

Then, under the conditions of Theorem 4.1.4 we have

$$\gamma_{\alpha,\beta}(T,d)\le K(p,\eta,\alpha)\Big(\sum_n\big(2^{n/\alpha}e_n(T,d)\big)^{\beta'}\Big)^{1/\beta'}\,.$$

The case of Theorem 4.1.4 is the case where $\beta'=\infty$. Theorem 4.1.11 allows e.g. to provide a purely geometrical (i.e. not using Gaussian processes arguments) proof of (2.115) as follows. We choose $\alpha=2, \beta=1, \beta'=p=2$ to obtain

$$\gamma_2(T,d)\le L\Big(\sum_n\big(2^{n/2}e_n(T,d)\big)^2\Big)^{1/2}\,. \tag{4.23}$$

Now (2.118) implies that for $n\ge 3$,

$$e_n(T,d)\le L\max_{k\le n-3} a_{2^k}2^{k-n}\le L\max_{k\le n} a_{2^k}2^{k-n}\,,$$

and since $e_n(T,d)\le e_0(T,d)\le \Delta(T,d)\le 2a_1$ this inequality holds for each $n\ge 0$. Combining with (4.23) yields

$$\gamma_2(T,d)\le L\Big(\sum_n\max_{k\le n}a_{2^k}^2 2^{2k-n}\Big)^{1/2}\,. \tag{4.24}$$

Now

$$\begin{aligned}\sum_n\max_{k\le n}a_{2^k}^2 2^{2k-n} &\le \sum_n\sum_{k\le n}a_{2^k}^2 2^{2k-n}=\sum_k\sum_{n\ge k}a_{2^k}^2 2^{2k-n}\\ &\le L\sum_k a_{2^k}^2 2^k\le L\sum_{i\ge 0}a_i^2\,,\end{aligned}$$

so that indeed (4.24) implies (2.115).

Proof of Theorem 4.1.11. For $n \geq 0$ we set

$$d(n) = \eta\Big(\frac{1}{4e_n(T,d)}\Big)^p .$$

We now implement the idea that $\theta(n)$ is a regularized version of this quantity. We define $a := p/(2\alpha)$, $b := 2p/\alpha$ ($= 4a$), and we set

$$\theta(n) := \min\Big(\inf_{k\geq n} d(k)2^{a(n-k)} , \inf_{k\leq n} d(k)2^{b(n-k)}\Big) .$$

We claim that

$$2^a\theta(n) \leq \theta(n+1) \leq 2^b\theta(n) . \tag{4.25}$$

For example, to prove the left-hand side, we note that

$$\begin{aligned}2^a \inf_{k\geq n} d(k)2^{a(n-k)} &\leq \inf_{k\geq n+1} d(k)2^{a(n+1-k)}\\ 2^b \inf_{k\leq n} d(k)2^{b(n-k)} &\leq \inf_{k\leq n} d(k)2^{b(n+1-k)}\end{aligned}$$

and we observe that $\theta(n+1)$ is the minimum of the right-hand sides of the two previous inequalities. Thus (4.14) holds for $\xi = \min(2, 2^a)$ and $r = \max(4, 2^{(b+1)/p})$ and by Theorem 4.1.10 we can find an increasing sequence $(\mathcal{A}_n)$ of partitions of T with $\operatorname{card}\mathcal{A}_n \leq N_{n+1}$ and

$$\sup_{t\in T}\sum_{n\geq 0}\theta(n)\Delta(A_n(t))^p \leq K(\alpha,p) . \tag{4.26}$$

Now we use (4.22) and Hölder's inequality to get

$$\Big(\sum_{n\geq 0}\big(\Delta(A_n(t))2^{n/\alpha}\big)^\beta\Big)^{1/\beta} \leq \Big(\sum_{n\geq 0}\theta(n)\Delta(A_n(t))^p\Big)^{1/p}\Big(\sum_{n\geq 0}\frac{2^{n\beta'/\alpha}}{\theta(n)^{\beta'/p}}\Big)^{1/\beta'} . \tag{4.27}$$

Defining $c := \beta'/\alpha$, we have $a\beta'/p = c/2$ and $b\beta'/p = 2c$, so that

$$\theta(n)^{-\beta'/p} \leq \sum_{k\geq n} d(k)^{-\beta'/p}2^{c(k-n)/2} + \sum_{k\leq n} d(k)^{-\beta'/p}2^{2c(k-n)} \tag{4.28}$$

and

$$\begin{aligned}\sum_{n\geq 0}\frac{2^{n\beta'/\alpha}}{\theta(n)^{\beta'/p}} &\leq \sum_{n,k;k\geq n} d(k)^{-\beta'/p}2^{c(k+n)/2} + \sum_{n,k;k\leq n} d(k)^{-\beta'/p}2^{c(2k-n)}\\ &\leq K(c)\sum_{k\geq 0} d(k)^{-\beta'/p}2^{ck}\end{aligned}$$

by performing the summation in n first. (The reader might wonder why the brutish bound (4.28), which replaces the maximum of a sequence by the sum of the terms allows to get a good bound. The basic reason is that in a geometric series, the sum of the terms is of the same order as the largest one.) Thus, recalling the value of $d(k)$,

$$\sum_{n\geq 0}\frac{2^{n\beta'/\alpha}}{\theta(n)^{\beta'/p}}\leq K(p,\beta,\eta)\sum_{k\geq 0}\bigl(2^{k/\alpha}e_k(T,d)\bigr)^{\beta'}\ .$$

Combining with (4.26) and (4.27) concludes the proof. □

4.2 Matchings

The rest of this chapter is devoted to the following problem. Consider N r.v.s $X_1,\ldots,X_N$ independently uniformly distributed in the unit cube $[0,1]^d$, where $d\geq 2$. Consider a typical realization of these points. How evenly distributed in $[0,1]^d$ are the points $X_1,\ldots,X_N$? To measure this, we will match the points $(X_i)_{i\leq N}$ with *non-random* "evenly distributed" points $(Y_i)_{i\leq N}$, that is, we will find a permutation π of $\{1,\ldots,N\}$ such that the points X_i and $Y_{\pi(i)}$ are "close". There are of course different ways to measure "closeness". For example one may wish that the sum of the distances $d(X_i,Y_{\pi(i)})$ be as small as possible (Section 4.3), that the maximum distance $d(X_i,Y_{\pi(i)})$ be as small as possible (Section 4.4), or one can use more complicated measures of "closeness" (Section 14.1). The case where $d=2$ is very special, and is the object of the present chapter. The case $d\geq 3$ will be studied in Chapter 15. The reader having never thought of the matter might think that the points $X_1,\ldots,X_N$ are very evenly distributed. A moment thinking reveals this is not quite the case, for example, with probability close to one, one is bound to find a little square of area about $N^{-1}\log N$ that contains no point X_i. This is a very local irregularity. In a somewhat informal manner one can say that this irregularity occurs at scale $\sqrt{\log N}/\sqrt{N}$. The specific feature of the case $d=2$ is that in some sense there are irregularities at all scales 2^{-j} for $1\leq j\leq L^{-1}\log N$, and that these are all of the same order. Of course, such a statement is by no means obvious at this stage. In the same direction, a rather deep fact about matchings is that

obstacles to matchings at different scales may combine in dimension 2 but not in dimension ≥ 3 . (4.29)

It is difficult to state a real theorem to this effect, but this is actually seen with great clarity in the proofs. The crucial estimates involve controlling sums (depending on a parameter), each term of representing a different scale. In dimension 2, many terms contribute to the final sum (which therefore results in the contribution of many different scales), while in higher dimension only a few terms contribute. (The case of higher dimension remains non-trivial because *which* terms contribute depend on the value of the parameter.) Of course these statements are very mysterious at this stage, but we expect that a serious study of the methods involved will gradually bring the reader to share this view.

What does it mean to say that the non-random points $(Y_i)_{i\le N}$ are evenly distributed? When N is a square, $N = n^2$, everybody will agree that the N points $(k/n, \ell/n), 1 \le k\,, \ell \le n$ are evenly distributed. More generally we will say that the non-random points $(Y_i)_{i\le N}$ are *evenly spread* if one can cover $[0,1]^2$ with N rectangles with disjoint interiors, such that each rectangle R has an area $1/N$, contains exactly one point Y_i, and is such that $R \subset B(Y_i, 10/\sqrt{N})$. To construct such points when N is not a square, one can simply cut $[0,1]^2$ into horizontal strips of width k/N, where k is about $\sqrt{N}$ (and depends on the strip), use vertical cuts to cut such a strip into k rectangles of area $1/N$, and put a point Y_i in each rectangle. There is a more elegant approach that dispenses from this slightly awkward construction. It is the concept of "transportation cost". One attributes mass $1/N$ to each point X_i, and one measures the "cost of transporting" the resulting probability measure to the uniform probability on $[0,1]^2$. (In the presentation one thus replaces the evenly spread points Y_i by a more canonical object, the uniform probability on $[0,1]^2$.) This approach does not make the proofs any easier, so we shall not use it despite its aesthetic appeal.

The basic tool to construct matchings is the following classical fact.

Proposition 4.2.1. *Consider a matrix* $C = (c_{ij})_{i,j\le N}$. *Let*

$$M(C) = \inf \sum_{i\le N} c_{i\pi(i)} \ ,$$

where the infimum is over all permutations π *of* $\{1,\dots,N\}$. *Then*

$$M(C) = \sup \sum_{i\le N} (w_i + w'_i) \ , \tag{4.30}$$

where the supremum is over all families $(w_i)_{i\le N}\,,\ (w'_i)_{i\le N}$ *that satisfy*

$$\forall i,j \le N\,,\ w_i + w'_j \le c_{ij} \ . \tag{4.31}$$

Thus, if c_{ij} is the cost of matching i with j, $M(C)$ is the minimal cost of a matching, and is given by the "duality formula" (4.30).

Proof. Let us denote by a the right-hand side of (4.30). If the families $(w_i)_{i\le N}\,,\ (w'_i)_{i\le N}$ satisfy (4.31), then for any permutation π of $\{1,\dots,N\}$, we have

$$\sum_{i\le N} c_{i\pi(i)} \ge \sum_{i\le N} (w_i + w'_i)$$

and taking the supremum over the values of w_i and w'_i we get

$$\sum_{i \le N} c_{i\pi(i)} \ge a \,,$$

so that $M(C) \ge a$.

The converse relies on the Hahn-Banach Theorem. Consider the subset $\mathcal{C}$ of $\mathbb{R}^{N \times N}$ that consists of the vectors $(x_{ij})_{i,j \le N}$ for which there exists numbers $(w_i)_{i \le N}$, and $(w'_i)_{i \le N}$ such that

$$\sum_{i \le N} (w_i + w'_i) > a \tag{4.32}$$

$$\forall i, j \le N \,,\; x_{ij} \ge w_i + w'_j \,. \tag{4.33}$$

Then, by definition of a, we have $(c_{ij})_{i,j \le N} \notin \mathcal{C}$. Since $\mathcal{C}$ is an open convex subset of $\mathbb{R}^{N \times N}$, we can separate the point $(c_{ij})_{i,j \le N}$ from $\mathcal{C}$ by a linear functional, i.e. we can find numbers $(p_{ij})_{i,j \le N}$ such that

$$\forall (x_{ij}) \in \mathcal{C} \,,\; \sum_{i,j \le N} p_{ij} c_{ij} < \sum_{i,j \le N} p_{ij} x_{ij} \,. \tag{4.34}$$

Since by definition of $\mathcal{C}$, and in particular (4.33), this remains true when one increases x_{ij}, we see that $p_{ij} \ge 0$, and because of the strict inequality in (4.34) we see that not all the numbers p_{ij} are 0. Thus there is no loss of generality to assume that $\sum_{i,j \le N} p_{ij} = N$. Consider families $(w_i)_{i \le N}$, $(w'_i)_{i \le N}$ that satisfy (4.32). Then if $x_{ij} = w_i + w'_j$, the point $(x_{ij})_{i,j \le N}$ belongs to $\mathcal{C}$ and using (4.34) for this point we obtain

$$\sum_{i,j \le N} p_{ij} c_{ij} \le \sum_{i,j \le N} p_{ij} (w_i + w'_j) \,. \tag{4.35}$$

If $(y_i)_{i \le N}$ are numbers with $\sum_{i \le N} y_i = 0$, we have

$$\begin{aligned} \sum_{i,j \le N} p_{ij} c_{ij} &\le \sum_{i,j \le N} p_{ij} (w_i + y_i + w'_j) \\ &\le \sum_{i,j \le N} p_{ij} (w_i + w'_j) + \sum_{i \le N} y_i \Big(\sum_{j \le N} p_{ij} \Big) \end{aligned} \tag{4.36}$$

as follows from (4.35), replacing w_i by $w_i + y_i$. This inequality holds whenever $\sum_{i \le N} y_i = 0$, so that

$$\sum_{i \le N} y_i = 0 \Rightarrow \sum_{i \le N} y_i \Big(\sum_{j \le N} p_{ij} \Big) = 0 \,,$$

and this forces in turn all the sums $\sum_{j \le N} p_{ij}$ to be equal. Since $\sum_{i,j \le N} p_{ij} = N$, we have $\sum_{j \le N} p_{ij} = 1$, for all i. Similarly, we have $\sum_{i \le N} p_{ij} = 1$ for all j, i.e. the matrix $(p_{ij})_{i,j \le N}$ is bistochastic. Thus (4.35) becomes

$$\sum_{i,j \le N} p_{ij} c_{ij} \le \sum_{i \le N} (w_i + w_i')$$

so that $\sum_{i,j \le N} p_{ij} c_{ij} \le a$. The set of bistochastic matrices is a convex set, so the infimum of $\sum_{i,j \le N} p_{ij} c_{ij}$ over this convex set is obtained at an extreme point. The extreme points are of the type $p_{ij} = \mathbf{1}_{\{\pi(i)=j\}}$ for a permutation π of $\{1, \ldots, N\}$ (a classical result known as Birkhoff's theorem), so that we can find such a permutation with $\sum_{i \le N} c_{i\pi(i)} \le a$. □

The following is a well-known, and rather useful, result of combinatorics.

Corollary 4.2.2 (Hall's Marriage Lemma). *Assume that to each $i \le N$ we associate a subset $A(i)$ of $\{1, \ldots, N\}$ and that, for each subset I of $\{1, \ldots, N\}$ we have*

$$\operatorname{card}\Big(\bigcup_{i \in I} A(i)\Big) \ge \operatorname{card} I \,. \tag{4.37}$$

Then we can find a permutation π of $\{1, \ldots, N\}$ for which

$$\forall i \le N \,,\ \pi(i) \in A(i) \,.$$

Proof. We set $c_{ij} = 0$ if $j \in A(i)$ and $c_{ij} = 1$ otherwise. Using the notations of Proposition 4.2.1, we aim to prove that $M(C) = 0$. Using (4.30), it suffices to show that given numbers $u_i (= -w_i)$, $v_i (= w_i')$ we have

$$\forall i, \forall j \in A(i) \,,\ v_j \le u_i \Rightarrow \sum_{i \le N} v_i \le \sum_{i \le N} u_i \,. \tag{4.38}$$

Adding a suitable constant, we may assume $v_i \ge 0$ and $u_i \ge 0$ for all i, and thus

$$\sum_{i \le N} u_i = \int_0^\infty \operatorname{card}\{i \le N \,;\ u_i \ge t\} \mathrm{d}t \tag{4.39}$$

$$\sum_{i \le N} v_i = \int_0^\infty \operatorname{card}\{i \le N \,;\ v_i \ge t\} \mathrm{d}t \,. \tag{4.40}$$

Given t, using (4.37) for $I = \{i \le N \,;\ u_i < t\}$ and since $v_j \le u_i$ if $j \in A(i)$, we obtain

$$\operatorname{card}\{j \le N \,;\ v_j < t\} \ge \operatorname{card}\{i \le N \,;\ u_i < t\}$$

and thus

$$\operatorname{card}\{i \le N \,;\ u_i \ge t\} \le \operatorname{card}\{i \le N \,;\ v_i \ge t\} \,.$$

Combining with (4.39) and (4.40) this proves (4.38). □

There are other proofs of Hall's Marriage Lemma, based on different ideas, see [2], § 2.

Another well-known application of Proposition 4.2.1 is the following "duality formula".

Proposition 4.2.3. *Consider points* $(X_i)_{i\le N}$ *and* $(Y_i)_{i\le N}$ *in a metric space* (T,d). *Then*

$$\inf_\pi \sum_{i\le N} d(X_i, Y_{\pi(i)}) = \sup_{f\in\mathcal{C}} \sum_{i\le N} (f(X_i) - f(Y_i)) \;, \tag{4.41}$$

where $\mathcal{C}$ *denotes the class of 1-Lipschitz functions on* (T,d), *i.e. functions* f *for which* $|f(x)-f(y)| \le d(x,y)$.

Proof. Given any permutation π and any 1-Lipschitz function f we have

$$\sum_{i\le N} f(X_i) - f(Y_i) = \sum_{i\le N} (f(X_i) - f(Y_{\pi(i)})) \le \sum_{i\le N} d(X_i, Y_{\pi(i)}) \;.$$

This proves the inequality $\ge$ in (4.41). To prove the converse, we use (4.30) with $c_{ij} = d(X_i, Y_j)$, so that

$$\inf_\pi \sum_{i\le N} d(X_i, Y_{\pi(i)}) = \sup \sum_{i\le N} (w_i + w'_i) \;, \tag{4.42}$$

where the supremum is over all families (w_i), (w'_i) for which

$$\forall i\,,\, j \le N\,,\; w_i + w'_j \le d(X_i, Y_j) \;. \tag{4.43}$$

Given a family $(w'_i)_{i\le N}$, consider the function

$$f(x) = \min_{j\le N} (-w'_j + d(x, Y_j)) \;. \tag{4.44}$$

It is 1-Lipschitz, since it is the minimum of functions which are themselves 1-Lipschitz. By definition we have $f(Y_j) \le -w'_j$ and by (4.43) for $i \le N$ we have $w_i \le f(X_i)$, so that

$$\sum_{i\le N} (w_i + w'_i) \le \sum_{i\le N} (f(X_i) - f(Y_i)) \;. \qquad \square$$

Corollary 4.2.2 and Proposition 4.2.3 are all we need in Sections 4.3 and 4.4, and both are fairly easy consequences of Proposition 4.2.1. Sadly, this ease of use of Proposition 4.2.1 is the exception rather than the rule. In Section 14.1 and in Chapter 15, we shall need other uses of Proposition 4.2.1, and these will require considerable efforts. Since the difficulty is not related to any further probabilistic considerations, but to the very nature of this proposition, we briefly explain it now. In a sense, it is the main difficulty in proving matching theorems beyond those of the next two sections.

Let consider points $(X_i)_{i\le N}$ and points $(Y_i)_{i\le N}$ in a set T, and let us try to match them so as to get a small value for

$$\sum_{i\le N} \psi(X_i, Y_{\pi(i)}) \;.$$

Here ψ is a function $T \times T \to \mathbb{R}^+$, which is a kind of "measure of distance" between the arguments, but it certainly does not look like a distance in general, and does not satisfy anything like a triangle inequality. It is natural to assume that $\psi(x,x) = 0$. We use (4.30) with $c_{ij} = \psi(X_i, Y_j)$, so that

$$\inf_\pi \sum_{i \le N} \psi(X_i, Y_{\pi(i)}) = \sup \sum_{i \le N} (w_i + w'_i) , \tag{4.45}$$

where the supremum is over all families (w_i) , (w'_i) for which

$$\forall i, j \le N , \; w_i + w'_j \le \psi(X_i, Y_j) . \tag{4.46}$$

Given a family $(w'_i)_{i \le N}$ and the points $(Y_j)_{j \le N}$, we are again led to consider the function

$$f(x) = \min_{j \le N} (-w'_j + \psi(x, Y_j)) , \tag{4.47}$$

so that (4.46) implies $w_i \le f(X_i)$ and (4.45) yields

$$\inf_\pi \sum_{i \le N} \psi(X_i, Y_{\pi(i)}) \le \sup \sum_{i \le N} (f(X_i) + w'_i) , \tag{4.48}$$

where the supremum is over all families (w'_i). Assume now that the points X_i form an i.i.d. sequence with distribution μ. To simplify, (and we shall always be able to reduce to this exact situation) since the points Y_i are uniformly spread, we assume that $N \int h \mathrm{d}\mu = \sum_{i \le N} h(Y_i)$ for any function h, and quite naturally we write

$$\sum_{i \le N} (f(X_i) + w'_i) = \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\mu) + \sum_{i \le N} (w'_i + f(Y_i)) . \tag{4.49}$$

Since we assume that $\psi(Y_i, Y_i) = 0$, the definition (4.47) of f shows that $f(Y_i) \le -w'_i$, and the last term on the right-hand side of (4.49) is negative, so that it has a chance to compensate the first term. Let us consider a number $A \ge 0$ and the class $\mathcal{H}(A)$ consisting of all functions of the type (4.47) for a certain family (w'_i) of numbers, and for which $\sum_{i \le N} f(Y_i) + w'_i \ge -A$. Then for such a function the last term in (4.49) is $\ge -A$. This limits how it can compensate the first term, so that there seems to be no other way than to bound

$$\sup_{f \in \mathcal{H}(A)} \Big| \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\mu) \Big| . \tag{4.50}$$

Generally speaking, the study of expressions of this type

$$\sup_{f \in \mathcal{F}} \Big| \sum_{i \le N} (f(X_i) - \int f \mathrm{d}\mu) \Big| \tag{4.51}$$

for a class of function $\mathcal{F}$ will be important in the present book, and in particular in Chapter 9. A bound on such a quantity is called a *discrepancy* *bound* because it bounds uniformly on $\mathcal{F}$ the "discrepancy" between the true measure $\int f \mathrm{d}\mu$ and the "empirical measure" $N^{-1}\sum_{i\le N} f(X_i)$. Finding such a bound simply requires finding a bound for the supremum of the process $(|Z_f|)_{f\in\mathcal{F}}$, where the r.v.s Z_f is given by

$$Z_f = \sum_{i\le N}\Big(f(X_i) - \int f \mathrm{d}\mu\Big)\,, \tag{4.52}$$

a topic at the very center of our attention. *Every* matching theorem proved in this book will be proved through a discrepancy bound.

Bounding a stochastic process, and in particular proving a discrepancy bound always ultimately requires in some form an understanding of the geometry of the index set, which the author simply lacks in general in the case of $\mathcal{H}(A)$. In the case of Proposition 4.2.3 the miracle is that all the information about the construction (4.47) is contained in the fact that f is 1-Lipschitz, but it is very unclear what happens already if e.g. $\psi(x,y) = d(x,y)^2$. When proving the matching theorems of Chapter 14 and Chapter 15 (and especially in that latter case) it will be a significant task to figure out what usable information one can get about the class $\mathcal{H}(A)$.

4.3 The Ajtai, Komlós, Tusnády Matching Theorem

Theorem 4.3.1 ([1]). *If the points $(Y_i)_{i\le N}$ are evenly spread and the points $(X_i)_{i\le N}$ are i.i.d. uniform on $[0,1]^2$, then (for $N\ge 2$)*

$$\mathsf{E}\inf_\pi \sum_{i\le N} d(X_i, Y_{\pi(i)}) \le L\sqrt{N\log N}\,, \tag{4.53}$$

where the infimum is over all permutations of $\{1,\dots,N\}$ and where d is the Euclidean distance.

The term $\sqrt{N}$ is just a scaling effect. There are N terms $d(X_i, Y_{\pi(i)})$ each of which should be about $1/\sqrt{N}$. The non-trivial part of the theorem is the factor $\sqrt{\log N}$. In Section 6.4 we shall show that (4.53) can be reversed, i.e.

$$\mathsf{E}\inf_\pi \sum_{i\le N} d(X_i, Y_{\pi(i)}) \ge \frac{1}{L}\sqrt{N\log N}\,. \tag{4.54}$$

We repeat that *every* matching theorem which we prove in this book is deduced from a "discrepancy bound", i.e. a bound on a quantity of the form (4.51). This should be expected after the discussion which ends Section 4.2. Let us state the "discrepancy bound" at the root of Theorem 4.3.1. Consider the class $\mathcal{C}$ of 1-Lipschitz functions on $[0,1]^2$, i.e. of functions f that satisfy

$$\forall x,y\in[0,1]^2\,,\ |f(x)-f(y)|\le d(x,y)\,,$$

where d denotes the Euclidean distance. We denote by λ the uniform measure on $[0,1]^2$.

Theorem 4.3.2. *We have*

$$\mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\le N}(f(X_i)-\int f\mathrm{d}\lambda)\Big|\le L\sqrt{N\log N}\,. \tag{4.55}$$

Research problem 4.3.3. Prove that the following limit

$$\lim_{N\to\infty}\frac{1}{\sqrt{N\log N}}\mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\le N}(f(X_i)-\int f\mathrm{d}\lambda)\Big|$$

exists.

Theorem 4.3.2 is obviously interesting in its own right, and we shall show soon how it is related to Theorem 4.3.1 through Proposition 4.2.3. Let us first discuss it. As already pointed out, we simply think of the left-hand side as $\mathsf{E}\sup_{f\in\mathcal{C}}|Z_f|$, where Z_f is the random variable of (4.52). The first task is to find nice tail properties for these r.v.s, here in the form of the celebrated Bernstein's inequality below. One then applies the methods of Chapter 2, in the present case in the form of Theorem 4.3.6 below. In the end (and because we are dealing with a deep fact) we shall have to prove some delicate "smallness" property of the class $\mathcal{C}$. In the present chapter, this smallness property will always ultimately be derived from the ellipsoid theorem, in the form of Corollary 4.1.7. In the case of Theorem 4.3.2, the (very beautiful) strategy for the hard part of the estimates relies on a kind of 2-dimensional version of Proposition 4.1.8 and is outlined on page 110.

Proof of Theorem 4.3.1. We recall (4.41), i.e.

$$\inf_{\pi}\sum_{i\le N}d(X_i,Y_{\pi(i)})=\sup_{f\in\mathcal{C}}\sum_{i\le N}(f(X_i)-f(Y_i))\,, \tag{4.56}$$

and we simply write

$$\sum_{i\le N}(f(X_i)-f(Y_i))\le\Big|\sum_{i\le N}(f(X_i)-\int fd\lambda)\Big|+\Big|\sum_{i\le N}(f(Y_i)-\int f\mathrm{d}\lambda)\Big|\,. \tag{4.57}$$

Next, we claim that

$$\Big|\sum_{i\le N}(f(Y_i)-\int f\mathrm{d}\lambda)\Big|\le L\sqrt{N}\,. \tag{4.58}$$

We recall that since $(Y_i)_{i\le N}$ are evenly spread one can cover $[0,1]^2$ with N rectangles R_i with disjoint interiors, such that each rectangle R_i has an area $1/N$ and is such that $Y_i\in R_i\subset B(Y_i,10/\sqrt{N})$. Consequently

$$\Big|\sum_{i\le N}(f(Y_i)-\int f\mathrm{d}\lambda)\Big|\le\sum_{i\le N}\Big|(f(Y_i)-N\int_{R_i}f\mathrm{d}\lambda)\Big|\,,$$

and since f is Lipschitz each term in the right-hand side is $\le L/\sqrt{N}$. This proves the claim.

Now, using (4.56) and taking expectation

$$\begin{aligned}\mathsf{E}\inf_{\pi}\sum_{i\le N} d(X_i, Y_{\pi(i)}) &\le L\sqrt{N} + \mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\le N}(f(X_i) - \int f\mathrm{d}\lambda)\Big| \\ &\le L\sqrt{N\log N}\end{aligned}$$

by (4.55). □

Let us now prepare for the proof of Theorem 4.3.2. The following fundamental classical result will allow us to control the tails of the r.v. Z_f of (4.52). It will be used many times.

Lemma 4.3.4 (Bernstein's inequality). *Let $(Y_i)_{i\ge 1}$ be independent r.v.s with $\mathsf{E}Y_i = 0$ and consider a number U with $|Y_i| \le U$ for each i. Then, for $v > 0$,*

$$\mathsf{P}\Big(\Big|\sum_{i\ge 1} Y_i\Big| \ge v\Big) \le 2\exp\Big(-\min\Big(\frac{v^2}{4\sum_{i\ge 1}\mathsf{E}Y_i^2}, \frac{v}{2U}\Big)\Big)\,. \tag{4.59}$$

Proof. For $|x| \le 1$, we have

$$|e^x - 1 - x| \le x^2\sum_{k\ge 2}\frac{1}{k!} = x^2(e-2) \le x^2$$

and thus, since $\mathsf{E}Y_i = 0$, for $U|\lambda| \le 1$, we have

$$|\mathsf{E}\exp\lambda Y_i - 1| \le \lambda^2\mathsf{E}Y_i^2\,.$$

Therefore $\mathsf{E}\exp\lambda Y_i \le 1 + \lambda^2\mathsf{E}Y_i^2 \le \exp\lambda^2\mathsf{E}Y_i^2$, and thus

$$\mathsf{E}\exp\lambda\sum_{i\ge 1} Y_i = \prod_{i\ge 1}\mathsf{E}\exp\lambda Y_i \le \exp\lambda^2\sum_{i\ge 1}\mathsf{E}Y_i^2\,.$$

Now, for $0 \le \lambda \le 1/U$ we have

$$\begin{aligned}\mathsf{P}\Big(\sum_{i\ge 1} Y_i \ge v\Big) &\le \exp(-\lambda v)\mathsf{E}\exp\lambda\sum_{i\ge 1} Y_i \\ &\le \exp\Big(\lambda^2\sum_{i\ge 1}\mathsf{E}Y_i^2 - \lambda v\Big)\,.\end{aligned}$$

If $Uv \le 2\sum_{i\ge 1}\mathsf{E}Y_i^2$, we take $\lambda = v/(2\sum_{i\ge 1}\mathsf{E}Y_i^2)$, obtaining a bound $\exp(-v^2/(4\sum_{i\ge 1}\mathsf{E}Y_i^2))$. If $Uv > 2\sum_{i\ge 1}\mathsf{E}Y_i^2$, we take $\lambda = 1/U$, and we note that

$$\frac{1}{U^2}\sum_{i\geq 1}\mathsf{E}Y_i^2 - \frac{v}{U} \leq \frac{Uv}{2U^2} - \frac{v}{U} \leq -\frac{v}{2U}\ ,$$

so that $\mathsf{P}(\sum_{i\geq 1} Y_i \geq v) \leq \exp(-\min(v^2/4\sum_{i\geq 1}\mathsf{E}Y_i^2, v/2U))$. Changing Y_i into $-Y_i$ we obtain the same bound for $P(\sum_{i\geq 1} Y_i \leq -v)$. □

Corollary 4.3.5. *For each $v > 0$ we have*

$$\mathsf{P}(|Z_f| \geq v) \leq 2\exp\left(-\min\left(\frac{v^2}{4N\|f\|_2^2}, \frac{v}{4\|f\|_\infty}\right)\right)\ , \tag{4.60}$$

where $\|f\|_p$ denotes the norm of f in $L_p(\lambda)$.

Proof. We use Bernstein's inequality with $Y_i = f(X_i) - \int f\mathrm{d}\lambda$ if $i \leq N$ and $Y_i = 0$ if $i > N$. We then observe that $\mathsf{E}Y_i^2 \leq \mathsf{E}f^2 = \|f\|_2^2$ and $|Y_i| \leq 2\sup|f| = 2\|f\|_\infty$. □

We can now state a general bound, from which we will deduce Theorem 4.3.2.

Theorem 4.3.6. *Consider a class $\mathcal{F}$ of functions on $[0,1]^2$ and assume that $0 \in \mathcal{F}$. Then*

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\leq N}(f(X_i) - \int f\mathrm{d}\lambda)\Big| \leq L\Big(\sqrt{N}\gamma_2(\mathcal{F}, d_2) + \gamma_1(\mathcal{F}, d_\infty)\Big)\ , \tag{4.61}$$

where d_2 and d_∞ are the distances induced on $\mathcal{F}$ by the norms of L^2 and L^∞ respectively.

Proof. Since $Z_f - Z_{f'} = Z_{f-f'}$, combining Corollary 4.3.5 with Theorem 2.2.23 we get, since $0 \in \mathcal{F}$, that

$$\begin{aligned}\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\leq N}(f(X_i) - \sum f\mathrm{d}\lambda)\Big| &\leq \mathsf{E}\sup_{f,f'\in\mathcal{F}}|Z_f - Z_{f'}| \\ &\leq L\big(\gamma_2(\mathcal{F}, 2\sqrt{N}d_2) + \gamma_1(\mathcal{F}, 4d_\infty)\big)\,.\end{aligned}$$

To conclude, we use that $\gamma_2(\mathcal{F}, 2\sqrt{N}d_2) = 2\sqrt{N}\gamma_2(\mathcal{F}, d_2)$ and $\gamma_1(\mathcal{F}, 4d_\infty) = 4\gamma_1(\mathcal{F}, d_\infty)$. □

To deduce Theorem 4.3.2 from Theorem 4.3.6, the difficulty is in controlling the ℓ^2 distance. It is not completely trivial to control the ℓ^∞ distance but there is plenty of room.

We recall that for a metric space (T, d), the covering number $N(T, d, \epsilon)$ denotes the smallest number of balls of radius ϵ that are needed to cover T. Theorem 4.3.2 is a prime example of a natural situation where using covering numbers does not yield the correct result. This is closely related to the fact that, as explained in Section 2.5, covering numbers do not describe well the size of ellipsoids. It is of course hard to formulate a theorem to the effect that covering numbers do not suffice, but the claim should make more sense after Exercise 4.3.11 below.

Exercise 4.3.7. Prove that for each $0 < \epsilon \le 1$

$$\log N(\mathcal{C}, d_2, \epsilon) \ge \frac{1}{L\epsilon^2} \ . \tag{4.62}$$

(Hint: Consider an integer $n \ge 0$, and divide $[0,1]^2$ into 2^{2n} equal squares of area 2^{-2n}. For every such square C consider a number $\varepsilon_C = \pm 1$. Consider then the function $f \in \mathcal{C}$ such that $f(x) = \varepsilon_C d(x, B)$ for $x \in C$, where B denotes the boundary of C. Prove that by appropriate choices of the signs ε_C one may find at least $\exp(2^{2n}/L)$ such functions which are at mutual distance $\ge 2^{-n}/L$. You are permitted to peek at Lemma 6.3.1 below.)

We shall not apply Theorem 4.3.6 to $\mathcal{F} = \mathcal{C}$ (it follows from Proposition 6.4.3 below that $\gamma_2(\mathcal{C}, d_2) = \infty$), but rather to a sufficiently large finite subset $\mathcal{F}$ of $\mathcal{C}$, for which we shall need the crucial estimate $\gamma_2(\mathcal{F}, d_2) \le L\sqrt{\log N}$, and we try to outline the strategy which yields this estimate. In an ideal world, we would not deal with the class $\mathcal{C}$, but with the smaller class $\mathcal{C}^*$ of functions of $\mathcal{C}$ which are zero on the boundary of $[0,1]^2$. As in Proposition 4.1.8, one may then parametrize $\mathcal{C}^*$ as a subset of a certain ellipsoid using the Fourier transform, and then Corollary 4.1.7 yields $\gamma_{2,2}(\mathcal{C}^*, d_2) \le L$. Finally the simple use of Cauchy-Schwarz inequality in (4.6) yields $\gamma_2(\mathcal{F}, d_2) \le L\sqrt{\log\log \operatorname{card} \mathcal{F}}$, which is the desired estimate. In real life we must unfortunately deal with the class $\mathcal{C}$. Even though $\mathcal{C}^*$ is "the main part" of $\mathcal{C}$, and even though there is plenty of room to deal with the "remaining part" of $\mathcal{C}$, this creates complications, and taking care of these requires about as much space and energy as the main argument.

Rather than the class $\mathcal{C}^*$ we shall actually use as "main part of $\mathcal{C}$" the class $\mathcal{C}_0$ consisting of functions $f : [0,1]^2 \to \mathbb{R}$ which are differentiable and satisfy

$$\sup\Big|\frac{\partial f}{\partial x}\Big| \le 1 \ ; \ \sup\Big|\frac{\partial f}{\partial y}\Big| \le 1$$

$$\int f \mathrm{d}\lambda = 0 \ ; \ \forall u \, , \ 0 \le u \le 1 \, , \ f(u,0) = f(u,1) \, , \ f(0,u) = f(1,u) \ . \tag{4.63}$$

The fact that the functions in $\mathcal{C}_0$ need not be 1-Lipschitz, but only $\sqrt{2}$-Lipschitz is not important.

The main ingredient in controlling the ℓ^2 distance is the following 2-dimensional version of Proposition 4.1.8 , where we use the functional $\gamma_{2,2}$ of (4.5), and where the underlying distance is the distance induced by $L^2([0,1]^2)$.

Proposition 4.3.8. *We have* $\gamma_{2,2}(\mathcal{C}_0, d_2) < \infty$*.*

Proof. We represent $\mathcal{C}_0$ as a subset of an ellipsoid using the Fourier transform. The Fourier transform associates to each function f on $L^2([0,1]^2)$ the complex numbers $c_{p,q}(f)$ given by

$$c_{p,q}(f) = \iint_{[0,1]^2} f(x_1, x_2) \exp 2i\pi(px_1 + qx_2) \mathrm{d}x_1 \mathrm{d}x_2 \ . \tag{4.64}$$

The Plancherel formula

$$\|f\|_2 = \Big(\sum_{p,q \in \mathbb{Z}} |c_{p,q}(f)|^2 \Big)^{1/2} \tag{4.65}$$

asserts that Fourier transform is an isometry, so that if

$$\mathcal{D} = \{(c_{p,q}(f))_{p,q \in \mathbb{Z}} \ ; \ f \in \mathcal{C}_0)\} ,$$

it suffices to show that $\gamma_{2,2}(\mathcal{D}, d) < \infty$ where d is the distance in the complex Hilbert space $\ell^2_{\mathbb{C}}(\mathbb{Z} \times \mathbb{Z})$. Using (4.64), integration by parts and (4.63), we get

$$-2i\pi p c_{p,q}(f) = c_{p,q}\Big(\frac{\partial f}{\partial x}\Big) .$$

Using (4.65) for $\partial f / \partial x$, and since $\|\partial f / \partial x\|_2 \le 1$ we get $\sum_{p,q \in \mathbb{Z}} p^2 |c_{p,q}(f)|^2 \le 1/4\pi^2$. Proceeding similarly for $\partial f / \partial y$, we get

$$\mathcal{D} \subset \mathcal{E} = \Big\{(c_{p,q}) \in \ell^2_{\mathbb{C}}(\mathbb{Z} \times \mathbb{Z}) \ ; \ c_{0,0} = 0 \, , \ \sum_{p,q \in \mathbb{Z}} (p^2 + q^2) |c_{p,q}|^2 \le 1 \Big\} .$$

We view each complex number $c_{p,q}$ as a pair $(x_{p,q}, y_{p,q})$ of real numbers, and $|c_{p,q}|^2 = x^2_{p,q} + y^2_{p,q}$, so that

$$\begin{aligned} \mathcal{E} = \ & \Big\{ \big((x_{p,q}), (y_{p,q})\big) \in \ell^2(\mathbb{Z} \times \mathbb{Z}) \times \ell^2(\mathbb{Z} \times \mathbb{Z}) \ ; \\ & \quad x_{0,0} = y_{0,0} = 0 \, , \ \sum_{p,q \in \mathbb{Z}} (p^2 + q^2)(x^2_{p,q} + y^2_{p,q}) \le 1 \Big\} . \end{aligned} \tag{4.66}$$

For $u \ge 1$, we have

$$\operatorname{card}\big\{(p,q) \in \mathbb{Z} \times \mathbb{Z} \ ; \ p^2 + q^2 \le u^2 \big\} \le (2u+1)^2 \le L u^2 .$$

We then deduce from Corollary 4.1.7 that $\gamma_{2,2}(\mathcal{E}, d) < \infty$. □

We now turn to the control in the supremum norm. In order to avoid repetition, we state a general principle (which was already known to Kolmogorov).

Lemma 4.3.9. *Consider a metric space (T, d) and assume that for certain numbers B and $\alpha \ge 1$ and each $\epsilon > 0$ we have*

$$N(T, d, \epsilon) \le \Big(\frac{B}{\epsilon}\Big)^{\alpha} . \tag{4.67}$$

Consider the set $\mathcal{B}$ of 1-Lipschitz functions f on T with $\|f\|_\infty \le B$. Then for each $\epsilon > 0$ we have

$$N(\mathcal{B}, d_\infty, \epsilon) \le \exp K \Big(\frac{B}{\epsilon}\Big)^{\alpha} , \tag{4.68}$$

where K depends only on α. In particular,

$$e_n(\mathcal{B}, d_\infty) \le K B 2^{-n/\alpha} . \tag{4.69}$$

Proof. By homogeneity we may and do assume that $B = 1$. Given $h \in \mathcal{B}$ and an integer k, consider the set

$$A = \left\{ f \in \mathcal{B} \,;\, \|f - h\|_\infty \le 2^{1-k} \right\} . \tag{4.70}$$

We first show that

$$N(A, d_\infty, 2^{-k}) \le \exp(K_0 2^{\alpha k}) \,. \tag{4.71}$$

Consider a subset C of A that is maximal with respect to the property that if $f_1, f_2 \in C$ and $f_1 \ne f_2$ then $d_\infty(f_1, f_2) > 2^{-k}$. Then each point of A is within distance $\le 2^{-k}$ of C, so $N(A, d_\infty, 2^{-k}) \le \operatorname{card} C$. From (4.67) there exists a subset U of T with $\operatorname{card} U \le 2^{\alpha(k+2)}$ such that each point of T is within distance 2^{-k-2} of a point of U. If $f_1, f_2 \in C$ and $f_1 \ne f_2$, there is $x \in T$ with $|f_1(x) - f_2(x)| > 2^{-k}$. Consider $y \in U$ with $d(x, y) \le 2^{-k-2}$. Then for $j = 1\,,\,2$ we have $|f_j(x) - f_j(y)| \le 2^{-k-2}$ so that $|f_1(y) - f_2(y)| > 2^{-k-1}$. Let C' be the image of C through the map $f \mapsto (f(x))_{x \in U}$, so that C' is a subset of $\mathbb{R}^U$. We have shown that any two distinct points of C' are at distance at least 2^{-k-1} of each other for the supremum norm. The cubes of edge length 2^{-k-2} centered at the points of C' have disjoint interiors, and (from (4.70)) are entirely contained in a certain cube of edge length $2^{2-k} + 2^{-k-1} = 18 \cdot 2^{-k-2}$, so that, by volume considerations, we have $\operatorname{card} C' \le 18^{\operatorname{card} U}$, and this proves (4.71).

We now prove by induction over $k \ge 1$ that

$$N(\mathcal{B}, d_\infty, 2^{1-k}) \le \exp(K_0 2^{\alpha k}) \,. \tag{4.72}$$

This certainly holds true for $k = 1$. For the induction step from k to $k+1$, we use the induction hypothesis to cover $\mathcal{B}$ by $\exp(K_0 2^{\alpha k})$ sets A of the type (4.70) and we use (4.71) for each of these sets. This completes the induction since $2 \cdot 2^{\alpha k} \le 2^{\alpha(k+1)}$. Finally, (4.68) follows from (4.72), and it implies (4.69) (see Exercise 2.2.13). □

We apply the previous lemma to $T = [0,1]^2$ and

$$\mathcal{C}_1 = \{ f \in \mathcal{C} \,;\, \|f\|_\infty \le 2 \} \,.$$

It should be obvious that $T = [0,1]^2$ satisfies (4.67) for $\alpha = 2$, so that (4.69) implies that for $n \ge 0$,

$$e_n(\mathcal{C}_1, d_\infty) \le L 2^{-n/2} \,. \tag{4.73}$$

Let us recall that the class $\mathcal{C}_0$ is defined just before (4.63) and observe (since the functions in $\mathcal{C}_0$ are $\sqrt{2}$-Lipschitz) that $\mathcal{C}_0 \subset \sqrt{2}\mathcal{C}_1$.

Proposition 4.3.10. *We have*

$$\mathsf{E} \sup_{f \in \mathcal{C}_0} \Big| \sum_{i \le N} f(X_i) \Big| \le L \sqrt{N \log N} \,. \tag{4.74}$$

Proof. Consider the largest integer m with $2^{-m} \geq 1/N$. By (4.73), and since $\mathcal{C}_0 \subset \sqrt{2}\mathcal{C}_1$, we may find a subset T of $\mathcal{C}_0$ with $\operatorname{card} T \leq N_m$ and

$$\forall f \in \mathcal{C}_0\,,\ d_\infty(f,T) \leq L2^{-m/2} \leq L/\sqrt{N}\,.$$

Thus

$$\mathsf{E}\sup_{f\in\mathcal{C}_0}\Big|\sum_{i\leq N} f(X_i)\Big| \leq \mathsf{E}\sup_{f\in T}\Big|\sum_{i\leq N} f(X_i)\Big| + L\sqrt{N}\,. \tag{4.75}$$

To prove (4.74) it suffices to show that

$$\mathsf{E}\sup_{f\in T}\Big|\sum_{i\leq N} f(X_i)\Big| \leq L\sqrt{N\log N}\,. \tag{4.76}$$

Proposition 4.3.8 and Lemma 4.1.3 imply $\gamma_2(T,d_2) \leq L\sqrt{m} \leq L\sqrt{\log N}$. Using Corollary 2.3.2 and (4.73), and since $e_n(T,d_\infty) = 0$ for $n \geq m$ yields $\gamma_1(T,d_\infty) \leq L2^{m/2} \leq L\sqrt{N}$. Thus (4.76) follows from Theorem 4.3.6 and this completes the proof. □

Exercise 4.3.11. Use Exercise 4.3.7 to convince yourself that covering numbers cannot yield better than the estimate $\gamma_2(T,d_2) \leq L\log N$.

We claimed that $\mathcal{C}_0$, which is somewhat smaller than $\mathcal{C}$ (see (4.63)), constitutes "the main part" of $\mathcal{C}$. We turn now to the control of the "remainder", for which there is plenty of room. This exceedingly un-exciting argument could be skipped at first reading. (Rather, the reader is advised to save her energy to think about Exercise 4.3.13 below.) We consider the class $\mathcal{C}_2$ of functions of the type

$$f(x_1,x_2) = x_1 g(x_2)$$

where $g : [0,1] \to \mathbb{R}$ is 1-Lipschitz, $g(0) = g(1)$ and $|g| \leq 1$.

Proposition 4.3.12. *We have*

$$\mathsf{E}\sup_{f\in\mathcal{C}_2}\Big|\sum_{i\leq N}(f(X_i) - \int f\mathrm{d}\lambda)\Big| \leq L\sqrt{N}\,.$$

Proof. First we want to bound $N(\mathcal{C}_2,d_\infty,\epsilon)$. For this we observe that for any two functions g and g^* we have $|x_1g(x_2) - x_1g^*(x_2)| \leq d_\infty(g,g^*)$. We then mimic the proof of the entropy estimate (4.68). The difference is that now we are dealing with functions on $[0,1]$ rather than $[0,1]^2$, and that it suffices to use about $1/\epsilon$ points of $[0,1]$ to approximate each point of this interval within distance ϵ. In this manner we obtain that for $\epsilon > 0$ we have $N(\mathcal{C}_2,d_\infty,\epsilon) \leq \exp(L/\epsilon)$ and hence $e_n(\mathcal{C}_2,d_\infty) \leq L2^{-n}$. Thus Corollary 2.3.2 implies $\gamma_2(\mathcal{C}_2,d_2) \leq \gamma_2(\mathcal{C}_2,d_\infty) \leq L$.

Consider now the largest integer m such that $2^{-m} \geq 1/N$. We choose $T \subset \mathcal{C}_2$ with $\operatorname{card} T \leq N_m$ and

$$\forall f \in \mathcal{C}_2\,,\; d_\infty(f,T) \le L2^{-m}\,.$$

As in the proof of Proposition 4.3.10, we get $\gamma_1(T,d_\infty) \le Lm$ and we conclude by Theorem 4.3.6, using an inequality similar to (4.75), with huge room to spare. □

Proof of Theorem 4.3.2. We first observe that in (4.55) the supremum is the same if we replace the class $\mathcal{C}$ of 1-Lipschitz functions by the class of differentiable 1-Lipschitz functions. For a function f on $[0,1]^2$, we set $\Delta = f(1,1) - f(1,0) - f(0,1) + f(0,0)$ and we decompose

$$f = f_1 + f_2 + f_3 + f_4\,, \tag{4.77}$$

where

$$\begin{aligned} f_4(x_1,x_2) &= x_1x_2\Delta \\ f_3(x_1,x_2) &= x_2(f(x_1,1) - f(x_1,0) - \Delta x_1) \\ f_2(x_1,x_2) &= x_1(f(1,x_2) - f(0,x_2) - \Delta x_2) \\ f_1 &= f - f_2 - f_3 - f_4\,. \end{aligned}$$

It is straightforward to check that $f_1(x_1,0) = f_1(x_1,1)$ and $f_1(0,x_2) = f_1(1,x_2)$, so that if f is 2-Lipschitz and differentiable, f_1 is L-Lipschitz, differentiable, and $f_1 - \int f_1 \mathrm{d}\lambda$ satisfies (4.63). We then write

$$\Big|\sum_{i\le N}\big(f(X_i) - \int f\mathrm{d}\lambda\big)\Big| \le \sum_{j\le 4}\mathcal{D}_j$$

where $\mathcal{D}_j = |\sum_{i\le N}(f_j(X_i) - \int f_j \mathrm{d}\lambda)|$. We then deduce from Proposition 4.3.10 that $\mathsf{E}\sup_{f\in\mathcal{C}}\mathcal{D}_1 \le L\sqrt{N\log N}$ and from Proposition 4.3.12 that $\mathsf{E}\sup_{f\in\mathcal{C}}\mathcal{D}_2 \le L\sqrt{N}$ and $\mathsf{E}\sup_{f\in\mathcal{C}}\mathcal{D}_3 \le L\sqrt{N}$. Since obviously $\mathsf{E}\sup_{f\in\mathcal{C}}\mathcal{D}_4 \le L\sqrt{N}$ this completes the proof. □

Exercise 4.3.13. Consider the space $T = \{0,1\}^{\mathbb{N}}$ provided with the distance $d(t,t') = 2^{-j/2}$, where $j = \min\{i \ge 1; t_i \ne t'_i\}$ for $t = (t_i)_{i\ge 1}$. This space somewhat resembles the unit square, in the sense that $N(T,d,\epsilon) \le L\epsilon^{-2}$. Prove that if $(X_i)_{i\le N}$ are i.i.d. uniformly distributed in T and $(Y_i)_{i\le N}$ are uniformly spread (in a manner which is left to the reader to define precisely) then

$$\frac{1}{L}\sqrt{N}\log N \le \mathsf{E}\inf_\pi \sum_{i\le N} d(X_i, Y_\pi(i)) \le L\sqrt{N}\log N\;, \tag{4.78}$$

where the infimum is of course over all one to one maps π from $\{1,\dots,N\}$ to itself. (Hint: for the upper bound, covering numbers suffice, e.g. in the form of (4.68). Probably the challenging lower bound cannot be proved before one has meditated over the methods of Section 6.4.)

4.4 The Leighton-Shor Grid Matching Theorem

Theorem 4.4.1 ([4]). *If the points* $(Y_i)_{i\le N}$ *are evenly spread and if* $(X_i)_{i\le N}$ *are i.i.d. uniform over* $[0,1]^2$*, then (for* $N \ge 2$*), with probability at least* $1 - L\exp(-(\log N)^{3/2}/L)$ *we have*

$$\inf_\pi \sup_{i\le N} d(X_i, Y_{\pi(i)}) \le L\frac{(\log N)^{3/4}}{\sqrt{N}}\ , \tag{4.79}$$

and thus

$$\mathsf{E}\inf_\pi \sup_{i\le N} d(X_i, Y_{\pi(i)}) \le L\frac{(\log N)^{3/4}}{\sqrt{N}}\ . \tag{4.80}$$

To deduce (4.80) from (4.79) one simply uses any matching in the (rare) event that (4.79) fails. We shall prove in Section 6.4 that the inequality (4.80) can be reversed.

A first simple idea is that to prove Theorem 4.4.1 we do not care about what happens at a scale smaller than $(\log N)^{3/4}/\sqrt{N}$. Consider the largest integer ℓ_1 with $2^{-\ell_1} \ge (\log N)^{3/4}/\sqrt{N}$ (so that in particular $2^{\ell_1} \le \sqrt{N}$. We divide $[0,1]$ into little squares of side $2^{-\ell_1}$. For each such square, we are interested in how many points (X_i) it contains, but we do not care where these points are located in the square. We shall (as is the case for each matching theorem) deduce Theorem 4.4.1 from a discrepancy theorem for a certain class of functions. What we really have in mind is the class of functions which are indicators of a union A of little squares of side $2^{-\ell_1}$, and such that the boundary of A has a given length. It turns out that we shall have to parametrize the boundaries of these sets by curves, so it is convenient to turn things around and to consider the class of sets A that are the interiors of curves of given length.

To make things precise, let us define the *grid* G of $[0,1]^2$ of mesh width $2^{-\ell_1}$ by

$$G = \left\{(x_1, x_2) \in [0,1]^2\ ;\ 2^{\ell_1}x_1 \in \mathbb{N}\ \ \textbf{or}\ \ 2^{\ell_1}x_2 \in \mathbb{N}\right\}\ .$$

A *vertex* of the grid is a point $(x_1, x_2) \in [0,1]^2$ with $2^{\ell_1}x_1 \in \mathbb{N}$ **and** $2^{\ell_1}x_2 \in \mathbb{N}$. An *edge* of the grid is the segment between two vertices that are at distance $2^{-\ell_1}$ of each other. A *square* of the grid is a square of side $2^{-\ell_1}$ whose edges are edges of the grid. Thus, an edge of the grid is a subset of the grid, but a square of the grid is not a subset of the grid.

A *curve* is the image of a continuous map $\varphi : [0,1] \to \mathbb{R}^2$. We say that the curve is a *simple curve* if it is one-to-one on $[0,1[$. We say that the curve is *traced on* G if $\varphi([0,1]) \subset G$, and that it is *closed* if $\varphi(0) = \varphi(1)$. If C is a closed simple curve in $\mathbb{R}^2$, the set $\mathbb{R}^2 \setminus C$ has two connected components. One of these is bounded. It is called the interior of C and is denoted by $\overset{o}{C}$.

The key ingredient to Theorem 4.4.1 is as follows.

Theorem 4.4.2. *With probability at least* $1-L\exp(-(\log N)^{3/2}/L)$*, the following occurs. Given any closed simple curve* C *traced on* G*, we have*

$$\Big|\sum_{i\le N}\big(\mathbf{1}_{\overset{o}{C}}(X_i)-\lambda(\overset{o}{C})\big)\Big| \le L\ell(C)\sqrt{N}(\log N)^{3/4}\,, \tag{4.81}$$

where $\lambda(\overset{o}{C})$ *is the area of* $\overset{o}{C}$ *and* $\ell(C)$ *is the length of* C*.*

It will be easier to discuss the following result, which concerns curves of given length going through a given vertex.

Proposition 4.4.3. *Consider a vertex* τ *of* G *and* $k\in\mathbb{Z}$*. Define* $\mathcal{C}(\tau,k)$ *as the set of closed simple curves traced on* G *that contain* τ *and have length* $\le 2^k$*. Then, if* $k\le \ell_1+2$*, with probability at least* $1-L\exp(-(\log N)^{3/2}/L)$*, for each* $C\in\mathcal{C}(\tau,k)$ *we have*

$$\Big|\sum_{i\le N}\big(\mathbf{1}_{\overset{o}{C}}(X_i)-\lambda(\overset{o}{C})\big)\Big| \le L2^k\sqrt{N}(\log N)^{3/4}\,. \tag{4.82}$$

It would be easy to control the left-hand side if one considered only curves with a simple pattern, such as boundaries of rectangles. The point however is that the curves we consider can be very complicated, and of course, the longer we allow them to be, the more so. Let us denote by $\mathcal{F}$ the class of functions of the type $\mathbf{1}_{\overset{o}{C}}$, where $C\in C(\tau,k)$. Then the left-hand side of (4.82) is

$$\sup_{\mathcal{F}}\Big|\sum_{i\le N}(f(X_i)-\int f\mathrm{d}\lambda)\Big|\,.$$

To bound this quantity we shall use again Bernstein's inequality, together with Theorem 2.2.28 which is tailored to yields bounds in probability rather than in expectation. The key point again is the control on the size of $\mathcal{F}$ for the distance of $L^2(\lambda)$. The basis for this is to parametrize curves by Lipschitz functions on the unit interval and to use Proposition 4.1.8. In contrast with the previous section, no complications due to secondary terms mar the beauty of the proof.

We first deduce Theorem 4.4.2 from Proposition 4.4.3.

Proof of Theorem 4.4.2. Since there are at most $(2^{\ell_1}+1)^2\le LN$ choices for τ, we can assume with probability at least

$$1-L(2^{\ell_1}+1)^2(2\ell_1+4)\exp(-(\log N)^{3/2}/L)\ge 1-L'\exp\big(-(\log N)^{3/2}/L'\big)$$

that (4.82) occurs for all choices of $C\in\mathcal{C}(\tau,k)$, for any τ and any k with $-\ell_1\le k\le \ell_1+2$.

Consider a simple curve C traced on G. Then, bounding the length of C by the total length of the edges of G, we have

$$2^{-\ell_1} \le \ell(C) \le 2(2^{\ell_1}+1) \le 2^{\ell_1+2}\,,$$

so if k is the smallest integer for which $\ell(C) \le 2^k$, then $-\ell_1 \le k \le \ell_1 + 2$, so that we can use (4.82) and since $2^k \le 2\ell(C)$ the proof is finished. □

Lemma 4.4.4. *We have* $\operatorname{card}\mathcal{C}(\tau,k) \le 2^{2^{k+\ell_1+1}} = N_{k+\ell_1+1}$.

Proof. A curve $C \in \mathcal{C}(\tau,k)$ consists of at most $2^{k+\ell_1}$ edges of G. If we move through C, at each vertex of G we have at most 4 choices for the next edge, so $\operatorname{card}\mathcal{C}(\tau,k) \le 4^{2^{k+\ell_1}} = N_{k+\ell_1+1}$. □

On the set of closed simple curves traced on G, we define the distance d_1 by $d_1(C,C') = \lambda(\overset{o}{C}\,\Delta\,\overset{o}{C}{}')$.

Proposition 4.4.5. *We have*

$$\gamma_{1,2}(\mathcal{C}(\tau,k),\, d_1) \le L2^{2k}\,. \tag{4.83}$$

This is the main ingredient of Proposition 4.4.3; we shall prove it later. The next lemma reveals how the exponent 3/4 occurs. It uses the fact that $\sqrt{d}$ is a distance whenever d is a distance.

Lemma 4.4.6. *Consider a finite metric space* (T,d) *with* $\operatorname{card}T \le N_m$. *Then*

$$\gamma_2(T,\sqrt{d}) \le m^{3/4}\gamma_{1,2}(T,d)^{1/2}\,. \tag{4.84}$$

Proof. Since T is finite there exists an admissible sequence $(\mathcal{A}_n)$ of T such that

$$\forall t \in T\,,\ \sum_{n\ge 0}(2^n\Delta(A_n(t),d))^2 \le \gamma_{1,2}(T,d)^2\,. \tag{4.85}$$

Without loss of generality we can assume that $A_m(t) = \{t\}$ for each t, so that in (4.85) the sum is over $n \le m-1$. Now

$$\Delta(A,\sqrt{d}) \le \Delta(A,d)^{1/2}$$

so that, using Hölder's inequality,

$$\begin{aligned}\sum_{0\le n\le m-1} 2^{n/2}\Delta(A_n(t),\sqrt{d}) &\le \sum_{0\le n\le m-1}(2^n\Delta(A_n(t),d))^{1/2}\\ &\le m^{3/4}\Big(\sum_{n\ge 0}\big(2^n\Delta(A_n(t),d)\big)^2\Big)^{1/4}\\ &\le m^{3/4}\gamma_{1,2}(T,d)^{1/2}\,,\end{aligned}$$

which concludes the proof. □

Of course the real secret of the exponent 3/4 is that we shall use Lemma 4.4.6 in situations where inequality (4.84) is basically an equality. This will become apparent only in Section 6.5, where we prove that Theorem 4.4.1 is in a sense optimal.

On the set of simple curves traced on G we consider the distance

$$d_2(C_1, C_2) = \sqrt{N}\Big\|\mathbf{1}_{\overset{o}{C}_1} - \mathbf{1}_{\overset{o}{C}_2}\Big\|_2 = \big(N d_1(C_1, C_2)\big)^{1/2} , \tag{4.86}$$

so that

$$\gamma_2(\mathcal{C}(\tau, k), d_2) \le \sqrt{N}\gamma_2(\mathcal{C}(\tau, k), \sqrt{d_1}) .$$

When $k \le \ell_1 + 2$ we have $m := k + \ell_1 + 1 \le L \log N$, so that combining Proposition 4.4.5 with Lemmas 4.4.4 and 4.4.6 we obtain

$$\gamma_2(\mathcal{C}(\tau, k), d_2) \le L2^k \sqrt{N}(\log N)^{3/4} . \tag{4.87}$$

Proof of Proposition 4.4.3. It relies on Theorem 2.2.28. On $T = \mathcal{C}(\tau, k)$ consider the process

$$X_C := \sum_{i \le N} (\mathbf{1}_{\overset{o}{C}}(X_i) - \lambda(\overset{o}{C})) ,$$

the distance d_2 given by (4.86) and distance δ given by $\delta(C, C') = 1$ if $C \ne C'$ and $\delta(C, C') = 0$ if $C = C'$. We have $X_C - X_{C'} = \sum_i Y_i$ where

$$Y_i = \mathbf{1}_{\overset{o}{C}}(X_i) - \mathbf{1}_{\overset{o}{C}'}(X_i) - \lambda(\overset{o}{C}) + \lambda(\overset{o}{C}') ,$$

so that $\|Y_i\|_\infty \le 4\delta(C, C')$ and $\sum_{i \le N} \mathsf{E}Y_i^2 \le d_2(C, C')^2$. It then follows from Bernstein's inequality (4.59) that for $u > 0$ the process (X_C) satisfies

$$\mathsf{P}(|X_C - X_{C'}| > u) \le \exp\Big(-\frac{1}{L}\min\Big(\frac{u^2}{d_2(C, C')^2}, \frac{u}{\delta(C, C')}\Big)\Big) .$$

Using an admissible sequence (A_n) such that $A_n(t) = \{t\}$ when $n = k + \ell_1 + 1$, Lemma 4.4.4 implies

$$\gamma_1(T, \delta) \le L \log \operatorname{card} C \le L2^{k+\ell_1} \le L2^k\sqrt{N} . \tag{4.88}$$

Moreover, since $e_n(T, \delta) \le 1$ and $e_n(T, \delta) = 0$ for $n \ge k + \ell_1 + 1$, we have

$$\sum_n e_n(T, \delta) \le k + \ell_1 + 1 . \tag{4.89}$$

Also, from (4.83) and (4.10) we obtain $e_n(T, d_1) \le L2^{2k}2^{-n}$, so that $e_n(T, d_2) \le L2^k 2^{-n/2}\sqrt{N}$ and

$$\sum_{n \ge 0} e_n(T, d_2) \le L2^k \sqrt{N} . \tag{4.90}$$

Now we use (2.64), (4.87) and (4.88) to (4.90) to obtain

$$\mathsf{P}\Big(\sup_{C,C'} |X_C - X_{C'}| \ge L\big(2^k\sqrt{N}(\log N)^{3/4} + u_1(k+\ell_1+1) + u_2 2^k\sqrt{N}\big)\Big)$$
$$\le \exp(-\min(u_2^2, u_1)) \,.$$

We now choose $u_1 = (\log N)^{3/2}$ and $u_2 = (\log N)^{3/4}$. Since $\ell_1 \le \log N$ and since $X_{C'} = 0$ when C' is the empty curve we obtain the desired bound. □

We turn to the proof of Proposition 4.4.5. It holds simply because the metric space $(\mathcal{C}(\tau,k), d_1)$ is a Lipschitz image of a subset of the set $\mathcal{L}$ of Proposition 4.1.8. In Lemma 4.4.7 we check the obvious fact that the functionals $\gamma_{\alpha,\beta}$ behave as expected under Lipschitz maps, and in Lemma 4.4.8 we construct an actual Lipschitz map from a subset of $\mathcal{L}$ onto $\mathcal{C}(\tau_k, d_1)$, a boring but completely elementary task.

Lemma 4.4.7. *Consider two metric spaces* (T,d) *and* (U,d')*. If* $f : (T,d) \to (U,d')$ *is onto and satisfies*

$$\forall x\,,\, y \in T\,,\; d'(f(x), f(y)) \le A d(x,y)$$

for a certain constant A*, then*

$$\gamma_{\alpha,\beta}(U,d') \le K(\alpha,\beta) A \gamma_{\alpha,\beta}(T,d)\,.$$

Proof. We proceed as in Theorem 2.7.5, (b). It is straight forward to extend the second proof of Theorem 2.3.1 (given on page 66) to the case of $\gamma_{\alpha,\beta}$. □

Lemma 4.4.8. *There exists a map* W *from a subset* T *of* $\mathcal{L}$ *onto* $\mathcal{C}(\tau,k)$ *which for any* $f_1, f_2 \in T$ *satisfies*

$$d_1(W(f_0), W(f_1)) \le L 2^{2k} \|f_0 - f_1\|_2 \,. \tag{4.91}$$

Proof. Consider the subset $\mathcal{L}^*$ of $\mathcal{L}$ consisting of the functions f for which $f(1/2) = 0$. To $f \in \mathcal{L}^*$ we associate the curve $W(f)$ traced out by the map

$$u \mapsto \bigg(\tau^1 + 2^{k+1} f\big(\frac{u}{2}\big), \tau^2 + 2^{k+1} f\big(\frac{u+1}{2}\big)\bigg)\,,$$

where $(\tau^1, \tau^2) = \tau$. A curve in $\mathcal{C}(\tau,k)$ can be parameterized, starting at τ and moving at speed 1 along each successive edges. It is therefore the range of a map of the type $t \mapsto (\tau^1 + f_1(t), \tau^2 + f_2(t))$ where f_1 and f_2 are Lipschitz maps from $[0, 2^k]$ to $\mathbb{R}$ with $f_1(0) = f_2(0) = f_1(2^k) = f_2(2^k) = 0$. Considering the function f on $[0,1]$ given by $f(u) = 2^{-k-1} f_1(2^{k+1}u)$ for $u \le 1/2$ and $f(u) = 2^{-k-1} f_2(2^{k+1}(u - 1/2))$ for $1/2 \le u \le 1$ proves that $\mathcal{C}(\tau,k) \subset W(\mathcal{L}^*)$. We set $T = W^{-1}(\mathcal{C}(\tau,k))$. Consider f_0 and f_1 in T and the map $h : [0,1]^2 \to [0,1]^2$ given by

$$h(u,v) = \Big(\tau^1 + 2^{k+1}\Big(vf_0\big(\frac{u}{2}\big) + (1-v)f_1\big(\frac{u}{2}\big)\Big)\,,$$
$$\tau^2 + 2^{k+1}\Big(vf_0\big(\frac{1+u}{2}\big) + (1-v)f_1\big(\frac{1+u}{2}\big)\Big)\Big)\,.$$

The area of $h([0,1]^2)$ is at most $\iint_{[0,1]^2} |Jh(u,v)|\mathrm{d}u\mathrm{d}v$, where Jh is the Jacobian of h, and a straightforward computation gives

$$Jh(u,v) = 2^{2k+1}\Big(\Big(vf_0'\big(\frac{u}{2}\big) + (1-v)f_1'\big(\frac{u}{2}\big)\Big)\Big(f_0\big(\frac{1+u}{2}\big) - f_1\big(\frac{1+u}{2}\big)\Big)$$
$$- \Big(vf_0'\big(\frac{1+u}{2}\big) + (1-v)f_1'\big(\frac{1+u}{2}\big)\Big)\Big(f_0\big(\frac{u}{2}\big) - f_1\big(\frac{u}{2}\big)\Big)\Big)\,,$$

so that, since $|f_0'| \le 1$, $|f_1'| \le 1$,

$$|Jh(u,v)| \le 2^{2k+1}\Big(\Big|f_0\big(\frac{u}{2}\big) - f_1\big(\frac{u}{2}\big)\Big| + \Big|f_0\big(\frac{1+u}{2}\big) - f_1\big(\frac{1+u}{2}\big)\Big|\Big)\,.$$

The Cauchy-Schwarz inequality implies

$$\iint |Jh(u,v)|\mathrm{d}u\mathrm{d}v \le L2^{2k}\|f_0 - f_1\|_2\,. \tag{4.92}$$

If x does not belong to the range of h, both curves $W(f_0)$ and $W(f_1)$ "turn the same number of times around x". This is because "the number of times the closed curve $u \mapsto h(u,v)$ turns around x" is then a continuous function of v, so that since it is integer valued, it takes the same value for $u = 0$ and $u = 1$. Consequently either $x \in \overset{o}{W}(f_0) \cap \overset{o}{W}(f_1)$ or $x \notin \overset{o}{W}(f_0) \cup \overset{o}{W}(f_1)$. Thus the range of h contains $\overset{o}{W}(f_0) \Delta \overset{o}{W}(f_1)$, and (4.92) implies (4.91). □

Proof of Proposition 4.4.5. Combine Proposition 4.1.8 with Lemmas 4.4.7 and 4.4.8. □

It remains to deduce Theorem 4.4.1 from Theorem 4.4.2. This is a purely deterministic argument, which is unrelated to any other material in the present book. The basic idea is very simple, and to keep it simple we describe it in slightly imprecise terms. Consider a union A of little squares of side $2^{-\ell_1}$ and the union A' of all the little squares that touch A. We want to prove that A' contains as many points Y_i as A contains points X_i, so that by Hall's Marriage Lemma each point X_i can be matched to a point Y_i in the same little square, or in a neighbor of it. Since the points Y_i are evenly spread the number of such points in A' is $N\lambda(A')$. There may be more than $N\lambda(A)$ points X_i in A, but (4.81) tells us that the excess number of points cannot be more than a proportion of the length ℓ of the boundary of A. The marvelous fact is that we may also expect that $\lambda(A') - \lambda(A)$ is also proportional to ℓ, so that we may hope that the excess number of points X_i in A

should not exceed $N(\lambda(A') - \lambda(A))$, proving the result. A slight problem is that the proportionality constant is not quite right to make the argument work, but this difficulty is bypassed simply by applying the same argument to a slightly coarser grid. A more serious problem is that when one tries to describe precisely what is meant by the previous argument, without waving hands but with complete proofs, one has to check a number of details. These are completely elementary, but require patience. We find that the Leighton-Shor Theorem deserves this effort, and we have written every step in full detail as well as we could.

As a last preparation for this effort, we say that a simple curve C traced on G is a *chord* if it is the range of $[0,1]$ by a continuous map φ where $\varphi(0)$ and $\varphi(1)$ belong to the boundary of $[0,1]^2$. If C is a chord, $]0,1[^2 \setminus C$ is the union of two regions R_1 and R_2, and (assuming without loss of generality that no point X_i belongs to G),

$$\sum_{i\le N}\big(\mathbf{1}_{R_1}(X_i) - \lambda(R_1)\big) = -\sum_{i\le N}\big(\mathbf{1}_{R_2}(X_i) - \lambda(R_2)\big) .$$

We define

$$\mathcal{D}(C) = \Big|\sum_{i\le N}(\mathbf{1}_{R_1}(X_i) - \lambda(R_1))\Big| = \Big|\sum_{i\le N}(\mathbf{1}_{R_2}(X_i) - \lambda(R_2))\Big| .$$

If C is a chord, "completing C by following the boundary of $[0,1]^2$" produces a closed simple curve C' on G such that either $R_1 = \overset{o}{C'}$ or $R_2 = \overset{o}{C'}$. The length we add along each side of the boundary is less than the length of the chord itself, so that $\ell(C') \le 3\ell(C)$. Thus, the following is a consequence of Theorem 4.4.2.

Theorem 4.4.9. *With probability at least* $1 - L\exp(-(\log N)^{3/2}/L)$*, for each chord* C *we have*

$$\mathcal{D}(C) \le L\ell(C)\sqrt{N}(\log N)^{3/4} . \tag{4.93}$$

Proof of Theorem 4.4.1. Consider a number $\ell_2 < \ell_1$, to be determined later, and the grid $G' \subset G$ of mesh width $2^{-\ell_2}$. (This is the slightly coarser grid we mentioned above.)

A union of squares of G' is called a *domain*. Given a domain R, we denote by R' the union of the squares of G' such one at least one of the 4 edges that form their boundary is entirely contained in R (recall that squares include their boundaries). The main argument is to establish that if (4.81) and (4.93) hold, and provided ℓ_2 has been chosen appropriately, then for any choice of R we have

$$N\lambda(R') \ge \operatorname{card}\{i \le N\ ;\ X_i \in R\} . \tag{4.94}$$

We will then conclude with Hall's Marriage Lemma. The basic idea to prove (4.94) is to reduce to the case where R is the closure of the interior of a simple closed curve minus a number of "holes" which are themselves the interiors of simple closed curves.

Let us say that a domain R is *decomposable* if $R = R_1 \cup R_2$ where R_1 and R_2 are non-empty unions of squares of G', and when every square of G' included in R_1 has at most one vertex belonging to R_2. (Equivalently, $R_1 \cap R_2$ is finite.) We can write $R = R_1 \cup \ldots \cup R_k$ where each R_j is undecomposable (i.e. not decomposable) and where any two of these sets have a finite intersection. This is obvious by writing R as the union of as many domains as possible, under the condition that the intersection of any of two of these domains is finite. Then each of them must be undecomposable.

We claim that

$$\frac{1}{4}\sum_{\ell\le k}\lambda(R'_\ell\backslash R_\ell) \le \lambda(R'\backslash R)\ . \tag{4.95}$$

To see this, let us set $\mathcal{S}_\ell = R'_\ell\backslash R_\ell$, so that by definition of R'_ℓ, $\mathcal{S}_\ell$ is the union of the squares $\mathcal{D}$ of G' that have at least one of the edges that form their boundary contained in R_ℓ but are not themselves contained in R_ℓ. Obviously we have $\mathcal{S}_\ell \subset R'$. When $\ell \ne \ell'$, the sets R_ℓ and $R_{\ell'}$ have a finite intersection, so that a square $\mathcal{D}$ contained in $\mathcal{S}_\ell$ cannot be contained in $R_{\ell'}$, since it has an entire edge contained in R_ℓ. Since $\mathcal{D}$ is not contained in R_ℓ either, it is not contained in R. Thus the interior of $\mathcal{D}$ is contained in $R'\backslash R$, and since this is true for any square $\mathcal{D}$ of $\mathcal{S}_\ell$ and any $\ell \le k$, we have

$$\lambda\Big(\bigcup_{\ell\le k}\mathcal{S}_\ell\Big) \le \lambda(R'\backslash R)\ .$$

Moreover, a given square $\mathcal{D}$ of G' can be contained in a set $\mathcal{S}_\ell$ for at most 4 values of ℓ (one for each of the edges of $\mathcal{D}$), so that

$$\sum_{\ell\le k}\lambda(R'_\ell \setminus R_\ell) = \sum_{\ell\le k}\lambda(\mathcal{S}_\ell) \le 4\lambda\Big(\bigcup_{\ell\le k}\mathcal{S}_\ell\Big)\ .$$

This proves (4.95).

To prove that (4.94) holds for any domain R, it suffices to prove that when R is an undecomposable domain we have (pessimistically)

$$\frac{N}{4}\lambda(R'\backslash R) \ge \operatorname{card}\{i \le N\ ;\ X_i \in R\} - N\lambda(R)\ . \tag{4.96}$$

Indeed, writing (4.96) for $R = R_\ell$, summing over $\ell \le k$ and using (4.95) implies (4.94).

We turn to the proof of (4.96) when R is an undecomposable domain. The boundary S of R is a subset of G'. Inspection of the cases shows that:

$$\text{If a vertex } \tau \text{ of } G' \text{ belongs to } S,\ \text{either 2 or 4 of the edges of } G' \text{ incident to } \tau \text{ are contained in } S. \tag{4.97}$$

Next we show that any subset S of G' that satisfies (4.97) is a union of closed simple curves, any two of them intersecting only at vertices of G'. (This is simply the decomposition into cycles of Eulerian graphs.) To see this, it suffices to construct a closed simple curve C contained in S, to remove C from S and to iterate, since $S \setminus C$ still satisfies (4.97). The construction goes as follows. Starting with an edge $\tau_1\tau_2$ in S, we find successively edges $\tau_2\tau_3, \tau_3\tau_4, \ldots$ with $\tau_k \neq \tau_{k-2}$, and we continue the construction until the first time $\tau_k = \tau_\ell$ for some $\ell \leq k-2$ (in fact $\ell \leq k-3$). Then the edges $\tau_\ell\tau_{\ell+1}, \tau_{\ell+1}\tau_{\ell+2}, \ldots, \tau_{k-1}\tau_k$ define a closed simple curve contained in S.

Thus the boundary of an undecomposable domain R is a union of closed simple curves $C_1, \ldots, C_k$, any two of them having at most a finite intersection.

We next show that for each ℓ, the set R is either contained in the closure C_ℓ^* of $\overset{o}{C}_\ell$ (so that C_ℓ is then the "outer boundary" of R) or else $\overset{o}{C}_\ell \cap R = \emptyset$ (in which case $\overset{o}{C}_\ell$ is "a hole" in R). Let us fix ℓ and assume otherwise for contradiction. Consider the domain R_1 which is the union of the squares of G' that are contained in R but not in C_ℓ^*, so that R_1 is not empty by hypothesis. Consider also the domain R_2 that is the union of the squares of G' contained in R whose interiors are contained in $\overset{o}{C}_\ell$. Then R_2 is not empty either. Given a square of G', and since $\overset{o}{C}_\ell$ is the interior of C_ℓ^*, either its interior is contained in $\overset{o}{C}_\ell$ or else the square is not contained in C_ℓ^*. This proves that $R = R_1 \cup R_2$. Next we show that the domains R_1 and R_2 cannot have an edge of the grid G' in common. Assuming for contradiction that such an edge exists, it is an edge of exactly 2 squares A and B of G'. One of these squares is a subset of R_1 and the other is a subset of R_2. Thus the edge must belong to C_ℓ for otherwise A and B would be "on the same side of C_ℓ" and they would both be subsets of R_1 or both subsets of R_2. Next, we observe that this edge cannot be on the boundary of R because both A and B are subsets of R. This contradicts the fact that C_ℓ is contained in the boundary of R, therefore proving that R_1 and R_2 cannot have an edge in common. Since $R = R_1 \cup R_2$, this in turn would imply that R is decomposable, contradicting our assumption.

Without loss of generality we assume that C_1 is the outer boundary of R, and that for $2 \leq \ell \leq k$ we have $R \cap \overset{o}{C}_\ell = \emptyset$. The goal now is to prove that

$$R = C_1^* \setminus \bigcup_{2 \leq \ell \leq k} \overset{o}{C}_\ell \,. \tag{4.98}$$

It is obvious that $R \subset C_1^* \setminus \bigcup_{2 \leq \ell \leq k} \overset{o}{C}_\ell$ so that we have to show that $D := (C_1^* \setminus \bigcup_{2 \leq \ell \leq k} \overset{o}{C}_\ell) \setminus R$ is empty. We assume for contradiction that D is not empty. Consider a square A of G' which is contained in D, and a square A' of G' which has an edge in common with A. First, we claim that $A' \subset C_1^*$. Otherwise, A and A' would have to be on different sides of C_1, which means that their common edge has to belong to C_1 and hence to the boundary of

R. This is impossible because neither A nor A' is then a subset of R. Indeed in the case of A' this is because we assume that $A' \not\subset C_1^*$, and in the case of A this is because we assume that $A \subset D$. Exactly the same argument shows that the interior of A' cannot be contained in $\overset{o}{C}_\ell$ for $2 \le \ell \le k$. Indeed then A and A' would be on different sides of C_ℓ so that their common edge would belong to C_ℓ and hence to the boundary of R, which is impossible since neither A nor A' is a subset of R. We have now shown that A and A' lie on the same side of each curve C_ℓ, so that their common edge cannot belong to the boundary of R, and since A is not contained in R this is not the case of A' either. Consequently the definition of D shows that $A' \subset D$, but since A was an arbitrary square contained in D, this is absurd, and completes the proof that $D = \emptyset$ and of (4.98).

Let $R_\ell^\sim$ be the union of the squares of G' that have at least one edge contained in C_ℓ. Thus, as in (4.95), we have

$$\sum_{\ell \le k} \lambda(R_\ell^\sim \backslash R) \le 4\lambda(R' \backslash R)$$

and to prove (4.96) it suffices (recalling that we assume that no point X_i belongs to G) to show that for each $1 \le \ell \le k$ we have

$$\left|\mathrm{card}\left\{i \le N\ ;\ X_i \in \overset{o}{C}_\ell\right\} - \lambda(\overset{o}{C}_\ell)\right| \le N 2^{-4} \lambda(R_\ell^\sim \backslash R)\ . \tag{4.99}$$

For $\ell \ge 2$, C_ℓ does not intersect the boundary of $[0,1]^2$. Each edge contained in C_ℓ is in the boundary of R. One of the 2 squares of G' that contain this edge is included in $R_\ell^\sim \backslash R$, and the other in R. Since a given square contained in $R_\ell^\sim \backslash R$ must arise in this manner from one of its 4 edges, we have

$$\lambda(R_\ell^\sim \backslash R) \ge \frac{1}{4} 2^{-\ell_2} \ell(C_\ell)\ . \tag{4.100}$$

On the other hand, (4.81) implies

$$\left|\mathrm{card}\left\{i \le N\ ;\ X_i \in \overset{o}{C}_\ell\right\} - \lambda(\overset{o}{C}_\ell)\right| \le L\ell(C_\ell)\sqrt{N}(\log N)^{3/4}\ ,$$

so that (4.99) follows provided

$$2^{-\ell_2} \ge \frac{2^6 L}{\sqrt{N}} (\log N)^{3/4}\ , \tag{4.101}$$

where L is the constant of (4.81).

When $\ell = 1$, (4.100) need not be true because parts of C_1 might be traced on the boundary of $[0,1]^2$. In that case we simply decompose C_1 in a union of chords and of parts of the boundary of $[0,1]^2$ to deduce (4.99) from (4.93).

Thus we have proved that (4.81) and (4.93) imply (4.94) provided that (4.101) holds. Next, for a domain R, we denote by R^* the set of points which are within distance $2^{-\ell_2}$ of R', and we show that, provided

$$2^{-\ell_2} \geq \frac{20}{\sqrt{N}} \tag{4.102}$$

we have

$$\operatorname{card}\{i \leq N\,;\; Y_i \in R^*\} \geq N\lambda(R')\,. \tag{4.103}$$

This is simply because since the sequence Y_i is widely spread, the points Y_i are centers of disjoint rectangles of area $1/N$ and diameter $\leq 20/\sqrt{N}$. There are at least $N\lambda(R')$ points Y_i such that the corresponding rectangle intersects R' (because the union of these rectangles cover R') and (4.102) implies that these little rectangles are entirely contained in R^*. Therefore (4.94) and (4.103) imply

$$\operatorname{card}\{i \leq N\,;\; Y_i \in R^*\} \geq \operatorname{card}\{i \leq N\,;\; X_i \in R\}\,. \tag{4.104}$$

Next, consider a subset I of $\{1, \ldots, N\}$ and let R be the domain that is the union of the squares of G' that contain at least a point X_i , $i \in I$. Then, using (4.104),

$$\operatorname{card} I \leq \operatorname{card}\{i \leq N\,;\; X_i \in R\} \leq \operatorname{card}\{i \leq N\,;\; Y_i \in R^*\}\,. \tag{4.105}$$

A point of R' is within distance $2^{-\ell_2}$ of a point of R. A point of R^* is within distance $2^{-\ell_2+1}$ of a point of R. A point of R is within distance $\sqrt{2}\cdot 2^{-\ell_2} \leq 2^{-\ell_2+1}$ of a point X_i with $i \in I$. Consequently each point of R^* is within distance $\leq 2^{-\ell_2+2}$ of a point X_i with $i \in I$. Therefore if we define

$$A(i) = \left\{j \leq N\,;\; d(X_i, Y_j) \leq 2^{-\ell_2+2}\right\},$$

we have proved that $\{j \leq N; Y_j \in R^*\} \subset \bigcup_{i\in I} A(i)$, and combining with (4.105) that

$$\operatorname{card} \bigcup_{i\in I} A(i) \geq \operatorname{card} I\,.$$

Hall's Marriage Lemma (Corollary 4.2.2) then shows that we can find a matching π for which $Y_{\pi(i)} \in A_i$ for any $i \leq N$, so that

$$\sup_{i\leq N} d(X_i, Y_{\pi(i)}) \leq 2^{-\ell_2+2} \leq \frac{L}{\sqrt{N}}(\log N)^{3/4}\,,$$

by taking for ℓ_2 the largest integer that satisfies (4.101) and (4.102). Since this is true whenever (4.81) and (4.93) occur, the proof of (4.79) is complete. □

4.5 Notes and Comments

The original proof of the Leighton-Shor theorem amount basically to perform by hand a kind generic chaining in this highly non-trivial case, an incredible tour de force. A first attempt was made in [6] to relate (an important consequence of) the Leighton-Shor theorem to general methods for bounding stochastic processes, but runs into technical complications. Then Coffman and Shor [3] introduced the use of Fourier transforms and brought to light the role of ellipsoids, after which it became clear that the structure of these ellipsoids plays a central part in these matching results, a point of view systematically expounded in [8].

Chapter 14 is a continuation of the present chapter. The more difficult material it contains is presented later for fear of scaring readers at this early stage. A notable feature of the result presented there is that ellipsoids do not suffice, a considerable source of complication.

One could wonder for which kind of probability distributions on the unit square Theorem 4.3.1 remains true. The intuition is that the uniform distribution considered in Theorem 4.3.1 is the "worst possible". This intuition is correct. This is proved in [7]. The proof is overall similar but one has to find an appropriate substitute for Fourier transforms. The situation is different for Theorem 4.4.1, as is shown by the trivial example of a distribution concentrated at exactly two points at distance d (where the reader will show that the best matching typically requires moving some of the random points for a distance d).

Methods similar to those of this chapter may be used to obtain non-trivial discrepancy theorems for various classes of functions, a topic which is investigated in [8]. Let us mention one such result. We denote by λ the uniform probability on the unit cube $[0,1]^3$, and by $(X_i)_{i\le N}$ independent uniformly distributed r.v.s valued in this unit cube.

Theorem 4.5.1. *Consider the class $\mathcal{C}$ of convex sets in $\mathbb{R}^3$. Then*

$$\mathsf{E}\sup_{C\in\mathcal{C}} |\operatorname{card}\{i\le N\,;\ X_i\in C\} - N\lambda(C)| \le L\sqrt{N}(\log N)^{3/4}\,.$$

The original results of [1] are proved using an interesting technique called the *transportation method*. A version of this method, which avoids many of the technical difficulties of the original approach is presented in [9]. With the notation of Theorem 4.3.1, it is proved in [9] (a stronger version of the fact) that with probability $\ge 9/10$ one has

$$\inf_\pi \frac{1}{N}\sum_{i\le N}\exp\Big(\frac{Nd(X_i,Y_{\pi(i)})^2}{K\log N}\Big)\le 2\,. \tag{4.106}$$

It has not been investigated whether this result could be obtained by the methods presented here.

Since $\exp x\ge x$, (4.106) implies that $\sum_{i\le N} d(X_i,Y_{\pi_i})^2\le \log N$ and hence using the Cauchy-Schwarz inequality

$$\sum_{i\le N} d(X_i,Y_{\pi(i)})\le L\sqrt{N\log N}\,. \tag{4.107}$$

Moreover it implies also

$$\max_{i \le N} d(X_i, Y_{\pi(i)}) \le L \log N / \sqrt{N} \,. \tag{4.108}$$

It does not seem known whether one can achieve simultaneously (4.107) and $\max_{i \le N} d(X_i, Y_{\pi(i)}) \le L(\log N)^{3/4}/\sqrt{N}$. In this circle of idea, see the ultimate matching conjecture on page 447.

For results about matching for unbounded distributions, see the work of J. Yukich [10].

References

1. Ajtai, M., Komlós, J., Tusnády, G.: On optimal matchings. Combinatorica **4**(4), 259–264 (1984)
2. Bollobas, B.: Combinatorics. Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability. Cambridge University Press, Cambridge (1986)
3. Coffman, E.G. Jr., Shor, P.W.: A simple proof of the $O(\sqrt{n}\log^{3/4} n)$ upright matching bound. SIAM J. Discrete Math. **4**(1), 48–57 (1991)
4. Leighton, T., Shor, P.: Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. Combinatorica **9**(2), 161–187 (1989)
5. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces, II. Function Spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 97. Springer, Berlin (1979). x+243 pp. ISBN: 3-540-08888-1
6. Rhee, W., Talagrand, M.: Exact bounds for the stochastic upward matching problem. Trans. Am. Math. Soc. **109**, 109–126 (1988)
7. Talagrand, M.: The Ajtai-Komlos-Tusnady matching theorem for general measures. In: Probability in a Banach Space. Progress in Probability, vol. 8, pp. 39–54. Birkhäuser, Basel (1992)
8. Talagrand, M.: Matching theorems and discrepancy computations using majorizing measures. J. Am. Math. Soc. **7**, 455–537 (1994)
9. Talagrand, M., Yukich, J.: The integrability of the square exponential transportation cost. Ann. Appl. Probab. **3**, 1100–1111 (1993)
10. Yukich, J.: Some generalizations of the Euclidean two-sample matching problem. In: Probability in Banach Spaces, vol. 8, Brunswick, ME, 1991. Progr. Probab., vol. 30, pp. 55–66. Birkhäuser, Boston (1992)

5. Bernoulli Processes

5.1 Boundedness of Bernoulli Processes

Arguably Gaussian r.v.s are the central object of Probability theory, but Bernoulli (= coin-flipping) r.v.s are also very useful. (We recall that ε is a Bernoulli r.v. if $\mathsf{P}(\varepsilon = \pm 1) = 1/2$.) In particular Bernoulli r.v.s are often involved when one deals with symmetric r.v.s. We have already used this procedure in Section 3.2.

Consider a subset T of $\ell^2 = \ell^2(\mathbb{N}^*)$, and i.i.d. Bernoulli r.v.s $(\varepsilon_i)_{i\geq 1}$. The Bernoulli process defined by T is the family $(X_t)_{t\in T}$ where $X_t = \sum_{i\geq 1} t_i \varepsilon_i$. We have explained on page 46 the fundamental importance of processes of this type for the study of random series of functions, and indeed the abstract results we shall prove in this chapter (and in particular Theorem 5.2.1 below) will be crucial for our understanding of random Fourier series with general coefficients in Chapter 7.

We set

$$b(T) := \mathsf{E} \sup_{t\in T} X_t = \mathsf{E} \sup_{t\in T} \sum_{i\geq 1} t_i \varepsilon_i \ , \tag{5.1}$$

a definition that mimics the case of Gaussian processes, where we defined

$$g(T) = \mathsf{E} \sup_{t\in T} \sum_{i\geq 1} t_i g_i .$$

We observe that $b(T) \geq 0$, that $b(T) \leq b(T')$ if $T \subset T'$, and that $b(T + t_0) = b(T)$, where $T + t_0 = \{t + t_0 ; t \in T\}$.

We would like to understand the value of $b(T)$ from the geometry of T, as we did in the case of Gaussian processes. As we already observed in Chapter 3, the subgaussian inequality (3.11) implies that if d denotes the distance in ℓ^2, then the process $(X_t)_{t\in T}$ satisfies the increment condition (1.4):

$$\mathsf{P}(|X_s - X_t| \geq u) \leq 2 \exp\Big(-\frac{u^2}{2d(s,t)^2}\Big) \ , \tag{1.4}$$

so that Theorem 2.2.22 implies

$$b(T) \leq L\gamma_2(T) \ , \tag{5.2}$$

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_5,

where we remind the reader that we often write $\gamma_2(T)$ instead of $\gamma_2(T,d)$ when d is the ℓ^2 distance. Since $\gamma_2(T) \le Lg(T)$ by Theorem 2.4.1, Bernoulli processes "are smaller than the corresponding Gaussian processes". This is also a consequence of the following avatar of (3.27), with the same proof.

Proposition 5.1.1. *We have*

$$b(T) \le \sqrt{\frac{\pi}{2}} g(T) . \tag{5.3}$$

Thus, we can bound a Bernoulli process by comparing it with a Gaussian process, or equivalently by using (5.2). There is however a completely different method to bound Bernoulli processes. We denote by $\|t\|_1 = \sum_{i\ge1} |t_i|$ the ℓ^1 norm of t. The following is trivial.

Proposition 5.1.2. *We have*

$$b(T) \le \sup_{t\in T} \|t\|_1 . \tag{5.4}$$

Thus, we have found two very different ways in which we can bound $b(T)$. To understand how different these was are, the reader can consider the following two cases: $T = \{u, 0\}$ where $u \notin \ell^1$ and T the unit ball of ℓ^1. The following definition and proposition formalize the idea that we can also use mixtures of the previous situations.

Definition 5.1.3. *For a subset T of ℓ^2, we set*

$$b^*(T) := \inf\Big\{\gamma_2(T_1) + \sup_{t\in T_2} \|t\|_1 \; ; \; T \subset T_1 + T_2\Big\} . \tag{5.5}$$

Here of course $T_1 + T_2$ is the Minkowski sum

$$T_1 + T_2 = \{t_1 + t_2 \; ; \; t_1 \in T_1, t_2 \in T_2\} . \tag{5.6}$$

Note that $X_{t_1+t_2} = X_{t_1} + X_{t_2}$ and hence

$$\sup_{t\in T_1+T_2} X_t \le \sup_{t\in T_1} X_t + \sup_{t\in T_2} X_t .$$

Taking expectation yields $b(T) \le b(T_1+T_2) \le b(T_1)+b(T_2)$. Combining with (5.2) and (5.4), we have proved the following.

Proposition 5.1.4. *We have*

$$b(T) \le Lb^*(T) . \tag{5.7}$$

It is natural to conjecture that the previous bound on $b(T)$ is sharp. This was known as the Bernoulli Conjecture. It took nearly 25 years to prove it.

Theorem 5.1.5 (The Bednorz-Latała theorem). *There exists a universal constant L such that given any subset T of ℓ^2 we have*

$$b^*(T) \le Lb(T) . \tag{5.8}$$

The proof of Theorem (5.1.5) will consist in describing a procedure to decompose each point $t \in T$ as a sum $T = t^1 + t^2$ where $\|t^2\|_1 \le Lb(T)$ and $T_1 = \{t^1; t \in T\}$ satisfies $\gamma_2(T_1) \le Lb(T)$. This procedure makes T naturally appear as a subset of a sum $T_1 + T_2$, even though T may be very different itself from such a sum. The intrinsic difficulty is that this decomposition is neither unique nor canonical. Another way to explain the difficulty is as follows. Consider a set T_1 with $\gamma_2(T_1) \le 1$, so that $b(T_1) \le L$. To each point t of T_1 let us associate a point $\varphi(t)$ with $\|\varphi(t)\|_1 \le 1$, and let $T = \{t + \varphi(t); t \in T_1\}$. Thus $b(T) \le L$. Now, we are only given the set T. How do we reconstruct the set T_1?

The next two sections introduce basic tools concerning Bernoulli processes. These tools will be much used later on, and also in the proof of Theorem 5.1.5. The rest of the chapter is then devoted to the completion of the proof of this result. This proof is not very long. However the whole approach involves several new and deep ideas, and a nearly magic way to fit them together. The understanding of all this will likely require a real effort from the reader. It should be most rewarding. If ever a proof in this book deserves to be called truly deep and beautiful, this is the one. Should, however, this proof turn out to be too difficult at first reading, the reader should not be discouraged. The remainder of the book depends little on Theorem 5.1.5 and on the material of the present chapter starting from Section 5.6.

5.2 Chaining for Bernoulli Processes

We cannot expect that an oracle will always reveal possible choices of the sets T_1 and T_2 that witness the inequality $b^*(T) \le Lb(T)$. We need a practical method to bound $b(T)$. The basic idea is to use chaining, and, along each chain, to use that, for any subset I of $\mathbb{N}^*$,

$$\mathsf{P}\Big(\Big|\sum_{i\ge 1} \varepsilon_i a_i\Big| \ge u + \sum_{i\in I} |a_i|\Big) \le \mathsf{P}\Big(\Big|\sum_{i\notin I} \varepsilon_i a_i\Big| \ge u\Big) ,$$

and then to use the subgaussian inequality to bound the last term. But how do we organize the argument efficiently, and if possible, optimally? This is the purpose of the next theorem. This theorem is the generalization of the generic chaining bound (2.45) to Bernoulli processes: it organizes the chaining in a optimal way, a statement which will be made precise in Proposition 5.2.5 below. This is not only a theoretical consideration: in Chapter 7 we shall demonstrate the sweeping effectiveness of this result when applied to the classical problem of convergence of random Fourier series.

The only negative point is that the theorem looks a bit complicated at first glance, because it involves several new (and fundamental) ideas. We need to state it before explaining it. We consider a number $r \geq 2$.

Theorem 5.2.1. *Consider a subset T of ℓ^2, and assume that $0 \in T$. Consider an admissible sequence of partitions $(\mathcal{A}_n)$ of T, and for $A \in \mathcal{A}_n$ consider a number $j_n(A) \in \mathbb{Z}$ with the following properties, where $u \geq 1$ is a parameter:*

$$A \in \mathcal{A}_n \ , \ B \in \mathcal{A}_{n-1} \ , \ A \subset B \Rightarrow j_n(A) \geq j_{n-1}(B) \ , \tag{5.9}$$

$$\forall\, x, y \in A \in \mathcal{A}_n \ , \ \sum_{i \geq 1} \big(r^{2j_n(A)} (x_i - y_i)^2 \big) \wedge 1 \leq u 2^n \ , \tag{5.10}$$

where $x \wedge y = \min(x, y)$. Then

$$b^*(T) \leq L \bigg(u \sup_{x \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(x))} + \sup_{x \in T} \sum_{i \geq 1} |x_i| \mathbf{1}_{\{2|x_i| \geq r^{-j_0(T)}\}} \bigg) \ . \tag{5.11}$$

Moreover if ε_i are independent Bernoulli r.v.s, for any $p \geq 1$ we have

$$\Big(\mathsf{E} \sup_{x \in T} \Big| \sum_{i \geq 1} x_i \varepsilon_i \Big|^p \Big)^{1/p} \leq K(p) u \sup_{x \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(x))} + L \sup_{x \in T} \sum_{i \geq 1} |x_i| \mathbf{1}_{\{2|x_i| \geq r^{-j_0(T)}\}} \ . \tag{5.12}$$

We first point out the following simple consequence.

Corollary 5.2.2. *Assume that moreover*

$$\forall\, x \in T \ , \ \|x\|_\infty < r^{-j_0(T)}/2 \ . \tag{5.13}$$

Then

$$b^*(T) \leq L u \sup_{x \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(x))} \ . \tag{5.14}$$

Proof. In (5.11) the second term in the right-hand side is identically zero. □

Let us now comment on Theorem 5.2.1. Condition (5.9) is a mild technical requirement. The central condition is (5.10). It could also be written

$$\forall\, x, y \in A \in \mathcal{A}_n \ , \ \sum_{i \geq 1} (x_i - y_i)^2 \wedge r^{-2j_n(A)} \leq u 2^n r^{-2j_n(A)} \ . \tag{5.15}$$

The point of writing (5.10) rather than (5.15) is simply that this is more in line with the generalizations of this statement that we shall study later.

Let us imagine that instead of condition (5.15) we had the stronger condition

$$\forall\, x, y \in A \in \mathcal{A}_n \ , \ \sum_{i \geq 1} (x_i - y_i)^2 \leq u 2^n r^{-2j_n(A)} \ , \tag{5.16}$$

and let us investigate what this would mean.

Proposition 5.2.3. *Consider a subset T of ℓ^2. Consider an admissible sequence of partitions $(\mathcal{A}_n)$ of T, and for $A \in \mathcal{A}_n$ consider a number $j_n(A) \in \mathbb{Z}$. Assume that (5.9) and (5.16) hold, where $u \geq 1$ is a parameter. Then*

$$\gamma_2(T) \leq \sqrt{u} \sup_{x \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(x))} . \tag{5.17}$$

Proof. It follows from (5.16) that

$$\Delta(A) \leq \sqrt{u} 2^{n/2} r^{-j_n(A)} , \tag{5.18}$$

and thus

$$\sup_{x \in T} \sum_{n \geq 0} 2^{n/2} \Delta(A_n(x)) \leq \sqrt{u} \sup_{x \in T} 2^n r^{-j_n(A_n(x))} . \qquad \square$$

This has the following converse.

Proposition 5.2.4. *Consider a subset T of ℓ^2. Then there exists an admissible sequence of partitions $(\mathcal{A}_n)$ of T, and for $A \in \mathcal{A}_n$ a number $j_n(A) \in \mathbb{Z}$ such that (5.9) and (5.16) hold for $u = 1$ and $\sup_{x \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(x))} \leq K(r)\gamma_2(T)$.*

Proof. Consider an admissible sequence of partitions $(\mathcal{A}_n)$ of T for which $\sup_x \sum_{n \geq 0} 2^{n/2} \Delta(A_n(x)) \leq 2\gamma_2(T, d)$. Define $j_n(A)$ as the largest integer $j \in \mathbb{Z}$ for which $\Delta(A) \leq 2^{n/2} r^{-j}$, so that $2^{n/2} r^{-j_n(A)} \leq r\Delta(A)$ and (5.9) is obviously satisfied. Moreover $\sup_{x \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(x))} \leq 2r\gamma_2(T)$. $\square$

The previous considerations show that we should think of the quantity $\sup_{x \in T} 2^n r^{-j_n(A_n(x))}$ as a substitute for the quantity $\sup_{x \in T} 2^{n/2} \Delta(A_n(x))$, so that the occurrence of this quantity in the right-hand side of (5.17) is not surprising. This method of controlling the size of a set A through the parameter $j_n(A)$ is motivated by one of the leading ideas of this work, that many processes require the use of "families of distances" to control them in an optimal manner, a topic that we will develop gradually. The point of Theorem 5.2.1 is that (5.10) is significantly weaker than condition (5.16), because it requires a much weaker control on the large values of $x_i - y_i$. It is of course difficult at this stage to really understand that this is a considerable gain. Some comments that might help may be found on page 280.

The following result shows that in some sense Corollary 5.2.2 is optimal.

Proposition 5.2.5. *Assume that $0 \in T \subset \ell^2$. Then we can find a sequence $(\mathcal{A}_n)$ of admissible partitions of T and for $A \in \mathcal{A}_n$ a number $j_n(A)$ such that conditions (5.13), (5.9) and (5.10) are satisfied for $u = 1$ and moreover*

$$\sup_{x \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(x))} \leq K(r) b^*(T) . \tag{5.19}$$

This will be proved in Section 16.3. Of course, the situation is the same as for the generic chaining bound for Gaussian processes. There is no magic wand to discover the proper choice of the partitions $\mathcal{A}_n$, and in specific situations this can be done only by understanding the underlying combinatorics.

Also, we should point out one of the (psychological) difficulties in discovering the proof of Theorem 5.1.5. Even though it turns out from Proposition 5.2.5 that one can find the partitions $\mathcal{A}_n$ such that (5.10) holds, when proving Theorem 5.1.5 it seems necessary to use partitions with a weaker property, which replaces the summation over all values of i in (5.10) by the summation over the values in an appropriate subset $\Omega_n(t)$ of $\mathbb{N}^*$, see (5.25) below.

In order to avoid repetition, we shall deduce Theorem 5.2.1 from a more general principle, which will cover all our future needs (and in particular is one of the keys to the proof of Theorem 5.1.5). This is going to look complicated at first sight, but we shall show almost immediately how to use it. We consider a measure space (Ω, μ) where μ is a σ-finite measure. It suffice for this chapter to consider the case where $\Omega = \mathbb{N}^*$ provided with the counting measure, but the proofs are not any simpler in this special situation.

Theorem 5.2.6. *Consider a countable set T of measurable functions on Ω, a number $r \geq 2$, and assume that $0 \in T$. Consider an admissible sequence of partitions $(\mathcal{A}_n)$ of T. For $t \in T$ and $n \geq 0$ consider an element $j_n(t) \in \mathbb{Z}$ and $\pi_n(t) \in T$. Assume that $\pi_0(t) = 0$ for each t and the following properties. First, the values of $j_n(t)$ and $\pi_n(t)$ depend only on $A_n(t)$:*

$$\forall s, t \in T \;,\; \forall n \geq 0 \;;\; s \in A_n(t) \Rightarrow j_n(s) = j_n(t) \;;\; \pi_n(s) = \pi_n(t) \;. \tag{5.20}$$

The sequence $(j_n(t))_{n \geq 1}$ is non-decreasing:

$$\forall t \in T \;,\; \forall n \geq 0 \;,\; j_{n+1}(t) \geq j_n(t) \;. \tag{5.21}$$

When going from n to $n+1$ the value of $\pi_n(t)$ can change only when the value of $j_n(t)$ increases:

$$\forall t \in T \;,\; \forall n \geq 0 \;,\; j_n(t) = j_{n+1}(t) \Rightarrow \pi_n(t) = \pi_{n+1}(t) \;. \tag{5.22}$$

When going from n to $n+1$, if the value of $j_n(t)$ increases, then $\pi_{n+1}(t) \in A_n(t)$:

$$\forall t \in T \;,\; \forall n \geq 0 \;,\; j_{n+1}(t) > j_n(t) \Rightarrow \pi_{n+1}(t) \in A_n(t) \;. \tag{5.23}$$

For $t \in T$ and $n \geq 0$ we define $\Omega_n(t) \subset \Omega$ as $\Omega_0(t) = \Omega$ if $n = 0$ and

$$\Omega_n(t) = \left\{ \omega \in \Omega \;;\; 0 \leq q < n \Rightarrow |\pi_{q+1}(t)(\omega) - \pi_q(t)(\omega)| \leq r^{-j_q(t)} \right\} \;. \tag{5.24}$$

Let us consider a parameter $u > 0$ and assume that

$$\forall t \in T \;,\; \forall n \geq 0 \;,\; \int_{\Omega_n(t)} \left(r^{2j_n(t)} (t(\omega) - \pi_n(t)(\omega))^2 \wedge 1 \right) \mathrm{d}\mu(\omega) \leq u 2^n \;. \tag{5.25}$$

Then we can write $T \subset T_1 + T_2 + T_3$ *where* $0 \in T_1$ *and*

$$\gamma_2(T_1, d_2) \le L\sqrt{u} \sup_{t \in T} \sum_{n \ge 0} 2^n r^{-j_n(t)} \tag{5.26}$$

$$\gamma_1(T_1, d_\infty) \le L \sup_{t \in T} \sum_{n \ge 0} 2^n r^{-j_n(t)} \tag{5.27}$$

$$\forall t \in T_2 \;,\; \|t\|_1 \le Lu \sup_{t \in T} \sum_{n \ge 0} 2^n r^{-j_n(t)} \;. \tag{5.28}$$

Moreover,

$$\forall t \in T_3 \,,\, \exists s \in T \,,\, |t| \le 5|s| \mathbf{1}_{\{2|s| \ge r^{-j_0(t)}\}} \;. \tag{5.29}$$

The term T_3 of the decomposition is of secondary importance, and will be easy to control. It is required because (5.25) says little about the functions $|s|\mathbf{1}_{\{|s| \ge r^{-j_0(T)}\}}$. The important statements are (5.26) to (5.28). The important case is where $u \ge 1$ although it changes nothing to the proof to assume only $u > 0$. (Incidentally Theorem 5.2.1 also holds for $u > 0$ at the expense of replacing u by $\sqrt{u}$ in (5.11) and (5.12).)

Let us first give a slightly simpler (and weaker) statement.

Theorem 5.2.7. *Consider a countable set* T *of measurable functions on* Ω, *a number* $r \ge 2$, *and assume that* $0 \in T$. *Consider an admissible sequence of partitions* $(\mathcal{A}_n)$ *of* T, *and for* $A \in \mathcal{A}_n$ *consider* $j_n(A) \in \mathbb{Z}$, *with the following properties, where* $u > 0$ *is a parameter*

$$A \in \mathcal{A}_n \,,\, B \in \mathcal{A}_{n-1} \,,\, A \subset B \Rightarrow j_n(A) \ge j_{n-1}(B) \;. \tag{5.30}$$

$$\forall \, s, t \in A \in \mathcal{A}_n \,,\, \int \left(r^{2j_n(A)}(s(\omega) - t(\omega))^2\right) \wedge 1 \mathrm{d}\mu(\omega) \le u2^n \;. \tag{5.31}$$

Then we can write $T \subset T_1 + T_2 + T_3$ *as in Theorem 5.2.6, where* $j_n(t) = j_n(A_n(t))$.

Condition (5.30) is a mild technical requirement. The central condition is (5.31). If it were replaced by the stronger condition

$$\forall \, s, t \in A \,,\, \int r^{2j_n(t)}(s(\omega) - t(\omega))^2 \mathrm{d}\mu(\omega) \le u2^n \;, \tag{5.32}$$

this would simply mean $\Delta(A, d_2) \le \sqrt{u} 2^{n/2} r^{-j_n(t)}$. Again, the point is that (5.31) requires a much weaker control of the large values of $s - t$ than (5.32).

Proof. We deduce this result from Theorem 5.2.6. We set $j_n(t) = j_n(A_n(t))$ and we define

$$p(n, t) = \inf\{p \ge 0 \;;\; j_n(t) = j_p(t)\},$$

so that $p(n,t) \le n$ and thus $A_{p(n,t)} \supset A_n(t)$. We define $t_T = 0$. For $A \in \mathcal{A}_n$, $n \ge 1$, we choose an arbitrary point t_A in A. We define

$$\pi_n(t) = t_B \quad \text{where } B = A_{p(n,t)}(t) \ ,$$

and we note that $\pi_0(t) = 0$. When $s \in A_n(t)$ we have $A_p(s) = A_p(t)$ for $p \le n$ and thus $p(n,s) = p(n,t)$ so that $\pi_n(s) = \pi_n(t)$. Also, if $j_{n+1}(t) = j_n(t)$ we have $p(n,t) = p(n+1,t)$, so that $\pi_n(t) = \pi_{n+1}(t)$. This proves that (5.20) to (5.22) hold. Moreover, when $j_n(t) > j_{n-1}(t)$ we have $p(n,t) = n$ so that $\pi_n(t) = t_A$ for $A = A_n(t)$, and thus $\pi_n(t) \in A_n(t) \subset A_{n-1}(t)$, and this proves (5.23). Finally, (5.31) used for $p = p(n,t)$ and $B = A_{p(n,t)}(t)$ reads

$$\forall s, s' \in B \ , \ \int r^{2j_p(B)} |s - s'|^2 \wedge 1 \mathrm{d}\mu \le u 2^p$$

and this is stronger than (5.25) since $j_p(B) = j_n(A) = j_n(t)$ and $\pi_n(t) = t_B \in B$. The proof is complete. □

Corollary 5.2.8. *Under the conditions of Theorem 5.2.1 we can write* $T \subset T_1 + T_2 + T_3$ *where*

$$\gamma_2(T_1, d_2) \le L\sqrt{u} \sup_{x \in T} \sum_{n \ge 0} 2^n r^{-j_n(A_n(x))} \ , \tag{5.33}$$

$$\gamma_1(T_1, d_\infty) \le L \sup_{x \in T} \sum_{n \ge 0} 2^n r^{-j_n(A_n(x))} \ , \tag{5.34}$$

$$\forall\, x \in T_2 \ , \ \|x\|_1 \le Lu \sup_{x \in T} \sum_{n \ge 0} 2^n r^{-j_n(A_n(x))} \ , \tag{5.35}$$

and

$$\forall\, x \in T_3 \ , \ \exists y \in T \ , \ \forall\, i \ge 1 \ , \ |x_i| \le 5|y_i| \mathbf{1}_{\{2|y_i| \ge r^{-j_0(T)}\}} \ . \tag{5.36}$$

Proof. This follows from Theorem 5.2.7 in the case $\Omega = \mathbb{N}^*$ and where μ is the counting measure. □

Proof of Theorem 5.2.1. It relies upon Corollary 5.2.8. Let us define $S = \sup_{x \in T} \sum_{n \ge 0} 2^n r^{-j_n(A_n(x))}$ and $S^* = \sup_{x \in T} \sum_{i \ge 1} |x_i| \mathbf{1}_{\{2|x_i| \ge r^{-j_0(T)}\}}$. Then (5.33) implies that $\gamma_2(T_1, d_2) \le L\sqrt{u} S$. Moreover $\|x\|_1 \le LuS$ for $x \in T_2$ by (5.35) while for $x \in T_3$ we have $\|x\|_1 \le LS^*$ by (5.36). Consequently, $\|x\|_1 \le L(uS + S^*)$ for $x \in T_2 + T_3$. This proves (5.11).

Let us then decompose $0 \in T$ as $0 = x_1 + y_1$ where $x_1 \in T_1$ and $y_1 = -x_1 \in T_2 + T_3$. Let $T_1' = T_1 - x_1, T_2' = T_2 + T_3 - y_1$, so that $\gamma_2(T_1', d_2) = \gamma_2(T_1, d_2) \le L\sqrt{u} S$, $0 \in T_1'$ and

$$x \in T_2' \Rightarrow \|x\|_1 \le L(uS + S^*) \ . \tag{5.37}$$

Moreover $T \subset T_1' + T_2'$.

To prove (5.12) it suffices to prove this inequality when in the left-hand side we replace T by either T_1' or T_2'. In the case of T_1' this follows from (5.33), since $0 \in T_1'$. In the case of T_2' this follows from (5.37) since

$$\Big(\mathsf{E}\sup_{x\in T_2'}\Big|\sum_{i\geq 1} x_i\varepsilon_i\Big|^p\Big)^{1/p} \leq \sup_{x\in T_2'}\sum_{i\geq 1}|x_i| = \sup_{x\in T_2'}\|x\|_1\ . \qquad \square$$

Although the following is not directly used in the proof of Theorem 5.2.6, it is connected in the simplest possible case to some of the underlying ideas of the proofs, and it might help the reader to spell out this simple interpolation principle. It decomposes a function of L^2 in its '"peaky" part and its "spread out" part.

Lemma 5.2.9. *Consider $f \in L^2$ and $u > 0$. Then we can write $f = f_1 + f_2$ where*

$$\|f_1\|_2 \leq \|f\|_2\ ,\ \|f_1\|_\infty \leq u\ \ ;\ \ \|f_2\|_2 \leq \|f\|_2\ ,\ \|f_2\|_1 \leq \frac{\|f\|_2^2}{u}\ . \tag{5.38}$$

Proof. We set $f_1 = f\mathbf{1}_{\{|f|\leq u\}}$, so that the first part of (5.38) is obvious. We set $f_2 = f\mathbf{1}_{\{|f|>u\}} = f - f_1$, so that

$$u\|f_2\|_1 = \int u|f|\mathbf{1}_{\{|f|>u\}}\mathrm{d}\mu \leq \int f^2\mathrm{d}\mu = \|f\|_2^2\ . \qquad \square$$

Proof of Theorem 5.2.6. The principle of the proof is, given $t \in T$, to write $t(\omega) = t^1(\omega) + t^2(\omega) + t^3(\omega)$ where one defines the values $t^1(\omega), t^2(\omega), t^3(\omega)$ from the values $\pi_n(t)(\omega)$, $n \geq 1$. A natural way to implement this strategy is to write the chaining identity

$$t = \sum_{n\geq 1}(\pi_n(t) - \pi_{n-1}(t))\ ,$$

and to use Lemma 5.2.9 for each of the increments $\pi_n(t) - \pi_{n-1}(t)$, with a suitable value of $u = u(t,n)$. The reader can find the proof of a special case of Theorem 5.2.7 along this line in the second proof of Theorem 9.1.9 on page 276. It might be helpful to look at this proof before or in parallel with the study of the present arguments.

At some point the author got the fancy idea that the approach outlined above was not clean, and that one should define $t^1(\omega)$ as $\pi_{n(\omega)}(t)(\omega)$ for a cleverly chosen value of $n(\omega)$. In retrospect, and despite considerable efforts, this does not make the proof any more intuitive. Maybe it is unavoidable that the proof is not very simple. There seems to be a genuine difficulty here: Theorem 5.2.1 is an immediate consequence of Theorem 5.2.7, and we have already mentioned that it has sweeping consequences. The reader might like to postpone reading the details of the present proof until she has found enough motivation through the subsequent applications of this principle.

Certainly we may assume that

$$\sup_{t\in T}\sum_{n\geq 0} 2^n r^{-j_n(t)} < \infty\ , \tag{5.39}$$

and in particular that

$$\forall t \in T \; , \; \lim_{n\to\infty} j_n(t) = \infty \; . \tag{5.40}$$

Let us also observe that using (5.20) and (5.25) we have

$$\forall t \in T,\, \forall n \geq 0,\, \forall s \in A_n(t),\, \int_{\Omega_n(t)} r^{2j_n(t)}(s(\omega) - \pi_n(t)(\omega))^2 \wedge 1 \mathrm{d}\mu(\omega) \leq u2^n. \tag{5.41}$$

For $t \in T$ and $\omega \in \Omega$, we define

$$m(t,\omega) = \inf\bigl\{n \geq 0 \; ; \; |\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| > r^{-j_n(t)}\bigr\}$$

if the set on the right is not empty and $m(t,\omega) = \infty$ otherwise. In words, this is the first place at which $\pi_n(\omega)$ and $\pi_{n+1}(\omega)$ differ significantly. Thus

$$n < m(t,\omega) \Rightarrow |\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| \leq r^{-j_n(t)} \; , \tag{5.42}$$

and we note from the definition (5.24) of $\Omega_n(t)$ that

$$\Omega_n(t) = \{m(t,\cdot) \geq n\} \; . \tag{5.43}$$

From (5.22), when $j_{n+1}(t) = j_n(t)$ we have $\pi_{n+1}(t) = \pi_n(t)$. Thus, using also (5.23) in the last implication,

$$\pi_{n+1}(t) \neq \pi_n(t) \Rightarrow j_{n+1}(t) \geq j_n(t) + 1 \Rightarrow \pi_{n+1}(t) \in A_n(t) \; . \tag{5.44}$$

Consequently for $m < m(t,\omega)$ we have

$$|\pi_{m+1}(t)(\omega) - \pi_m(t)(\omega)| \leq r^{-j_m(t)} \mathbf{1}_{\{j_{m+1}(t) > j_m(t)\}} \; .$$

Since $r \geq 2$, we deduce from (5.42) that if $n < m(t,\omega)$ then

$$\sum_{n \leq m < m(t,\omega)} |\pi_{m+1}(t)(\omega) - \pi_m(t)(\omega)| \leq \sum_{j \geq j_n(t)} r^{-j} \leq 2r^{-j_n(t)} \; . \tag{5.45}$$

Let us define t^1 by $t^1(\omega) = \pi_{m(t,\omega)}(t)(\omega)$ if $m(t,\omega) < \infty$ and $t^1(\omega) = \lim_{n\to\infty} \pi_n(t)(\omega)$ if $m(t,\omega) = \infty$. The limit exists from (5.45) and (5.40), and since $\pi_0(t) = 0$, using (5.45) with $n = 0$ we have

$$|t^1(\omega)| \leq 2r^{-j_0(T)} \; . \tag{5.46}$$

We define $T_1 = \{t^1 ; t \in T\}$. For $n \geq 0$, we define t^1_n by

$$t^1_n(\omega) = \pi_{n \wedge m(t,\omega)}(t)(\omega) \; ,$$

so that

$$\forall \omega \; , \; t^1(\omega) = \lim_{n\to\infty} t^1_n(\omega) \; . \tag{5.47}$$

We aim now to show that if $U_n = \{t_n^1 \; ; \; t \in T\}$, then $\operatorname{card} U_n \le N_n$. When $s \in A_n(t)$ we have $\pi_n(s) = \pi_n(t)$ by (5.20). Since then $A_q(s) = A_q(t)$ for $q \le n$, we also have $\pi_q(s) = \pi_q(t)$ for such values of q. The definition of $m(t,\omega)$ shows that for any n', the points $\pi_q(t)$ for $0 \le q \le n'$ entirely determine whether or not it is true that $m(t,\omega) < n'$. Consequently, when $s \in A_n(t)$ we have $n \wedge m(t,\omega) = n \wedge m(s,\omega)$ for each ω, so that $t_n^1 = s_n^1$. This proves that $\operatorname{card} U_n \le \operatorname{card} \mathcal{A}_n \le N_n$.

The next goal is to prove that the sets U_n are in a sense approximations of the set T_1, both for the ℓ^2 and the ℓ^∞ norm. We note that $t^1(\omega) - t_n^1(\omega) = 0$ if $n \ge m(t,\omega)$, and by (5.45) that if $n < m(t,\omega)$, then

$$|t^1(\omega) - t_n^1(\omega)| \le \sum_{n \le m < m(t,\omega)} |\pi_{m+1}(t)(\omega) - \pi_m(t)(\omega)| \le 2r^{-j_n(t)} \,.$$

Thus $\|t^1 - t_n^1\|_\infty \le 2r^{-j_n(t)}$, and hence $d_\infty(t^1, U_n) \le 2r^{-j_n(t)}$. Thus (5.27) follows from Theorem 2.3.1 with $\alpha = 1$.

We turn to the proof of (5.26). We observe that

$$t_{n+1}^1 - t_n^1 = (\pi_{n+1}(t) - \pi_n(t)) \mathbf{1}_{\{m(t,\cdot) > n\}} \,.$$

Indeed, if $m(t,\omega) \le n$ then $t_n^1(\omega) = t_{n+1}^1(\omega) = \pi_{m(t,\omega)}(\omega)$ while if $m(t,\omega) > n$ then $t_n^1(\omega) = \pi_n(t)(\omega)$ and $t_{n+1}^1(\omega) = \pi_{n+1}(t)(\omega)$. By definition of $m(t,\omega)$ we have $|\pi_{n+1}(t) - \pi_n(t)| \le r^{-j_n(t)}$ whenever $m(t,\cdot) > n$ and also $\Omega_n(t) = \{m(t,\cdot) > n\}$ by (5.43). Therefore,

$$|t_{n+1}^1 - t_n^1| \le |\pi_{n+1}(t) - \pi_n(t)| \mathbf{1}_{\{|\pi_{n+1}(t) - \pi_n(t)| \le r^{-j_n(t)}\} \cap \Omega_n(t)} \,.$$

Now if the right-hand side above is not 0 then $\pi_{n+1}(t) \in A_n(t)$ by (5.44) so that using (5.41) in the fourth line,

$$\begin{aligned}
&\|t_{n+1}^1 - t_n^1\|_2^2 \\
&\le \int_{\Omega_n(t)} |\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)|^2 \mathbf{1}_{\{|\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| \le r^{-j_n(t)}\}} \mathrm{d}\mu(\omega) \\
&\le \int_{\Omega_n(t)} \big(\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)\big)^2 \wedge r^{-2j_n(t)} \mathrm{d}\mu(\omega) \\
&\le u 2^n r^{-2j_n(t)} \,.
\end{aligned}$$

Thus $\|t_{n+1}^1 - t_n^1\|_2 \le \sqrt{u} 2^{n/2} r^{-j_n(t)}$ and (5.39) implies that the sequence (t_n^1) is a Cauchy sequence in ℓ^2, so that it converges to its limit, which is t^1 from (5.47), and hence $\lim_{n\to\infty} \|t^1 - t_n^1\|_2 = 0$. Consequently

$$\begin{aligned}
d_2(t^1, U_n) &\le \|t^1 - t_n^1\|_2 \le \lim_{q\to\infty} \|t_q^1 - t_n^1\|_2 \\
&\le \sum_{m \ge n} \|t_{m+1}^1 - t_m^1\|_2 \le \sqrt{u} \sum_{m \ge n} 2^{m/2} r^{-j_m(t)} \,. \qquad (5.48)
\end{aligned}$$

Since

$$\sum_{n\geq 0} 2^{n/2} \sum_{m\geq n} 2^{m/2} r^{-j_m(t)} \leq \sum_{m\geq 0} 2^{m/2} r^{-j_m(t)} \sum_{n\leq m} 2^{n/2} \leq L \sum_{m\geq 0} 2^m r^{-j_m(t)} ,$$

we conclude by Theorem 2.3.1 again that (5.26) holds.

For $t \in T$, define $\Xi(t) = \{\omega\ ;\ |t(\omega)| \leq r^{-j_0(T)}/2\}$ and $t^3 = (t - t^1)\mathbf{1}_{\Xi(t)^c}$. Since for $\omega \in \Xi(t)^c$ we have $|t(\omega)| \geq r^{-j_0(T)}/2$ and since $|t^1(\omega)| \leq 2r^{-j_0(T)}$ by (5.46), we have $|t^3| \leq 5|t|\mathbf{1}_{\Xi(t)^c}$, so that the set $T_3 = \{t^3\ ;\ t \in T\}$ satisfies (5.29).

We set $t^2 := t - t^1 - t^3 = (t - t^1)\mathbf{1}_{\Xi(t)}$, $T_2 = \{t^2; t \in T\}$, and we turn to the proof of (5.28). We define

$$r(t, \omega) = \inf\Big\{n \geq 0\ ;\ |\pi_{n+1}(t)(\omega) - t(\omega)| \geq \frac{1}{2} r^{-j_{n+1}(t)}\Big\}$$

if the set on the right is not empty and $r(t, \omega) = \infty$ otherwise. Thus,

$$n < r(t, \omega) \Rightarrow |\pi_{n+1}(t)(\omega) - t(\omega)| < \frac{1}{2} r^{-j_{n+1}(t)} . \tag{5.49}$$

Consequently, for $0 \leq n < r(t, \omega)$,

$$\begin{aligned} |\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| &\leq |\pi_{n+1}(t)(\omega) - t(\omega)| + |\pi_n(t)(\omega) - t(\omega)| \\ &\leq \frac{1}{2} r^{-j_{n+1}(t)} + |\pi_n(t)(\omega) - t(\omega)| . \end{aligned} \tag{5.50}$$

When $n > 0$ we use (5.49) for $n - 1$ to obtain

$$|\pi_n(t)(\omega) - t(\omega)| \leq r^{-j_n(t)}/2 . \tag{5.51}$$

Let us fix $\omega \in \Xi(t)$. Then this inequality still holds true for $n = 0$ since

$$|\pi_0(t)(\omega) - t(\omega)| = |t(\omega)| \leq r^{-j_0(T)}/2 = r^{-j_0(t)}/2 .$$

Thus for $0 \leq n < r(t, \omega)$, we have

$$|\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)| \leq \frac{1}{2}(r^{-j_{n+1}(t)} + r^{-j_n(t)}) \leq r^{-j_n(t)} ,$$

and consequently $r(t, \omega) \leq m(t, \omega)$. When $r(t, \omega) = \infty$ then $m(t, \omega) = \infty$ so that, recalling (5.40), $t(\omega) = \lim_{n\to\infty} \pi_n(t)(\omega) = t^1(\omega)$. Therefore we have proved that

$$t^2(\omega) = t(\omega) - t^1(\omega) = \sum_{n\geq 0} (t(\omega) - t^1(\omega))\mathbf{1}_{\{r(t,\omega)=n\}} . \tag{5.52}$$

Now, when $n = r(t, \omega)$, we have $m(t, \omega) \geq n$ and, using (5.45), $|\pi_n(t)(\omega) - t^1(\omega)| \leq 2r^{-j_n(t)}$. Consequently, using also (5.51),

$$|t(\omega) - t^1(\omega)| \leq |t(\omega) - \pi_n(t)(\omega)| + |\pi_n(t)(\omega) - t^1(\omega)| \leq 3r^{-j_n(t)} \ . \quad (5.53)$$

Let us define $t_n^2 = (t - t^1)\mathbf{1}_{\{r(t,\cdot)=n\}\cap \Xi(t)}$, so that $t^2 = \sum_{n\geq 0} t_n^2$ by (5.52) and by (5.53) it holds $\|t_n^2\|_1 \leq Lr^{-j_n(t)}\mu(\{\omega; r(t,\omega) = n\} \cap \Xi(t))$.

To finish the proof we show that $\mu(\{\omega; r(t,\omega) = n\} \cap \Xi(t)) \leq Lu2^n$. Since for $\omega \in \Xi(t)$ we have $r(t,\omega) \leq m(t,\omega)$, using (5.43) we get $\{\omega; r(t,\omega) = n\} \cap \Xi(t) \subset \Omega_n(t)$ and therefore

$$\begin{aligned}\mu(\{\omega\ ;\ r(t,\omega) = n\} \cap \Xi(t)) \leq \mu(\{\omega\ ;\ r(t,\omega) = n\} \cap \Omega_{n+1}(t))\\ + \mu(\Omega_n(t) \setminus \Omega_{n+1}(t)) \ . \quad (5.54)\end{aligned}$$

Now, since $|\pi_{n+1}(t)(\omega) - t(\omega)| \geq r^{-j_{n+1}(t)}/2$ when $r(t,\omega) = n$, we have

$$\begin{aligned}&\tfrac{1}{4}\mu\left(\{\omega\ ;\ r(t,\omega) = n\} \cap \Omega_{n+1}(t)\right)\\ &\leq \int_{\Omega_{n+1}(t)} r^{2j_{n+1}(t)}(\pi_{n+1}(t)(\omega) - t(\omega))^2 \wedge 1 \mathrm{d}\mu(\omega)\\ &\leq u2^{n+1} ,\end{aligned}$$

using (5.41) for $n+1$ rather than n in the last inequality. Since $|\pi_n(t)(\omega) - \pi_{n+1}(t)(\omega)| \geq r^{-j_n(t)}$ for $\omega \in \Omega_n(t) \setminus \Omega_{n+1}(t)$ and since $\pi_{n+1}(t) \in A_n(t)$ by (5.44) we have again from (5.41)

$$\mu(\Omega_n(t) \setminus \Omega_{n+1}(t)) \leq \int_{\Omega_n(t)} r^{2j_n(t)}|\pi_{n+1}(t)(\omega) - \pi_n(t)(\omega)|^2 \wedge 1 \mathrm{d}\mu(\omega) \leq u2^n ,$$

and the proof is complete. □

5.3 Fundamental Tools for Bernoulli Processes

We start by a simple fact.

Lemma 5.3.1. *For a subset T of ℓ^2 we have*

$$\Delta(T, d_2) \leq Lb(T) \ . \quad (5.55)$$

Proof. Assuming without loss of generality that $0 \in T$, we have

$$\forall t \in T\ ,\ b(T) \geq \mathsf{E}\max\Big(0, \sum_{i\geq 1} \varepsilon_i t_i\Big) = \frac{1}{2}\mathsf{E}\Big|\sum_{i\geq 1} \varepsilon_i t_i\Big| \geq \frac{1}{L}\|t\|_2 \ ,$$

using symmetry in the equality and Khinchin's inequality (3.12) in the last inequality. This proves (5.55). □

The first fundamental fact about Bernoulli processes is a "concentration of measure" result, which should be compared with Lemma 2.4.7.

Theorem 5.3.2. *Consider a subset $T \subset \ell^2$ and assume that for a certain $t_0 \in \ell_2$ we have $T \subset B(t_0, \sigma)$. Consider numbers $(a(t))_{t \in T}$ and let M be a median of the r.v. $\sup_{t \in T}(\sum_i \varepsilon_i t_i + a(t))$. Then*

$$\forall u > 0 \ , \ \mathsf{P}\Big(\Big|\sup_{t\in T}\big(\sum_{i\geq 1} \varepsilon_i t_i + a(t)\big) - M\Big| \geq u\Big) \leq 4\exp\Big(-\frac{u^2}{4\sigma^2}\Big) \ . \tag{5.56}$$

In particular,

$$\Big|\mathsf{E}\sup_{t\in T}\big(\sum_{i\geq 1} \varepsilon_i t_i + a(t)\big) - M\Big| \leq L\sigma \ , \tag{5.57}$$

and also

$$\forall u > 0 \ , \ \mathsf{P}\Big(\Big|\sup_{t\in T}\sum_{i\geq 1} \varepsilon_i t_i - b(T)\Big| \geq u\Big) \leq L\exp\Big(-\frac{u^2}{L\sigma^2}\Big) \ . \tag{5.58}$$

This theorem has a short and almost magic proof, which can be found e.g in [8], in [4] or in [3]. We do not reproduce this proof for a good reason: the reader must face the fact that if she intends to become really competent about the area of probability theory with which we are dealing here, she must learn more about concentration of measure, and that this is better done by looking at the previous references rather than just at the proof of Theorem 5.3.2.

When all the coefficients t_i are small (say, compared to $\|t\|_2$), the r.v. $\sum_{i\geq 1} t_i \varepsilon_i$ resembles a Gaussian r.v., by the central limit theorem. Therefore one expects that when, in some sense, the set T is small for the ℓ^∞ norm, $g(T)$ (or, equivalently, $\gamma_2(T)$) is not too much larger than $b(T)$. This is the main idea of the next section, as well as the main idea of the following result, which is our second fundamental tool about Bernoulli processes.

Theorem 5.3.3. *Consider $t_1, \ldots, t_m$ in ℓ^2, and assume that*

$$\ell \neq \ell' \Rightarrow \|t_\ell - t_{\ell'}\|_2 \geq a \ . \tag{5.59}$$

Assume moreover that

$$\forall \ell \leq m \ , \ \|t_\ell\|_\infty \leq b \ . \tag{5.60}$$

Then

$$\mathsf{E}\sup_{\ell \leq m}\sum_{i\geq 1} \varepsilon_i t_{\ell,i} \geq \frac{1}{L}\min\Big(a\sqrt{\log m}, \frac{a^2}{b}\Big) \ . \tag{5.61}$$

This should be compared with Lemma 2.4.2, which in the present language asserts that

$$\mathsf{E}\sup_{\ell \leq m}\sum_{i\geq 1} g_i t_{\ell,i} \geq \frac{a}{L_1}\sqrt{\log m} \ . \tag{5.62}$$

This inequality will be the basis of the proof.

Exercise 5.3.4. Convince yourself that in (5.61) the term a^2/b is of the correct order. (Hint: remember that $\sum_i \varepsilon_i t_{\ell,i} \le \sum_i |t_{\ell,i}|$. Look for examples where $t_{\ell,i} \in \{0, b\}$.)

The main step of the proof of Theorem 5.3.3 is as follows.

Proposition 5.3.5. *The conclusion of Theorem 5.3.3 holds true if we assume moreover that $\|t_\ell\|_2 \le 2a$ for each $\ell \le m$.*

Proof. We recall the constant L_1 of (5.62). By Lemma 5.3.1 we observe that $\mathsf{E}\sup_{\ell\le m}\sum_{i\ge1} t_{\ell,i}\varepsilon_i \ge a/L$, so that it suffices to prove (5.61) when $a/b \ge L_1\sqrt{\log 2}$, for otherwise (5.61) holds automatically provided the constant L is large enough. Consider the largest integer $N \le m$ for which

$$L_1\sqrt{\log N} \le \frac{a}{b}\,. \tag{5.63}$$

Then $N \ge 2$, and, distinguishing whether $N = m$ or not, we obtain

$$a\sqrt{\log N} \ge \frac{1}{L}\min\Big(a\sqrt{\log m}, \frac{a^2}{b}\Big)\,. \tag{5.64}$$

The plan is to prove that

$$\mathsf{E}\sup_{\ell\le N}\sum_{i\ge1}\varepsilon_i t_{\ell,i} \ge \frac{1}{L}a\sqrt{\log N}\,, \tag{5.65}$$

so that the result follows by combining with (5.64). The argument is related to that of Theorem 3.2.12. Let us consider a parameter $c > 0$ and define $\xi_i = g_i \mathbf{1}_{\{|g_i|>c\}}$ and $\xi_i' = g_i\mathbf{1}_{\{|g_i|\le c\}}$. Thus, using (5.62),

$$\frac{a}{L_1}\sqrt{\log N} \le \mathsf{E}\sup_{\ell\le N}\sum_{i\ge1} g_i t_{\ell,i} \le \mathsf{E}\sup_{\ell\le N}\sum_{i\ge1}\xi_i' t_{\ell,i} + \mathsf{E}\sup_{\ell\le N}\sum_{i\ge1}\xi_i t_{\ell,i}\,. \tag{5.66}$$

Copying the argument of (3.25) shows that

$$\mathsf{E}\sup_{\ell\le N}\sum_{i\ge1}\xi_i' t_{\ell,i} \le c\mathsf{E}\sup_{\ell\le N}\sum_{i\ge1}\varepsilon_i t_{\ell,i}\,.$$

Therefore (5.66) shows that to prove (5.65) it suffices to prove that if c is large enough we have

$$\mathsf{E}\sup_{\ell\le N}\sum_{i\ge1}\xi_i t_{\ell,i} \le \frac{a}{2L_1}\sqrt{\log N}\,. \tag{5.67}$$

Consider a parameter $\lambda > 0$. Since the r.v. ξ_i is symmetric, we have

$$\varphi_c(\lambda) := \frac{1}{\lambda^2}(\mathsf{E}\exp\lambda\xi_i - 1) = \frac{1}{2\lambda^2}\mathsf{E}(\exp\lambda\xi_i + \exp(-\lambda\xi_i) - 2)\,,$$

and

$$\mathsf{E}\exp\lambda\xi_i = 1 + \lambda^2\varphi_c(\lambda) \le \exp(\lambda^2\varphi_c(\lambda)) \ . \tag{5.68}$$

Since the function $x \mapsto x^{-2}(\exp x + \exp(-x) - 2)$ increases on $\mathbb{R}^+$, this is also the case of φ_c. Also it is obvious from dominated convergence that $\lim_{c\to\infty}\varphi_c(x) = 0$ for each x. We fix c large enough that

$$\varphi_c(16) \le V := 2^{-9}L_1^{-2} \tag{5.69}$$

and we proceed to prove (5.67) for this value of c. For each $t \in \ell^2$ we have, using that φ_c increases on $\mathbb{R}^+$,

$$\mathsf{E}\exp\lambda\xi_i t_i \le \exp(\lambda^2 t_i^2\varphi_c(|t_i|\lambda)) \le \exp(\lambda^2 t_i^2\varphi_c(\|t\|_\infty\lambda)) \ ,$$

and thus

$$\mathsf{E}\exp\lambda\sum_{i\ge1}\xi_i t_i = \prod_{i\ge1}\mathsf{E}\exp\lambda\xi_i t_i \le \exp(\lambda^2\|t\|_2^2\varphi_c(\|t\|_\infty\lambda)) \ .$$

In particular, since $\|t_\ell\|_2^2 \le 4a^2$ and $\|t_\ell\|_\infty \le b$, whenever $\lambda \le 16/b$ we get

$$\mathsf{E}\exp\Big(\lambda\sum_{i\ge1}\xi_i t_{\ell,i}\Big) \le \exp(4a^2\lambda^2 V) \ .$$

Using the inequality $\mathsf{P}(Z \ge x) \le \exp(-\lambda x)\mathsf{E}\exp(\lambda Z)$ for $\lambda, x \ge 0$, we then obtain

$$\mathsf{P}\Big(\sum_{i\ge1}\xi_i t_{\ell,i} \ge x\Big) \le \exp(-\lambda x + 4a^2\lambda^2 V) \ ,$$

and then

$$\mathsf{P}\Big(\sup_{\ell\le N}\sum_{i\ge1}\xi_i t_{\ell,i} \ge x\Big) \le N\exp(-\lambda x + 4a^2\lambda^2 V) \ .$$

Consequently, for any $y > 0$ we have

$$\begin{aligned}\mathsf{E}\sup_{\ell\le N}\sum_{i\ge1}\xi_i t_{\ell,i} \le \mathsf{E}\max\Big(\sup_{\ell\le N}\sum_{i\ge1}\xi_i t_{\ell,i}, 0\Big) &= \int_0^\infty \mathsf{P}\Big(\sup_{\ell\le N}\sum_{i\ge1}\xi_i t_{\ell,i} \ge x\Big)\mathrm{d}x \\ &\le y + \int_y^\infty N\exp(-\lambda x + 4a^2\lambda^2 V)\mathrm{d}x \\ &= y + \frac{N}{\lambda}\exp(-\lambda y + 4a^2\lambda^2 V) \ . \end{aligned} \tag{5.70}$$

Let us now make the choice

$$y = \frac{a\sqrt{\log N}}{4L_1} \ ; \ \lambda = \frac{y}{8a^2V} \ .$$

This choice is legitimate because, since $V = 2^{-9}/L_1^2$, we have

$$\lambda = \frac{\sqrt{\log N}}{2^5 L_1 a V} = \frac{16 L_1}{a}\sqrt{\log N} \le \frac{16}{b}$$

from (5.63). Moreover

$$-y\lambda + 4a^2\lambda^2 V = -\frac{y^2}{16a^2 V} = -\frac{a^2 \log N}{2^8 L_1^2 V a^2} = -2\log N \ .$$

Then (5.70) yields

$$\mathsf{E}\sup_{\ell \le N}\sum_{i\ge 1}\xi_i t_{\ell,i} \le y + \frac{1}{N\lambda} = \frac{a\sqrt{\log N}}{4L_1}\Big(1 + \frac{1}{4N\log N}\Big) \ ,$$

and since $N \ge 2$ we have indeed proved (5.67) and completed the proof. □

Proof of Theorem 5.3.3. The proof relies on a simple iteration procedure. As in the proof of Proposition 5.3.5, it suffices to consider the case where $\sqrt{\log m} \le a/b$. Let $T = \{t_1, \dots, t_m\}$. Consider a point $t \in T$ and an integer $k \ge -1$. Assume that in the ball $B(t, 2^{k+1}a)$ we can find points $u_1, \dots, u_N$ with $d(u_\ell, u_{\ell'}) \ge 2^k a$ whenever $\ell \ne \ell'$. We can then use Proposition 5.3.5 for the points $u_1 - t, \dots, u_N - t$, with $2^k a$ instead of a and $2b$ instead of b to obtain, using that $\sqrt{\log N} \le \sqrt{\log m} \le a/b$ in the last inequality,

$$b(T) \ge \frac{1}{L}\min\Big(2^k a\sqrt{\log N}, \frac{2^{2k}a^2}{2b}\Big) \ge \frac{1}{L}2^k a\sqrt{\log N} \ .$$

Thus $N \le M_k := \exp(L 2^{-2k} b(T)^2/a^2)$. Consequently every ball in T of radius $2^{k+1}a$ can be covered by at most M_k balls of radius $2^k a$. Iteration of this result shows that T can be covered by at most $\prod_{k\ge -1} M_k$ balls of radius $a/2$. Since $t_\ell \notin B(t_{\ell'}, a/2)$ for $\ell \ne \ell'$ we have $m \le \prod_{k \ge -1} M_k \le \exp(L b(T)^2/a^2)$, i.e. $b(T) \ge a\sqrt{\log m}/L$. □

Our last fundamental result is a comparison principle. Let us say that a map θ from $\mathbb{R}$ to $\mathbb{R}$ is a contraction if $|\theta(s) - \theta(t)| \le |s - t|$ for each $s, t \in \mathbb{R}$.

Theorem 5.3.6. *For $i \ge 1$ consider contractions θ_i with $\theta_i(0) = 0$. Then for each (finite) subset T of ℓ^2 we have*

$$\mathsf{E}\sup_{t\in T}\sum_{i\ge 1}\varepsilon_i\theta_i(t_i) \le b(T) = \mathsf{E}\sup_{t\in T}\sum_{i\ge 1}\varepsilon_i t_i \ . \tag{5.71}$$

A more general comparison result may be found in [5], Theorem 2.1. We give here only the simpler proof of the special case (5.71) that we need.

Proof. The purpose of the condition $\theta_i(0) = 0$ is simply to ensure that $(\theta_i(t_i)) \in \ell^2$ whenever $(t_i) \in \ell^2$. A simple approximation procedure shows that it suffices to show that for each N we have

$$\mathsf{E}\sup_{t\in T}\sum_{1\le i\le N}\varepsilon_i\theta_i(t_i)\le \mathsf{E}\sup_{t\in T}\sum_{1\le i\le N}\varepsilon_i t_i\ .$$

By iteration it suffices to show that $\mathsf{E}\sup_{t\in T}\sum_{1\le i\le N}\varepsilon_i t_i$ decreases when t_1 is replaced by $\theta_1(t_1)$. By conditioning on $\varepsilon_2,\varepsilon_3,\ldots,\varepsilon_N$ it suffices to prove that for a subset T of $\mathbb{R}^2$ and a contraction θ we have

$$\mathsf{E}\sup_{t\in T}(\varepsilon_1\theta(t_1)+t_2)\le \mathsf{E}\sup_{t\in T}(\varepsilon_1 t_1+t_2)\ .$$

To prove this it suffices to show that for $s,s'\in T$ we have

$$\theta(s'_1)+s'_2-\theta(s_1)+s_2\le 2\mathsf{E}\sup_{t\in T}(\varepsilon_1 t_1+t_2)\ . \tag{5.72}$$

Now,

$$2\mathsf{E}\sup_{t\in T}(\varepsilon_1t_1+t_2)\ge\max(s'_1+s'_2-s_1+s_2,s_1+s_2-s'_1+s'_2)=s_2+s'_2+|s'_1-s_1|\ ,$$

so that (5.72) simply follows from the fact that $\theta(s'_1)-\theta(s_1)\le|s'_1-s_1|$ since θ is a contraction. □

5.4 Control in ℓ^∞ Norm

The main result of this section is as follows.

Theorem 5.4.1. *There exists a universal constant L such that for any subset T of ℓ^2 we have*

$$\gamma_2(T)\le L\big(b(T)+\sqrt{b(T)\gamma_1(T,d_\infty)}\big)\ . \tag{5.73}$$

We leave it as an exercise to the reader to prove that this result is actually a consequence of Theorems 5.1.5 and 16.4.12. It is of interest however to give a direct proof, so that we can learn to use some of the basic ideas of the previous section before we plunge in the proof of Theorem 5.1.5.

Corollary 5.4.2. *We have*

$$b(T)\ge\frac{1}{L}\min\left(\gamma_2(T),\frac{\gamma_2(T)^2}{\gamma_1(T,d_\infty)}\right)\ . \tag{5.74}$$

Proof. Denoting by L^* the constant of (5.73), if $b(T)\le\gamma_2(T)/(2L^*)$ then (5.73) implies

$$\gamma_2(T)\le\gamma_2(T)/2+L^*\sqrt{b(T)\gamma_1(T,\delta_\infty)}\ ,$$

hence $b(T)\ge\gamma_2(T)^2/4(L^*)^2\gamma_1(T,d_\infty)$. □

Exercise 5.4.3. Find examples of situations where $\gamma_1(T, d_\infty) \ge \gamma_2(T)$ and $b(T)$ is of order $\gamma_2(T)^2/\gamma_1(T, d_\infty)$. (Hint: try cases where $t_i \in \{0, 1\}$ for each i and each t.)

Our main tool is as follows. It should be compared with Proposition 2.4.9.

Proposition 5.4.4. *There exists constants L_1 and L_2 with the following properties. Consider numbers $a, b, \sigma > 0$, vectors $t_1, \dots, t_m \in \ell^2$, that satisfy (5.59) and (5.60). For $\ell \le m$ consider sets H_ℓ with $H_\ell \subset B_2(t_\ell, \sigma)$. Then*

$$b\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge \frac{1}{L_1} \min\Big(a\sqrt{\log m}, \frac{a^2}{b}\Big) - L_2 \sigma \sqrt{\log m} + \min_{\ell \le m} b(H_\ell) . \quad (5.75)$$

The proof is identical to that of Proposition 2.4.9, if one replaces Lemmas 2.4.2 and 2.4.7 respectively by Theorem 5.3.3 and Theorem 5.3.2.

Corollary 5.4.5. *There exists a constant L_0 with the following property. Consider a set D with $\Delta(D, d_\infty) \le 2a/\sqrt{\log m}$, and points $t_\ell \in D$ that satisfy (5.59). Consider moreover sets $H_\ell \subset B_2(t_\ell, a/L_0)$. Then*

$$b\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge \frac{a}{L_0}\sqrt{\log m} + \min_{\ell \le m} b(H_\ell) . \quad (5.76)$$

Proof. We observe that without loss of generality we may assume $t_1 = 0$, so that $\|t_\ell\|_\infty \le b = 4a/\sqrt{\log m}$ for all $\ell \le m$ and (5.75) used for $\sigma = a/L_0$ gives

$$b\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge \frac{1}{4L_1} a\sqrt{\log m} - \frac{aL_2}{L_0}\sqrt{\log m} + \min_{\ell \le m} b(H_\ell) ,$$

so that if $L_0 \ge 8L_1L_2$ and $L_0 \ge 8L_1$ we get (5.76). □

Proof of Theorem 5.4.1. We consider an integer $\tau \ge 1$ to be specified later, and an admissible sequence of partitions $(\mathcal{D}_n)$ of T such that

$$\sup_{t \in T} \sum_{p \ge 0} 2^p \Delta(D_p(t), d_\infty) \le 2\gamma_1(T, d_\infty) . \quad (5.77)$$

The proof will rely on the application of Theorem 2.7.2 to the functionals

$$F_n(A) = \sup\big\{b(A \cap D) + U_n(D) ,\, D \in \mathcal{D}_{n+\tau} ,\, A \cap D \ne \emptyset\big\} ,$$

where

$$U_n(D) = \sup_{t \in D} \sum_{p \ge n} 2^p \Delta(D_{p+\tau}(t), d_\infty) .$$

Let us observe right away that $U_{n+1}(D) \le U_n(D)$ and that $U_n(D)$ is an increasing function of D.

We now check that the functionals F_n satisfy the growth condition of Definition 2.7.1 for a suitable value of the parameters. Consider $m = N_{n+\tau+1}$ and points $t_1, \ldots, t_m$ of T such that

$$\ell \neq \ell' \Rightarrow \|t_\ell - t_{\ell'}\|_2 \geq a\,, \tag{5.78}$$

and consider sets $H_\ell \subset B_2(t_\ell, a/r)$, where $r = 4L_0$, $L_0 \geq 1$ being the constant of Corollary 5.4.5.

Consider $c < \min_{\ell \leq m} F_{n+1}(H_\ell)$, so that by definition of F_n for each ℓ we can find a set $D_\ell \in \mathcal{D}_{n+\tau+1}$ such that $H_\ell \cap D_\ell \neq \emptyset$ and

$$b(H_\ell \cap D_\ell) + U_{n+1}(D_\ell) > c\,. \tag{5.79}$$

Each of the m sets D_ℓ is contained in one of the sets of $\mathcal{D}_{n+\tau}$. Since $m = N_{n+\tau+1} = N_{n+\tau}^2 \geq N_{n+\tau} \cdot \operatorname{card}\mathcal{D}_{n+\tau}$, by the pigeon hole principle we can find $D \in \mathcal{D}_{n+\tau}$ such that the set

$$I = \{\ell \leq m\,;\ D_\ell \subset D\}$$

satisfies $\operatorname{card} I \geq N_{n+\tau}$. The definition of F_n implies

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq b\Big(D \cap \bigcup_{\ell \in I} H_\ell\Big) + U_n(D)\,. \tag{5.80}$$

Now, for each $\ell \in I$, we have

$$U_n(D) = 2^n \Delta(D, d_\infty) + U_{n+1}(D) \geq 2^n \Delta(D, d_\infty) + U_{n+1}(D_\ell)\,. \tag{5.81}$$

Case 1. We have $\Delta(D, d_\infty) \geq a 2^{-n/2}$. Then, (5.80) and (5.81) show that if ℓ_0 is an arbitrary element of I, we have, using (5.79) for $\ell = \ell_0$ in the last line,

$$\begin{aligned} F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) &\geq 2^{n/2} a + b(D_{\ell_0} \cap H_{\ell_0}) + U_{n+1}(D_{\ell_0}) \\ &\geq 2^{n/2} a + c\,, \end{aligned}$$

and thus

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq 2^{n/2} a + \inf_{\ell \leq m} F_{n+1}(H_\ell)\,. \tag{5.82}$$

Case 2. We have $\Delta(D, d_\infty) \leq a 2^{-n/2}$, and thus $\Delta(D, d_\infty) \leq a/\sqrt{\log N_n}$. We select an arbitrary subset J of I with $\operatorname{card} J = N_n$. For $\ell \in J$ we choose arbitrarily $u_\ell \in H_\ell \cap D_\ell \subset D$, so that, since $H_\ell \subset B_2(t_\ell, a/r)$, we have $H_\ell \subset B_2(u_\ell, 2a/r) = B_2(u_\ell, a/(2L_0))$ since $r = 4L_0$. We observe that, since $r \geq 4$, by (5.78) we have $d_2(u_\ell, u_{\ell'}) \geq a/2$ for $\ell \neq \ell'$.

We use Corollary 5.4.5 with $m = N_n$, $H_\ell \cap D_\ell$ instead of H_ℓ, $a/2$ instead of a and u_ℓ instead of t_ℓ to obtain

$$b\Big(D \cap \bigcup_{\ell \in I} H_\ell\Big) \ge b\Big(\bigcup_{\ell \in J} (H_\ell \cap D_\ell)\Big)$$
$$\ge \frac{a}{2L_0}\sqrt{\log N_n} + \inf_{\ell \in J} b(H_\ell \cap D_\ell) \,.$$

Combining with (5.80) and (5.81) we get

$$\begin{aligned} F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) &\ge \frac{2^{n/2} a}{L} + \inf_{\ell \in J}(b(H_\ell \cap D_\ell) + U_n(D)) \\ &\ge \frac{2^{n/2} a}{L} + \inf_{\ell \in J}(b(H_\ell \cap D_\ell) + U_n(D_\ell)) \\ &\ge \frac{2^{n/2} a}{L} + \inf_{\ell \in J} F_{n+1}(H_\ell) \\ &\ge \frac{2^{n/2} a}{L} + \inf_{\ell \le m} F_{n+1}(H_\ell) \,. \end{aligned} \tag{5.83}$$

Thus, this relation holds, whichever of the preceding cases occur. That is, we have proved that the growth condition of Definition 2.7.1 holds with $\theta(n) = 2^{n/2}/L$, $\tau + 1$ instead of τ and $\beta = 1$ and we can apply Theorem 2.7.2 for these values of the parameters. By definition of F_0 we have

$$F_0(T) \le b(T) + U_0(T)$$

and by (5.77) we have $2^\tau U_0(T) \le 2\gamma_1(T, d_\infty)$, so that

$$F_0(T) \le b(T) + 2^{-\tau+1}\gamma_1(T, d_\infty) \,.$$

Since $\Delta(T, d_2) \le Lb(T)$ by (5.55), we deduce from Lemma 2.3.5 and Theorem 2.7.2 that

$$\gamma_2(T) \le L 2^{\tau/2}\big(b(T) + 2^{-\tau}\gamma_1(T, d_\infty)\big)$$

and Theorem 5.4.1 follows by optimization over $\tau \ge 1$. □

5.5 Latała's Principle

The following crucial result was first proved in [2], but it was not obvious at the time how important this is. The simpler proof presented here comes from [1].

Proposition 5.5.1. *There exists a constant L_1 with the following property. Consider a subset T of ℓ^2 and a subset J of $\mathbb{N}^*$. Assume that for certain numbers $c, \sigma > 0$ and that for an integer m the following holds:*

$$\forall\, s, t \in T \;,\; \sum_{i \in J} (s_i - t_i)^2 \le c^2 \;, \tag{5.84}$$

$$t \in T \Rightarrow \|t\|_\infty < \frac{\sigma}{\sqrt{\log m}} \; . \tag{5.85}$$

Then provided

$$c \le \frac{\sigma}{L_1} \; , \tag{5.86}$$

we can find $m' \le m+1$ *and a partition* $(A_\ell)_{\ell \le m'}$ *of* T *such that for each* $\ell \le m'$ *we have* **either**

$$\exists t^\ell \in T \; , \; A_\ell \subset B(t^\ell, \sigma) \; , \tag{5.87}$$

or else

$$b_J(A_\ell) := \mathsf{E} \sup_{t \in A_\ell} \sum_{i \in J} \varepsilon_i t_i \le b(T) - \frac{\sigma}{L}\sqrt{\log m} \; . \tag{5.88}$$

In this statement there are two distances involved, the canonical distance $d^2(s,t) = \sum_{i \ge 1}(s_i - t_i)^2$ and the smaller distance $d_J^2(s,t) = \sum_{i \in J}(s_i - t_i)^2$. In (5.84) we assume that the diameter of T is small for the *small* distance d_J. We then produce these sets A_ℓ satisfying (5.87) with a small diameter for the *large* distance d. This will turn out to be the key to Theorem 5.1.5, although it will probably take some time for the reader to understand how Proposition 5.5.1 fits into the picture. For the time being it suffices to understand that this proposition allows to split T into not too many pieces on which extra information is gained.

Proof. Certainly we may assume that T cannot be covered by m balls of the type $B(t,\sigma)$. For $t \in T$ set

$$Y_t = \sum_{i \in J} \varepsilon_i t_i \; ; \; Z_t = \sum_{i \notin J} \varepsilon_i t_i \; , \tag{5.89}$$

and define

$$\alpha = \inf_{t^1, \ldots, t^m \in T} \mathsf{E} \sup_{t \in T \setminus \cup_{\ell \le m} B(t^\ell, \sigma)} Y_t \; .$$

To prove the theorem we shall prove that provided the constant L_1 of (5.86) is large enough we have

$$\alpha \le b(T) - \frac{\sigma}{L}\sqrt{\log m} \; . \tag{5.90}$$

Indeed, consider points $t^1, \ldots, t^m$ such that $\mathsf{E}\sup_{t \in T \setminus \cup_{\ell \le m} B(t^\ell, \sigma)} Y_t \le b(T) - \sigma\sqrt{\log m}/L$. The required partition is obtained by taking $A_\ell \subset B(t^\ell, \sigma)$ for $\ell \le m$ and $A_{m+1} = T \setminus \cup_{\ell \le m} B(t^\ell, \sigma)$.

We turn to the proof of (5.90). By definition of α, given points $t^1, \ldots, t^k \in T$ with $k \le m$, the r.v.

$$W = \sup_{t \in T \setminus \cup_{\ell \le k} B(t^\ell, \sigma)} Y_t \quad \text{satisfies} \quad \mathsf{E} W \ge \alpha \; . \tag{5.91}$$

Moreover, using (5.84) and (5.58) it satisfies

$$\forall u > 0 \ , \ \mathsf{P}(|W - \mathsf{E}W| \ge u) \le L \exp\Big(-\frac{u^2}{Lc^2}\Big) \ . \tag{5.92}$$

Let us consider independent copies $(Y_t^k)_{t\in T}$ of the process $(Y_t)_{t\in T}$ (which are also independent of the r.v.s $(\varepsilon_i)_{i\ge 1}$) and a small number $\epsilon > 0$. First, we consider $W_1 := \sup_{t\in T} Y_t^1$ and we select a point $t^1 \in T$ (depending on the r.v.s Y_t^1) with

$$Y^1_{t^1} \ge W_1 - \epsilon \ . \tag{5.93}$$

Next, we let $W_2 = \sup_{t\in T\setminus B(t^1,\sigma)} Y_t^2$ and we find t^2 such that

$$Y^2_{t^2} \ge W_2 - \epsilon \ . \tag{5.94}$$

We proceed in this manner until we construct a last point t^m. The next goal is to bound from above and from below the quantity

$$S := \mathsf{E} \max_{k\le m} (Y^k_{t^k} + Z_{t^k}) \ . \tag{5.95}$$

To find a bound from below, we write

$$\max_{k\le m}(Y^k_{t^k} + Z_{t^k}) \ge \max_{k\le m}(W_k + Z_{t^k}) - \epsilon \ge \min_{k\le m} W_k + \max_{k\le m} Z_{t^k} - \epsilon \ . \tag{5.96}$$

Now, using (5.91) given the points $t^1, \ldots, t^{k-1}$ implies that $\mathsf{E}W_k \ge \alpha$, because the process (Y_t^k) is independent of $t^1, \ldots, t^{k-1}$. Using (5.92) we obtain that for all $u > 0$ we have $\mathsf{P}(W_k \le \alpha - u) \le L\exp(-u^2/(Lc^2))$, so that proceeding as in (2.87) we get

$$\mathsf{E} \min_{k\le m} W_k \ge \alpha - Lc\sqrt{\log m} \ . \tag{5.97}$$

Next, denoting by E_{J^c} expectation in the r.v.s $(\varepsilon_i)_{i\in J^c}$ only, we prove that

$$\mathsf{E}_{J^c} \max_{k\le m} Z_{t^k} \ge \frac{1}{L}\sigma\sqrt{\log m} \ . \tag{5.98}$$

For this we observe that for $s, t \in T$ with $\|s - t\|_2 \ge \sigma$ then, using (5.84),

$$\sum_{i\notin J}(s_i - t_i)^2 = \sum_{i\ge 1}(s_i - t_i)^2 - \sum_{i\in J}(s_i - t_i)^2 \ge \sigma^2 - c^2 \ge (\sigma/2)^2 \ .$$

Thus (5.98) follows from (5.61). Taking expectation in (5.96) and letting $\epsilon \to 0$ we have proved that

$$S \ge \alpha + \Big(\frac{\sigma}{L} - Lc\Big)\sqrt{\log m} \ . \tag{5.99}$$

Consider now some numbers $(a(t))_{t\in T}$ and a median M' of the process $\sup_{t\in T}(Y_t + a(t))$. Using (5.56), (5.84) we obtain

$$\forall u > 0 \ , \ \mathsf{P}\Big(\Big|\sup_{t\in T}(Y_t + a(t)) - M'\Big| \geq u\Big) \leq 4\exp\Big(-\frac{u^2}{4c^2}\Big) \ .$$

Proceeding as in (2.87) we get

$$\mathsf{E}\max_{k\leq m}\sup_{t\in T}(Y_t^k + a(t)) \leq M' + Lc\sqrt{\log m} \ ,$$

and since by (5.57) we have $|M' - \mathsf{E}\sup_{t\in T}(Y_t + a(t))| \leq Lc$ we finally obtain

$$\mathsf{E}\max_{k\leq m}\sup_{t\in T}(Y_t^k + a(t)) \leq \mathsf{E}\sup_{t\in T}(Y_t + a(t)) + Lc\sqrt{\log m} \ . \tag{5.100}$$

In particular, since Y_t^k does not depend on the r.v.s $(\varepsilon_i)_{i\in J^c}$, denoting now by E^{J^c} expectation given the r.v.s $(\varepsilon_i)_{i\in J^c}$, (5.100) implies

$$\mathsf{E}^{J^c}\max_{k\leq m}\sup_{t\in T}(Y_t^k + Z_t) \leq \mathsf{E}^{J^c}\sup_{t\in T}(Y_t + Z_t) + Lc\sqrt{\log m} \ .$$

Taking expectation yields

$$S \leq \mathsf{E}\max_{k\leq m}\sup_{t\in T}(Y_t^k + Z_t) \leq b(T) + Lc\sqrt{\log m} \ ,$$

and combining with (5.99) we obtain

$$\alpha + \Big(\frac{\sigma}{L} - Lc\Big)\sqrt{\log m} \leq b(T) + Lc\sqrt{\log m} \ ,$$

so that indeed (5.90) holds true provided the constant L_1 of (5.86) is large enough. □

5.6 Chopping Maps and Functionals

One of the most successful ideas about Bernoulli processes is that of chopping maps. The basic idea is to replace the individual r.v.s $\varepsilon_i x_i$ by a sum $\sum \varepsilon_{i,j} x_{i,j}$ where $\varepsilon_{i,j}$ are independent Bernoulli r.v.s and where $x_{i,j}$ are "small pieces of x_i". It is then easier to control the ℓ^∞ norm of the new process.

Given $u \leq v \in \mathbb{R}$ we define the function $\varphi_{u,v}$ as the unique continuous function for which $\varphi_{u,v}(0) = 0$, which is constant for $x \leq u$ and $x \geq v$ and has slope 1 between these values. Thus

$$\varphi_{u,v}(x) = \min(v, \max(x,u)) - \min(v, \max(u,0)) \ . \tag{5.101}$$

Consequently $|\varphi_{u,v}(x)| \leq v - u$, and $|\varphi_{u,v}(x) - \varphi_{u,v}(y)| \leq |x-y|$, with equality when $u \leq x, y \leq v$.

It is very useful to note that if $u_1 \leq u_2 \leq \cdots \leq u_k$, then

$$\varphi_{u_1,u_k}(x) = \sum_{1\leq \ell < k}\varphi_{u_\ell,u_{\ell+1}}(x) \ . \tag{5.102}$$

This is simply because both the left-hand side and the right-hand side are continuous, constant for $x \le u_1$ and $x \ge u_k$, have slope 1 between these values, and take the value 0 at 0.

Given a finite subset G of $\mathbb{R}$ we define

$$G^- := \{u \in G\ ;\ \exists v \in G\ ,\ u < v\}\ ,$$

and for $u \in G^-$ we define $u^+ = \inf\{v \in G\ ;\ u < v\}$. It will always be implicitly assumed that $\operatorname{card} G \ge 2$ so that $G^- \ne \emptyset$. The following is then obvious but also essential.

Lemma 5.6.1. *For each $x, y \in \mathbb{R}$ and each finite set G, we have*

$$\sum_{u \in G^-} |\varphi_{u,u^+}(x) - \varphi_{u,u^+}(y)| \le |x - y|\ . \tag{5.103}$$

Moreover there is equality if $\min G \le x, y \le \max G$.

In particular, since $\varphi_{u,u^+}(0) = 0$, we have

$$\sum_{u \in G^-} |\varphi_{u,u^+}(x)| \le |x|\ . \tag{5.104}$$

In the remainder of this chapter we consider independent Bernoulli r.v.s $\varepsilon_{x,i}$ for $x \in \mathbb{R}$ and $i \in \mathbb{N}^*$. These are also assumed to be independent of all other Bernoulli r.v.s considered, in particular the ε_i.

Consider now for $i \ge 1$ a finite set $G_i \subset \mathbb{R}$. For $t \in \ell^2$ we consider the r.v.

$$X_t(G_i, i) := \sum_{u \in G_i^-} \varepsilon_{u,i} \varphi_{u,u^+}(t_i)\ . \tag{5.105}$$

That is, the value t_i is "chopped" into the potentially smaller pieces $\varphi_{u,u^+}(t_i)$. We do this for all values of i. We write $\mathcal{G} = (G_i)_{i \ge 1}$ and we consider the r.v.

$$X_t(\mathcal{G}) := \sum_{i \ge 1} X_t(G_i, i) = \sum_{i \ge 1} \sum_{u \in G_i^-} \varepsilon_{u,i} \varphi_{u,u^+}(t_i)\ . \tag{5.106}$$

Observe that the series converges in L^2 if $t \in \ell^2$ thanks to (5.104). In this manner to a Bernoulli process $(X_t)_{t \in T}$ we associate a new Bernoulli process $(X_t(\mathcal{G}))_{t \in T}$. Another way to express this which we shall use at times is that we replace T by the set

$$\left\{ (\varphi_{u,u^+}(t_i))_{i \in \mathbb{N}^*, u \in G_i^-}\ ;\ t \in T \right\}\ ,$$

so that the index set $\mathbb{N}^*$ has been replaced by the larger index set $\{(i, u); i \in \mathbb{N}^*, u \in G_i^-\}$. It follows from (5.103) and the inequality $\sum a_k^2 \le (\sum |a_k|)^2$ that the canonical distance $d_{\mathcal{G}}$ associated to the new process satisfies

$$d_{\mathcal{G}}(s,t)^2 = \sum_{i\geq 1, u\in G_i^-} (\varphi_{u,u^+}(s_i) - \varphi_{u,u^+}(t_i))^2 \leq d(s,t)^2 = \sum_{i\geq 1}(s_i - t_i)^2 \,. \tag{5.107}$$

The problem is that the reverse inequality is by no means true, and that a set can very well be of small diameter for $d_{\mathcal{G}}$ but not for d. This is in a sense the main difficulty in using chopping maps. We shall discover soon how brilliantly Bednorz and Latała bypassed this difficulty using Proposition 5.5.1.

The following is fundamental.

Proposition 5.6.2. *For any family $\mathcal{G}$ and any finite set $T \subset \ell^2$ we have*

$$\mathsf{E}\sup_{t\in T} X_t(\mathcal{G}) \leq b(T) = \mathsf{E}\sup_{t\in T}\sum_{i\geq 1}\varepsilon_i t_i \,. \tag{5.108}$$

Proof. The families $(\varepsilon_{x,i})$ and $(\varepsilon_i\varepsilon_{x,i})$ have the same distribution, so that

$$\begin{aligned}\mathsf{E}\sup_{t\in T} X_t(\mathcal{G}) &= \mathsf{E}\sup_{t\in T}\sum_{i\geq 1, u\in G_i^-}\varepsilon_{u,i}\varphi_{u,u^+}(t_i)\\ &= \mathsf{E}\sup_{t\in T}\sum_{i\geq 1, u\in G_i^-}\varepsilon_i\varepsilon_{u,i}\varphi_{u,u^+}(t_i)\\ &= \mathsf{E}\Big(\mathsf{E}_\varepsilon\sup_{t\in T}\sum_{i\geq 1}\varepsilon_i\theta_i(t_i)\Big)\,, \end{aligned}\tag{5.109}$$

where $\theta_i(x) = \sum_{u\in G_i^-}\varepsilon_{u,i}\varphi_{u,u^+}(x)$, and where E_ε means averaging only in $(\varepsilon_i)_{i\geq 1}$. We note that θ_i is a contraction, since

$$|\theta_i(x) - \theta_i(y)| \leq \sum_{u\in G_i^-}|\varphi_{u,u^+}(x) - \varphi_{u,u^+}(y)| \leq |x-y|$$

by (5.103). The key point is (5.71) which implies

$$\mathsf{E}_\varepsilon\sup_{t\in T}\sum_{i\geq 1}\varepsilon_i\theta_i(t_i) \leq \mathsf{E}\sup_{t\in T}\sum_{i\geq 1}\varepsilon_i t_i = b(T)\,.$$

Combining with (5.109) finishes the proof. □

Chopping maps were invented to prove the following, which illustrates well their power. We denote by B_1 the unit ball of ℓ^1.

Proposition 5.6.3. *There exists a constant L such that for each subset T of ℓ^2 we have, for $\epsilon > 0$*

$$\epsilon\sqrt{\log N(T, \epsilon B_2 + Lb(T)B_1)} \leq Lb(T)\,,$$

where $N(T,C)$ is the smallest number of translates of C that can cover T.

Proof. Consider $c > 0$, and the map $\Psi_c : \ell^2 = \ell^2(\mathbb{N}^*) \to \ell^2(\mathbb{N}^* \times \mathbb{Z})$ given by $\Psi_c(t) = ((\varphi_{\ell c,(\ell+1)c}(t_i))_{(i,\ell)})$. Applying in turn (an obvious adaptation of) Proposition 5.6.2 and Theorem 5.3.3 yields

$$b(T) \geq b(\Psi_c(T)) \geq \frac{1}{L}\min\Big(\epsilon\sqrt{\log N(\Psi_c(T), \epsilon B_2)}, \frac{\epsilon^2}{c}\Big)\,, \tag{5.110}$$

because if $m \leq N(\Psi_c(T), \varepsilon B_2)$, we can find points $(t_\ell)_{\ell \leq m}$ in $\Psi_c(T)$ with $\|t_\ell - t_{\ell'}\| \geq \epsilon$ for $\ell \neq \ell'$, and since $\|t\|_\infty \leq c$ for $t \in \Psi_c(T)$. Thus if we choose $c = \epsilon^2/(2Lb(T))$ where L is as in (5.110) we get

$$b(T) \geq \min\Big(\frac{1}{L}\epsilon\sqrt{\log N(\Psi_c(T), \epsilon B_2)}, 2b(T)\Big)\,,$$

so that $Lb(T) \geq \epsilon\sqrt{\log N(\Psi_c(T), \epsilon B_2)}$. It then follows from (5.111) below that

$$\Psi_c(x) \in \Psi_c(y) + \epsilon B_2 \Rightarrow x \in y + 2\epsilon B_2 + \frac{4\epsilon^2}{c} B_1$$

and therefore

$$N(T, 2\epsilon B_2 + 8Lb(T)B_1) = N(T, 2\epsilon B_2 + \frac{4\epsilon^2}{c} B_1) \leq N(\Psi_c(T), \epsilon B_2)\,. \qquad \square$$

The following exercise helps explain the nice behavior of chopping maps with respect to interpolation between ℓ^2, ℓ^∞ and ℓ^1 norms. It will be used in Chapter 11. The elementary proof is better left to the reader. There is no reason to believe that the constants are optimal, they are just a reasonable choice.

Exercise 5.6.4. Prove that for $x, y \in \mathbb{R}$ we have

$$|x-y|^2 \mathbf{1}_{\{|x-y|<c\}} + c|x-y|\mathbf{1}_{\{|x-y|\geq c\}} \leq 4\sum_{\ell\in\mathbb{Z}} |\varphi_{c\ell,c(\ell+1)}(x) - \varphi_{c\ell,c(\ell+1)}(y)|^2. \tag{5.111}$$

and

$$\sum_{\ell\in\mathbb{Z}} |\varphi_{c\ell,c(\ell+1)}(x) - \varphi_{c\ell,c(\ell+1)}(y)|^2 \leq |x-y|^2\mathbf{1}_{\{|x-y|<c\}} + 2c|x-y|\mathbf{1}_{\{|x-y|\geq c\}}\,. \tag{5.112}$$

Exercise 5.6.5. Deduce Proposition 5.6.3 from Theorem 5.1.5. (Hint: Observe that whenever $T_2 \subset aB_1$ we have

$$N(T_1 + T_2, \varepsilon B_2 + aB_1) \leq N(T_1, \varepsilon B_2)$$

and use the Sudakov minoration Lemma 2.4.2.)

We are going to be confronted by the following situation. For $i \geq 1$ we consider finite sets $G_i \subset G'_i$. Letting $\mathcal{G} = (G_i)_{i\geq 1}$ and $\mathcal{G}' = (G'_i)_{i\geq 1}$, we want to compare the processes $(X_t(\mathcal{G}))_t$ and $(X_t(\mathcal{G}'))_t$. We start by comparing the associated distances.

Proposition 5.6.6. *(a) Assume that for a certain integer q*

$$\forall i \in \mathbb{N}^* \ , \ \forall u \in G_i^- \ , \ \mathrm{card}([u, u^+[\cap G'_i) \le q \ , \tag{5.113}$$

where u^+ is the successor of u in G_i. Then

$$d_{\mathcal{G}} \le \sqrt{q} d_{\mathcal{G}'} \ . \tag{5.114}$$

(b) Assume that

$$\forall i \in \mathbb{N}^* \ , \ \min G_i = \min G'_i \ , \ \max G_i = \max G'_i \ . \tag{5.115}$$

Then

$$d_{\mathcal{G}'} \le d_{\mathcal{G}} \ . \tag{5.116}$$

Proof. Throughout the proof we write u an element of G_i^- and u^+ it successor in G_i; and v an element of $G_i'^-$ and v^+ its successor in G'_i. Thus, for $s, t \in T$ we have

$$d_{\mathcal{G}}(s,t)^2 = \sum_{i \ge 1} \sum_{u \in G_i^-} (\varphi_{u,u^+}(s_i) - \varphi_{u,u^+}(t_i))^2 \tag{5.117}$$

and

$$d_{\mathcal{G}'}(s,t)^2 = \sum_{i \ge 1} \sum_{v \in G_i'^-} (\varphi_{v,v^+}(s_i) - \varphi_{v,v^+}(t_i))^2 \ . \tag{5.118}$$

Given $i \in \mathbb{N}^*$ and $u \in G_i^-$ let us define the set $G_{i,u} = G_i'^- \cap [u, u^+[$. As u varies in G_i^-, one thus obtains disjoint subsets of $G_i'^-$. The union of these sets is $G_i'^-$ exactly when $\min G_i = \min G'_i$ and $\max G_i = \max G'_i$.

Next, we observe from (5.102) that for any $u \in G_i^-$,

$$|\varphi_{u,u^+}(s_i) - \varphi_{u,u^+}(t_i)| = \sum_{v \in G_{i,u}} |\varphi_{v,v^+}(s_i) - \varphi_{v,v^+}(t_i)| \ . \tag{5.119}$$

Thus, using the inequality $(\sum_{k \le q} a_k)^2 \le q \sum_{k \le q} a_k^2$, and since under condition (5.113) we have $\mathrm{card}\, G_{i,u} \le q$, we get then

$$d_{\mathcal{G}}(s,t)^2 \le q \sum_{i \ge 1} \sum_{u \in G_i^-} \sum_{v \in G_{i,u}} (\varphi_{v,v^+}(s_i) - \varphi_{v,v^+}(t_i))^2 \le q d_{\mathcal{G}'}(s,t)^2 \ ,$$

and we have proved (5.114). Next, using again (5.119) as well as the inequality $(\sum_k |a_k|)^2 \ge \sum a_k^2$ we obtain

$$d_{\mathcal{G}}(s,t)^2 \ge \sum_{i \ge 1} \sum_{u \in G_i^-} \sum_{v \in G_{i,u}} (\varphi_{v,v^+}(s_i) - \varphi_{v,v^+}(t_i))^2 \ ,$$

and we have observed that under (5.115) the right-hand side is exactly $d_{\mathcal{G}'}(s,t)^2$, so that we have proved (5.116) as well. ☐

Proposition 5.6.7. *Under (5.115) we have*

$$\mathsf{E}\sup_{t\in T} X_t(\mathcal{G}') \le \mathsf{E}\sup_{t\in T} X_t(\mathcal{G})\ . \tag{5.120}$$

Proof. This is a consequence of Proposition 5.6.2, and the only difficulty is in the notation. Let us start with a simple observation. To lighten notation for $u < u'$ let $\theta(u,u') = -\min(u', \max(u,0))$ so that (5.101) means $\varphi_{u,u'}(x) = \min(u', \max(x,u)) + \theta(u,u')$. Next, we observe that

$$u \le v \le v' \le u' \Rightarrow \varphi_{v,v'}(x) = \varphi_{v+\theta(u,u'),v'+\theta(u,u')}(\varphi_{u,u'}(x))\ . \tag{5.121}$$

Indeed, the right-hand side is zero for $x = 0$, is constant until $\varphi_{u,u'}(x)$ reaches the value $v + \theta(u,u')$, i.e.until $x = v$, then has a slope 1 and is constant after $\varphi_{u,u'}(x)$ passes the value $v' + \theta(u,u')$, i.e.after $x = v'$.

Let us start the main argument. Consider the set $J = \{(i,u)\ ;\ i \in \mathbb{N}^*, u \in G_i^-\}$. Let

$$T' = \{(\varphi_{u,u^+}(t_i))_{(i,u)\in J}\ ;\ t \in T\}$$

so that

$$\mathsf{E}\sup_{t\in T} X_t(\mathcal{G}) = \mathsf{E}\sup_{s\in T'} \sum_{j\in J} \varepsilon_j s_j\ ,$$

where $\varepsilon_j = \varepsilon_{u,i}$ for $j = (i,u) \in J$. For such a j let us define the set

$$G_j^* = (G_i' \cap [u,u^+]) + \theta(u,u^+)\ ,$$

so that, recalling the sets $G_{i,u} = G_i' \cap [u, u^+[$, we have

$$G_j^{*-} = G_{i,u} + \theta(u,u^+) \tag{5.122}$$

when $j = (i,u)$.

Denoting by $(\varepsilon^*_{x,j})_{x\in\mathbb{R},j\in J}$ a new sequence of independent Bernoulli r.v.s, it follows from Proposition 5.6.2 that

$$\mathsf{E}\sup_{s\in T'} \sum_{j\in J} \sum_{v\in G_j^{*-}} \varepsilon^*_{v,j}\varphi_{v,v^+}(s_j) \le \mathsf{E}\sup_{s\in T'} \sum_{j\in J} \varepsilon_j s_j = \mathsf{E}\sup_{t\in T} X_t(\mathcal{G})\ . \tag{5.123}$$

Recalling that $j = (i,u)$, the left-hand side of (5.123) is then, using (5.122) in the first line, the definition of T' in the second line, and (5.121) in the third line,

$$\begin{aligned}
&\mathsf{E}\sup_{s\in T'} \sum_{i\ge 1} \sum_{u\in G_i^-} \sum_{v\in G_{i,u}} \varepsilon^*_{v,(i,u)}\varphi_{v+\theta(u,u^+),v^++\theta(u,u^+)}(s_i) \\
&= \mathsf{E}\sup_{t\in T} \sum_{i\ge 1} \sum_{u\in G_i^-} \sum_{v\in G_{i,u}} \varepsilon^*_{v,(i,u)}\varphi_{v+\theta(u,u^+),v^++\theta(u,u^+)}(\varphi_{u,u^+}(t_i)) \\
&= \mathsf{E}\sup_{t\in T} \sum_{i\ge 1} \sum_{u\in G_i^-} \sum_{v\in G_{i,u}} \varepsilon^*_{v,(i,u)}\varphi_{v,v^+}(t_i)\ .
\end{aligned} \tag{5.124}$$

Now the sequence $(\varepsilon^*_{v,(i,u)})$ is simply an independent sequence $(\varepsilon_{v,i})$ so the last expression is $\mathsf{E}\sup_{t\in T}\sum_{i\geq 1}\sum_{u\in G_i^-}\sum_{v\in G_{i,u}}\varepsilon_{v,i}\varphi_{v,v^+}(t_i)$. Moreover (5.115) ensures that $G_i'^- = \cup_{u\in G_i^-} G_{i,u}$ so that

$$\mathsf{E}\sup_{t\in T}\sum_{i\geq 1}\sum_{u\in G_i^-}\sum_{v\in G_{i,u}}\varepsilon_{v,i}\varphi_{v,v^+}(t_i) = \mathsf{E}\sup_{t\in T}\sum_{i\geq 1}\sum_{v\in G_i'^-}\varepsilon_{v,i}\varphi_{v,v^+}(t_i)\ . \quad (5.125)$$

Since this last quantity is $\mathsf{E}\sup_{t\in T} X_t(\mathcal{G}')$, this concludes the proof. □

We are now ready to define the functionals which we will use to prove Theorem 5.1.5. Of course, the motivation for these definitions will become only gradually clear. These functionals depend on four parameters, two integers $k \leq h \in \mathbb{Z}$, (yes, h denotes an **integer**), a point $w \in \ell^2$ and a subset I of $\mathbb{N}^*$. We fix an integer $r \geq 2$, which will be chosen later on. First, for $x \in \mathbb{R}$ and $k \in \mathbb{Z}$ we define the set

$$G(x,k) = \{pr^{-k}\ ;\ p \in \mathbb{Z}\ ,\ |pr^{-k} - x| \leq 4r^{-k}\}\ , \quad (5.126)$$

and we observe right away that $\operatorname{card} G(x,k) \leq 9$. We also observe that

$$\min G(x,k) \leq x - 3r^{-k} \leq x \leq x + 3r^{-k} \leq \max G(x,k)\ . \quad (5.127)$$

Next, given $k \leq h \in \mathbb{Z}$ and $x \in \mathbb{R}$, we define the set

$$G(x,k,h) = \{pr^{-h}\ ;\ p \in \mathbb{Z}\ ,\ \min G(x,k) \leq pr^{-h} \leq \max G(x,k)\}\ , \quad (5.128)$$

and we observe that

$$\min G(x,h,k) = \min G(x,h)\ ;\ \max G(x,h,k) = \max G(x,h)\ . \quad (5.129)$$

We note that $G(x,k) = G(x,k,k)$ (so that $\operatorname{card} G(x,k,k) \leq 9$) and that $G(x,k,h)$ increases with h. In words, $G(x,k,h)$ consists of about $9 \cdot 2^{h-k}$ points evenly spaced (with a spacing of 2^{-h}) roughly centered on the point x. The purpose of the parameter $w \in \ell^2$ is that for the i coordinate we will use the sets $G(w_i,k,h)$ which are roughly centered around w_i.

Definition 5.6.8. *For a set $T \subset \ell^2$, integers $k \leq h$, a point $w \in \ell^2$ and a subset I of $\mathbb{N}^*$ we define*

$$F(T,I,w,k,h) = \mathsf{E}\sup_{t\in T}\sum_{i\in I}\sum_{u\in G(w_i,k,h)^-}\varepsilon_{u,i}\varphi_{u,u^+}(t_i)\ . \quad (5.130)$$

We denote by $\Delta(T,I,w,k,h)$ the diameter of T for the corresponding distance,

$$\Delta(T,I,w,k,h)^2 = \sup_{s,t\in T}\sum_{i\in I}\sum_{u\in G(w_i,k,h)^-}(\varphi_{u,u^+}(s_i) - \varphi_{u,u^+}(t_i))^2\ . \quad (5.131)$$

Even if we forget to mention it again, when writing these expressions it is always assumed that $h \geq k$. We note that decreasing I and increasing h decreases the number of terms in the summation (5.130). This opens the door to the use of Proposition 5.5.1.

Let us first point out some regularity properties of these functionals.

Lemma 5.6.9. *If $I' \subset I \subset \mathbb{N}^*$, $k' \geq k$ and $h' \geq h$ then*

$$F(T, I', w, k', h') \leq F(T, I, w, k, h) \tag{5.132}$$

and

$$\Delta(T, I', w, k', h') \leq \Delta(T, I, w, k, h) . \tag{5.133}$$

Proof. That $F(T, I, w, k, h)$ is an increasing function of I follows from Jensen's inequality, by moving the expectation over the r.v.s $\varepsilon_{u,i}$ for $i \in I' \setminus I$ inside the supremum rather than outside. That $F(T, I, w, k, h)$ is a decreasing function of k' follows similarly since for $k \leq k' \leq h$ it holds that $G(w_i, k', h)^- \subset G(w_i, k, h)^-$, by moving inside the supremum expectation with respect to the r.v.s $\varepsilon_{u,i}$ for $u \in G(w_i, k, h)^- \setminus G(w_i, k', h)^-$. That $F(T, I, w, k, h)$ is a decreasing function of h follows from Proposition 5.6.7 and (5.129). The statements concerning $\Delta(T, I, w, k, h)$ are easier, using now (5.116). □

Another key idea is that the distances corresponding to the functionals (5.24) relate well to the distance considered in Theorem 5.2.6 (in the case where μ is the counting measure). We formulate this now.

Lemma 5.6.10. *Consider $x, y, z \in \mathbb{R}$ and assume that $|y - x| \leq 2r^{-k}$. Then*

$$|y - z|^2 \wedge r^{-2h} \leq 2 \sum_{u \in G(x,k,h)^-} (\varphi_{u,u^+}(y) - \varphi_{u,u^+}(z))^2 . \tag{5.134}$$

Proof. First we reduce to the case where $|y - z| \leq r^{-h}$. To do this, assume for example that $y \leq z \leq x$. Then replacing y by $\max(y, z - r^{-h})$ does not change the left-hand side and decreases the right-hand side. In a second step we use observe from (5.127) that $\min G(x, k, h) \leq y, z \leq \max G(x, k, h)$. Then (5.103) implies that

$$|y - z| = \sum_{u \in G(x,k,h)^-} |\varphi_{u,u^+}(y) - \varphi_{u,u^+}(z)| .$$

Now, since $|y - z| \leq r^{-h}$, there are at most two non-zero terms in the right-hand side, and $(a + b)^2 \leq 2(a^2 + b^2)$. □

We now state and prove the key step in the proof of Theorem 5.1.5.

Proposition 5.6.11. *There exists a constant L_2 with the following property. Consider $w, w' \in \ell^2$, a set $I \subset \mathbb{N}^*$ and integers $k \le h$. Consider a subset T of ℓ^2 such that $\Delta(T, I, w, k, h+2) \le c$. Assume that*

$$c \le \frac{\sigma}{L_2} \; ; \; r^{-h-1}\sqrt{\log m} \le \sigma \; . \tag{5.135}$$

Let

$$I' = \{i \in I \; ; \; |w_i - w'_i| \le 2r^{-k}\} \; . \tag{5.136}$$

Then we can find $m' \le m+1$ and a partition $(A_\ell)_{\ell \le m'}$ of T such that for each $\ell \le m'$ we have **either**

$$\Delta(A_\ell, I, w, k, h+1) \le \sigma \tag{5.137}$$

or else

$$F(A_\ell, I', w', h+2, h+2) \le F(T, I, w, k, h+1) - \frac{\sigma}{L}\sqrt{\log m} \; . \tag{5.138}$$

The fundamental point of this result is that the hypothesis on T involves a control of $\Delta(T, I, w, k, h+2)$ rather than of the larger quantity $\Delta(T, I, w, k, h+1)$. In words, each piece produced by this decomposition is either is such that its diameter for the large distance is small, or else its properly measured size for the functionals has decreased.

Proof. There is no loss of generality to assume for notational convenience that $I = \mathbb{N}^*$. For $i \in \mathbb{N}^*$ consider the sets

$$G_i = G(w_i, k, h+1) \; ,$$

and $\mathcal{G} = (G_i)_{i \ge 1}$. For $i \in \mathbb{N}^* \setminus I'$ let $G'_i = G_i$ and for $i \in I'$ define $G'_i = G_i \cup G(w'_i, h+2, h+2)$ and $\mathcal{G}' = (G'_i)_{i \ge 1}$. The central object of the proof is the process $(X_t(\mathcal{G}'))_{t \in T}$. The magic of this proof is that neither this process nor the distances it induces are exactly what we need, but rather are related to the quantities of interest through three inequalities which all turn out to be in the right direction.

First we observe that since $r \ge 2$ and $h \ge k$,

$$G(w'_i, h+2, h+2) \subset [w'_i - 4r^{-h-2}, w'_i + 4r^{-h-2}] \subset [w'_i - r^{-k}, w'_i + r^{-k}]$$

and since $|w_i - w'_i| \le 2r^{-k}$ for $i \in I'$ it follows then that $G(w'_i, h+2, h+2) \subset [w_i - 3 \cdot r^{-k}, w_i + 3 \cdot r^{-k}]$ and consequently from (5.127) that the sets G_i and G'_i satisfy (5.115). Therefore by (5.120) (used for I rather than N^*) we have

$$\mathsf{E} \sup_{t \in T} X_t(\mathcal{G}') \le \mathsf{E} \sup_{t \in T} X_t(\mathcal{G}) = F(T, I, w, k, h+1) \; . \tag{5.139}$$

Next, since $\operatorname{card} G'_i = \operatorname{card} G(w'_i, h+2, h+2) \le 9$ it is obvious that the sets G_i and G'_i satisfy (5.113) with $q = 16$ (and even with $q = 10$). Thus Proposition 5.6.6 implies that $d_{\mathcal{G}} \le 4 d_{\mathcal{G}'}$.

Finally, since $|w_i - w'_i| \le 2r^{-k}$ for $i \in I'$ we have

$$|pr^{-h-2} - w'_i| \le 4r^{-h-2} \Rightarrow |pr^{-h-2} - w_i| \le 2r^{-k} + 4r^{-h-2} \le 3r^{-k}$$

so that $G(w'_i, h+2, h+2) \subset G(w_i, k, h+2)$. In fact the points of the left-hand set are consecutive points of the right-hand set, and hence obviously

$$\Delta(T, I', w', h+2, h+2) \le \Delta(T, \mathbb{N}^*, w, k, h+2) \le c\,. \tag{5.140}$$

The proof now simply consists in applying Proposition 5.5.1 with $\sigma' = \sigma/8$ instead of σ with the set of indices

$$J^* = \{(i,u)\ ;\ i \in \mathbb{N}^*, u \in G_i'^-\}$$

instead of $\mathbb{N}^*$ and J given by

$$J = \{(i,u)\ ;\ i \in I', u \in G(w'_i, h+2, h+2)^-\}\ ,$$

to the set $T' \subset \ell^2(J^*)$ given by

$$T' = \{(\varphi_{u,u^+}(t_i))_j\ ;\ t \in T\ ,\ j = (i,u) \in J^*\}\ .$$

Thus, with the notation of that proposition,

$$b(T') = \mathsf{E} \sup_{t \in T} X_t(\mathcal{G}')$$

and, for $A \subset T$,

$$b_J(A) = F(A, I', w', h+2, h+2)\ ,$$

while $d_{\mathcal{G}'}$ is what corresponds to the ℓ^2 distance in Proposition 5.5.1. Now (5.140) implies (5.84). Moreover (5.85) holds since $r^{-h-1}\sqrt{\log m} \le \sigma$ by (5.135). Also, (5.86), i.e. $c \le \sigma'/L_1$ follows from $c \le \sigma/L_2$ provided that $L_2 = 8L_1$. Thus we can indeed apply Proposition 5.5.1. Those sets A_ℓ it produces satisfying (5.87) have a diameter $\le 2\sigma'$ for the distance $d_{\mathcal{G}'}$, so that they have a diameter $\le 8\sigma' = \sigma$ for the distance $d_{\mathcal{G}}$, which is exactly the distance used in computing the diameter in (5.137).

Moreover (5.88) means precisely that

$$F(A_\ell, I', w', h+2, h+2) \le \mathsf{E} \sup_{t \in T} X_t(\mathcal{G}') - \frac{\sigma}{L}\sqrt{\log m}\,. \tag{5.141}$$

Therefore (5.138) is a direct consequence of (5.141) and (5.139). □

5.7 The Decomposition Lemma

Besides Proposition 5.6.11, we need another decomposition principle, which is very similar to what we did in the Gaussian case. Here of course Δ denotes the diameter for the ℓ^2 distance.

Lemma 5.7.1. *There exists a universal constant L_3 with the following property. Consider a set $T \subset \ell^2$ and $b, c > 0$. Assume that $\|t\|_\infty \le b$ for all $t \in T$. Consider m with $b\sqrt{\log m} \le c$. Then we can find $m' \le m$ and a partition $(A_\ell)_{\ell \le m'}$ of T such that for each $\ell \le m'$ we have* **either**

$$\forall D \subset A_\ell \; ; \; \Delta(D) \le \frac{c}{L_3} \Rightarrow b(D) \le b(T) - \frac{c}{L}\sqrt{\log m} \tag{5.142}$$

or else

$$\Delta(A_\ell) \le c \; . \tag{5.143}$$

Proof. The proof is identical to the proof of Lemma 2.6.2, using now Corollary 5.4.5 with $a = b\sqrt{\log m}/2$. □

The following is an immediate consequence of Lemma 5.7.1, used with $b = r^{-h}$ and $J = \{(i, u); i \in I, u \in G(w_i, k, h)^-\}$ instead of $\mathbb{N}^*$.

Corollary 5.7.2. *Consider a set $T \subset \ell^2$ and $w \in \ell^2$. Consider $I \subset \mathbb{N}^*$, $c > 0$ and integers $k \le h$. Assume that $r^{-h}\sqrt{\log m} \le c$. Then we can find $m' \le m$ and a partition $(A_\ell)_{\ell \le m'}$ of T such that for each $\ell \le m'$ we have* **either**

$$\begin{aligned} &\forall D \subset A_\ell \; ; \; \Delta(D, I, w, k, h) \le \frac{c}{L_3} \\ &\Rightarrow F(D, I, w, k, h) \le F(T, I, w, k, h) - \frac{c}{L}\sqrt{\log m} \end{aligned} \tag{5.144}$$

or else

$$\Delta(A_\ell, I, w, k, h) \le c \; . \tag{5.145}$$

We can now state and prove the basic tool to construct partitions.

Lemma 5.7.3. *There exists a number n_0 with the following property. Consider an integer $n \ge n_0$. Consider a set $T \subset \ell^2$, a point $w \in \ell^2$, a subset $I \subset \mathbb{N}^*$, integers $k \le j$. Then we can find $m \le N_n$ and a partition $(A_\ell)_{\ell \le m}$ such that for each $\ell \le m$ we have* **either** *of the following properties:*
(a) We have

$$\begin{aligned} &D \subset A_\ell \; ; \; \Delta(D, I, w, k, j+2) \le \frac{1}{L_4} 2^{(n+1)/2} r^{-j-1} \Rightarrow \\ &F(D, I, w, k, j+2) \le F(T, I, w, k, j+2) - \frac{1}{L} 2^n r^{-j-1} \; , \end{aligned} \tag{5.146}$$

or
(b) There exists $w_\ell \in T$ such that for $I_\ell = \{i \in I; |w_i - w_{\ell,i}| \le 2r^{-k}\}$ we have

$$F(A_\ell, I_\ell, w_\ell, j+2, j+2) \le F(T, I, w, k, j+1) - \frac{1}{L} 2^n r^{-j-1} \; , \tag{5.147}$$

and in particular, using (5.132),

$$F(A_\ell, I_\ell, w_\ell, j+2, j+2) \le F(T, I, w, k, j) - \frac{1}{L} 2^n r^{-j-1} , \qquad (5.148)$$

or else

(c)

$$\Delta(A_\ell, I, w, k, j+1) \le 2^{n/2} r^{-j-1} . \qquad (5.149)$$

Here cases (a) and (c) are as in the Gaussian case. In case (c) the set is of small diameter. In case (a) the small diameter subsets of the set have a small value for the functional. This should not come as a surprise since these cases are produced by the application of Lemma 5.7.1 which is identical to what is done in the Gaussian case. The really new feature in case (b).

Proof. First we apply Corollary 5.7.2 with the same value of k, with $h = j+2$, $c = 2^{n/2} r^{-j-1}/L_2$, where L_2 occurs in Proposition 5.6.11, and where m is the largest integer $m \le N_{n-1}$ such that $r^{-j-2}\sqrt{\log m} \le c$. This produces pieces $(C_\ell)_{\ell \le m'}$ with $m' \le m$ which satisfy either (5.144) or (5.145). Now, if n_0 is large enough we have $\sqrt{\log m} \ge 2^{n/2}/L$ and therefore $c\sqrt{\log m} \ge 2^n r^{-j-1}/L$. Moreover, letting $L_4 = \sqrt{2} L_2 L_3$, we have $c/L_3 = 2^{(n+1)/2} r^{-j-1}/L_4$ so that the pieces which satisfy (5.144) also satisfy (5.146). The other pieces C_ℓ satisfy (5.145) i.e.$\Delta(C_\ell, I, w, k, j+2) \le c$. We split each such piece by applying Proposition 5.6.11 with the same value of k, with $h = j$, $c = 2^{n/2} r^{-j-1}/L_2$, with $\sigma = 2^{n/2} r^{-j-1}$, with $w' = w_\ell$ any point of C_ℓ, and with now m the largest integer $m \le N_{n-1} - 1$ such that $r^{-j-1}\sqrt{\log m} \le \sigma$. Thus if n_0 is large enough we have $\sqrt{\log m} \ge 2^{n/2}/L$ and therefore $\sigma\sqrt{\log m} \ge 2^n r^{-j-1}/L$. Each of the resulting pieces satisfies either (5.147) or (5.149). The total number of pieces produced is $\le N_{n-1}^2 = N_n$. □

We are now ready to prove the basic partition result by iterating Proposition 5.7.3 in the spirit of the proof of Theorem 2.6.1. A remarkable new feature of this construction is that the functionals we use depend on the set we partition. We fix an integer $\kappa \ge 3$ with $2^{\kappa/2} \ge 2L_4$ and we set $r = 2^\kappa$ (so that r is now a universal constant ≥ 8). One feature of the construction is that when we produce a set satisfying (5.148) we do not wish to split it for a while. So, we assign to each set A a "counter" $p(A)$ that tells us how many steps ago this set was produced satisfying (5.148). If $(\mathcal{A}_n)$ is an increasing sequence of partitions, for $A \in \mathcal{A}_n$ and $n > 0$ we denote by A' the unique element of $\mathcal{A}_{n-1}$ which contains A.

Consider a set $T \subset \ell^2$ with $0 \in T$. By induction over $n \ge 0$ we construct an increasing sequence $(\mathcal{A}_n)$ of partitions of T, with $\operatorname{card} \mathcal{A}_n \le N_n$. For $A \in \mathcal{A}_n$ we construct a set $I_n(A) \subset \mathbb{N}^*$, a point $w_n(A) \in \ell^2$ and integers $k_n(A) \le j_n(A) \in \mathbb{Z}$, $0 \le p_n(A) \le 4\kappa - 1$, such that for each $n \ge 0$,

$$\forall A \in \mathcal{A}_n \; , \; p_n(A) = 0 \Rightarrow \Delta(A, I_n(A), w_n(A), k_n(A), j_n(A)) \le 2^{n/2} r^{-j_n(A)} . \qquad (5.150)$$

This condition is essential. When $n > 0$ and $p_n(A) > 0$, it is replaced by the following substitute:

$$\forall A \in \mathcal{A}_n\ ,\ p_n(A) > 0 \Rightarrow$$
$$\Delta(A, I_n(A), w_n(A), k_n(A), j_n(A)) \le 2^{(n-p_n(A))/2} r^{-j_n(A)+2}\ . \tag{5.151}$$

The next two conditions, for $n > 0$, are mild regularity conditions:

$$\forall A \in \mathcal{A}_n\ ,\ j_{n-1}(A') \le j_n(A) \le j_{n-1}(A') + 2 \tag{5.152}$$

$$\forall A \in \mathcal{A}_n\ ,\ p_n(A) = 0 \Rightarrow j_n(A) \le j_{n-1}(A') + 1\ . \tag{5.153}$$

The next property says that "k, I, w do not change unless $p_n(A) = 1$":

$$\forall A \in \mathcal{A}_n\ ,\ p_n(A) \ne 1 \Rightarrow k_n(A) = k_{n-1}(A')\ ;$$
$$I_n(A) = I_{n-1}(A')\ ;\ w_n(A) = w_{n-1}(A')\ . \tag{5.154}$$

For $n \ge 0$ let us define

$$F_n(A) := F(A, I_n(A), w_n(A), k_n(A), j_n(A))\ . \tag{5.155}$$

Then we also have, for $n \ge 1$,

$$p_n(A) = 1 \Rightarrow F_n(A) \le F_{n-1}(A') - \frac{1}{L} 2^n r^{-j_n(A)}\ , \tag{5.156}$$

$$p_n(A) = 1 \Rightarrow w_n(A) \in A' \tag{5.157}$$

and

$$p_n(A) = 1 \Rightarrow I_n(A) = \left\{ i \in I_{n-1}(A')\ ;\ |w_n(A)_i - w_{n-1}(A')_i| \le 2r^{-k_{n-1}(A')} \right\}\ . \tag{5.158}$$

Finally, the (all important) *last requirement* is as follows. If $n > n_0$ and if $p_n(A) = 0$, **either** we have $p_{n-1}(A') = 4\kappa - 1$ **or** $j_n(A) = j_{n-1}(A') + 1$ **or else**

$$D \subset A\ ,\ \Delta(D, I_n(A), w_n(A), k_n(A), j_n(A) + 2) \le \frac{1}{L_4} 2^{n/2} r^{-j_n(A)-1} \Rightarrow$$
$$F(D, I_n(A), w_n(A), k_n(A), j_n(A) + 2) \le$$
$$F(A, I_n(A), w_n(A), k_n(A), j_n(A) + 2) - \frac{1}{L} 2^n r^{-j_n(A)-1}\ . \tag{5.159}$$

A fundamental property of the previous construction is that for $n \ge 1$ we have

$$F_n(A) \le F_{n-1}(A')\ . \tag{5.160}$$

First, if $p_n(A) \ne 1$ this follows form (5.152), (5.154) and (5.132). Next, if $p_n(A) = 1$ this is a consequence of (5.156).

To start the construction, for $n \le n_0$ we set $\mathcal{A}_n = \{T\}$, $I_n(T) = \mathbb{N}^*$, $w_n(T) = 0$, $p_n(T) = 0$, and $k_n(T) = j_n(T) = j_0$, where j_0 satisfies $\Delta(T) \le r^{-j_0}$. Conditions (5.150) and (5.152) to (5.158) are satisfied, while the last requirement of the construction is automatically satisfied since $n \le n_0$.

Assuming that the construction has been done up to $\mathcal{A}_n$, we fix an element B of $\mathcal{A}_n$, and we proceed as follows.

First, if $1 \le p_n(B) < 4\kappa - 1$, we do not split B, i.e. we decide that $B \in \mathcal{A}_{n+1}$, and we simply set $p_{n+1}(B) = p_n(B)+1$, $I_{n+1}(B) = I_n(B), w_{n+1}(B) = w_n(B), k_{n+1}(B) = k_n(B), j_{n+1}(B) = j_n(B)$, and all our required conditions hold. In the case of (5.151) this is because $(n+1) - p_{n+1}(B) = n - p_n(B)$. And the last requirement is automatically satisfied since $p_{n+1}(B) \neq 0$.

Next, if $p_n(B) = 4\kappa - 1$, we do not split B either, but we set $p_{n+1}(B) = 0$, $I_{n+1}(B) = I_n(B)$, $w_{n+1}(B) = w_n(B)$, $j_{n+1}(B) = j_n(B)$, $k_{n+1}(B) = k_n(B)$. To prove that (5.150) holds for B and $n+1$, we write, using (5.151) for B and n in the second line,

$$\begin{aligned}
&\Delta(B, I_{n+1}(B), w_{n+1}(B), j_{n+1}(B), k_{n+1}(B)) \\
&= \Delta(B, I_n(B), w_n(B), j_n(B), k_n(B)) \\
&\le 2^{(n-4\kappa+1)/2} r^{-j_n(B)+2} = 2^{(n+1)/2} r^{-j_{n+1}(B)} ,
\end{aligned}$$

since $2^{-2\kappa} = r^{-2}$. It is then obvious that all our requirements are satisfied (the last one holds since $p_n(B') = p_n(B) = 4\kappa - 1$).

Finally, if $p_n(B) = 0$, we split B in at most N_n pieces using Lemma 5.7.3, with $I = I_n(B), w = w_n(B)$, $j = j_n(B)$ and $k = k_n(B)$. There are three cases to consider.

First, we are in case (a) of this lemma, the piece A produced by the lemma has property (5.146). We define $p_{n+1}(A) = 0$. We then set

$$I_{n+1}(A) = I_n(B),\ w_{n+1}(A) = w_n(B),\ j_{n+1}(A) = j_n(B),\ k_{n+1}(A) = k_n(B) . \tag{5.161}$$

All our conditions are satisfied. In the case of the last requirement this is because (5.146) implies (5.159) (for $n+1$ rather than n).

Second, we are in case (b) of the lemma, the piece $A = A_\ell$ produced has property (5.147). We set $p_{n+1}(A) = 1$, and we define

$$j_{n+1}(A) = j_n(B) + 2 = k_{n+1}(A) .$$

We define $w_{n+1}(A) = w_\ell \in B = A'$ and

$$I_{n+1}(A) = \{i \in I_n(B) \ ; \ |w_{n+1}(A)_i - w_n(B)_i| \le 2r^{-k_n(B)}\} .$$

It is then obvious that all the required conditions hold. Indeed (5.156) follows from (5.148), and the last requirement is automatically satisfied since $p_{n+1}(A) \neq 0$.

Finally, we are in case (c) of the lemma, and the piece we produce has property (5.149). We then set $p_{n+1}(A) = 0$, $j_{n+1}(A) = j_n(B) + 1$ and we define

$$I_{n+1}(A) = I_n(A) \ , \ w_{n+1}(A) = w_n(B) \ , \ k_{n+1}(A) = k_n(B) .$$

All our conditions are satisfied. In the case of the last requirement this is because $j_{n+1}(A) = j_n(B) + 1$. This finishes the construction.

Given $t \in T$ and $n \geq 0$, define then $j_n(t) := j_n(A_n(t))$. The fundamental property is as follows.

Proposition 5.7.4. *We have*

$$\forall t \in T \ , \ \sum_{n \geq 0} 2^n r^{-j_n(t)} \leq L(r^{-j_0} + b(T)) \ . \tag{5.162}$$

To prepare for the proof, let us fix $t \in T$ and define $j(n) := j_n(t) = j_n(A_n(t))$ and $a(n) = 2^n r^{-j(n)}$. Let us first observe that $\sum_{n \leq n_0} a(n) \leq L r^{-j_0}$ so that it suffices to bound $\sum_{n \geq n_0} a(n)$. Let us recall (5.155) and define

$$F(n) := F_n(A_n(t)) \geq 0 \ .$$

As a consequence of (5.160) the sequence $(F(n))_{n \geq 0}$ is non-increasing, and of course $F(0) \leq b(T)$.

Let us define

$$J_0 = \{n_0\} \cup \{n \geq n_0 \ ; \ j(n+1) > j(n)\} \ ,$$

which we enumerate as $J_0 = \{n_0, n_1, \ldots, \}$. We observe that $j(n) = j(n_k + 1)$ for $n_k + 1 \leq n \leq n_{k+1}$.

Let us further define

$$C^* := \left\{k \geq 0 \ ; \ \forall k' \geq 0 \ , \ a(n_k) \geq 2^{-|k-k'|} a(n_{k'})\right\} \ .$$

Once we know that the sequence $(a(n))$ is bounded, it will suffice, using Lemma 2.6.3 twice, to bound $\sum_{k \in C^*} a(n)$, observing first that $a(n+1) = 2a(n)$ for $n \notin J_0$ to obtain that $\sum_{n \geq n_0} a(n) \leq L \sum_{k \geq 0} a(n_k)$, and using the lemma again to obtain that $\sum_{k \geq 0} a(n_k) \leq L \sum_{k \in C^*} a(n_k)$.

Let us further define $p(n) := p_n(A_n(t))$. A good part of the argument is contained in the following fact.

Lemma 5.7.5. *Consider $k \in C^*$ with $k \geq 1$ and assume that*

$$n_k - 1 \leq n \leq n_{k+1} + 1 \Rightarrow p(n) = 0 \ . \tag{5.163}$$

Then

$$a(n_k) \leq L(F(n_k) - F(n_{k+2})) \ . \tag{5.164}$$

Proof. Let us first observe that by (5.153), for $n \in J_0$ we have $j(n+1) = j(n) + 1$ when $p(n+1) = 0$, so that by (5.163) we have

$$j(n_{k+1} + 1) = j(n_{k+1}) + 1 = j(n_k + 1) + 1 = j(n_k) + 2 \ . \tag{5.165}$$

We also observe that we cannot have $n_k - 1 \in J_0$ because then $n_k - 1 = n_{k-1}$ and $j(n_{k-1}) < j(n_k)$ so that $a(n_{k-1}) \geq r a(n_k)/2 > 2a(n_k)$ and the choice

$k' = k - 1$ violates the fact that $k \in C^*$. Therefore $j(n_k - 1) = j(n_k)$. A key point is that we have also $p(n_k - 1) = p(n_k) = 0$, by (5.163). The last requirement of the construction then shows that (5.159) holds true for $n = n_k$, which is the main ingredient of the proof.

Setting $n^* = n_{k+1} + 1$ we recall (5.165): $j(n^*) = j(n_k) + 2$. Now, by definition of C^*, we have $a(n_k) \geq a(n_{k+1})/2$ so that

$$2^{n_k} r^{-j(n_k)} \geq 2^{-1} \cdot 2^{n_{k+1}} r^{-j(n_{k+1})} ,$$

and thus, using again that $j(n_{k+1}) = j(n_k) + 1$,

$$2^{n_{k+1} - n_k} \leq 2r = 2^{\kappa+1}$$

and therefore $n_{k+1} - n_k \leq \kappa + 1$, so that $n^* \leq n_k + \kappa + 2$. Using that $2^{\kappa/2} \geq 2L_4$ and $r = 2^\kappa$, we get

$$2^{n^*/2} r^{-j(n^*)} \leq 2^{(2+\kappa)/2} 2^{n_k/2} r^{-j(n_k)-2} \leq \frac{1}{L_4} 2^{n_k/2} r^{-j(n_k)-1} . \tag{5.166}$$

Now, using (5.150) for n^* instead of n and $A = A_{n^*}(t)$ yields

$$\begin{aligned} &\Delta\big(A_{n^*}(t), I_{n^*}(A_{n^*}(t)), w_{n^*}(A_{n^*}(t)), k_{n^*}(A_{n^*}(t)), j_{n^*}(A_{n^*}(t))\big) \\ &\leq 2^{n^*/2} r^{-j(n^*)} \leq \frac{1}{L_4} 2^{n_k/2} r^{-j(n_k)-1} . \end{aligned} \tag{5.167}$$

Let us write $D = A_{n^*}(t)$ and $A = A_{n_k}(t)$. We recall that $j(n^*) = j(n_k) + 2$ i.e. $j_{n^*}(A_{n^*}(t)) = j_{n_k}(A_{n_k}(t)) + 2$ and we observe that from (5.154) we have

$$w_{n^*}(A_{n^*}(t)) = w_{n_k}(A) \; ; \; I_{n^*}(A_{n^*}(t)) = I_{n_k}(A) \; ; \; k_{n^*}(A_{n^*}(t)) = k_{n_k}(A) . \tag{5.168}$$

Therefore (5.167) yields

$$\Delta(D, I_{n_k}(A), w_{n_k}(A), k_{n_k}(A), j_{n_k}(A) + 2) \leq \frac{1}{L_4} 2^{n_k/2} r^{-j_{n_k}(A)-1} .$$

We can then use (5.159) for $n = n_k$. Using again (5.168) this implies

$$F(n^*) \leq F(n_k) - \frac{1}{L} 2^{n_k} r^{-j_{n_k}(t)} ,$$

i.e. $a(n_k) \leq L(F(n_k) - F(n^*)) \leq L(F(n_k) - F(n_{k+2}))$. □

Proof of Proposition 5.7.4. Let us examine the set of values of n for which $p(n) > 0$. By construction, if $p(n) = 1$ then $j(n) = j(n-1) + 2$ so that $n - 1 \in J_0$. Also by construction, if $n \in J_0$ we have $p(n) = 0$ and $p(n + 1) \in \{0, 1\}$. Moreover, if $p(n) = 1$, then for $n \leq n' \leq n + 4\kappa - 2$ we have $p(n') = n' - n + 1 = p(n' - 1) + 1$ and also $p(n + 4\kappa - 1) = 0$. In particular:

$$\text{The set } \{n \; ; \; p(n) > 0\} \text{ consists of disjoint intervals of cardinality } 4\kappa - 1, \text{ each of them starting to the right of a point of } J_0 . \tag{5.169}$$

Let us set

$$J_1 = \{n \geq n_0 \; ; \; p(n+1) = 1\} \; ,$$

so that for $n \geq n_0, n \in J_0 \setminus J_1$ we have $p(n+1) = 0$. Using (5.132) and the definitions, (5.156) implies that for $n \in J_1$,

$$a(n) \leq L(F(n) - F(n+1)) \; , \tag{5.170}$$

and thus

$$\sum_{n \in J_1} a(n) \leq L \sum_{n \geq n_0} (F(n) - F(n+1)) \leq LF(n_0) \leq Lb(T) \; . \tag{5.171}$$

In particular, if $C_1 := \{k \geq 0; n_k \in J_1\}$ we have

$$\sum_{k \in C_1} a(n_k) \leq Lb(T) \; . \tag{5.172}$$

Let us now define $C_2 := \{k \geq 0; n_{k+1} \in J_1\}$. Then (5.152) implies

$$a(n_{k+1}) = 2^{n_{k+1}} r^{-j(n_{k+1})} \geq 2^{n_k} r^{-j(n_k)-2} = a(n_k) r^{-2}$$

and therefore (5.172) implies

$$\sum_{k \in C_2} a(n_k) \leq Lb(T) \; . \tag{5.173}$$

Next let us define $C_3 = \{k \geq 0; p(n_k - 1) > 0\}$. Since $p(n_k) = 0$ we then have $p(n_k - 1) = 4\kappa - 1$. Recalling that by construction $p(n+1) = p(n) + 1$ when $1 \leq p(n) \leq 4\kappa - 2$, for $k \in C_3$ we have $p(n_k - 4\kappa + 1) = 1$ i.e. $n_k - 4\kappa \in J_1$. Moreover since obviously $a(n+1) \leq 2a(n)$ it follows from (5.171) that

$$\sum_{k \in C_3} a(n_k) \leq Lb(T) \; . \tag{5.174}$$

Finally let $C_4 = C^* \setminus (C_1 \cup C_2 \cup C_3)$. Thus for $k \in C_4$ we have $p(n_k - 1) = p(n_k + 1) = p(n_{k+1} + 1) = 0$. Now, the structure of the set $\{n; p(n) > 0\}$ that we described in (5.169) implies that (5.163) holds. Therefore (and taking into account the case $k = 0$ for which Lemma 5.7.5 does not apply) (5.164) implies that $\sum_{k \in C_4} a(n_k) \leq L(r^{-j_0} + b(T))$. Combining with (5.172) to (5.174) this show that $\sum_{k \in C^*} a(n_k) \leq L(r^{-j_0} + b(T))$.

It remains only to prove that the use of Lemma 2.6.3 is legitimate by proving that $a(n) \leq L(r^{-j_0} + b(T))$ for $n > n_0$. This is done by a much simplified version of the previous arguments. We observe that by (5.170) that this is the case if either $p(n-1) > 0$ or $p(n) > 0$ (since then J_1 contains a point which is not much less than n). Since $r > 4$ we have $a(n-1) > 2a(n)$ when $j(n) > j(n-1)$. Since $a(n_0) \leq Lr^{-j_0}$ it thus it suffices to consider the case where $j(n-1) = j(n)$ and $p(n-1) = p(n) = 0$. But then the last requirement of the construction shows that n satisfies (5.159). In particular taking there D reduced to a single point shows that $a(n) \leq LF(n_0) \leq Lb(T)$. □

Proof of Theorem 5.1.5. First we use Lemma 5.3.1 to find j_0 such that $r^{-j_0} \le Lb(T)$ while $|t_i| < r^{-j_0}/2$ for $t \in T$ and $i \in \mathbb{N}$. We then perform the above decomposition with this value of j_0, and we obtain $\sup_{t\in T} \sum_{n\ge 0} 2^n r^{-j_n(t)} \le Lb(T)$. The plan is to use (5.162) together with Theorem 5.2.6 when $\Omega = \mathbb{N}$ and μ is the counting measure. The choice of j_0 implies that $T_3 = \{0\}$. We define the elements $\pi_n(A)$ for $A \in A_n$ recursively as follows. We choose $\pi_0(T) = 0$. When $A \in \mathcal{A}_{n+1}$ and when $j_{n+1}(A) = j_n(A')$ we simply take $\pi_{n+1}(A) = \pi_n(A')$. When $p_{n+1}(A) = 0$ and $j_{n+1}(A) = j_n(A') + 1$ we take for $\pi_{n+1}(A)$ any point of A. Finally, when $p_{n+1}(A) = 1$ we take for $\pi_{n+1}(A) = w_{n+1}(A) \in A'$ using (5.157). Thus (5.23) holds, while (5.20) to (5.22) should be obvious by construction. Let us consider the set

$$J_n(t) = \left\{ i \in \mathbb{N}^* \ ; \ \forall q < n \ , \ |\pi_q(t)_i - \pi_{q+1}(t)_i| \le r^{-j_q(t)} \right\} .$$

To apply Theorem 5.2.6 it suffices to prove (5.25) for $u = L$ i.e. that

$$\forall t \in T \ , \ \forall n \ge 0 \ , \ \sum_{i\in J_n(t)} (t_i - \pi_n(t)_i)^2 \wedge r^{-2j_n(t)} \le L2^n r^{-2j_n(t)} \ . \tag{5.175}$$

Let us define $k_n(t) := k_n(A_n(t))$ and $w_n(t) := w_n(A_n(t))$ and prove the inequality

$$i \in J_{n+1}(t) \Rightarrow |\pi_{n+1}(t)_i - w_n(t)_i| \le 2r^{-k_n(t)} \ . \tag{5.176}$$

To see this, let $J' = \{0\} \cup \{n; p_n(t) = 1\}$. Then by construction we have $\pi_n(t) = w_n(t)$ for $n \in J'$. Given n let us consider the largest $n' \in J'$ with $n' \le n$. Then by (5.154) we have $w_n(t) = w_{n'}(t) = \pi_{n'}(t)$ and $k_n(t) = k_{n'}(t)$, while for $i \in J_{n+1}(t)$ we have

$$|\pi_{n+1}(t)_i - \pi_{n'}(t)_i| \le \sum_{j \ge j_{n'}(t)} r^{-j} \le 2r^{-j_{n'}(t)} \le 2r^{-k_{n'}(t)} = 2r^{-k_n(t)} \ . \tag{5.177}$$

Thus we have proved (5.176). Next we prove that $J_n(t) \subset I_n(t) := I_n(A_n(t))$ by induction over n. For the induction from n to $n+1$, the result is obvious when $p_{n+1}(t) \ne 1$, for then $I_{n+1}(t) = I_n(t) \supset J_n(t) \supset J_{n+1}(t)$. On the other hand, when $p_{n+1}(t) = 1$ then by construction $\pi_{n+1}(t) = w_{n+1}(t)$ and (5.176) implies that for $i \in J_{n+1}(t)$ we have $|w_{n+1}(t)_i - w_n(t)_i| \le 2r^{-k_n(t)}$. Thus (5.158) concludes the proof that $J_{n+1}(t) \subset I_{n+1}(t)$.

Now, the proof of (5.177) shows that $|\pi_n(t)_i - \pi_{n'}(t)_i| \le 2r^{-k_n(t)}$ for $i \in J_n(t)$. Therefore the proof of (5.176) shows that we have also

$$i \in J_n(t) \Rightarrow |\pi_n(t)_i - w_n(t)_i| \le 2r^{-k_n(t)} \ .$$

Hence since $J_n(t) \subset I_n(t)$, (5.175) follows from (5.150) and (5.134). Thus we have proved (5.25) and we can indeed apply Theorem 5.2.6 to obtain the required decomposition. □

5.8 Notes and Comments

An important result about empirical processes is Ossiander's bracketing theorem that we shall prove in Section 9.1. This theorem was proved somewhat later than Dudley's theorem because it requires a genuinely new idea. We feel somehow that Theorem 5.2.1 succeeds in carrying out the idea of Ossiander's theorem to a general setting, and this might explain why it is successful. Not surprisingly, it will be very easy to deduce Ossiander's theorem from Theorem 5.2.1, see page 280. It is an interesting story that the author proved Theorem 5.2.7 (in an essentially equivalent form) as early as [7], and in the exact form presented here in [9], but did not understand then its potential as a chaining theorem. The version of this work at the time the author received [1] contained only Theorem 5.2.7, with a proof very similar to the proof of Theorem 5.2.6 which we present here.

There exists a rather different proof of Proposition 5.3.5 which is given in [6]. Probably the proof of [6] is more elegant and deeper than the proof we give here, but the latter has the extra advantage to show the connection between Proposition 5.3.5 and the Marcus-Pisier theorem, Theorem 3.2.12.

The present paragraph assumes that the reader is familiar with the material of the next chapter. In the case of a metric space (T,d), one knows how to identify simple structures (trees), the presence of which provides a lower bound on $\gamma_2(T,d)$. One then can dream of identifying geometric structures inside a set $T \subset \ell^2$, which would provide lower bounds for $b(T)$ of the correct order. Maybe this is a dream which is impossible to achieve. Not the least remarkable feature of the Bednorz-Latała proof of Theorem 5.1.5 is that it completely bypasses this problem.

References

1. Bednorz, W., Latała, R.: On the suprema of Bernoulli processes. C. R. Math. Acad. Sci. Paris **351**(3–4), 131–134 (2013)
2. Latała, R.: On the boundedness of Bernoulli processes on thin sets. Electron. Commun. Probab. **13**, 175–186 (2008)
3. Ledoux, M.: The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001). x+181 pp. ISBN 0-8218-2864-9
4. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
5. Talagrand, M.: Regularity of infinitely divisible processes. Ann. Probab. **21**, 362–432 (1993)
6. Talagrand, M.: The supremum of certain canonical processes. Am. J. Math. **116**, 283–325 (1994)
7. Talagrand, M.: Constructions of majorizing measures Bernoulli processes and cotype. Geom. Funct. Anal. **4**, 660–717 (1994)

8. Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. IHÉS **81**, 73–205 (1995)
9. Talagrand, M.: The Generic Chaining. Springer Monographs in Mathematics. Springer, Berlin (2005). viii+222 pp. ISBN: 3-540-24518-9

6. Trees and the Art of Lower Bounds

6.1 Introduction

The concept of tree presented in Section 6.2 is historically important in the type of results presented in this work. The author discovered many of the results he presents while thinking in terms of trees. One knows now how to present these results and their proofs without ever mentioning trees, and arguably in a more elegant fashion, so that trees are not used explicitly elsewhere in this book. However it might be too early to dismiss this concept, at least as an instrument of discovery. Let the reader judge by herself!

In Section 6.3 we present a lower bound for $\gamma_2(\mathcal{E})$ for certain ellipsoids $\mathcal{E}$, without using Proposition 2.5.1, but rather some simple combinatorics. This "exercise" is a preparation to the more delicate methods by which we prove in the rest of the chapter that the upper bounds on matchings of Chapter 5 cannot be improved.

6.2 Trees

In this section we describe different ways to measure the size of a metric space. We shall show that they are all equivalent to the functional $\gamma_2(T,d)$. It is possible to consider more general notions corresponding to other functionals considered in the book, but for simplicity we consider only the case of γ_2.

In a nutshell, a tree is a certain structure that requires a "lot of space" to be constructed, so that a metric space that contains large trees needs itself to be large. At the simplest level, it already takes some space to construct in a set A sets $B_1, \ldots, B_n$ which are appropriately separated from each other. This is even more so if the sets $B_1, \ldots, B_n$ are themselves large (for example because they themselves contain many sets far from each other). Trees are a proper formulation of the iteration of this idea. The basic use of trees is to measure the size of a metric space by the size of the largest tree (of a certain type) which it contains. Different types of trees yield different measures of size.

A *tree* $\mathcal{T}$ of a metric space (T,d) is a *finite* collection of subsets of T with the following two properties.

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_6,

$$\text{Given } A, B \text{ in } \mathcal{T}\,, \text{ if } A \cap B \neq \emptyset\,, \text{ then either } A \subset B \text{ or else } B \subset A\,. \tag{6.1}$$

$$\mathcal{T} \text{ has a largest element}\,. \tag{6.2}$$

The important condition here is (6.1), and (6.2) is just for convenience.

If $A, B \in \mathcal{T}$ and $B \subset A$, $B \neq A$, we say that B is a *child* of A if

$$C \in \mathcal{T}\,,\, B \subset C \subset A \Rightarrow C = B \text{ or } C = A\,. \tag{6.3}$$

We denote by $c(A)$ the number of children of A. Since our trees are finite, some of their sets will have no children. It is convenient to "shrink these sets to a single point", so we will consider only trees with the following property

$$\text{If } A \in \mathcal{T} \text{ and } c(A) = 0\,, \text{ then } A \text{ contains exactly one point}\,. \tag{6.4}$$

A fundamental property of trees is as follows. Consider trees $\mathcal{T}_1, \ldots, \mathcal{T}_m$ and for $1 \leq \ell \leq m$ let A_ℓ be the largest element of $\mathcal{T}_\ell$. Assume that the sets A_ℓ are disjoint, and consider a set $A \supset \bigcup_{\ell \leq m} A_\ell$. Then the collection of subsets of T consisting of A and of $\bigcup_{\ell \leq m} \mathcal{T}_\ell$ is a tree. The proof is straightforward. This fact allows one to construct iteratively more and more complicated (and larger) trees.

An important structure in a tree is a *branch*. A sequence $A_0, A_1, \ldots, A_k$ is a branch if $A_{\ell+1}$ is a child of A_ℓ, and if moreover A_0 is the largest element of $\mathcal{T}$ while A_k has no child. Then by (6.4) the set A_k is reduced to a single point t, and $A_0, \ldots, A_k$ are exactly those elements of $\mathcal{T}$ which contain t. So in order to describe the branches of $\mathcal{T}$ it is convenient to introduce the set

$$S_{\mathcal{T}} = \{t \in T\;;\; \{t\} \in \mathcal{T}\}\,, \tag{6.5}$$

which we call the "support" of $\mathcal{T}$. Thus by considering all the sets $\{A \in \mathcal{T}; t \in A\}$ as t varies in $S_{\mathcal{T}}$ we obtain all the branches of $\mathcal{T}$.

We now quantify our desired property that the children of a given set should be far from each other in an appropriate sense. A *separated* tree is a tree $\mathcal{T}$ such that to each A in $\mathcal{T}$ with $c(A) \geq 1$ is associated an integer $s(A) \in \mathbb{Z}$ with the following properties. First,

$$\text{If } B_1 \text{ and } B_2 \text{ are distinct children of } A, \text{then } d(B_1, B_2) \geq 4^{-s(A)}\,. \tag{6.6}$$

Here of course $d(B_1, B_2) = \inf\{d(x_1, x_2); x_1 \in B_1, x_2 \in B_2\}$. We observe that in (6.6) we make no restriction on the diameter of the children of A. (Such restrictions will however occur in the other notion of tree that we consider later.) Second, we will also make the following purely technical assumption:

$$\text{If } B \text{ is a child of } A\,, \text{ then } s(B) > s(A)\,. \tag{6.7}$$

Although this is not obvious now, the meaning of this condition is that $\mathcal{T}$ contains no sets which are obviously irrelevant for the measure of its size.

To measure the size of a separated tree T we introduce its *depth*, i.e.

$$d(\mathcal{T}) := \inf_{t \in S_{\mathcal{T}}} \sum_{t \in A \in \mathcal{T}} 4^{-s(A)} \sqrt{\log c(A)}\ . \tag{6.8}$$

Here and below we make the convention that the summation does not include the term $A = \{t\}$ (for which $c(A) = 0$). We observe that in (6.8) we have the *infimum* over $t \in S_{\mathcal{T}}$. In words a tree is large if it is large along *every* branch. We can then measure the size of T by

$$\sup\{d(\mathcal{T})\ ;\ \mathcal{T} \text{ separated tree} \subset T\}\ . \tag{6.9}$$

The notion of separated tree we just considered is but one of many possible notions of trees. As it turns out, this notion of separated tree does not seem fundamental. Rather, the quantity (6.9) is used as a convenient intermediate technical step to prove the equivalence of several more important quantities. Let us now consider now another notion of trees, which is more restrictive (and apparently much more important). An *organized* tree is a tree $\mathcal{T}$ such that to each $A \in \mathcal{T}$ with $c(A) \geq 1$ are associated an integer $j = j(A) \in \mathbb{Z}$, a point $t \in T$ and points $t_1, \ldots, t_{c(A)} \in B(t, 4^{-j})$ with the properties that

$$1 \leq \ell < \ell' \leq c(A) \Rightarrow d(t_\ell, t_{\ell'}) \geq 4^{-j-1}$$

and that each ball $B(t_\ell, 4^{-j-2})$ contains exactly one child of A. This should be compared to Definition 2.3.8 for $r = 4$.

If B_1 and B_2 are distinct children of A in an organized tree, then

$$d(B_1, B_2) \geq 4^{-j(A)-2}\,, \tag{6.10}$$

so that an organized tree is also a separated tree, with $s(A) = j(A) + 2$, but the notion of organized tree is more restrictive. (For example we have no control over the diameter of the children of A in a separated tree.)

We define the depth $d'(\mathcal{T})$ of an organized tree by

$$d'(\mathcal{T}) = \inf_{t \in S_{\mathcal{T}}} \sum_{t \in A \in \mathcal{T}} 4^{-j(A)} \sqrt{\log c(A)}\ .$$

Another way to measure the size of T is then

$$\sup\{d'(\mathcal{T})\ ;\ \mathcal{T} \text{ organized tree} \subset T\}\ . \tag{6.11}$$

If we simply view an organized tree $\mathcal{T}$ as a separated tree using (6.10), then $d(\mathcal{T}) = d'(\mathcal{T})/16$ (where $d(\mathcal{T})$ is the depth of $\mathcal{T}$ as a separated tree). Thus we have shown the following.

Proposition 6.2.1. *We have*

$$\sup\{d'(\mathcal{T})\ ;\ \mathcal{T} \text{ organized tree}\} \leq 16 \sup\{d(\mathcal{T})\ ;\ \mathcal{T} \text{ separated tree}\}\ . \tag{6.12}$$

The next result provides the fundamental connection between trees and the functional γ_2.

Proposition 6.2.2. *We have*

$$\gamma_2(T,d) \le L \sup\{d'(\mathcal{T})\;;\;\mathcal{T} \text{ organized tree}\}\;. \tag{6.13}$$

Proof. We consider the functional

$$F_n(A) = F(A) = \sup\{d'(\mathcal{T})\;;\;\mathcal{T} \subset A\,,\,\mathcal{T} \text{ organized tree}\}\,,$$

where we write $\mathcal{T} \subset A$ as a shorthand for "$\forall B \in \mathcal{T}\,,\, B \subset A$".

The main part of the argument is to prove that the growth condition (2.147) holds true when $r = 4$, $\theta(n) = 2^{n/2-2}$, $\beta = 1$, $\tau = 1$, and a is of the type r^{-j-1}. For this consider $n \ge 0$ and $m = N_{n+1}$. Consider $j \in \mathbb{Z}\,,\, t \in T$ and $t_1, \cdots, t_m \in B(t, 4^{-j})$ with

$$1 \le \ell < \ell' \le m \Rightarrow d(t_\ell\,,\,t_{\ell'}) \ge 4^{-j-1}\;.$$

Consider sets $H_\ell \subset B(t_\ell, 4^{-j-2})$ and $c < \min_{\ell \le m} F(H_\ell)$. Consider, for $\ell \le m$ a tree $\mathcal{T}_\ell \subset H_\ell$ with $d'(\mathcal{T}_\ell) > c$ and denote by A_ℓ its largest element. Then it should be obvious that the tree $\mathcal{T}$ consisting of $C = \bigcup_{\ell \le m} H_\ell$ (its largest element) and the union of the trees $\mathcal{T}_\ell\,,\,\ell \le m$, is organized (with $j(C) = j$, and $A_1, \ldots, A_m$ as children of C). Moreover $S_{\mathcal{T}} = \bigcup_{\ell \le m} S_{\mathcal{T}_\ell}$.

Consider $t \in S_{\mathcal{T}}$, and let ℓ with $t \in S_{\mathcal{T}_\ell}$. Then

$$\begin{aligned}\sum_{t \in A \in \mathcal{T}} 4^{-j(A)}\sqrt{\log c(A)} &= 4^{-j}\sqrt{\log m} + \sum_{t \in A \in \mathcal{T}_\ell} 4^{-j(A)}\sqrt{\log c(A)} \\ &\ge 4^{-j}\sqrt{\log m} + d'(\mathcal{T}_\ell) \ge 4^{-j}\sqrt{\log m} + c\;.\end{aligned}$$

Since $\sqrt{\log m} \ge 2^{n/2}$, this proves the growth condition (2.147).

In the course of the proof of Theorem 2.7.2 we have noted that this theorem holds true as soon as the growth condition (2.147) holds true when a is of the type r^{-j-1}, and we have just proved that this is the case (for $r = 4$, $\theta(n) = 2^{n/2-2}$, $\beta = 1$ and $\tau = 1$). To prove (6.13) we then apply Lemma 2.3.5 and Theorem 2.7.2. To control the diameter of T, we simply note that if $s, t \in T$, and j is the largest integer with $4^{-j} \ge d(s,t)$, then the tree $\mathcal{T}$ consisting of $T, \{t\}, \{s\}$, is organized with $j = j(T)$ and $c(T) = 2$, so $d'(\mathcal{T}) \ge 4^{-j}\sqrt{\log 2}$. □

For a probability measure μ on a metric space (T,d), with countable support, we define for each $t \in T$ the quantity

$$I_\mu(t) = \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon = \int_0^{\Delta(T)} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon\;.$$

The equality follows from the fact that $\mu(B(t,\epsilon)) = 1$ when $B(t,\epsilon) = T$, so that then the integrand is 0.

Proposition 6.2.3. *Given a metric space (T,d) we can find on T a probability measure μ, supported by a countable subset of T, and such that*

$$\sup_{t\in T} I_\mu(t) = \sup_{t\in T}\int_0^\infty \sqrt{\log\frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon \le L\gamma_2(T,d)\ . \tag{6.14}$$

A probability measure μ on (T,d) such that the left-hand side of (6.14) is usefully small is called a majorizing measure. The idea of this somewhat unsatisfactory name is that such a measure can be used to "majorize" the processes on T. The (in)famous theory of majorizing measures used the quantity

$$\inf_\mu \sup_{t\in T} I_\mu(t) \tag{6.15}$$

as a measure of the size of the metric space (T,d), where the infimum is over all choices of the probability measure μ. Even though this method is in the end equivalent to the use of the functional γ_2, its use is technically quite more challenging, so there seems to be no longer any use for this method in the present context. In other contexts majorizing measures remain useful, and we shall consider integrals such as the left-hand side of (6.14) (but with different functions of $\mu(B(t,\epsilon))$ e.g. in (13.155).

Proof. Consider an admissible sequence $(\mathcal{A}_n)$ with

$$\forall t\in T\,,\ \sum_{n\ge 0} 2^{n/2}\Delta(A_n(t)) \le 2\gamma_2(T,d)\ .$$

Let us now pick a point $t_{n,A}$ in each set $A\in\mathcal{A}_n$, for each $n\ge 0$. Since $\operatorname{card}\mathcal{A}_n\le N_n$, there is a probability measure μ on T, supported by a countable set, and satisfying $\mu(\{t_{n,A}\})\ge 1/(2^nN_n)$ for each $n\ge 0$ and each $A\in\mathcal{A}_n$. Then,

$$\forall n\ge 1\,,\ \forall A\in\mathcal{A}_n\,,\ \mu(A)\ge\mu(\{t_{n,A}\})\ge\frac{1}{2^nN_n}\ge\frac{1}{N_n^2}$$

so that given $t\in T$ and $n\ge 1$,

$$\begin{aligned}\epsilon>\Delta(A_n(t)) &\Rightarrow \mu(B(t,\epsilon))\ge\frac{1}{N_n^2}\\ &\Rightarrow \sqrt{\log\frac{1}{\mu(B(t,\epsilon))}}\le 2^{n/2+1}\ .\end{aligned} \tag{6.16}$$

Now, since μ is a probability, $\mu(B(t,\epsilon)) = 1$ for $\epsilon>\Delta(T)$, and then $\log(1/\mu(B(t,\epsilon))) = 0$. Thus

$$I_\mu(t) = \int_0^\infty \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon = \sum_{n\geq 1} \int_{\Delta(A_n(t))}^{\Delta(A_{n-1}(t))} \sqrt{\log \frac{1}{\mu(B(t,\epsilon))}}\mathrm{d}\epsilon$$
$$\leq \sum_{n\geq 1} 2^{n/2+1}\Delta(A_{n-1}(t)) \leq L\gamma_2(T,d)$$

using (6.16). □

Proposition 6.2.4. *If μ is a probability measure on T (supported by a countable set) and $\mathcal{T}$ is a separated tree on T, then*

$$d(\mathcal{T}) \leq L \sup_{t\in T} I_\mu(t) .$$

This completes the proof that the four "measures of the size of T" considered in this section, namely (6.9), (6.11), (6.15) and $\gamma_2(T,d)$ are indeed equivalent.

Proof. The basic observation is as follows. The sets

$$B(C, 4^{-s(A)-1}) = \{x \in T\ ;\ d(x,C) < 4^{-s(A)-1}\}$$

are disjoint as C varies over the children of A (as follows from (6.6)), so that one of them has measure $\leq c(A)^{-1}$.

We then proceed in the following manner, constructing recursively an appropriate branch of the tree. This is a typical and fundamental way to proceed when working with trees. We start with the largest element A_0 of $\mathcal{T}$. We then select a child A_1 of A_0 with $\mu(B(A_1, 4^{-s(A_0)-1})) \leq 1/c(A_0)$, and a child A_2 of A_1 with $\mu(B(A_2, 4^{-s(A_1)-1})) \leq 1/c(A_1)$, etc., and continue this construction as long as we can. It ends only when we reach a set of $\mathcal{T}$ that has no child, and hence by (6.4) is reduced to a single point t which we now fix. For any set A with $t \in A \in \mathcal{T}$, by construction we have

$$\mu(B(t, 4^{-s(A)-1})) \leq \frac{1}{c(A)}$$

so that

$$4^{-s(A)-2}\sqrt{\log c(A)} \leq \int_{4^{-s(A)-2}}^{4^{-s(A)-1}} \sqrt{\frac{1}{\log \mu(B(t,\epsilon))}}\mathrm{d}\epsilon . \tag{6.17}$$

By (6.7) the intervals $]4^{-s(A)-2}, 4^{-s(A)-1}[$ are disjoint for different sets A with $t \in A \in \mathcal{T}$, so summation of the inequalities (6.17) yields

$$\frac{1}{16}d(\mathcal{T}) \leq \sum_{t\in A\in\mathcal{T}} 4^{-s(A)-2}\sqrt{\log c(A)} \leq \int_0^\infty \sqrt{\frac{1}{\log \mu(B(t,\epsilon))}}\mathrm{d}\epsilon = I_\mu(t) . \quad \square$$

In the rest of this chapter, we will implicitly use the previous method of "selecting recursively the branch of the tree we follow" to prove lower bounds without mentioning trees.

The following two exercises provide two more "measures of size" which are equivalent to the four measures of size considered above.

Exercise 6.2.5 ([1]). For a metric space (T, d), define

$$\delta_2(T,d) = \sup_{\mu} \inf_{t \in T} I_\mu(t) \ ,$$

where the supremum is taken over all probability measures μ on T. (The reader observes of course that the infimum and the supremum are not as in (6.15).) Prove that $\delta_2(T,d)$ is equivalent to $\gamma_2(T,d)$. (Hint: To prove that $\gamma_2(T,d) \le L\delta_2(T,d)$ one proves that the functionals $F_n(A) := \delta_2(A,d)$ satisfy the growth condition of Definition 2.3.10. For this, given probability measures μ_ℓ on each of the pieces H_ℓ, one uses the probability measure $\mu = m^{-1}\sum_{\ell \le m} H_\ell$. To prove that $\delta_2(T,d) \le L\gamma_2(T,d)$, given a probability measure μ on T and an admissible sequence $(\mathcal{A}_n)$ of partitions of T, one recursively constructs a decreasing sequence $A_n \in \mathcal{A}_n$ such that $\mu(A_n)$ is as large as possible, and one estimates from above $I_\mu(t)$ where $A_n = A_n(t)$.)

Exercise 6.2.6. For a metric space (T, d) define

$$\chi_2(T,d) = \sup_{\mu} \inf \int \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \mathrm{d}\mu(t) \ ,$$

where the infimum is taken over all admissible sequences and the supremum over all probability measures. Prove that this measure of size is equivalent to $\gamma_2(T,d)$. It is obvious that $\chi_2(T,d) \le \gamma_2(T,d)$, but the converse is far from trivial. (Hint: use the functionals $F_n(A) = \inf_\mu \int \sum_{k \ge n} 2^{k/2} \Delta(A_k(t)) \mathrm{d}\mu(t)$, where the infimum is over all admissible sequences, and the supremum is over all probability measures supported by A. Prove that these functionals satisfy the growth condition of (2.7.1) for $\tau = 3, \beta = 1, r = 4$ and $\theta(n) = 2^{n/2}/L$. Given the probability measures μ_ℓ on the pieces H_ℓ use the probability measure $\mu = m^{-1}\sum_{\ell \le N_m} \mu_\ell$ on their union. The key point is to prove that for any admissible sequence one has $\int \Delta(A_n(t)) \mathrm{d}\mu(t) \ge a/L$. If this is too difficult, came back to this exercise after you study the proof of Lemma 8.1.7 below.)

Exercise 6.2.7. For a tree $\mathcal{T}$, recall the definition (6.5) of $S_\mathcal{T}$. Prove that there is unique "canonical" probability measure μ on $S_\mathcal{T}$, defined by the property that all the children of a given set $A \in \mathcal{T}$ have the same probability. Prove that for each admissible sequence $(\mathcal{A}_n)$ of T one has

$$\int \sum_{n \ge 0} 2^{n/2} \Delta(A_n(t)) \mathrm{d}\mu(t) \ge \frac{1}{L} d(\mathcal{T}) \ .$$

(You might find this challenging.) Explain why this provides another proof of the result of the previous exercise. Prove that this same probability measure provides a useful lower bound on the quantity $\delta_2(T,d)$ of Exercise 6.2.5.

6.3 A Toy Lower Bound

Consider $N \geq 1$ and the ellipsoid

$$\mathcal{E} = \Big\{(x_i)_{1\leq i\leq N}\ ;\ \sum_{i\leq N} i x_i^2 \leq 1\Big\} \subset \mathbb{R}^N . \tag{6.18}$$

Combining Proposition 2.5.1 with the generic chaining bound (2.45) we obtain

$$\gamma_2(\mathcal{E}) \geq \frac{1}{L}\sqrt{\log N}\ , \tag{6.19}$$

where of course $\gamma_2(\mathcal{E}) = \gamma_2(\mathcal{E}, d)$ where d is the Euclidean distance. This fact is closely connected to the lower bounds we shall prove in the next two sections. Here we have obtained it using the magic of Gaussian processes. This will not be available in the next two sections, and we will have to use geometry. To prepare for these forthcoming proofs we shall now give a direct geometric proof of (6.19), by explicitly constructing a separated tree $\mathcal{T}$ in $\mathcal{E}$ for which $d(\mathcal{T}) \geq \sqrt{\log N}/L$.

The following classical result is used in many constructions.

Lemma 6.3.1. *For each integer n the set $\{-1,1\}^n$ contains a subset V with* $\operatorname{card} V \geq \exp(n/8)$ *such that any two distinct elements of V differ in at least $n/4$ coordinates.*

Proof. This is a standard counting argument. Consider the uniform probability P on $\{-1,1\}^n$, so that for P the coordinates functions ε_i are i.i.d. Bernoulli r.v.s. Then, using the subgaussian inequality (3.2.2) we obtain

$$\mathsf{P}\Big(\sum_{i\leq n}(1-\varepsilon_i) \leq \frac{n}{2}\Big) = \mathsf{P}\Big(\sum_{i\leq n}\varepsilon_i \geq \frac{n}{2}\Big) \leq \exp\Big(-\Big(\frac{n}{2}\Big)^2\frac{1}{2n}\Big) = \exp\Big(-\frac{n}{8}\Big) .$$

Consequently the proportion of elements of $\{-1,1\}^n$ that differ in at most $n/4$ coordinates from a given element is $\leq \exp(-n/8)$. In terms of the Hamming distance (i.e. the proportion of coordinates where two elements differ), this means that balls of radius $1/4$ have probability $\leq \exp(-n/8)$. A maximal subset of points at mutual distances $\geq 1/4$ must then be of cardinality $\geq \exp(n/8)$ since the balls of radius $1/4$ centered at these points cover $\{-1,1\}^n$. □

Let us fix an integer c. It will be chosen later, but we should think of it as a universal constant. (We will end up by taking $c = 6$.) Consider also the largest integer r which is a power of 16 and for which $2^{c(r+1)} \le N$ (so that $r \simeq \log N$ for N large enough, the only case we have to consider). Let us denote as usual by e_i the canonical basis of $\mathbb{R}^N$. For $1 \le k \le r$ we use Lemma 6.3.1 with $n = 2^{ck}$ to construct a set W_k of vectors of the type

$$w = 2^{-ck} \sum_{2^{ck} \le i < 2^{ck+1}} z_i e_i \tag{6.20}$$

with $z_i = \pm 1$,

$$\operatorname{card} W_k \ge \exp(2^{ck-3}) \ , \tag{6.21}$$

and (since $|z_i - z'_i| \in \{0, 2\}$)

$$w, w' \in W_k \ , \ w \ne w' \Rightarrow d(w, w') = \|w - w'\|_2 \ge 2^{-ck}\sqrt{n} = 2^{-ck/2} \ . \tag{6.22}$$

The elements of W_k have a non-zero i-th coordinate only for i in the interval $[2^{ck}, 2^{ck+1}[$. The purpose of the parameter c is to ensure that these intervals are sufficiently disjoint from each other so that consecutive stages of the construction do not interfere too much with each other.

For $x \in \mathbb{R}^N$ let us define $\|x\|_{\mathcal{E}}$ by $\|x\|_{\mathcal{E}}^2 := \sum_{i \le N} i x_i^2$. It is then straightforward from (6.20) that

$$w \in W_k \Rightarrow \|w\|_2 \le 2^{-ck/2} \ ; \ \|w\|_{\mathcal{E}} \le 2 \ . \tag{6.23}$$

Consider then the functions of the type

$$f_q = \frac{1}{2\sqrt{r}} \sum_{k \le q} w_k \ , \tag{6.24}$$

where q is an integer $\le r$ and $w_k \in W_k$. We observe from the second part of (6.23) that $f_q \in \mathcal{E}$. We recall that $B(f, a)$ denotes a ball for the ℓ^2 norm. We claim (if c is properly chosen) that the collection $\mathcal{T}$ of sets of the following type:

- the set $\mathcal{E}$,
- the sets $B_q(f_q) := B(f_q, \frac{1}{\sqrt{r}} 2^{-c(q+1)/2}) \cap \mathcal{E}$ for all values of $1 \le q \le r - 1$ and of f_q,
- the sets $\{f_r\}$ for all possible choices of f_r as in (6.24),

is the tree we are looking for. The purpose of the unimportant third class of sets is just to satisfy (6.4).

Given f_q, for $w \in W_{q+1}$ let us define

$$f^w := f_q + \frac{1}{2\sqrt{r}} w \ .$$

Since $\|w\|_2 \le 2^{-c(q+1)/2}$, it is straightforward (if $c \ge 2$) that

$$B\Big(f^w, \frac{1}{\sqrt{r}} 2^{-c(q+2)/2}\Big) \subset B\Big(f_q, \frac{1}{\sqrt{r}} 2^{-c(q+1)/2}\Big) . \tag{6.25}$$

Consider $w' \in W_{q+1}$, $w' \neq w$. By (6.22) we have

$$d(f^w, f^{w'}) \geq 2^{-c(q+1)/2}/2\sqrt{r} = 2^{-c(q+1)/2-1}/\sqrt{r} .$$

Assuming $c = 6$, we have $2^{-c(q+2)/2} = 2^{-c(q+1)/2-3}$ and then we obtain the following separation condition:

$$d\bigg(B\Big(f^w, \frac{1}{\sqrt{r}} 2^{-c(q+2)/2}\Big), B\Big(f^{w'}, \frac{1}{\sqrt{r}} 2^{-c(q+2)/2}\Big)\bigg) \geq \frac{1}{\sqrt{r}} 2^{-c(q+1)/2-2} . \tag{6.26}$$

It is not difficult to check that (6.25) and (6.26) suffice to imply that $\mathcal{T}$ is indeed a tree. Since the notation is heavy, and since we are soon going to give an alternate approach, we leave this task to the reader and we turn to the control of the size of $\mathcal{T}$. Given $q < r$ and f_q as in (6.24) the sets $B_{q+1}(f^w)$ for $w \in W_{q+1}$ are the children of the set $A := B_q(f_q)$, where we recall that $B_q(f_q) = B(f_q, 2^{-c(q+1)/2}/\sqrt{r}) \cap \mathcal{E}$. It follows from (6.21) that the number $c(A)$ of children of A is at least $2^{2^{c(q+1)-3}}$ and (6.26) implies (6.6) with $4^{-s(A)} = 2^{-c(q+1)/2-3}/\sqrt{r}$. (Observe that $s(A)$ is an integer because r is a power of 16.) Consequently by definition (6.8) we get indeed that $d(\mathcal{T}) \geq \sqrt{r}/L$, and since r is about $\log N$ this completes our geometrical proof of (6.19). Let us also observe that this quantity about $\sqrt{r}$ is obtained from r contributions of size about $1/\sqrt{r}$. This helps to explain why the use of the Cauchy-Schwarz inequality in Chapter 4 is successful.

The reader may of course wonder how one discovers the previous construction. Being able to do this is really what we mean when we speak of "understanding the geometry of the situation".

In the next two sections, we will use a similar construction, but there is an extra difficulty which we try now to describe at a high level. Imagine that the points of $\mathcal{E}$ we try to construct represent the Fourier coefficients of a function. We would like this function to be Lipschitz. A control in norm $\|\cdot\|_{\mathcal{E}}$ (or a similar norm) of these Fourier coefficients is then going to be necessary for this, but it is not sufficient. Having constructed a Lipschitz function f_q, we will not be able to consider all possible values of $w \in W_{q+1}$ to form a function $f_{q+1} = f^w = f_q + w/(2\sqrt{s})$, because such a function need not be Lipschitz, but only some of them. The set of w which we may use for this purpose depends on f_q. The notation becomes complicated, and we shall write the argument without any mention of trees. The tree is nonetheless implicit, and the main argument recursively determines one of its branches. We now detail this argument in the setting of the previous construction (thereby providing a new geometrical proof of (6.19)). The argument is simply, given f_q, to prove that (provided now $c = 10$), we can find w in W_{q+1} for which

$$\gamma_2\Big(\mathcal{E}\cap B\Big(f_q,\frac{1}{\sqrt{r}}2^{-c(q+1)/2}\Big)\Big)\geq\frac{1}{L\sqrt{r}}+\gamma_2\Big(\mathcal{E}\cap B\Big(f^w,\frac{1}{\sqrt{r}}2^{-c(q+2)/2}\Big)\Big)\,. \tag{6.27}$$

Let us prove that this is possible. For $w \neq w'$, it follows from (6.22) that

$$d(f^w,f^{w'})\geq a:=\frac{1}{\sqrt{r}}2^{-c(q+1)/2-2}\,,$$

while it follows from the first part of (6.23) that $d(f_q, f^w) \leq 2a$. Moreover, for $c = 10$ we have $2^{-c(q+2)/2}/\sqrt{r} = a/8$, so that

$$B\Big(f^w,\frac{1}{\sqrt{r}}2^{-c(q+2)/2}\Big)=B(f^w,\frac{a}{8})\,.$$

We proved in Theorem 2.3.15 that the functional γ_2 satisfies the growth condition of Definition 2.3.10 with $c^* = 1/4$, and the parameter r of this definition (which is not the same as the parameter we use here) equal to 8. We can then apply this growth condition with a as above and $n = c(q+1)-3$ (see (6.22)) to obtain

$$\gamma_2\Big(\mathcal{E}\cap B\Big(f_q,\frac{1}{\sqrt{r}}2^{-c(q+1)/2}\Big)\Big)\geq\frac{1}{L\sqrt{r}}+\min_{w\in W}\gamma_2\Big(\mathcal{E}\cap B\Big(f^w,\frac{1}{\sqrt{r}}2^{-c(q+2)/2}\Big)\Big)\,.$$

We then choose $f_{q+1} = f^w$ where $w \in W$ is a value that achieves the minimum in the right-hand side. This completes the proof of (6.27).

We write the inequality (6.27) for $f^w = f_{q+1}$, and we sum over $1 \leq q \leq r-1$ to obtain (when $r \geq 2$)

$$\begin{aligned}\gamma_2(\mathcal{E})&\geq\gamma_2\Big(\mathcal{E}\cap B\Big(f_1,\frac{1}{\sqrt{r}}2^{-c}\Big)\Big)\\&\geq\frac{\sqrt{r}}{L}+\gamma_2\Big(\mathcal{E}\cap B\Big(f_r,\frac{1}{\sqrt{r}}2^{-c(r+1)/2}\Big)\Big)\geq\frac{\sqrt{r}}{L}\,,\end{aligned}$$

which is the desired result.

Exercise 6.3.2. Prove that the ellipsoid $\mathcal{E} = \{(x_i)_{i\leq N}; \sum_{1\leq i\leq N} i^{2/\alpha}x_i^2 \leq 1\}$ satisfies $\gamma_\alpha(\mathcal{E}) \geq \sqrt{\log N}/K$, where $\alpha \geq 1$.

6.4 Lower Bound for Theorem 4.3.2

Recalling that $\mathcal{C}$ denotes the class of functions that are 1-Lipschitz on the unit square, we shall prove the following, where $(X_i)_{i\leq N}$ are i.i.d. in $[0,1]^2$.

Theorem 6.4.1. *We have*

$$\mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\leq N}\Big(f(X_i)-\int f\mathrm{d}\lambda\Big)\Big|\geq\frac{1}{L}\sqrt{N\log N}\,. \tag{6.28}$$

In particular it follows from (4.58) that if the points Y_i are evenly spread then (provided $N \geq L$),

$$\mathsf{E}\sup_{f\in\mathcal{C}}\Big|\sum_{i\leq N}(f(X_i)-f(Y_i))\Big| \geq \frac{1}{L}\sqrt{N\log N}\ ,$$

so (4.41) implies that the expected cost of matching the points X_i and the points Y_i is at least $\sqrt{N\log N}/L$.

Let us now explain the basic idea of our approach. Consider standard Gaussian r.v.s (g_i) and suppose that instead of the left-hand side of (6.28) we want a lower bound for $\mathsf{E}\sup_{f\in\mathcal{C}}|\sum_{i\leq N} g_i f(X_i)|$. Let us think that we first fix the r.v.s X_i. Then, denoting by $\mathcal{D}$ a subset of $\mathcal{C}$ and by E_g expectation in the r.v.s g_i only, Theorem 2.4.1 implies

$$\mathsf{E}_g\sup_{f\in\mathcal{D}}\Big|\sum_{i\leq N} g_i f(X_i)\Big| \geq \frac{1}{L}\gamma_2(\mathcal{D},d_X)\ , \tag{6.29}$$

where d_X is the "empirical distance" on $\mathcal{D}$ defined by

$$d_X(f,f')^2=\sum_{i\leq N}(f(X_i)-f'(X_i))^2\ . \tag{6.30}$$

Let us now assume that

$$\forall\, f,f'\in\mathcal{D}\ ,\ d_X(f,f')\geq\frac{\sqrt{N}}{2}\|f-f'\|_2\ . \tag{6.31}$$

Then Theorem 2.7.5 (b) implies

$$\gamma_2(\mathcal{D},d_X)\geq\frac{\sqrt{N}}{2}\gamma_2(\mathcal{D})\ , \tag{6.32}$$

and (6.29) yields

$$\mathsf{E}_g\sup_{f\in\mathcal{D}}\Big|\sum_{i\leq N} g_i f(X_i)\Big| \geq \frac{\sqrt{N}}{L}\gamma_2(\mathcal{D})\ . \tag{6.33}$$

The strategy is then to look for a set $\mathcal{D}$ with $\gamma_2(\mathcal{D})\geq\sqrt{\log N}/L$ for which (6.31) holds with probability close to 1. We will then use (5.74) to show that in fact one even has, with obvious notation and probability close to 1

$$\mathsf{E}_\varepsilon\sup_{f\in\mathcal{D}}\Big|\sum_{i\leq N}\varepsilon_i f(X_i)\Big| \geq \frac{\sqrt{N}}{L}\gamma_2(\mathcal{D})\ , \tag{6.34}$$

which is basically the bound we want. To find $\mathcal{D}$, one must of course understand the combinatorics of the situation, which are somewhat similar to what happens in Section 6.2. It may help to point out from the start that (6.31) is really easy to satisfy (although actually writing every single detail will take some space). To see this, let us divide $[0,1]^2$ into little squares of side 2^{-p}. Then, as soon as 2^{-2p} is large compared to $\log N/N$, each such little square C contains about $N\lambda(C)$ points X_i. The simple argument goes as follows. The r.v.s $Z_i = \mathbf{1}_C(X_i) - \lambda(C)$ are centered, satisfy $|Z_i| \le 1$ and $\mathsf{E}Z_i^2 \le \lambda(C)$, so that Bernstein's inequality (4.59) implies

$$\mathsf{P}\Big(\Big|\sum_{i\le N} Z_i\Big| \ge \frac{1}{2}N\lambda(C)\Big) \le 2\exp(-N\lambda(C)/L) = 2\exp(-N2^{-2p}) \;,$$

and indeed with probability $\ge 1/2$ for each such C we have $|\sum_{i\le N} Z_i| \le N\lambda(C)/2$ and hence

$$\frac{1}{2}N\lambda(C) \le \operatorname{card}\{i \le N \;;\; X_i \in C\} \le \frac{3}{2}N\lambda(C) \tag{6.35}$$

provided $N2^{-2p}$ is large compared to $\log N$. Then (6.31) will automatically hold "provided we can describe the class $\mathcal{D}$ without having to look at a scale finer than 2^{-p}". There is all the room in the world in this argument, because the class we shall construct can actually be described without looking at a scale finer than, say, $2^{-p/10}$.

We now start the proof itself. Consider a parameter $c \in \mathbb{N}$. This parameter will be chosen later, and from now on we should think of it as of a universal constant. We also consider a number $r \in \mathbb{N}$ to be chosen later (about $\log N$). For $1 \le k \le r$ and $1 \le \ell \le 2^{ck}$ we consider the function $f'_{k,\ell}$ on $[0,1]$ defined as follows:

$$f'_{k,\ell}(x) = \begin{cases} 0 & \text{unless } x \in [(\ell-1)2^{-ck}, \ell 2^{-ck}[\\ \frac{1}{2\sqrt{r}} & \text{for } x \in [(\ell-1)2^{-ck}, (\ell-1/2)2^{-ck}[\\ -\frac{1}{2\sqrt{r}} & \text{for } x \in [(\ell-1/2)2^{-ck}, \ell 2^{-ck}[\;. \end{cases} \tag{6.36}$$

We define

$$f_{k,\ell}(x) = \int_0^x f'_{k,\ell}(y)\mathrm{d}y \;. \tag{6.37}$$

We now list a few useful properties of these functions. In these formulas $\|.\|_2$ denotes the norm in $L^2([0,1])$, etc. The proofs of these assertions are completely straightforward and better left to the reader.

Lemma 6.4.2. *The following holds true:*

$$f'_{k,\ell}(x) = 0 \textit{ unless } x \in [(\ell-1)2^{-ck}, \ell 2^{-ck}[\;. \tag{6.38}$$

$$\textit{The family } (f'_{k,\ell}) \textit{ is orthogonal in } L^2([0,1]) \;. \tag{6.39}$$

$$\|f'_{k,\ell}\|_2^2 = \frac{1}{4r}2^{-ck} \;. \tag{6.40}$$

$$\|f'_{k,\ell}\|_1 = \frac{1}{2\sqrt{r}} 2^{-ck} . \tag{6.41}$$

$$\|f_{k,\ell}\|_1 = \frac{1}{8\sqrt{r}} 2^{-2ck} . \tag{6.42}$$

$$\|f'_{k,\ell}\|_\infty = \frac{1}{2\sqrt{r}} \; ; \; \|f_{k,\ell}\|_\infty = \frac{1}{4\sqrt{r}} 2^{-ck} . \tag{6.43}$$

$$\|f_{k,\ell}\|_2^2 = \frac{1}{48r} 2^{-3ck} . \tag{6.44}$$

To prove Theorem 6.4.1 we will use the class of functions on $[0,1]^2$ of the type

$$f = \frac{\sqrt{r}}{16} \sum_{k \le r} 2^{ck} \sum_{\ell,\ell' \le 2^{ck}} z_{k,\ell,\ell'} f_{k,\ell} \otimes f_{k,\ell'} , \tag{6.45}$$

where $z_{k,\ell,\ell'} \in \{0,1,-1\}$ and where $f \otimes g(x,y) = f(x)g(y)$. The sum $\sum_{k \le r}$ define f as a sum of r functions f_k. Since $f_{k,\ell} \otimes f_{k,\ell'}$ is zero outside the little square $[(\ell-1)2^{-ck}, \ell 2^{-ck}[\times [(\ell'-1)2^{-ck}, \ell' 2^{-ck}[$, and since these little squares are disjoint as ℓ and ℓ' vary, the function f_k is easy to visualize. It is convenient to think of f_k as quite larger that f_{k+1}, and hence as quite larger than $\sum_{k'>k} f_{k'}$. (This is literally true only when the coefficients $z_{k,\ell,\ell'}$ are ± 1.) How much larger is governed by the number c: the larger c, the larger is f_k relative to f_{k+1}. Let us denote by $\mathcal{D}$ the class of functions f that are of the type (6.45), and that are also 1-Lipschitz. We think of $\mathcal{D}$ as a subset of the Hilbert space $L^2([0,1]^2)$. The central part of the argument is as follows.

Proposition 6.4.3. *We can choose c being a universal constant such that*

$$\gamma_2(\mathcal{D}) \ge \frac{\sqrt{r}}{L} . \tag{6.46}$$

One obstacle is that functions of the type (6.45) are not always 1-Lipschitz. It shall require some care to ensure that we properly choose the coefficients $z_{k,\ell,\ell'}$ to ensure that we construct only functions that are 1-Lipschitz. The next two lemmas prepare for this.

Lemma 6.4.4. *A function f given by (6.45) satisfies*

$$\left\| \frac{\partial f}{\partial x} \right\|_2 \le 2^{-7} . \tag{6.47}$$

Proof. First we write

$$\frac{\partial f}{\partial x}(x,y) = \frac{\sqrt{r}}{16} \sum_{k \le r} 2^{ck} \sum_{\ell,\ell' \le 2^{ck}} z_{k,\ell,\ell'} f'_{k,\ell}(x) f_{k,\ell'}(y) .$$

Using (6.39) and (6.40) we obtain, since $z_{k,\ell,\ell'}^2 \le 1$,

$$\int\Big(\frac{\partial f}{\partial x}\Big)^2 \mathrm{d}x = \frac{r}{(16)^2}\sum_{k\le r} 2^{2ck}\sum_{\ell,\ell'\le 2^{ck}} z^2_{k,\ell,\ell'}\|f'_{k,\ell}\|_2^2 f_{k,\ell'}(y)^2$$
$$\le \frac{1}{2^{10}}\sum_{k\le r} 2^{ck}\sum_{\ell,\ell'\le 2^{ck}} f_{k,\ell'}(y)^2 \ .$$

Integrating in y and using (6.44) yields

$$\Big\|\frac{\partial f}{\partial x}\Big\|_2^2 \le \frac{1}{2^{10}}\sum_{k\le r}\frac{1}{48r} \le 2^{-14} \ .$$

□

Lemma 6.4.5. *Consider a function of the type*

$$f = \frac{\sqrt{r}}{16}\sum_{k\le q} 2^{ck}\sum_{\ell,\ell'\le 2^{ck}} z_{k,\ell,\ell'} f_{k,\ell}\otimes f_{k,\ell'} \ ,$$

where $z_{k,\ell,\ell'}\in\{0,1,-1\}$. *Then*

$$\Big|\frac{\partial^2 f}{\partial x\partial y}\Big| \le \frac{2^{cq}}{2^5\sqrt{r}} \ . \tag{6.48}$$

Proof. We note that

$$\frac{\partial^2 f}{\partial x\partial y} = \frac{\sqrt{r}}{16}\sum_{k\le q} 2^{ck}\sum_{\ell,\ell'\le 2^{ck}} z_{k,\ell,\ell'} f'_{k,\ell}\otimes f'_{k,\ell'} \ ,$$

and we note from the second part of (6.43) that since the functions $(f'_{k,\ell})_{\ell\le 2^{ck}}$ have disjoint supports, that second sum is $\le 1/(4r)$ at every point. Also, $\sum_{k\le q} 2^{ck}\le 2^{cq+1}$. □

Proof of Proposition 6.4.3. By induction over $q\ge 0$, for $0\le q\le r$ we will construct functions $f_q\in\mathcal{D}$ with

$$f_q = \frac{\sqrt{r}}{16}\sum_{k\le q} 2^{ck}\sum_{\ell,\ell'\le 2^{ck}} z_{k,\ell,\ell'} f_{k,\ell}\otimes f_{k,\ell'} \ , \tag{6.49}$$

where $z_{k,\ell,\ell'}\in\{0,1,-1\}$, in such a manner that $f_0=0$ and that for $q\le r-1$

$$\gamma_2\Big(\mathcal{D}\cap B\Big(f_q,\frac{1}{2^7\sqrt{r}}2^{-c(q+1)}\Big)\Big) \ge \frac{1}{L\sqrt{r}} + \gamma_2\Big(\mathcal{D}\cap B\Big(f_{q+1},\frac{1}{2^7\sqrt{r}}2^{-c(q+2)}\Big)\Big) \ . \tag{6.50}$$

Summation of these inequalities over $0\le q\le r-1$ yields the result as in Section 6.3.

Given f_q, to construct f_{q+1} we need to construct the coefficients $z_{q+1,\ell,\ell'}$. First, we need to ensure that f_{q+1} is 1-Lipschitz, which is the really new part of the argument. For this let us consider the little squares of the type

$$[(\ell-1)2^{-c(q+1)}, \ell 2^{-c(q+1)}[\times[(\ell'-1)2^{-c(q+1)}, \ell' 2^{-c(q+1)}[\tag{6.51}$$

for $\ell, \ell' \in \mathbb{N}$ and $1 \le \ell, \ell' \le 2^{c(q+1)}$, so that there are $2^{2c(q+1)}$ such squares. To ensure that f_{q+1} is 1-Lipschitz, it suffices to ensure that it is 1-Lipschitz on each square (6.51). Let us say that the square (6.51) is *dangerous* if it contains a point for which either $|\partial f_q/\partial x| \ge 1/2$ or $|\partial f_q/\partial y| \ge 1/2$. (The danger is that on this square f_{q+1} might not be 1-Lipschitz.) We observe from the definition that all functions $f'_{k,\ell}$ for $k \le q$ are constant on the squares (6.51). So on such a square the quantity $\partial f_q/\partial x$ does not depend on x. Moreover it follows from (6.48) that if (x,y) and (x,y') belong to the same square (6.51) then

$$\Big|\frac{\partial f_q}{\partial x}(x,y) - \frac{\partial f_q}{\partial x}(x,y')\Big| \le |y-y'|\frac{2^{cq-5}}{\sqrt{r}} \le \frac{2^{-c-5}}{\sqrt{r}} \;.$$

In particular if a square (6.51) contains a point at which $|\partial f_q/\partial x| \ge 1/2$, then at each point of this square we have $|\partial f_q/\partial x| \ge 1/4$. Consequently (6.47) implies, with room to spare, that at most 1/2 of the squares (6.51) are dangerous. For these squares, we choose $z_{q+1,\ell,\ell'} = 0$, so that on these squares $f_{q+1} = f_q$ will be 1-Lipschitz. Let us say that a square (6.51) is *safe* if it is not dangerous, so that at each point of a safe square we have $|\partial f_q/\partial x| \le 1/2$ and $|\partial f_q/\partial x| \le 1/2$. Now (6.43) implies

$$\Big|\frac{\partial f_{q+1}}{\partial x} - \frac{\partial f_q}{\partial x}\Big| = \Big|\frac{\sqrt{r}}{16}2^{c(q+1)} \sum_{\ell,\ell' \le 2^{c(q+1)}} z_{q+1,\ell,\ell'} f'_{q+1,\ell} \otimes f_{q+1,\ell'}\Big| \le \frac{1}{2^7\sqrt{r}}$$

and

$$\Big|\frac{\partial f_{q+1}}{\partial y} - \frac{\partial f_q}{\partial y}\Big| = \Big|\frac{\sqrt{r}}{16}2^{c(q+1)} \sum_{\ell,\ell' \le 2^{c(q+1)}} z_{q+1,\ell,\ell'} f_{q+1,\ell} \otimes f'_{q+1,\ell'}\Big| \le \frac{1}{2^7\sqrt{r}} \;,$$

so we are certain than on a safe square we have $|\partial f_{q+1}/\partial x| \le 1/\sqrt{2}$ and $|\partial f_{q+1}/\partial y| \le 1/\sqrt{2}$, and hence that f_{q+1} is 1-Lipschitz. Let us denote by S the collection of pairs (ℓ, ℓ') such that the corresponding square (6.51) is safe, so $\operatorname{card} S \ge 2^{2c(q+1)-1}$. Lemma 6.3.1 produces a subset V of $\{-1,1\}^S$ with

$$\operatorname{card} V \ge 2^{2^{2c(q+1)-4}} = N_{2c(q+1)-4} \;, \tag{6.52}$$

and such that

$$\text{any two distinct elements } b \text{ and } b' \text{ of } V \text{ differ it at least } 2^{2c(q+1)-3} \text{ places} \;. \tag{6.53}$$

Any $b \in V$ is a family $(b_{\ell,\ell'})_{(\ell,\ell')\in S}$. For such a b we define

$$f^b = f_q + \frac{\sqrt{r}}{16}2^{c(q+1)} \sum_{\ell,\ell' \in S} b_{\ell,\ell'} f_{q+1,\ell} \otimes f_{q+1,\ell'} \;.$$

We will show that it works to take $f_{q+1} = f^b$ for an appropriate choice of b. First, using (6.43),

$$\|f^b - f_q\|_2 \le \|f^b - f_q\|_\infty \le \frac{\sqrt{r}}{16} 2^{c(q+1)} \frac{1}{16r} 2^{-2c(q+1)} \le \frac{1}{2^8\sqrt{r}} 2^{-c(q+1)} \; . \quad (6.54)$$

In particular, if $c \ge 1$,

$$B\Big(f^b, \frac{1}{2^7\sqrt{r}} 2^{-c(q+2)}\Big) \subset B\Big(f_q, \frac{1}{2^7\sqrt{r}} 2^{-c(q+1)}\Big) \; . \quad (6.55)$$

Also, if $b \ne b'$, using (6.53) and (6.44),

$$\begin{aligned} \|f^b - f^{b'}\|_2^2 &\ge \frac{r}{2^8} 2^{2c(q+1)} \sum_S (b_{\ell,\ell'} - b'_{\ell,\ell'})^2 \|f_{k,\ell}\|^2 \|f_{k,\ell'}\|^2 \\ &\ge \frac{r}{2^8} 2^{2c(q+1)} 4 \cdot 2^{2c(q+1)-3} \Big(\frac{1}{48r}\Big)^2 2^{-6c(q+1)} \\ &\ge \frac{1}{2^{22}r} 2^{-2c(q+1)} \; , \end{aligned}$$

and therefore

$$\|f^b - f^{b'}\| \ge a := \frac{1}{2^{11}\sqrt{r}} 2^{-c(q+1)} \; . \quad (6.56)$$

We now choose c such that

$$\frac{1}{2^7\sqrt{r}} 2^{-c(q+2)} \le \frac{a}{8} \; ,$$

i.e.

$$\frac{1}{2^7} 2^{-c} \le \frac{1}{8 \cdot 2^{11}} \; , \quad (6.57)$$

e.g. $c = 7$. Then

$$B\Big(f^b, \frac{1}{2^7\sqrt{r}} 2^{-c(q+2)}\Big) \subset B\Big(f^b, \frac{a}{8}\Big) \; .$$

We proved in Theorem 2.3.15 that the quantity γ_2 satisfies the growth condition of Definition 2.3.10 with $c^* = 1/4$, and the parameter r of this definition (with again is not the same as the parameter we use here) equal to 8. We can then apply this growth condition with $a = 2^{-c(q+1)-11}/\sqrt{r}$ and $n = 2c(q+1) - 5$ (see (6.52)) to obtain

$$\gamma_2\Big(\mathcal{D} \cap B\Big(f_q, \frac{1}{2^7\sqrt{r}} 2^{-c(q+1)}\Big)\Big) \ge \frac{1}{L\sqrt{r}} + \min_{b \in V} \gamma_2\Big(\mathcal{D} \cap B\Big(f^b, \frac{1}{2^7\sqrt{r}} 2^{-c(q+2)}\Big)\Big) \; .$$

We then choose $f_{q+1} = f^b$ where $b \in V$ is a value that achieves the minimum in the left-hand side. This completes the proof of (6.50) and of the proposition. □

Let us now prepare for the proof of Theorem 6.4.1. We fix N and we choose r as the largest for which $2^{cr} \le N^{1/100}$, so that $r \ge \log N/L$. We may assume that N is sufficiently large.

Lemma 6.4.6. *Condition (6.31) occurs with probability* $\geq 1/2$.

Proof. Since there is huge room to spare, let us be a bit informal in the proof. Divide $[0,1]^2$ in small subsquares of area about $L_2(\log N)/N$. Then if L_2 is large enough, with probability 1/2 each of these subsquares A contains at least $N\lambda(A)/2$ points X_i, where λ denotes the 2-dimensional Lebesgue measure, as we proved in (6.35). Assume that we are in this situation. We then claim that if f is a 1-Lipschitz function, with $|f| \leq 1$, then

$$\int f^2 \mathrm{d}\lambda \leq \frac{2}{N} \sum_{i \leq N} f(X_i)^2 + \frac{L\sqrt{\log N}}{\sqrt{N}} \,. \tag{6.58}$$

First we observe that f^2 is 2-Lipschitz since $|f^2(x) - f^2(y)| = |f(x) - f(y)||f(x) + f(y)| \leq 2|f(x) - f(y)|$. Then (6.58) holds because for each little square A we have $\max_A f^2 - \min_A f^2 \leq L\sqrt{\log N}/\sqrt{N}$ since f^2 is 2-Lipschitz and A is of diameter $\leq L\sqrt{\log N}/\sqrt{N}$. Indeed,

$$\int_A f^2 \mathrm{d}\lambda \leq \lambda(A) \max_A f^2 \leq \frac{L\sqrt{\log N}}{\sqrt{N}} \lambda(A) + \lambda(A) \min_A f^2 \,,$$

and moreover

$$\lambda(A) \min_A f^2 \leq \frac{2}{N} \operatorname{card}\{i \leq N \,;\, X_i \in A\} \min_A f^2 \leq \frac{2}{N} \sum_{i \leq N} f(X_i)^2 \mathbf{1}_{\{X_i \in A\}} \,.$$

Consequently,

$$\int_A f^2 \mathrm{d}\lambda \leq \frac{2}{N} \sum_{i \leq N} f(X_i)^2 \mathbf{1}_{\{X_i \in A\}} + \frac{L\sqrt{\log N}}{\sqrt{N}} \lambda(A) \,,$$

and summation over A yields (6.58). Using (6.58) for $(f - f')/2$ rather than f, and using the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ yields

$$\|f - f'\|_2 \leq \sqrt{2} \Big(\frac{1}{N} \sum_{i \leq N} (f(X_i) - f'(X_i))^2 \Big)^{1/2} + L \Big(\frac{\log N}{N} \Big)^{1/4} \,. \tag{6.59}$$

Next, we show that for f, f' in $\mathcal{D}$, $f \neq f'$, we have

$$\|f - f'\|_2 \geq \frac{1}{L\sqrt{r}} 2^{-2cr} \,. \tag{6.60}$$

To see this, consider (with obvious notation) the smallest integer q for which there exists ℓ^1 and ℓ^2 with $z_{q,\ell^1,\ell^2} \neq z'_{q,\ell^1,\ell^2}$. Let C be the square

$$C = [(\ell^1 - 1)2^{-cq}, \ell^1 2^{-cq}[\times [(\ell^2 - 1)2^{-cq}, \ell^2 2^{-cq}[\,.$$

Let

$$\overline{f} = \frac{\sqrt{r}}{16} \sum_{k>q} 2^{ck} \sum_{\ell,\ell' \le 2^{ck}} z_{k,\ell,\ell'} f_{k,\ell} \otimes f_{k,\ell'} \ ,$$

and define $\overline{f}'$ similarly. The definition of q implies

$$f - f' = \frac{\sqrt{r}}{16} 2^{cq} \sum_{\ell,\ell' \le 2^{cq}} (z_{q,\ell,\ell'} - z'_{q,\ell,\ell'}) f_{q,\ell} \otimes f_{q,\ell'} + \overline{f} - \overline{f}' \ ,$$

and therefore, since $|z_{q,\ell^1,\ell^2} - z'_{q,\ell^1,\ell^2}| \ge 1$,

$$\|f - f'\|_2 \ge \|(f - f')\mathbf{1}_C\|_2 \ge \frac{\sqrt{r}}{16} 2^{cq} \|f_{q,\ell^1} \otimes f_{q,\ell^2}\|_2 - \|\overline{f}\mathbf{1}_C\|_2 - \|\overline{f}'\mathbf{1}_C\|_2 \ .$$

We note now that (6.44) implies $\|f_{k,\ell} \otimes f_{k,\ell'}\|_2 = 1/(48r)2^{-3ck}$, and in particular $\|f_{q,\ell} \otimes f_{q,\ell'}\|_2 = 1/(48r)2^{-3cq}$. Next we use the bound

$$\|\overline{f}\mathbf{1}_C\|_2 \le \sum_{k>q} \frac{\sqrt{r}}{16} 2^{ck} \Big\| \sum_{\ell,\ell'} f_{k,\ell} \otimes f_{k,\ell'} \mathbf{1}_C \Big\|_2 \ .$$

Since the functions $f_{k,\ell} \otimes f_{k,\ell'}$ have disjoint support as ℓ and ℓ' vary and since only $2^{2c(k-q)}$ such functions are relevant in the sum $\sum_{\ell,\ell'} f_{k,\ell} \otimes f_{k,\ell'} \mathbf{1}_C$ we obtain, using also (6.44),

$$\Big\| \sum_{\ell,\ell'} f_{k,\ell} \otimes f_{k,\ell'} \mathbf{1}_C \Big\|_2 = \frac{2^{-3ck}}{48r} 2^{c(k-q)}$$

and thus

$$\|\overline{f}\mathbf{1}_C\|_2 \le \frac{\sqrt{r}}{16} \frac{1}{48r} 2^{-cq} \sum_{k>q} 2^{-ck}.$$

Combining these estimates we obtain

$$\|f - f'\| \ge \frac{\sqrt{r}}{16} \frac{1}{48r} \Big(2^{-2cq} - 2 \cdot 2^{-cq} \sum_{k>q} 2^{-ck} \Big) \ge \frac{L}{\sqrt{r}} 2^{-2cq} \ge \frac{L}{\sqrt{r}} 2^{-2cr} \ ,$$

and we have proved (6.60). Since 2^{-cr} is about $N^{-1/100}$, for large N (6.60) implies

$$L \Big(\frac{\log N}{N} \Big)^{1/4} \le \frac{1}{10} \|f - f'\|_2 \ ,$$

and combining with (6.59) this completes the proof. □

Proof of Theorem 6.4.1. When (6.31) holds, using (6.32) and (6.46), we get

$$\gamma_2(\mathcal{D}, d_X) \ge \frac{\sqrt{N}}{L} \sqrt{r} \ . \tag{6.61}$$

Now, Lemma 4.3.9 implies that $e_n(\mathcal{D}, d_\infty) \le L2^{-n/2}$. Moreover, since $\operatorname{card}\mathcal{D} \le 2^{L2^{2cr}}$, we have $e_n(\mathcal{D}, d_\infty) = 0$ for $n \ge 2cr + L$. Then Corollary 2.3.2 implies that $\gamma_1(\mathcal{D}, d_\infty) \le L2^{cr} \le L\sqrt{N}$ by the choice of r and since $c = 7$. We now appeal to (5.74) to obtain that if E_ε denotes expectation in the r.v.s ε_i only,

$$\mathsf{E}_\varepsilon \sup_{f\in\mathcal{C}} \sum_{i\le N} \varepsilon_i f(X_i) \ge \mathsf{E}_\varepsilon \sup_{f\in\mathcal{D}} \sum_{i\le N} \varepsilon_i f(X_i) \ge \frac{1}{L}\sqrt{rN} \,. \tag{6.62}$$

Since the probability corresponding to the choice of the r.v.s X_i that this holds is $\ge 1/2$, taking expectation (and since $\mathsf{E}_\varepsilon \sup_{f\in\mathcal{C}} \sum_{i\le N} \varepsilon_i f(X_i) \ge 0$ because $0 \in \mathcal{C}$) we obtain

$$\mathsf{E} \sup_{f\in\mathcal{C}} \sum_{i\le N} \varepsilon_i f(X_i) \ge \frac{1}{L}\sqrt{rN} \,. \tag{6.63}$$

It remains to use a "desymmetrization argument". For $f \in \mathcal{C}$ we have $|\int f \mathrm{d}\lambda| \le 1$, and thus (6.63) implies (for N large enough)

$$\mathsf{E} \sup_{f\in\mathcal{C}} \sum_{i\le N} \varepsilon_i \Big(f(X_i) - \int f \mathrm{d}\lambda \Big) \ge \frac{1}{L}\sqrt{rN} \,.$$

Let us denote by E^ε expectation *given* the r.v.s ε_i, and let $I = \{i \le N; \varepsilon_i = 1\}, J = \{i \le N; \varepsilon_i = -1\}$, so that

$$\begin{aligned}\mathsf{E}^\varepsilon \sup_{f\in\mathcal{C}} \sum_{i\le N} \varepsilon_i \Big(f(X_i) - \int f \mathrm{d}\lambda \Big) \le \mathsf{E}^\varepsilon \sup_{f\in\mathcal{C}} \Big| \sum_{i\in I} \Big(f(X_i) - \int f \mathrm{d}\lambda \Big) \Big| \\ + \mathsf{E}^\varepsilon \sup_{f\in\mathcal{C}} \Big| \sum_{i\in J} \Big(f(X_i) - \int f \mathrm{d}\lambda \Big) \Big| \,.\end{aligned}$$

Both terms on the right-hand side are $\le \mathsf{E} \sup_{f\in\mathcal{C}} |\sum_{i\in N} (f(X_i) - \int f \mathrm{d}\lambda)|$ by Jensen's inequality, and consequently $\sqrt{rN}/L \le \mathsf{E} \sup_{f\in\mathcal{C}} |\sum_{i\in N} (f(X_i) - \int f \mathrm{d}\lambda)|$. □

6.5 Lower Bound for Theorem 4.4.1

Theorem 6.5.1. *If the points $(X_i)_{i\le N}$ are i.i.d. uniform over $[0,1]^2$ and the points $(Y_i)_{i\le N}$ are evenly spread, then*

$$\mathsf{E} \inf_\pi \max_{i\le N} d(X_i, Y_{\pi(i)}) \ge \frac{(\log N)^{3/4}}{L\sqrt{N}} \,. \tag{6.64}$$

The proof will be strikingly similar in spirit to that of Theorem 6.4.1. The main step is as follows.

Theorem 6.5.2. *Denoting by $\mathcal{C}$ the class of functions $f : [0,1] \to [0,1]$ such that $f(0) = f(1) = 1/2$, $|f'| \le 1$, and by $\mathcal{S}$ the class of subgraphs*

$$S(f) = \{(x,y) \in [0,1]^2 \; ; \; y \le f(x)\} \; ,$$

of the functions in $\mathcal{C}$ then

$$\mathsf{E} \sup_{S \in \mathcal{S}} \Big| \sum_{i \le N} (\mathbf{1}_S(X_i) - \lambda(S)) \Big| \ge \frac{1}{L} \sqrt{N} (\log N)^{3/4} \; . \tag{6.65}$$

Proof of Theorem 6.5.1. We first observe that for any 1-Lipschitz function h we have

$$\operatorname{card}\{i \le N \; ; \; Y_i \in S(h)\} \le N\lambda(S(h)) + L\sqrt{N} \; . \tag{6.66}$$

This is because by definition of an evenly spread family, each point Y_i belongs to a small rectangle R_i of area $1/N$ and of diameter $\le 10/\sqrt{N}$, and a pessimistic bound for the left-hand side above is the number of such rectangles that meet $S(h)$. These rectangles are entirely contained in the set of points within distance $L/\sqrt{N}$ of $S(h)$, and since h is 1-Lipschitz, this set has area $\le S(h) + L\sqrt{N}$, hence the bound (6.66).

Consider now fix now $f \in \mathcal{C}$ and consider $\epsilon > 0$. Since f is 1-Lipschitz, the ϵ-neighborhood $S_\epsilon(f)$ of $S(f)$ in $[0,1]^2$ is contained in $S(f + 2\epsilon)$. Indeed, if (x,y) is within distance ϵ of $(x', y') \in S(f)$, then $y \le |y - y'| + y' \le \epsilon + f(x') \le 2\epsilon + f(x)$ since $|f(x') - f(x)| \le |x - x'| \le \epsilon$. In particular $\lambda(S(f + 2\epsilon)) \le \lambda(S(f)) + 2\epsilon$ and (6.66) implies

$$\operatorname{card}\{i \le N \; ; \; Y_i \in S_\epsilon(f)\} \le N(\lambda(S(f) + 2\epsilon) + L\sqrt{N} \; . \tag{6.67}$$

Let us consider $S = S(f) \in \mathcal{S}$ and let

$$D = \sum_{i \le N} \mathbf{1}_S(X_i) - N\lambda(S) = \operatorname{card}\{i \le N \; ; \; X_i \in S\} - N\lambda(S) \; .$$

Assume first $D > L\sqrt{N}$ and let $\epsilon = (D - L\sqrt{N})/(4N)$. Then

$$\operatorname{card}\{i \le N \; ; \; X_i \in S(f)\} = N\lambda(S(f)) + D \ge N(\lambda(S(f) + 4\epsilon) + L\sqrt{N} \; .$$

Consequently (6.67) implies

$$\operatorname{card}\{i \le N \; ; \; Y_i \in S_\epsilon(f)\} < \operatorname{card}\{i \le N \; ; \; X_i \in S(f)\} \; ,$$

and therefore any matching must pair at least one point $X_i \in S(f)$ with a point $Y_j \notin S_\epsilon(f)$, so that

$$\max_{i \le N} d(X_i, Y_{\pi(i)}) \ge \epsilon = \frac{D}{4N} - \frac{L\sqrt{N}}{4N} \; .$$

Proceeding in a similar manner when $D < -L\sqrt{N}$ we show that

$$\max_{i \le N} d(X_i, Y_{\pi(i)}) \ge \epsilon = \frac{|D|}{4N} - \frac{L\sqrt{N}}{4N} .$$

Taking the supremum over $S = S(f) \in \mathcal{S}$ yields

$$\max_{i \le N} d(X_i, Y_{\pi(i)}) \ge \frac{1}{4N} \sup_{S \in \mathcal{S}} \Big| \sum_{i \le N} (\mathbf{1}_S(X_i) - \lambda(S)) \Big| - \frac{L\sqrt{N}}{4N} .$$

Taking expectation and using (6.65) finishes the proof. □

Recalling the functions $f_{k,\ell}$ of (6.37), consider the functions of the type

$$f = \frac{1}{2} + \sum_{1 \le k \le r} \sum_{1 \le \ell \le 2^{ck}} x_{k,\ell} f_{k,\ell} , \tag{6.68}$$

where $x_{k,\ell} \in \{-1, 0, 1\}$. Then $f(0) = f(1) = 1/2$. Consider the class $\mathcal{D}$ of functions of this type for which $|f'| \le 1$, and observe that then f is valued in $[0, 1]$. The plan is to prove that for an appropriate choice of c and r then (6.65) already holds for $\mathcal{S}$ the class of subgraphs $S(f)$ for $f \in \mathcal{D}$.

Not all the functions of the type (6.68) satisfy $|f'| \le 1$, and we gather first some information that will allow us to satisfy this condition.

Lemma 6.5.3. *A function f of the type (6.68) satisfies*

$$\int_0^1 f'(x)^2 \mathrm{d}x \le \frac{1}{4} . \tag{6.69}$$

Proof. Using (6.37) and (6.39) we obtain

$$\int_0^1 f'(x)^2 \mathrm{d}x \le \sum_{k \le r} \sum_{\ell \le 2^{ck}} x_{k,\ell}^2 \frac{1}{4r} 2^{-ck} \le \frac{1}{4} .$$ □

The following is an immediate consequence of (6.42).

Lemma 6.5.4. *We have*

$$\Big\| \sum_{k \le r} \sum_{\ell \le 2^{ck}} x_{k,\ell} f_{k,\ell} \Big\|_1 \le \frac{1}{8\sqrt{r}} \sum_{k \le r} 2^{-2ck} \sum_{\ell \le 2^{ck}} |x_{k,\ell}| . \tag{6.70}$$

Moreover if $|x_{k,\ell}| \le 1$ and $x_{k,\ell} = 0$ for $k \le q$ then

$$\Big\| \sum_{k \le r} \sum_{\ell \le 2^{ck}} x_{k,\ell} f_{k,\ell} \Big\|_1 \le \frac{1}{4\sqrt{r}} 2^{-c(q+1)} . \tag{6.71}$$

The main ingredient in the proof of Theorem 6.5.2 is as follows, where we view $\{\mathbf{1}_{S(f)}; f \in \mathcal{D}\}$ as a subset of $L^2 = L^2([0,1]^2)$.

Proposition 6.5.5. *If c is appropriately chosen then*

$$\gamma_2(\{\mathbf{1}_{S(f)}\ ;\ f \in \mathcal{D}\}) \geq \frac{r^{3/4}}{L}\ . \tag{6.72}$$

Proof. For $f \in \mathcal{D}$ let us define

$$D(f,\rho) = \{\mathbf{1}_{S(h)}\ ;\ h \in \mathcal{D}\ ,\ \|h-f\|_1 \leq \rho\}\ . \tag{6.73}$$

Let us observe that since we are dealing here with indicators of sets,

$$D(f,\rho) = \{\mathbf{1}_{S(h)}\ ;\ h \in \mathcal{D}\ ,\ \|\mathbf{1}_{S(h)} - \mathbf{1}_{S(f)}\|_2 \leq \sqrt{\rho}\} \tag{6.74}$$

is a ball of radius $\sqrt{\rho}$ in L^2. By induction over q (and provided that c has been correctly chosen) we construct functions

$$f_q = \sum_{k \leq q} \sum_{\ell \leq 2^{ck}} x_{k,\ell} f_{k,\ell}\ , \tag{6.75}$$

where $x_{k,\ell} \in \{-1,0,1\}$ and $f_q \in \mathcal{D}$, such that

$$\gamma_2\Big(D\Big(f_q, \frac{1}{4\sqrt{r}}2^{-c(q+1)}\Big)\Big) \geq \frac{1}{Lr^{1/4}} + \gamma_2\Big(D\Big(f_{q+1}, \frac{1}{4\sqrt{r}}2^{-c(q+2)}\Big)\Big)\ . \tag{6.76}$$

Summation of these inequalities proves (6.75). For the construction we proceed as follows. We observe that f'_q is constant on the intervals $I_\ell = [(\ell-1)2^{-c(q+1)}, \ell 2^{-c(q+1)}[$. We denote by J the set of integers $\ell \leq 2^{c(q+1)}$ such that $|f'_q| \leq 1/\sqrt{2}$ on I_ℓ. (The intervals I_ℓ for $\ell \in J$ correspond to the "safe" small squares in the proof of Proposition 6.4.3.) Then $\int_{I_\ell} f'_q(x)^2 \mathrm{d}x \geq 2^{-c(q+1)-1}$ for $\ell \notin J$. Then Lemma 6.5.3 yields

$$\operatorname{card} J \geq 2^{c(q+1)-1}\ . \tag{6.77}$$

We appeal to Lemma 6.3.1 to find a subset V of $\{0,1\}^J$ with

$$\operatorname{card} V \geq 2^{2^{c(q+1)-1}-3} \geq 2^{2^{c(q+1)-2}} = N_{c(q+1)-2}\ , \tag{6.78}$$

such that any two distinct elements of V differ in at least $\operatorname{card} J/4$ coordinates. For $b \in V$ we define

$$f^b = f_q + \sum_{\ell \in J} b_\ell f_{q+1,\ell}\ ,$$

and we are going to show that one may choose for f_{q+1} one of the functions f^b for $b \in V$. From Lemma 6.5.4 we observe that

$$\|f^b - f_q\|_1 \leq \frac{1}{8\sqrt{r}} 2^{-c(q+1)} \tag{6.79}$$

and therefore

$$D\Big(f^b, \frac{1}{4\sqrt{r}} 2^{-c(q+2)}\Big) \subset D\Big(f_q, \frac{1}{4\sqrt{r}} 2^{-c(q+1)}\Big) . \tag{6.80}$$

Also, for $b \neq b'$ we have, using (6.42),

$$\|f^b - f^{b'}\|_1 \geq \frac{1}{8\sqrt{r}} 2^{-2c(q+1)} 2 \cdot \frac{1}{4} \operatorname{card} J \geq a^2 := \frac{1}{2^5\sqrt{r}} 2^{-c(q+1)} . \tag{6.81}$$

Keeping (6.74) in mind we now choose c so that

$$\Big(\frac{1}{4\sqrt{r}} 2^{-c(q+2)}\Big)^{1/2} \leq \frac{a}{8} ,$$

i.e. $2^{-c/2} \leq 2^{-3/2}/8$ (e.g. $c = 9$). We then appeal to Theorem 2.3.15 with $a = 2^{-5/2} r^{-1/4} 2^{-c(q+1)/2}$ and $n = c(q+1) - 3$ to obtain

$$\gamma_2\Big(D\Big(f_q, \frac{1}{4\sqrt{r}} 2^{-c(q+1)}\Big)\Big) \geq \frac{1}{Lr^{1/4}} + \min_{b \in V} \gamma_2\Big(D\Big(f^b, \frac{1}{4\sqrt{r}} 2^{-c(q+2)}\Big)\Big) . \tag{6.82}$$

To obtain (6.76) we then choose f_{q+1} as f^b for a value of b that gives the minimum in the right-hand side. □

Proof of Theorem 6.5.2. Given N we choose again r as the largest for which $2^{cr} \leq N^{1/100}$, so that $r \geq \log N/L$. First we prove that with probability $\geq 1/2$ we have

$$\forall\, f, h \in \mathcal{D} \,,\ \|f - h\|_1 = \int |\mathbf{1}_{S(f)} - \mathbf{1}_{S(h)}| \mathrm{d}\lambda \leq \frac{4}{N} \sum_{i \leq N} |\mathbf{1}_{S(f)}(X_i) - \mathbf{1}_{S(h)}(X_i)| . \tag{6.83}$$

To see this we divide the unit square in little subsquares of area about $L \log N/N$ where L is large enough that, with probability $\geq 1/2$, (and using of course Bernstein's inequality) each of these subsquares C contains at least $N\lambda(C)/2$ points X_i. We then estimate pessimistically from below

$$\frac{1}{N} \sum_{i \leq N} |\mathbf{1}_{S(f)}(X_i) - \mathbf{1}_{S(h)}(X_i)|$$

by the number of points X_i that are contained in little squares C that are entirely contained in the domain $\mathcal{S}$ between the graphs of f and of h. The number of such squares is at least N times the area of the region $\mathcal{S}'$ consisting of the points of $\mathcal{S}$ that are at distance $\geq L\sqrt{\log N}/\sqrt{N}$ of either the graph of f or the graph of g, and

$$\lambda(\mathcal{S}') \geq \lambda(\mathcal{S}) - \frac{L\sqrt{\log N}}{\sqrt{N}} = \|f - h\|_1 - \frac{L\sqrt{\log N}}{\sqrt{N}} .$$

In this manner we obtain

$$\frac{1}{N}\sum_{i\le N}|\mathbf{1}_{S(f)}(X_i)-\mathbf{1}_{S(h)}(X_i)|\ge\frac{1}{2}\|f-h\|_1-\frac{L\sqrt{\log N}}{\sqrt{N}}\ . \tag{6.84}$$

Since the rest of the proof is nearly identical to the case of Theorem 6.4.1 we provide only the outline and leave the details to the reader. First, we prove that

$$\forall\, f,h\in\mathcal{D}\ ,\ f\ne h\Rightarrow\|f-h\|_1\ge\frac{1}{16\sqrt{r}}2^{-3cr}\ ,$$

by proceeding very much as in the proof of (6.60). Combining with (6.84) this proves (6.83).

We recall the distance d_X of (6.30). We use Theorem 2.7.5 (b) to deduce from (6.72) that, with probability at least 1/2,

$$\gamma_2(\{\mathbf{1}_{S(f)}\ ;\ f\in\mathcal{D}\}\ ,\ d_X)\ge\frac{r^{3/4}\sqrt{N}}{L}\ .$$

Next, we use Corollary 2.3.2, and the fact that $\operatorname{card}\mathcal{D}\le 3^{2^{cr}}\le 2^{2^{cr+1}}$ to obtain

$$\gamma_1(\{\mathbf{1}_{S(f)}\ ;\ f\in\mathcal{D}\}\ ,\ d_\infty)\le L2^{cr/2}\ ,$$

and since $c=9$ this is much smaller than $\sqrt{N}$. We then deduce from (5.74) that if E_ε denotes expectation in the r.v.s ε_i only, with probability $\ge 1/2$ we have

$$\mathsf{E}_\varepsilon\sup_{S\in\mathcal{S}}\Big|\sum_{i\le N}\varepsilon_i\mathbf{1}_S(X_i)\Big|\ge\frac{1}{L}\sqrt{N}r^{3/4}\ .$$

We then take expectation and conclude with the same "desymmetrization" argument as in the proof of Theorem 6.4.1. □

Reference

1. Mendel, M., Naor, A.: Ultrametric subsets with large Hausdorff dimension. Invent. Math. **192**(1), 1–54 (2013)

7. Random Fourier Series and Trigonometric Sums, II

7.1 Introduction

The topic of random Fourier series illustrates well the impact of abstract methods, and it might be useful to provide an (extremely brief) history of the topic.

In a series of papers in 1930 and 1932 R. Paley and A. Zygmund [6], [7], [8] raised (among other similar problems) the question of the uniform convergence of the series

$$\sum_{k\geq 1} a_k \varepsilon_k \exp(ikx) \tag{7.1}$$

uniformly over $x \in [0, 2\pi]$, where a_k are real numbers and ε_k are independent Bernoulli r.v.s (and of course here i is not a summation index but $i^2 = -1$). Considering the numbers s_p defined by $s_p^2 = \sum_{2^p \leq n < 2^{p+1}} a_n^2$ they prove in particular the necessity of the condition $\sum_p s_p < \infty$. Later, R. Salem and A. Zygmund [9] proved that if the sequence (s_p) is non-increasing the condition $\sum_p s_p < \infty$ suffices for uniform convergence of the random Fourier series. The combination of these two results is remarkably sharp, but certainly does not settle the problem of the convergence of the series (7.1).

It belonged to M. Marcus and G. Pisier to find necessary and sufficient conditions for uniform convergence, along lines which have already been largely been explained in Chapter 3. This requires, as a major conceptual step, the Dudley-Fernique characterization of boundedness of stationary Gaussian processes. The conditions of Marcus and Pisier are of the type $\gamma_2([0, 2\pi], d) < \infty$ for a certain distance d, and it is a non-trivial task (which is thoroughly performed in [5]) to show that they improve on the "classical" results of Paley, Salem and Zygmund. The results of [5] cover not only the case of series of the type (7.1) but more general cases such as the series

$$\sum_{k\geq 1} a_k \xi_k \exp(ikx) \tag{7.2}$$

where the independent symmetric r.v.s ξ_k satisfy $\sum_k \mathsf{E}\xi_k^2/(\mathsf{E}|\xi_k|) < \infty$ (and many other situations). They still certainly however do not settle the problem of the uniform convergence of the series (7.2) in full generality.

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_7,

In this chapter we (basically) complete the program outlined in Section 3.2 of finding upper and lower bounds of the same order for the quantities $\mathsf{E}\|\sum_i \xi_i \chi_i\|$ where χ_i are characters and ξ_i are independent symmetric r.v.s. As a consequence we obtain necessary and sufficient conditions for the convergence of random Fourier series in a very general setting (and in particular the series (7.2)). These characterizations are in essence of the same nature as the results of Marcus and Pisier. Unfortunately this means that is not always immediate to apply them in concrete situations, but we will illustrate at length how this can be done. Fulfilling this program requires both technical and conceptual advances compared to the work of Section 3.2. The greatest technical challenge is to perform chaining in an essentially optimal way. Fortunately, it has already been addressed in Theorem 5.2.1, and this will be our first opportunity to demonstrate the power of this result. The conceptual advance is the idea of "families of distances" which already appeared implicitly in Chapter 5. It is one of the central themes of this entire work, and we turn to it now.

7.2 Families of Distances

Not all random processes of interest satisfy a condition as simple as (1.4) or even (2.50). In certain natural situations, a precise description of the increments of a process cannot be achieved using only one or two distances, but requires using a "family of distances." This discovery has led to the possibility of describing exactly when certain large families of processes are bounded. Quite interestingly, once the first surprise is passed and the right setting has been found, it turns out that working with a family of distances is not more difficult than working with a single distance.

Let us first explain this concept of "families of distances". For each $j \in \mathbb{Z}$ we consider a non-negative function $\varphi_j(s,t)$ on $T \times T$. Despite the fact that we use the convenient terminology "family of distances" the map φ_j is not usually a distance, but will be very often the *square* of a distance. In this case, using the triangle inequality for $\sqrt{\varphi_j}$, for $s, t_1, t_2 \in T$ and the inequality $(a+b)^2 \le 2(a^2+b^2)$ we have

$$\varphi_j(t_1,t_2) \le 2(\varphi_j(t_1,s) + \varphi_j(s,t_2)) .$$

We will always assume that φ_j is symmetric,

$$\varphi_j(s,t) = \varphi_j(t,s) .$$

Often (and in particular in this chapter) the sequence (φ_j) will be non-decreasing:

$$\varphi_j(s,t) \le \varphi_{j+1}(s,t) .$$

We have developed ways to measure the "size" of T when T is provided with a single distance d, such as Dudley's entropy integral $\sum_{n\ge0} 2^{n/2} e_n(T,d)$

or the functional $\gamma_2(T,d)$, and our purpose is to develop similar notions when T is provided with "a family of distances". In this chapter we deal with a "homogeneous setting" so that entropy numbers suffice (as exemplified in Theorem 3.1.1) and we shall explain only how to properly define the notion that corresponds to entropy integrals. (Later, in Chapter 10 we shall learn how properly generalize the quantity $\gamma_2(T,d)$.) Let us define (assuming these numbers exist)

$$j_0 = \sup\bigl\{j \in \mathbb{Z}\ ;\ \forall\, s,t \in T\ ,\ \varphi_j(s,t) \le 1\bigr\}\ , \tag{7.3}$$

and for $n \ge 1$ let us define j_n as the largest integer j for which "there exists a 2^n-net of size N_n for the distance φ_j", i.e.

$$\begin{aligned} j_n = \sup\bigl\{&j \in \mathbb{Z}\ ;\ \exists U \subset T\ ,\ \operatorname{card} U \le N_n = 2^{2^n}\ ,\\ &\forall\, t \in T\ ,\ \exists s \in U\ ,\ \varphi_j(t,s) \le 2^n\bigr\}\ . \end{aligned} \tag{7.4}$$

We note that the sequence (j_n) is increasing. To make sense out of these definitions, let us consider the case where, for a given distance d on T one has

$$\varphi_j(s,t) = 2^{2j} d(s,t)^2\ . \tag{7.5}$$

Then, obviously,

$$j_0 = \sup\{j \in \mathbb{Z}\ ;\ 2^{2j}\Delta(T,d)^2 \le 1\}\ ,$$

so that

$$2^{-j_0-1} \le \Delta(T,d) \le 2^{-j_0}\ . \tag{7.6}$$

Also, it should be obvious from (7.4) that for $n \ge 1$

$$2^{n/2-j_n-1} \le e_n(T,d) \le 2^{n/2-j_n}\ .$$

Consequently,

$$\sum_{n\ge 0} 2^{n-j_n-1} \le \sum_{n\ge 0} 2^{n/2} e_n(T,d) \le \sum_{n\ge 0} 2^{n-j_n}\ .$$

Thus, in the present situation, the quantity

$$\sum_{n\ge 0} 2^{n-j_n} \tag{7.7}$$

is basically Dudley's integral. It turns out that in the more general situation where (7.5) may not hold, this quantity is a useful generalization of Dudley's integral.

Let us also make two simple observations. First, we are not going to change much if we find it convenient to replace the definition (7.3) of j_0 by, for example,

$$j_0 = \sup\{j \in \mathbb{Z} \; ; \; \forall\, s,t \in T \; , \; \varphi_j(s,t) \le 1/4\} \; .$$

Second, if we are in a translation-invariant situation, and in particular when T is a compact group with Haar measure μ and unit element 0, we may then define j_n as

$$j_n = \sup\{j \in \mathbb{Z} \; ; \; \mu(\{s \; ; \; \varphi_j(s,0) \le 2^n\}) \ge 2^{-2^n} = N_n^{-1}\} \; . \tag{7.8}$$

In a special setting in Theorem 3.1.1 we proved that (7.8) is essentially the same as (7.4). In this chapter we shall use only the definition (7.8), but our first task is to prove that in the present general setting this is still basically the same as (7.4). This argument is absolutely central to the whole approach, so it is worth repeating it even though it is exactly the same as in case of Theorem 3.1.1. Let us denote by j_n^* the integer (7.8) and by j_n the integer (7.4). We assume that each function φ_j is the square of distance and is translation invariant, i.e. $\varphi_j(t_1,t_2) = \varphi_j(t_1+s, t_2+s)$ for each $t_1,t_2,s \in T$. We then prove that

$$j_n \le j_n^* \le j_{n+2} \; . \tag{7.9}$$

First, consider the set U as in (7.4) for $j = j_n$. Then $T \subset \bigcup_{s\in U}\{t \in T; \varphi_j(s,t) \le 2^n\}$, so that $1 \le \sum_{s\in U} \mu(\{t \in T; \varphi_j(s,t) \le 2^n\})$. Now, by translation invariance,

$$\mu(\{t \in T \; ; \; \varphi_j(s,t) \le 2^n\}) = \mu(\{t \in T \; ; \; \varphi_j(0,t) \le 2^n\}) \; ,$$

and thus $1 \le \operatorname{card} U \cdot \mu(\{t \in T; \varphi_j(0,t) \le 2^n\})$ so that $\mu(\{s \in T; \varphi_j(s,0) \le 2^n\}) \ge N_n^{-1}$ and $j = j_n \le j_n^*$.

It is the other direction which is fundamental. It is based on the "volume argument" that we have used many times, and which we state as follows.

Lemma 7.2.1. *Consider a subset B of T. Then there exists a subset U of T with* $\operatorname{card} U \le 1/\mu(B)$ *such that whenever $t \in T$ we can find $s \in U$ with $t \in s + B - B$, where $B - B = \{t_1 - t_2; t_1, t_2 \in B\}$.*

Proof. Consider U as large as possible so that the sets $s + B$ are disjoint for $s \in U$. Since $\mu(s+B) = \mu(B)$ we have $\operatorname{card} U \cdot \mu(B) \le 1$. The maximality of U implies that for each $t \in T$ there exists $s \in U$ for which $(t+B) \cap (s+B) \ne \emptyset$. Then $t \in s + B - B$. □

We continue the proof of (7.9). Consider then $j \le j_n^*$ and

$$B = \{s \in T \; ; \; \varphi_j(s,t) \le 2^n\} \; ,$$

so that $\mu(B) \ge 1/N_n$. Consider the set U provided by Lemma 7.2.1. Observe that if $t \in s + B - B$, then for some $t_1, t_2 \in B$ we have $t = s + t_1 - t_2$ and thus, using that φ_j is the square of a distance and translation invariance

$$\begin{aligned} \varphi_j(t,s) &= \varphi_j(t-s,0) = \varphi_j(t_1 - t_2, 0) = \varphi_j(t_1,t_2) \\ &\le 2(\varphi_j(t_1,0) + \varphi_j(0,t_2)) \le 2^{n+2} \; , \end{aligned}$$

so that $\varphi_j(s,t) \le 2^{n+2}$. Since $\operatorname{card} U \le N_n \le N_{n+2}$ this proves that $j = j_n^* \le j_n + 2$ and (7.9).

7.3 Statement of Main Results

As in Section 3.2, our purpose is not to prove the most general possible results, but to illustrate the concepts and the methods in a setting which is not obscured by technicalities. Therefore we assume as in Section 3.2 that T is a compact Abelian group with Haar measure μ. On the other hand, since our results seem to provide a final answer to a number of old questions, we will develop several versions of the basic result. We denote by G the set of characters on T.

Consider independent r.v.s Z_i valued in $\mathbb{C}G$. That is, $Z_i = \xi_i \chi_i$ where ξ_i is a complex-valued r.v. and χ_i is a random character. Please note that we do not assume that ξ_i and χ_i are independent r.v.s. The crucial property is

$$\forall\, s, t \in T\ ,\ |Z_i(s) - Z_i(t)| = |Z_i(s-t) - Z_i(0)|\ , \tag{7.10}$$

which holds since it holds for characters.

Our purpose is to study random trigonometric sums of the type $\sum_{i\ge 1} \varepsilon_i Z_i$ where ε_i are independent Bernoulli r.v.s, independent of the Z_i. For the time being these sums will always be finite, so the convergence of the series poses no problem. We set

$$X_t = \sum_{i\ge 1} \varepsilon_i Z_i(t)\ . \tag{7.11}$$

We aim to find upper and lower bounds for the quantity

$$\mathsf{E} \sup_{t\in T} |X_t| = \mathsf{E}\Big\|\sum_{i\ge 1} \varepsilon_i Z_i\Big\|\ , \tag{7.12}$$

where $\|\cdot\|$ denotes the supremum norm in the space of continuous functions on T. The reason why we consider sums of the type $\sum_{i\ge 1} \varepsilon_i Z_i$ rather than $\sum_i Z_i$ is that this amounts to consider sums of the type $\sum_i Z_i$ where the r.v.s Z_i are independent *symmetric*. We refer the reader to e.g. Proposition 8.1.5 of [1] for a detailed study of the "symmetrization procedure" in the setting of random Fourier series, a procedure showing that the symmetric case is the important one.

In order to avoid trivial situations we assume the following, where $\mathbf{1}$ denotes the unit of G, i.e. the character such that $\mathbf{1}(s) = 1$ for each s:

$$\forall\, i\ ,\ Z_i \notin \mathbb{C}\mathbf{1}\ a.s. \tag{7.13}$$

The leading idea of our approach is that, given the randomness of the Z_i, then $X_t = \sum_{i\ge 1} \varepsilon_i a_i \chi_i(t)$ where a_i are complex numbers and χ_i are characters. We can then use (3.22) to obtain lower bounds, and chaining to obtain upper bounds.

To give a more concrete example to which our setting applies, let us consider complex numbers a_i, only finitely many of which are not 0 and symmetric real-valued r.v.s ξ_i. We assume that $a_i = 0$ if $\chi_i = 1$. The problem is then to obtain upper and lower bounds for

$$\mathsf{E}\Big\|\sum_{i\geq 1} a_i \xi_i \chi_i\Big\| . \tag{7.14}$$

This is the special case of (7.12) where $Z_i = a_i \xi_i \chi_i$ is a r.v. valued in $\mathbb{C}G$. These r.v.s are independent symmetric (since it is the case for the r.v.s ξ_i), so that $\sum_{i\geq 1} Z_i$ and $\sum_{i\geq 1} \varepsilon_i Z_i$ have the same distribution.

We provided a partial answer to the question of bounding the quantity (7.14) in Section 3.2 under the condition that $\mathsf{E}\xi_i^2 < \infty$ for each i. Here we shall **not** assume that $\mathsf{E}\xi_i^2 < \infty$. The r.v.s ξ_i might have "fat tails" and the size of these tails governs the size of the quantity (7.14).

For $s, t \in T$ and $u \geq 0$ we consider the quantities

$$\varphi(s,t,u) = \sum_{i\geq 1} \mathsf{E}(|u(Z_i(s) - Z_i(t))|^2 \wedge 1) , \tag{7.15}$$

where $x \wedge 1 = \min(x, 1)$.

Given a number $r \geq 2$, for $j \in \mathbb{Z}$ we define

$$\varphi_j(s,t) = \varphi(s,t,r^j) . \tag{7.16}$$

Thus φ_j is the square of a translation-invariant distance on T. This "family of distances" is appropriate to estimate the quantity (7.12). For the purposes of this section it suffices to consider the case $r = 2$. Other values of r are useful for related purposes, so for consistency we allow the case $r > 2$, but at first reading there is no reason not to assume that $r = 2$ (which changes nothing to the proofs). We observe that $\varphi_{j+1} \geq \varphi_j$.

We also observe that since $|Z_i(t)| = |Z_i(0)|$,

$$\varphi_j(s,t) \leq \mathsf{E}\sum_{i\geq 1} |2r^j Z_i(0)|^2 \wedge 1 ,$$

and since the sum is finite it follows from dominated convergence that there exists j for which

$$\sup_{s,t\in T} \varphi_j(s,t) \leq \frac{1}{4} . \tag{7.17}$$

Our first result is a lower bound for the sum $\|\sum_{i\geq 1} \varepsilon_i Z_i\|$. It basically states that this sum is typically as large as the "entropy integral" (7.7) (computed of course for the previously defined "family of distances").

Theorem 7.3.1. *There exists a number $\alpha_0 > 0$ with the following property. According to (7.17) we may define*

$$j_0 = \sup\big\{j \in \mathbb{Z} \; ; \; \forall s,t \in T \; ; \; \varphi_j(s,t) \leq 1/4\big\} , \tag{7.18}$$

and, for $n \geq 1$,

$$j_n = \sup\{j \in \mathbb{Z}\ ;\ \mu(\{s\ ;\ \varphi_j(s,0) \leq 2^n\}) \geq 2^{-2^n} = N_n^{-1}\}\ . \tag{7.19}$$

(It may happen that $j_n = \infty$.) Then

$$\mathsf{P}\Big(\|\sum_{i\geq 1} \varepsilon_i Z_i\| > \frac{1}{K}\sum_{n\geq 0} 2^n r^{-j_n}\Big) \geq \alpha_0\ , \tag{7.20}$$

where K depends on r only.

The constant 1/4 in (7.17) is simply a convenient choice. Let us now investigate a possible converse to Theorem 7.3.1. Since the quantities j_n say nothing about the large values of Z_i, we cannot expect that the "entropy integral" $\sum_{n\geq 0} 2^n r^{-j_n}$ will control the tails of the r.v. $\|\sum_{i\geq 1}\varepsilon_i Z_i\|$. However, as the following expresses, we control the size of $\|\sum_{i\geq 1}\varepsilon_i Z_i\|$ as soon as we control the "entropy integral" $\sum_{n\geq 0} 2^n r^{-j_n}$ *and* the size of the single r.v. $\sum_{i\geq 1}\varepsilon_i Z_i(0)$.

Theorem 7.3.2. *For $n \geq 0$ consider numbers $j_n \in \mathbb{Z}$. Assume that*

$$\forall\, s,t \in T\ ,\ \varphi_{j_0}(s,t) \leq \frac{1}{4} \tag{7.21}$$

and

$$\mu(\{s\ ;\ \varphi_{j_n}(s,0) \leq 2^n\}) \geq 2^{-2^n} = N_n^{-1}\ . \tag{7.22}$$

Then, for any $p \geq 1$, we have

$$\Big(\mathsf{E}\|\sum_{i\geq 1}\varepsilon_i Z_i\|^p\Big)^{1/p} \leq K\Big(\sum_{n\geq 0} 2^n r^{-j_n} + \Big(\mathsf{E}|\sum_{i\geq 1}\varepsilon_i Z_i(0)|^p\Big)^{1/p}\Big)\ , \tag{7.23}$$

where K depends only on r and p.

Here for clarity we give only a statement that measures the size of the r.v. $\sum_{i\geq 1}\varepsilon_i Z_i(0)$ through its moments, but other statements using weaker ways to control the size of this variable are possible and in fact necessary to prove Theorem 7.3.4 below. Such a statement will be given in Lemma 7.6.5 below.

A surprising fact is that Theorem 7.3.2 is already of interest in the case where $Z_i = a_i\chi_i$ where a_i is a complex number and χ_i is a character. This situation was investigated in detail in Chapter 3, but Theorem 7.3.2 provides new information even in that case. This is part of an intriguing circle of facts and questions which will be detailed later (on page 221).

Together with Theorem 7.3.1, Theorem 7.3.2 allows upper and lower bounds for $(\mathsf{E}\|\sum_{i\geq 1}\varepsilon_i Z_i\|^p)^{1/p}$ that are of the same order. Let us state the result in the case of (7.14). From now on, K denotes a number that depends only on r and p, and that need not be the same on each occurrence.

Theorem 7.3.3. *Assume that the r.v.s ξ_i are independent symmetric. If the numbers j_n are as in Theorem 7.3.1, then, for each $p \geq 1$,*

$$\frac{1}{K}\Big(\sum_{n\geq 0} 2^n r^{-j_n} + \Big(\mathsf{E}\Big|\sum_{i\geq 1} a_i\xi_i\Big|^p\Big)^{1/p}\Big) \leq \Big(\mathsf{E}\Big\|\sum_{i\geq 1} a_i\xi_i\chi_i\Big\|^p\Big)^{1/p}$$
$$\leq K\Big(\sum_{n\geq 0} 2^n r^{-j_n} + \Big(\mathsf{E}\Big|\sum_{i\geq 1} a_i\xi_i\Big|^p\Big)^{1/p}\Big) . \tag{7.24}$$

Not the least remarkable feature of this result is that it assumes nothing (beyond independence and symmetry) on the r.v.s ξ_i.

Estimates as in Theorem 7.3.3 open wide the door to convergence theorems. We consider now independent r.v.s $(Z_i)_{i\geq 1}$ with $Z_i \in \mathbb{C}G$ and we study the convergence of the series $\sum_{i\geq 1}\varepsilon_i Z_i$, where of course ε_i are independent Bernoulli r.v.s independent of the randomness of the Z_i (so that the notation $\sum_{i\geq 1}\varepsilon_i Z_i$ no longer denotes a sum with finitely many terms). In this theorem we take $r = 2$.

Theorem 7.3.4. *The series $\sum_{i\geq 1}\varepsilon_i Z_i$ converges a.s (in the Banach space of continuous functions on T provided with the uniform norm) if and only if the following occurs. There exists j_0 such that*

$$\forall\ s,t\in T\ ,\ \sum_{i\geq 1}\mathsf{E}(|2^{j_0}(Z_i(s)-Z_i(t))|^2\wedge 1)\leq 1\ , \tag{7.25}$$

and for $n\geq 1$ there exists j_n for which

$$\mu\Big(\Big\{s\in T\ ;\ \sum_{i\geq 1}\mathsf{E}(|2^{j_n}(Z_i(s)-Z_i(0))|^2\wedge 1)\leq 2^n\Big\}\Big)\geq\frac{1}{N_n}\ , \tag{7.26}$$

and

$$\sum_{n\geq 1} 2^{n-j_n} < \infty\ . \tag{7.27}$$

Moreover, when these conditions are satisfied, for each $p\geq 1$ we have

$$\mathsf{E}\Big\|\sum_{i\geq 1}\varepsilon_i Z_i\Big\|^p<\infty \Leftrightarrow \mathsf{E}\Big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\Big|^p<\infty\ .$$

Explicit examples of application of these abstract theorems will be given in Section 7.7.

7.4 Proofs, Lower Bounds

Let us repeat that our approach will be to work given the r.v.s Z_i. Then $Z_i = a_i\chi_i$ where a_i is a complex number and χ_i is a character. It is therefore natural to consider the random distance

$$d_\omega(s,t)^2 = \sum_i |Z_i(s) - Z_i(t)|^2 (= \sum_i |a_i|^2 |\chi_i(s) - \chi_i(t)|^2) \,. \tag{7.28}$$

It is understood in this notation that the letter ω alludes to the given random choice of the Z_i. The proof of Theorem 7.3.1 has then two steps.

- First we prove that, given the randomness of the Z_i, the quantity

$$\|\sum_i \varepsilon_i Z_i\| = \|\sum_i \varepsilon_i a_i \chi_i\|$$

  is often of order about $\gamma_2(T, d_\omega)$.
- Then we prove that $\gamma_2(T, d_\omega)$ is often of order $\sum_{n\ge 0} 2^n r^{-j_n}$.

The first part of this program is performed by the following.

Lemma 7.4.1. *Consider complex numbers a_i and characters χ_i of T. Then*

$$\mathsf{P}\Big(\|\sum_{i\ge 1} \varepsilon_i a_i \chi_i\| \ge \frac{1}{L}\gamma_2(T,d)\Big) \ge \frac{1}{L} \,, \tag{7.29}$$

where the distance d is given by $d(s,t)^2 = \sum_{i\ge 1} |a_i|^2 |\chi_i(s) - \chi_i(t)|^2$.

Proof. The proof relies on the classical Paley-Zygmund inequality (sometimes called also the second moment method): for a r.v. $X \ge 0$,

$$\mathsf{P}\Big(X \ge \frac{1}{2}\mathsf{E}X\Big) \ge \frac{1}{4}\frac{(\mathsf{E}X)^2}{\mathsf{E}X^2} \,. \tag{7.30}$$

We then simply combine this inequality with (3.9) and (3.22). □

Exercise 7.4.2. Prove (7.30). (Hint: let $A = \{X \ge \mathsf{E}X/2\}$. Show that $\mathsf{E}X/2 \le \mathsf{E}(X\mathbf{1}_A) \le (\mathsf{E}X^2\mathsf{P}(A))^{1/2}$.)

The main step of the proof of Theorem 7.3.1 is to perform the second part of the program, to show that $\gamma_2(T, d_\omega)$ is typically as large as $\sum_{n\ge 0} 2^{n-j_n}$. An essential tool is the following fact, which provides an exponential control of certain deviations of special sums of independent r.v.s from their means. Of course, much more general and sharper results exist in the same direction, but the simple form we provide suffices for our needs.

Lemma 7.4.3. *Consider independent r.v.s $(W_i)_{i\ge 1}$, with $0 \le W_i \le 1$.*
(a) If $4A \le \sum_{i\ge 1} \mathsf{E}W_i$, then

$$\mathsf{P}\Big(\sum_{i\ge 1} W_i \le A\Big) \le \exp(-A) \,.$$

(b) If $A \ge 4\sum_{i\ge 1} \mathsf{E}W_i$, then

$$\mathsf{P}\Big(\sum_{i\ge 1} W_i \ge A\Big) \le \exp\Big(-\frac{A}{2}\Big) \,.$$

Proof. (a) Since $1-x \le e^{-x} \le 1-x/2$ for $0 \le x \le 1$, we have

$$\mathsf{E}\exp(-W_i) \le 1 - \frac{\mathsf{E}W_i}{2} \le \exp\Big(-\frac{\mathsf{E}W_i}{2}\Big)$$

and thus

$$\mathsf{E}\exp\Big(-\sum_{i\ge 1} W_i\Big) \le \exp\Big(-\frac{1}{2}\sum_{i\ge 1}\mathsf{E}W_i\Big) \le \exp(-2A) .$$

We conclude with the inequality $\mathsf{P}(Z \le A) \le \exp A\ \mathsf{E}\exp(-Z)$.
(b) Observe that $1+x \le e^x \le 1+2x$ for $0 \le x \le 1$, so, as before,

$$\mathsf{E}\exp\sum_{i\ge 1} W_i \le \exp 2\sum_{i\ge 1}\mathsf{E}W_i \le \exp\frac{A}{2}$$

and we use now that $\mathsf{P}(Z \ge A) \le \exp(-A)\mathsf{E}\exp Z$. □

Before we apply this result, let us recall a simple classical lower bound. If ε_i are independent Bernoulli r.v.s and b_i are complex numbers, then

$$\mathsf{P}\Big(\Big|\sum_{i\ge 1}\varepsilon_i b_i\Big| \ge \frac{1}{L}\Big(\sum_{i\ge 1}|b_i|^2\Big)^{1/2}\Big) \ge \frac{1}{L} . \tag{7.31}$$

To see this we observe that by the subgaussian inequality the r.v. $X = |\sum_{i\ge 1}\varepsilon_i b_i|^2$ satisfies $\mathsf{E}X^2 \le L(\sum_{i\ge 1}|b_i|^2)^2 = L(\mathsf{E}X)^2$, and we then apply the Paley-Zygmund inequality (7.30).

As a first illustration of the use of Lemma 7.4.3 we prove the following, which in a sense is a vast generalization of (7.31). The proof already reveals the main flavor of the proof of Theorem 7.3.1.

Lemma 7.4.4. *Consider independent complex-valued r.v.s U_i and independent Bernoulli r.v.s ε_i that are independent of the r.v.s U_i. Assume that for some number $w > 0$ we have*

$$\sum_{i\ge 1}\mathsf{E}(|wU_i|^2 \wedge 1) \ge \frac{1}{4}. \tag{7.32}$$

Then

$$\mathsf{P}\Big(\Big|\sum_{i\ge 1}\varepsilon_i U_i\Big| \ge \frac{1}{Lw}\Big) \ge \frac{1}{L} . \tag{7.33}$$

Proof. We use Lemma 7.4.3 (a) with $W_i = |wU_i|^2 \wedge 1$ and $A = 1/16$ to obtain

$$\mathsf{P}\Big(\sum_{i\ge 1}|wU_i|^2 \wedge 1 \ge \frac{1}{16}\Big) \ge \frac{1}{L} ,$$

so that in particular

$$\mathsf{P}\Big(\sum_{i\geq 1}|U_i|^2 \geq \frac{1}{16w^2}\Big) \geq \frac{1}{L}\ ,$$

and the conclusion follows using (7.31) given the randomness of $(U_i)_{i\geq 1}$. □

At this point we can try to explain in words the central idea of the (rest of the) proof of Theorem 7.3.1. If a point $s \in T$ satisfies

$$\varphi_{j_n+1}(s,0) \geq 2^n\ , \tag{7.34}$$

Lemma 7.4.3 (a) shows that for most of the choices of the randomness of the Z_i, we will have $\sum_{i\geq 1}|r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1 \geq 2^{n-2}$, and thus in particular $d_\omega(s,0) \geq 2^{n/2-1}r^{-j_n-1}$. The definition of j_n shows that all but very few of the points s satisfy (7.34). Thus for most of the choices of the randomness of the Z_i there will be only few points in T which satisfy $d_\omega(s,0) \leq 2^{n/2-1}r^{-j_n-1}$, and this certainly contributes to make $\gamma_2(T,d_\omega)$ large. (To be precise, it should contribute by about $2^n r^{-j_n-1}$ to this quantity). Using this information for many values of n at the same time carries the day. What the explanation in words does not reflect is that all the estimates fall very nicely into place.

Proof of Theorem 7.3.1. We assume that for a certain number M we have

$$\mathsf{P}\Big(\Big\|\sum_{i\geq 1}\varepsilon_i Z_i\Big\| > M\Big) \leq \alpha_0\ . \tag{7.35}$$

Our goal is to prove that

$$\sum_{n\geq 0} 2^n r^{-j_n} \leq KM\ , \tag{7.36}$$

where K depends on r only.

The value of α_0 will be determined later, but we assume right away that

$$\alpha_0 < \frac{1}{2L_0}\ ,$$

where L_0 is the constant of (7.33).

The first part of the argument is somewhat auxiliary. Its goal is to control the value of j_0. Since

$$\Big|\sum_{i\geq 1}\varepsilon_i(Z_i(s) - Z_i(t))\Big| \leq 2\Big\|\sum_i \varepsilon_i Z_i\Big\|\ ,$$

(7.35) implies

$$\forall\ s,t \in T\ ,\ \mathsf{P}\Big(\Big|\sum_{i\geq 1}\varepsilon_i(Z_i(s) - Z_i(t))\Big| \geq 2M\Big) \leq \alpha_0 < \frac{1}{2L_0}\ . \tag{7.37}$$

Consequently (7.33) fails when $U_i = Z_i(s) - Z_i(t)$ and $1/(L_0 w) = 2M$, i.e. $w = 1/(2ML_0)$, and therefore Lemma 7.4.4 implies

$$\forall\, s,t \in T\;,\; \sum_{i\ge 1} \mathsf{E}(|w(Z_i(s) - Z_i(t))|^2 \wedge 1) < \frac{1}{4}\;. \tag{7.38}$$

Let j^* be the largest integer with $r^{j^*} \le w$, so that (7.38) implies

$$\forall\, s,t \in T\;,\; \varphi_{j^*}(s,t) < \frac{1}{4}\;.$$

Consequently, $j_0 \ge j^*$. Also, the definition of j^* implies $r^{-j^*} \le r/w \le LrM$, so that $r^{-j_0} \le r^{-j^*} \le LrM$, the required control of j_0.

We now start the central argument. We show how to bound the value of $\gamma_2(T, d_\omega)$ from below. Consider an integer $n_0 \ge 5$ and let us assume that $j_n < \infty$ for $n \le n_0$. Our goal is to prove that the event

$$\sum_{5\le n\le n_0} 2^n r^{-j_n} \le Lr\gamma_2(T, d_\omega) \tag{7.39}$$

has probability $\ge 3/4$. The definition of j_n implies

$$\mu(\{s\;;\; \varphi_{j_n+1}(s,0) \le 2^n\}) < N_n^{-1}\;. \tag{7.40}$$

Consider $s \in T$ with

$$\varphi_{j_n+1}(s,0) = \sum_{i\ge 1} \mathsf{E}(|r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1) \ge 2^n\;.$$

Then, using Lemma 7.4.3 (a) with $W_i = |r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1$ and $A = 2^{n-2}$ we obtain

$$\mathsf{P}\Big(\sum_{i\ge 1} |r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1 \le 2^{n-2}\Big) \le e^{-2^{n-2}} \le N_{n-2}^{-1}\;. \tag{7.41}$$

Now, using that by (7.40) we integrate on a set of measure $\le 1/N_n$ a quantity ≤ 1 in the first integral below, and using that from (7.41) the integrand is $\le N_{n-2}$ in the second integral below, we obtain

$$\begin{aligned}
&\mathsf{E}\mu\Big(\Big\{s \in T\;;\; \sum_{i\ge 1} |r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1 \le 2^{n-2}\Big\}\Big)\\
&= \int_T \mathsf{P}\Big(\sum_{i\ge 1} |r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1 \le 2^{n-2}\Big)\mathrm{d}\mu(s)\\
&= \int_{\{\varphi_{j_n+1}(s,0)<2^n\}} + \int_{\{\varphi_{j_n+1}(s,0)\ge 2^n\}}\\
&\le \frac{1}{N_n} + \frac{1}{N_{n-2}} < \frac{2}{N_{n-2}}\;.
\end{aligned} \tag{7.42}$$

That is, the r.v. $Y = \mu(\{s \in T\ ;\ \sum_{i\geq 1} |r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1 \leq 2^{n-2}\})$ satisfies $\mathsf{E}Y < 2/N_{n-2}$, and by Markov's inequality it then satisfies $P(Y < 1/N_{n-3}) \geq 1 - 2N_{n-3}/N_{n-2} = 1 - 2/N_{n-3}$. Thus the event Ω_n defined by

$$\mu\Big(\Big\{s \in T\ ;\ \sum_{i\geq 1} |r^{j_n+1}(Z_i(s) - Z_i(0))|^2 \wedge 1 \leq 2^{n-2}\Big\}\Big) < \frac{1}{N_{n-3}} \qquad (7.43)$$

satisfies $\mathsf{P}(\Omega_n) \geq 1 - 2/N_{n-3}$. Consequently, the event

$$\Omega = \cap_{5\leq n\leq n_0} \Omega_n$$

satisfies $\mathsf{P}(\Omega) \geq 3/4$. Moreover, since

$$\sum_{i\geq 1} |r^{j_n+1}(Z_i(s) - Z_i(t))|^2 \wedge 1 \leq \sum_{i\geq 1} |r^{j_n+1}(Z_i(s) - Z_i(t))|^2 = r^{2j_n+2} d_\omega(s,t)^2\ ,$$

(7.43) yields

$$\mu\big(\big\{s \in T\ ;\ d_\omega(s,0) \leq r^{-j_n-1} 2^{n/2-1}\big\}\big) < \frac{1}{N_{n-3}}\ .$$

It follows that when Ω occurs the number $\epsilon_n = \epsilon_n(\omega)$ as in (3.1) satisfies $\epsilon_{n-3}(\omega) \geq r^{-j_n-1}2^{n/2-1}$ and (3.2) proves that (7.39) holds for $\omega \in \Omega$, and hence with probability $\geq 3/4$.

Having obtained the main information we are ready to conclude. We observe that (7.35) means

$$\mathsf{E}\mathsf{P}_\varepsilon\Big(\Big\|\sum_{i\geq 1} \varepsilon_i Z_i\Big\| \geq M\Big) \leq \alpha_0\ ,$$

so that since $\mathsf{P}(\Omega) \geq 3/4$ we can fix $\omega \in \Omega$ for which

$$\mathsf{P}_\varepsilon\Big(\Big\|\sum_{i\geq 1} \varepsilon_i Z_i\Big\| \geq M\Big) \leq 2\alpha_0\ . \qquad (7.44)$$

Given the r.v.s Z_i, and since $Z_i(t) \in \mathbb{C}G$, the sum $\sum_{i\geq 1} \varepsilon_i Z_i$ is of the type $\sum_{i\geq 1} a_i \varepsilon_i \chi_i$, where a_i is a complex number and χ_i is a character, so that (7.44) reads

$$\mathsf{P}_\varepsilon\Big(\Big\|\sum_{i\geq 1} a_i \varepsilon_i \chi_i\Big\| \geq M\Big) \leq 2\alpha_0\ . \qquad (7.45)$$

Consequently if we assume also that $2\alpha_0$ is less than the constant $1/L$ in the right-hand side of (7.29), this inequality implies that $\gamma_2(T, d_\omega) \leq LM$, and since (7.39) holds for $\omega \in \Omega$, we get

$$\sum_{5\leq n\leq n_0} 2^n r^{-j_n} \leq Lr\gamma_2(T, d_\omega) \leq LrM\ .$$

Since $r^{-j_0} \le LrM$, and since the sequence (j_n) is non-decreasing we finally get

$$\sum_{0 \le n \le n_0} 2^n r^{-j_n} \le LrM \ ,$$

and since n_0 is arbitrary the proof is complete. □

7.5 Proofs, Upper Bounds

We start the proof of Theorem 7.3.2. The main step is as follows, where we recall that E_ε denotes expectation in the r.v.s ε_i only.

Theorem 7.5.1. *For $n \ge 0$ consider numbers $j_n \in \mathbb{Z}$, and consider a parameter $v \ge 1$. Assume that*

$$\forall s, t \in T \ , \ \varphi_{j_0}(s,t) \le \frac{v}{4} \tag{7.46}$$

$$\forall\, n \ge 1 \ , \ \mu(\{s \ ; \ \varphi_{j_n}(s,0) \le v2^n\}) \ge 2^{-2^n} = N_n^{-1} \ . \tag{7.47}$$

Then for each $p \ge 1$ we can write

$$\Big(\mathsf{E}_\varepsilon \sup_{s \in T} \Big|\sum_{i \ge 1} \varepsilon_i (Z_i(s) - Z_i(0))\Big|^p\Big)^{1/p} \le Y_1 + Y_2 \ , \tag{7.48}$$

where

$$(\mathsf{E} Y_1^p)^{1/p} \le K(r,p)\sqrt{v} \sum_{n \ge 0} 2^n r^{-j_n} \ , \tag{7.49}$$

and

$$Y_2 \le K(r) \sum_{i \ge 1} |Z_i(0)| \mathbf{1}_{\{|Z_i(0)| \ge r^{-j_0}\}} \ . \tag{7.50}$$

The statement of this result will be less surprising if we keep in mind that it will ultimately follow from (5.12). It will then be a separate task to learn how to control the term Y_2.

We start the preparations for the proof of Theorem 7.5.1. We set $A_0 = T$ and for $n \ge 1$ we set

$$A_n = \{s \in T \ ; \ \varphi_{j_n}(s,0) \le 2^n\} \ ,$$

so that by (7.22) we have $\mu(A_n) \ge 1/N_n$. We consider a parameter $u \ge 1$.

Lemma 7.5.2. *There exists a constant L with the following property. For each $n \ge 0$ consider the* **random** *subset B_n of A_n defined as follows:*

$$B_n = B_{n,u} := \Big\{ s \in A_n \ ; \ \sum_{i \ge 1} |r^{j_n}(Z_i(s) - Z_i(0))|^2 \wedge 1 \le u2^{n+2} \Big\} \ . \tag{7.51}$$

Then for $u \ge L$ the event $\Omega_n(u)$ defined by $\mu(B_n) \ge 3\mu(A_n)/4$ satisfies

$$\mathsf{P}(\Omega_n(u)) \ge 1 - 4\exp(-u2^{n+1}) \ . \tag{7.52}$$

Proof. It follows from Lemma 7.4.3 (b), used with $A = u2^{n+2}$ that

$$s \in A_n \Rightarrow \mathsf{P}\Big(\sum_{i\geq 1} |r^{j_n}(Z_i(s) - Z_i(0))|^2 \wedge 1 \geq u2^{n+2}\Big) \leq \exp(-u2^{n+1}) \ . \quad (7.53)$$

Consequently if $\delta_n = \exp(-u2^{n+1})$, then $\mathsf{E}\mu(A_n \backslash B_n) \leq \delta_n \mu(A_n)$, so that $\mathsf{P}(\mu(A_n \backslash B_n) \geq \mu(A_n)/4) \leq 4\delta_n$by Markov's inequality. Therefore the event $\Omega_n(u)$ defined by $\mu(B_n) \geq 3\mu(A_n)/4$ satisfies (7.52). □

Proof of Theorem 7.5.1. We may assume that $\sum_{n\geq 0} 2^n r^{-j_n} < \infty$ for there is nothing to prove otherwise. Also, without loss of generality we may assume that the sequence (j_n) is non-decreasing. Then $\lim_{n\to\infty} j_n = \infty$. Given $u \geq 1$ we recall the sets $B_n = B_{n,u}$ and the event $\Omega_n(u)$ of Lemma 7.5.2. Consider the event

$$\Omega^{(k)} = \bigcap_{n\geq 0} \Omega_n(kv) \ , \quad (7.54)$$

so that for $k \geq 1$,

$$\mathsf{P}(\Omega^{(k)}) \geq 1 - L\exp(-k) \ . \quad (7.55)$$

We assume that $\Omega^{(k)}$ occurs, i.e that the sets

$$B_n = \Big\{s \in T \ ; \ \sum_{i\geq 1} |r^{j_n}(Z_i(s) - Z_i(0))|^2 \wedge 1 \leq kv2^{n+2}\Big\} \ , \quad (7.56)$$

satisfy

$$\forall n \geq 0 \ , \ \mu(B_n) \geq \frac{3}{4}\mu(A_n) \ . \quad (7.57)$$

We start the main chaining argument. This argument takes place given the randomness of the Z_i, so it helps to think of these as being fixed until further notice. We then have to control the supremum of a Bernoulli process, and the plan is to use Theorem 5.2.1 which organizes in an optimal manner the chaining for such processes.

The first part of the chaining argument will use translation invariance in a crucial manner and the fact that the sets B_n are not too small to construct an appropriate sequence of covering of T by translates of the sets $C_n := B_n - B_n$. If follows from Lemma 7.2.1 that we can find a subset T_n of T with $\operatorname{card} T_n \leq 1/\mu(B_n)$ and $T \subset T_n + C_n$. Since $\mu(B_0) \geq 3/4$ we have $\operatorname{card} T_0 = 1$ and for $n > 1$ since $\mu(B_n) \geq 3\mu(A_n)/4 \geq 1/(2N_n)$ we have $\operatorname{card} T_n \leq 2N_n$. Now, the definition of B_n and the fundamental property (7.10) show that $B_n = -B_n$, so that $C_n = B_n + B_n$. Consider $s \in C_n$, so that $s = t + t'$ for $t, t' \in C_n$. Thus $|r^{j_n}(Z_i(s) - Z_i(0))| \leq a + b$, where $a = |r^{j_n}(Z_i(t+t') - Z_i(t)| = |r^{j_n}(Z_i(t') - Z_i(0))|$ and $b = |r^{j_n}(Z_i(t) - Z_i(0))|$. Using the inequalities $(a+b) \wedge 1 \leq a \wedge 1 + b \wedge 1$ and $|a+b|^2 \leq 2|a|^2 + 2|b|^2$ we obtain

$$s \in C_n = B_n + B_n \Rightarrow \sum_{i\geq 1} |r^{j_n}(Z_i(s) - Z_i(0))|^2 \wedge 1 \leq kv2^{n+4} \ ,$$

and proceeding similarly,

$$s, s' \in t + C_n \Rightarrow \sum_{i \geq 1} |r^{j_n}(Z_i(s) - Z_i(s'))|^2 \wedge 1 \leq kv2^{n+6} \ . \tag{7.58}$$

In summary, we have succeeded to cover T by the sets $t+C_n$ for $t \in T_n$. There are not too many of these sets, and they are somewhat small by (7.58). This provides the appropriate information about the "smallness" of T. Consider the set $T^* \subset \ell^2$ of all sequences of the type $(Z_i(s) - Z_i(0))_{i \geq 1}$ for $s \in T$, so that $0 \in T^*$. The introduction of this set is motivated by the obvious identity

$$\sup_{s \in T} \Big| \sum_{i \geq 1} \varepsilon_i (Z_i(s) - Z_i(0)) \Big| = \sup_{x \in T^*} \Big| \sum_{i \geq 1} \varepsilon_i x_i \Big| \ . \tag{7.59}$$

Our goal is to apply Theorem 5.2.1 to the set T^* and for this purpose we construct the appropriate sequence of partitions. The idea of course is to use the smallness information we have about T and carry it to T^*. Starting with $\mathcal{B}_0 = \{T^*\}$ we construct inductively an increasing sequence $(\mathcal{B}_n)$ of partitions of T^* with the properties

$$\operatorname{card} \mathcal{B}_n \leq N_{n+2} \ , \tag{7.60}$$

$$B \in \mathcal{B}_n \Rightarrow \forall\ x, y \in B \ , \ \sum_{i \geq 1} |r^{j_n}(x_i - y_i)|^2 \wedge 1 \leq kv2^{n+6} \ . \tag{7.61}$$

For this, we simply observe from (7.58) that T^* can be covered by a family of $\leq 2N_n$ sets with the property (7.61), so we can achieve that each element of $\mathcal{B}_n$ satisfies (7.61) simply by partitioning each set of $\mathcal{B}_{n-1}$ into at most $2N_n$ pieces that satisfy (7.61), and (7.60) then follows by induction from the fact that $2N_{n+1}N_n \leq N_{n+2}$. We consider the admissible sequence $(\mathcal{A}_n)$ on T^* given by $\mathcal{A}_n = \{T^*\}$ for $n \leq 3$ and $\mathcal{A}_n = \mathcal{B}_{n-2}$ for $n \geq 3$. Then the conditions of Theorem 5.2.1 are satisfied with $j_n(A) = j_0 - 2$ if $A \in \mathcal{A}_n$ for $n \leq 2$, $j_n(A) = j_{n-2} - 2$ if $A \in \mathcal{A}_n$ for $n \geq 3$, and with $u = Lkv$. Consequently (5.12) yields

$$\Big(\mathsf{E}_\varepsilon \sup_{x \in T^*} \Big| \sum_{i \geq 1} x_i \varepsilon_i \Big|^p \Big)^{1/p} \leq K(r,p) \sqrt{kv} \sum_{n \geq 0} 2^n r^{-j_n} + K(r) \sup_{x \in T^*} \sum_{i \geq 1} |x_i| \mathbf{1}_{\{2|x_i| \geq r^{-j_0(T^*)}\}} \ . \tag{7.62}$$

This is the basic step, and we now explain how to control the last term above. For $x = (Z_i(s) - Z_i(0))_{i \geq 1} \in T^*$, we have, since $r \geq 2$ and since $j_0(T^*) = j_0 - 2$,

$$\sum_{i \geq 1} |x_i| \mathbf{1}_{\{2|x_i| \geq r^{-j_0(T^*)}\}} \leq \sum_{i \geq 1} |x_i| \mathbf{1}_{\{|x_i| \geq 2r^{-j_0(T^*)-2}\}} = \sum_{i \geq 1} |x_i| \mathbf{1}_{\{|x_i| \geq 2r^{-j_0}\}} \ .$$

Now $|x_i| = |Z_i(s) - Z_i(0)| \le |Z_i(s)| + |Z_i(0)| = 2|Z_i(0)|$, so that

$$\sum_{i\ge 1} |x_i| \mathbf{1}_{\{|x_i|\ge 2r^{-j_0}\}} \le 2 \sum_{i\ge 1} |Z_i(0)| \mathbf{1}_{\{|Z_i(0)|\ge r^{-j_0}\}}$$

and therefore

$$\sum_{i\ge 1} |x_i| \mathbf{1}_{\{2|x_i|\ge r^{-j_0(T^*)}\}} \le 2 \sum_{i\ge 1} |Z_i(0)| \mathbf{1}_{\{|Z_i(0)|\ge r^{-j_0}\}} .$$

Consequently (7.62) yields

$$\Big(\mathsf{E}_\varepsilon \sup_{x\in T^*} \Big|\sum_{i\ge 1} x_i \varepsilon_i\Big|^p\Big)^{1/p} \le K(r,p)\sqrt{kv} \sum_{n\ge 0} 2^n r^{-j_n} + K(r) \sum_{i\ge 1} |Z_i(0)| \mathbf{1}_{\{|Z_i(0)|\ge r^{-j_0}\}} ,$$

i.e., using (7.59) we have proved that for $\omega \in \Omega^{(k)}$ we have

$$\Big(\mathsf{E}_\varepsilon \sup_{s\in T} \Big|\sum_{i\ge 1} \varepsilon_i (Z_i(s) - Z_i(0))\Big|^p\Big)^{1/p} \le K(r,p)\sqrt{kv} \sum_{n\ge 0} 2^n r^{-j_n} + K(r) \sum_{i\ge 1} |Z_i(0)| \mathbf{1}_{\{|Z_i(0)|\ge r^{-j_0}\}} . \tag{7.63}$$

This finishes the main chaining argument.

We no longer assume that the r.v.s Z_i are fixed, and we define

$$Y_2 = K(r) \sum_{i\ge 1} |Z_i(0)| \mathbf{1}_{\{|Z_i(0)|\ge r^{-j_0}\}} ,$$

$$k(\omega) = \inf\{k \in \mathbb{N}^* \; ; \; \omega \in \Omega^{(k)}\}$$

and

$$Y_1(\omega) = K(r,p)\sqrt{k(\omega) v} \sum_{n\ge 0} 2^n r^{-j_n} ,$$

so that (7.63) implies

$$\Big(\mathsf{E}_\varepsilon \sup_{s\in T} \Big|\sum_{i\ge 1} \varepsilon_i (Z_i(s) - Z_i(0))\Big|^p\Big)^{1/p} \le Y_1 + Y_2 .$$

Moreover since $\{\sqrt{k(\omega)} \le \sqrt{k}\} \subset \Omega^{(k)}$ by definition of $k(\omega)$ and since the sequence $(\Omega^{(k)})$ increases, (7.55) implies that $\mathsf{E} k(\omega)^{p/2} \le K(r,p)$. This proves (7.49), which concludes the proof. □

To complete the proof of Theorem 7.3.2 we need to learn how to control $\mathsf{E} Y_2^p$. The basic reason we shall succeed is that typically not too many of the r.v.s $|Z_i(0)| \mathbf{1}_{\{|Z_i(0)|\ge r^{-j_0}\}}$ will be non-zero, and our first goal is to prove this. We start with a simple fact.

Lemma 7.5.3. *For each character $\chi \neq \mathbf{1}$ we have*

$$\int |\chi(s) - 1|^2 \mathrm{d}\mu(s) \geq \frac{3}{4} \,. \tag{7.64}$$

Proof. Consider $t \in T$ and observe that by translation invariance of μ,

$$\int |\chi(s+t) - 1|^2 \mathrm{d}\mu(s) = \int |\chi(s) - 1|^2 \mathrm{d}\mu(s) \,.$$

Since, for any $s \in T$,

$$|\chi(t) - 1|^2 = |\chi(s+t) - \chi(s)|^2 \leq 2|\chi(s+t) - 1|^2 + 2|\chi(s) - 1|^2 \,,$$

we obtain

$$\sup_{t \in T} |\chi(t) - 1|^2 \leq 4 \int |\chi(s) - 1|^2 \mathrm{d}\mu(s) \,.$$

Since we assume that $\chi \neq \mathbf{1}$ we can find t with $\chi(t) \neq 1$. Replacing if necessary t by $-t$ we can assume that $\chi(t) = \exp(i\theta)$ where $0 < \theta \leq \pi$. Then for some integer k we have $2\pi/3 \leq k\theta \leq 4\pi/3$ so that $|\chi(kt) - 1| \geq \sqrt{3}$, and hence $\sup_{t \in T} |\chi(t) - 1|^2 \geq 3$. □

Lemma 7.5.4. *For any $j \in \mathbb{Z}$ we have*

$$\sum_{i \geq 1} \mathsf{E}(|r^j Z_i(0)|^2 \wedge 1) \leq 2 \sup_{s,t \in T} \sum_{i \geq 1} \mathsf{E}(|r^j (Z_i(s) - Z_i(t))|^2 \wedge 1) \,. \tag{7.65}$$

In particular if j_0 satisfies (7.21) (i.e. (7.46) for $v = 1$) then

$$\sum_{i \geq 1} \mathsf{E}(|r^{j_0} Z_i(0)|^2 \wedge 1) \leq \frac{1}{2} \,. \tag{7.66}$$

One should stress the interesting nature of this statement: a control on the size of the *differences* $Z_i(s) - Z_i(t)$ implies a control of the size of $Z_i(0)$. The hypothesis (7.13) that $Z_i \notin \mathbb{C}\mathbf{1}$ a.e. is essential here. Let us also note that (7.66) implies that $\sum_i \mathsf{P}(|Z_i(0)| \geq r^{-j_0}) \leq 1/2$, so that in particular (as promised earlier) typically not too many of the r.v.s $|Z_i(0)|\mathbf{1}_{\{|Z_i(0)| \geq r^{-j_0}\}}$ can be non zero at the same time.

Proof. Assume that for a certain number C,

$$\forall\, s \in T \,, \sum_{i \geq 1} \mathsf{E}\big((r^{2j}|Z_i(s) - Z_i(0)|^2) \wedge 1\big) \leq C \,. \tag{7.67}$$

Since $Z_i \in \mathbb{C}G$, we have $Z_i(s) = \chi(s) Z_i(0)$ for a certain character χ, and since by (7.13) $\chi \neq \mathbf{1}$ a.e., (7.64) implies that a.e.

$$\int |Z_i(s) - Z_i(0)|^2 \mathrm{d}\mu(s) \geq \frac{3}{4} |Z_i(0)|^2 \,.$$

Using the concavity of the function $x \mapsto x \wedge 1$ and integrating (7.67) with respect to s we obtain

$$\sum_{i\geq 1} \mathsf{E}\Big((\frac{3}{4} r^{2j} |Z_i(0)|^2) \wedge 1 \Big) \leq C \ ,$$

and in particular

$$\sum_{i\geq 1} \mathsf{E}\Big((|r^j Z_i(0)|^2) \wedge 2 \Big) \leq 2C \ . \qquad \square$$

We need now two elementary facts. To lighten notation in the next few pages, $K = K(p)$ denotes a number depending on p only.

Lemma 7.5.5. *Consider independent centered complex valued r.v.s θ_i with*

$$\sum_{i\geq 1} \mathsf{P}(\theta_i \neq 0) \leq 1/2 \ .$$

Then, for each $p \geq 1$ we have

$$\sum_{i\geq 1} \mathsf{E}|\theta_i|^p \leq K \mathsf{E}\Big|\sum_{i\geq 1} \theta_i\Big|^p \ . \tag{7.68}$$

The intuition here is simply that there is no much cancellation in the sum $\sum_i \theta_i$ because the typical number of non-zero values of θ_i is about 1.

Proof. Assume first that θ_i is real-valued. We then prove by induction on n that

$$\sum_{i\leq n} \mathsf{E}|\theta_i|^p \leq 2\mathsf{E}\Big|\sum_{i\leq n} \theta_i\Big|^p \ . \tag{7.69}$$

It is obvious that (7.69) holds for $n = 1$. Assuming it holds for n, consider $\Omega_n = \{\exists i \leq n, \theta_i \neq 0\}$. Then $\mathsf{P}(\Omega_n) \leq 1/2$ by hypothesis, and

$$\mathsf{E}\Big|\sum_{i\leq n+1} \theta_i\Big|^p = \mathsf{E}\mathbf{1}_{\Omega_n}\Big|\sum_{i\leq n+1} \theta_i\Big|^p + \mathsf{E}\mathbf{1}_{\Omega_n^c}\Big|\sum_{i\leq n+1} \theta_i\Big|^p \ . \tag{7.70}$$

Now, since θ_{n+1} is independent of both Ω_n and $\sum_{i\leq n} \theta_i$, Jensen's inequality implies

$$\mathsf{E}\mathbf{1}_{\Omega_n}\Big|\sum_{i\leq n+1} \theta_i\Big|^p \geq \mathsf{E}\mathbf{1}_{\Omega_n}\Big|\sum_{i\leq n} \theta_i\Big|^p = \mathsf{E}\Big|\sum_{i\leq n} \theta_i\Big|^p \ , \tag{7.71}$$

and since for $i \leq n$ we have $\theta_i = 0$ on Ω_n^c,

$$\mathsf{E}\mathbf{1}_{\Omega_n^c}\Big|\sum_{i\leq n+1} \theta_i\Big|^p = \mathsf{E}\mathbf{1}_{\Omega_n^c}|\theta_{n+1}|^p = \mathsf{P}(\Omega_n^c)\mathsf{E}|\theta_{n+1}|^p \geq \frac{1}{2}\mathsf{E}|\theta_{n+1}|^p \ ,$$

using independence in the second equality. Combining with (7.70) and (7.71) and using the induction hypothesis, this proves (7.69) when θ_i is real-valued. Using (7.69) separately for the real and imaginary parts then proves (7.68). $\square$

Lemma 7.5.6. *Consider independent r.v.s* $\eta_i \ge 0$ *with* $\sum_{i\ge 1} \mathsf{P}(\eta_i > 0) \le 1$. *Then for each* $p \ge 1$,

$$\mathsf{E}\Big(\sum_{i\ge 1} \eta_i\Big)^p \le K \sum_{i\ge 1} \mathsf{E}\eta_i^p \ . \tag{7.72}$$

Again, the intuition here is that the typical number of non-zero values of η_i is about 1, so that $(\sum_{i\ge 1} \eta_i)^p$ is not much larger than $\sum_{i\ge 1} \eta_i^p$

Proof. The starting point of the proof is the inequality

$$(a+b)^p \le a^p + K(a^{p-1}b + b^p) \ , \tag{7.73}$$

where $a, b \ge 0$. This is elementary, by distinguishing the cases $b \le a$ and $b \ge a$. Let $S_n = \sum_{i\le n} \eta_i$, so that using (7.73) for $a = S_n$ and $b = \eta_{n+1}$ and taking expectation we obtain

$$\mathsf{E}S_{n+1}^p \le \mathsf{E}S_n^p + K(\mathsf{E}S_n^{p-1}\eta_{n+1} + \mathsf{E}\eta_{n+1}^p) \ . \tag{7.74}$$

Let $a_n = \mathsf{P}(\eta_n > 0)$. From Hölder's inequality we get

$$\mathsf{E}S_n^{p-1} \le (\mathsf{E}S_n^p)^{(p-1)/p} \ ; \ \mathsf{E}\eta_{n+1} \le a_{n+1}^{(p-1)/p}(\mathsf{E}\eta_{n+1}^p)^{1/p} \ .$$

Using independence then implies

$$\mathsf{E}S_n^{p-1}\eta_{n+1} = \mathsf{E}S_n^{p-1}\mathsf{E}\eta_{n+1} \le (\mathsf{E}S_n^p)^{(p-1)/p} a_{n+1}^{(p-1)/p}(\mathsf{E}\eta_{n+1}^p)^{1/p} \ .$$

Now, for numbers $a, b > 0$ we have $a^{(p-1)/p}b^{1/p} \le a + b$ and consequently

$$\mathsf{E}S_n^{p-1}\eta_{n+1} \le a_{n+1}\mathsf{E}S_n^p + \mathsf{E}\eta_{n+1}^p \ .$$

Combining with (7.74) yields

$$\mathsf{E}S_{n+1}^p \le \mathsf{E}S_n^p(1 + Ka_{n+1}) + K\mathsf{E}\eta_{n+1}^p \le (\mathsf{E}S_n^p + K\mathsf{E}\eta_{n+1}^p)(1 + Ka_{n+1}) \ .$$

In particular we obtain by induction on n that

$$\mathsf{E}S_n^p \le K\Big(\sum_{i\le n} \mathsf{E}\eta_i^p\Big) \prod_{i\le n}(1 + Ka_i) \ ,$$

which concludes the proof since $\sum_{i\ge 1} a_i \le 1$ by hypothesis. □

Finally, we observe a general fact. Combining the subgaussian inequality (3.11) with (2.22) yields the following, called Khinchin's inequality: For complex numbers (a_i),

$$\mathsf{E}\Big|\sum_{i\ge 1} \varepsilon_i a_i\Big|^p \le K(p)\Big(\sum_{i\ge 1} |a_i|^2\Big)^{1/2} \ . \tag{7.75}$$

Our next result provides the required control of $\mathsf{E}Y_2^p$.

Proposition 7.5.7. *Let us consider the r.v.s*

$$\eta_i := |Z_i(0)|\mathbf{1}_{\{|Z_i(0)|\geq r^{-j_0}\}} \,. \tag{7.76}$$

Then under (7.66) for each $p \geq 1$ we have

$$\Big(\mathsf{E}\big(\sum_{i\geq 1}\eta_i\big)^p\Big)^{1/p} \leq K\Big(r^{-j_0} + \Big(\mathsf{E}\big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\big|^p\Big)^{1/p}\Big) \,. \tag{7.77}$$

Proof. Let us define $\theta_i := Z_i(0)\mathbf{1}_{\{|Z_i(0)|\geq r^{-j_0}\}}$ and $\theta_i' := Z_i(0) - \theta_i = Z_i(0)\mathbf{1}_{\{|Z_i(0)|<r^{-j_0}\}}$. First, Khinchin's inequality (7.75) implies

$$\mathsf{E}_\varepsilon\big|\sum_{i\geq 1}\varepsilon_i\theta_i'\big|^p \leq K\Big(\sum_{i\geq 1}|\theta_i'|^2\Big)^{p/2} \,. \tag{7.78}$$

Consider the r.v.s $W_i = r^{2j_0}\theta_i'^2$, so that $0 \leq W_i \leq 1$ and $\sum_{i\geq 1}\mathsf{E}W_i \leq 1/2$ by (7.66). Lemma 7.4.3 (b) provides the estimate $\mathsf{P}(\sum_{i\geq 1}W_i \geq t) \leq \exp(-t/2)$ for $t \geq 1$ and as in (2.18) this implies $\mathsf{E}\big(\sum_{i\geq 1}W_i\big)^{p/2} \leq K$. Consequently taking expectation in (7.78) yields

$$\mathsf{E}\big|\sum_{i\geq 1}\varepsilon_i\theta_i'\big|^p \leq Kr^{-j_0 p} \,,$$

and therefore $(\mathsf{E}|\sum_{i\geq 1}\varepsilon_i\theta_i'|^p)^{1/p} \leq Kr^{-j_0}$. Since $\theta_i = Z_i(0) - \theta_i'$ it follows that

$$\Big(\mathsf{E}\big|\sum_{i\geq 1}\varepsilon_i\theta_i\big|^p\Big)^{1/p} \leq Kr^{-j_0} + K\Big(\mathsf{E}\big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\big|^p\Big)^{1/p} \,. \tag{7.79}$$

On the other hand, when $\theta_i \neq 0$ we have $|r^{j_0}Z_i(0)|^2 \wedge 1 = 1$, so that, again from (7.66),

$$\sum_{i\geq 1}\mathsf{P}(\theta_i \neq 0) \leq \frac{1}{2} \,, \tag{7.80}$$

and since $\eta_i = |\theta_i|$, combining (7.79), (7.68) (used for $\varepsilon_i\theta_i$ rather than θ_i) and (7.72) completes the proof. □

Proof of Theorem 7.3.2. We use Theorem 7.5.1 with $v = 1$. We raises (7.48) to the power p, we use that $(Y_1+Y_2)^p \leq K(Y_1^p+Y_2^p)$ and we take expectation. It follows from (7.66) that we can use (7.77) to control $\mathsf{E}Y_2^p$. □

Let us now investigate the content of Theorem 7.3.3 in a now classical case, discovered by M. Marcus and G. Pisier [5]. Assume that for complex numbers a_i we have $\xi_i = a_i\theta_i$ where the r.v.s θ_i are symmetric and satisfy, for a certain number $1 < p < 2$

$$\mathsf{P}(|\theta_i| \geq u) \leq Cu^{-p} \ . \tag{7.81}$$

Let us define the distance d_p on T by

$$d_p(s,t)^p = \sum_{i\geq 1} |a_i(\chi_i(s) - \chi_i(t))|^p \ .$$

With the notation of Theorem 7.3.1 we have the following, where K denotes a constant depending only on C, r and p, and $1/p + 1/q = 1$.

Proposition 7.5.8. *Under the preceding conditions we have*

$$\sum_{n\geq 0} 2^n r^{-j_n} \leq K\gamma_q(T, d_p) \ .$$

Proof. Using (2.10) in the first line and (7.81) in the third line we obtain that for $v \neq 0$,

$$\begin{aligned} \mathsf{E}(|v\theta_i|^2 \wedge 1) &= \int_0^1 \mathsf{P}(|v\theta_i|^2 \geq t)\mathrm{d}t \\ &= \int_0^1 \mathsf{P}\Big(|\theta_i| \geq \frac{t^{1/2}}{|v|}\Big)\mathrm{d}t \\ &\leq \int_0^1 C\frac{|v|^p}{t^{p/2}}\mathrm{d}t \\ &= K_0(C,p)|v|^p \ . \end{aligned} \tag{7.82}$$

Consider for $n \geq 0$ the numbers ϵ_n as in Theorem 3.1.1, for the distance d_p, so that

$$\mu(\{d_p(s,0) \leq \epsilon_n\}) \geq N_n^{-1} \ , \tag{7.83}$$

and $\sum_{n\geq 0} \epsilon_n 2^{n/q} \leq K_0(C,p)\gamma_q(T,d_p)$ by (3.4). Now, using (7.82) for $v = a_i r^j(\chi_i(s) - \chi_i(0))$ and summing over i implies

$$\varphi_j(s,0) \leq K_0(C,p) r^{jp} d_p(s,0)^p \ . \tag{7.84}$$

Consider for each n the largest integer k_n such that $K_0(C,p)(r^{k_n}\epsilon_n)^p \leq 2^n$, so that

$$d_p(s,0) \leq \epsilon_n \Rightarrow \varphi_{k_n}(s,0) \leq 2^n \ .$$

Then (7.83) yields

$$\mu(\{\varphi_{k_n}(s,0) \leq 2^n\}) \geq N_n^{-1} \ ,$$

and therefore $k_n \leq j_n$ by definition of j_n. Moreover, by definition of k_n we have $K_0(C,p)(r^{k_n+1}\epsilon_n)^p > 2^n$ and therefore $r^{-k_n} \leq K2^{-n/p}\epsilon_n$, so that

$$\sum_{n\geq 0} 2^n r^{-j_n} \leq \sum_{n\geq 0} 2^n r^{-k_n} \leq LK \sum_{n\geq 0} 2^{n/q}\epsilon_n \ . \qquad \square$$

In the case where

$$\forall u \geq 1 \; , \; \mathsf{P}(|\theta_i| \geq u) \geq \frac{1}{C} u^{-p} \; , \tag{7.85}$$

it is known (from the work of M. Marcus and G. Pisier [5]) that the following converse to Proposition 7.5.8 is true (and we prove it in Exercise 8.1.6 below).

$$\gamma_q(T, d_p) \leq K \mathsf{E} \Big\| \sum_{i \geq 1} a_i \theta_i \chi_i \Big\| \; . \tag{7.86}$$

Interestingly, it does not seem obvious how to deduce this from Theorem 7.3.3. This deserves some detailed comments. In retrospect, basically the same question arose earlier. Consider complex numbers a_i and the distance d given by $d(s,t)^2 = \sum_{i \geq 1} |a_i(\chi_i(s) - \chi_i(t))|^2$. Consider the numbers j_n defined as in the case of Theorem 7.3.2 for the sum $\sum_{i \geq 1} a_i \varepsilon_i \chi_i$. Then this theorem implies

$$\mathsf{E} \Big\| \sum_{i \geq 1} a_i \varepsilon_i \chi_i \Big\| \leq L \sum_{n \geq 0} 2^n r^{-j_n} + L \Big(\sum_{i \geq 1} |a_i|^2 \Big)^{1/2} \; ,$$

On the other hand, it follows from (3.30) that

$$\gamma_2(T, d) \leq L \mathsf{E} \Big\| \sum_{i \geq 1} a_i \varepsilon_i \chi_i \Big\| \; .$$

Moreover since $|Z_i(0)| = |a_i|$, (7.66) implies $(\sum_{i \geq 1} |a_i|^2)^{1/2} \leq L r^{-j_0}$. Consequently,

$$\gamma_2(T, d) \leq L \sum_{n \geq 0} 2^n r^{-j_n} \; . \tag{7.87}$$

This inequality does not seem obvious to prove directly (i.e. without using trigonometric sums) either. In particular, it would be very surprising if such a proof did not use Theorem 5.2.1 (or and equivalent principle). In some sense, one might be tempted to say that the upper bound in Theorem 7.3.3 is so effective that it looks smaller than the more traditional lower bounds such as (7.86).

7.6 Proofs, Convergence

In this part for simplicity we use only the case $r = 2$. After the hard work of proving inequalities has been completed, the proof of Theorem 7.3.4 involves only "soft arguments". In order to avoid repetition we separate a part of the argument that will be used again later. The following is a version of Theorem 7.3.1 adapted to infinite sums. We recall the number α_0 of this theorem.

Lemma 7.6.1. *Consider an independent sequence $(Z_i)_{i\geq 1}$ with $Z_i \in \mathbb{C}G$, and let $S_k = \sum_{1\leq i\leq k} \varepsilon_i Z_i$, where of course the ε_i Bernoulli r.v.s are independent of the Z_i. Assume that for each k we have*

$$\mathsf{P}(\|S_k\| \geq M) \leq \alpha_0 \; . \tag{7.88}$$

For $j \in \mathbb{Z}$ define as usual

$$\varphi_j(s,t) = \sum_{i\geq 1} \mathsf{E}(|2^j(Z_i(s) - Z_i(t))|^2 \wedge 1) \; . \tag{7.89}$$

Then we can find integers $(j_n)_{n\geq 0}$ such that

$$\forall\, s,t \in T \; , \; \varphi_{j_0}(s,t) \leq 1/4 \tag{7.90}$$

$$\mu(\{s \; ; \; \varphi_{j_n}(s,0) \leq 2^n\}) \geq N_n^{-1} \; , \tag{7.91}$$

and

$$\sum_{n\geq 0} 2^{n-j_n} \leq LM \; . \tag{7.92}$$

Proof. Let us define

$$\varphi_{k,j}(s,t) = \sum_{i\leq k} \mathsf{E}(|2^j(Z_i(s) - Z_i(t))|^2 \wedge 1) \; ,$$

so that

$$\varphi_j(s,t) = \lim_{k\to\infty} \varphi_{k,j}(s,t) \; . \tag{7.93}$$

Using Theorem 7.3.1 for $r = 2$ implies that for each k we can find numbers $(j_{k,n})_{n\geq 0}$ for which

$$\forall\, s,t \in T \; ; \; \varphi_{k,j_{k,0}}(s,t) \leq \frac{1}{4} \; ,$$

and, for $n \geq 0$,

$$\mu(\{s \; ; \; \varphi_{k,j_{k,n}}(s,0) \leq 2^n\}) \geq N_n^{-1}$$

such that the following holds:

$$\sum_{n\geq 0} 2^{n-j_{k,n}} \leq LM \; . \tag{7.94}$$

The conclusion will then follow by a straightforward limiting argument that we detail now. Consider first any sequence (j_n^*) such that $\sum_{n\geq 0} 2^{n-j_n^*} \leq KM$. Without loss of generality we may assume that $j_{k,n} \leq j_n^*$ simply by replacing $j_{k,n}$ by $\min(j_{k,n}, j_n^*)$. Also, (7.94) shows that for each n, $j_{k,n}$ stays bounded below independently of k. Thus we can find a sequence $(k(q))$ with $k(q) \to \infty$ such that for each n, $j_n = \lim_{q\to\infty} j_{k(q),n}$ exists. By taking a further subsequence if necessary, we may assume that

$$0 \le n \le q \Rightarrow j_{k(q),n} = j_n \ .$$

Consequently

$$\forall\, s,t \in T\ ;\ \varphi_{k(q),j_0}(s,t) \le \frac{1}{4}\ , \tag{7.95}$$

and

$$\mu(\{s\ ;\ \varphi_{k(q),j_n}(s,0) \le 2^n\}) \ge N_n^{-1} \tag{7.96}$$

for $1 \le n \le q$, while, from (7.94),

$$\sum_{0\le n\le q} 2^{n-j_n} = \sum_{0\le n\le q} 2^{n-j_{k(q),n}} \le LM\ .$$

Letting $q \to \infty$ proves that $\sum_{n\ge 0} 2^{n-j_n} \le LM$. On the other hand, (7.93) implies $\varphi_j(s,t) = \lim_{q\to\infty} \varphi_{k(q),j}(s,t)$. Together with (7.95) and (7.96) this proves that

$$\forall\, s,t \in T\ ;\ \varphi_{j_0}(s,t) \le \frac{1}{4}\ ,$$

and for each n,

$$\mu(\{s\ ;\ \varphi_{j_n}(s,0) \le 2^n\}) \ge N_n^{-1}\ . \qquad \square$$

To prove convergence of a series of independent symmetric r.v.s we shall use the following.

Lemma 7.6.2. *Consider independent symmetric Banach space valued r.v.s W_i. Then the series $\sum_{i\ge 1} W_i$ converges a.s. if and only if it is a Cauchy sequence in measure, i.e.*

$$\forall\, \delta > 0\ ,\ \exists k_0\ ,\ k_0 \le k \le n \Rightarrow \mathsf{P}\Big(\|\sum_{k\le i\le n} W_i\| \ge \delta\Big) \le \delta\ . \tag{7.97}$$

Proof. It suffice to prove that (7.97) implies convergence. Let $S_k = \sum_{i\le k} W_i$. Then the Lévy inequality

$$\mathsf{P}\Big(\sup_{k\le n} \|S_k\| \ge a\Big) \le 2\mathsf{P}(\|S_n\| \ge a)$$

(see [2], page 47, equation (2.6)) implies

$$\mathsf{P}\Big(\sup_{k} \|S_k\| \ge a\Big) \le 2 \sup_n \mathsf{P}(\|S_n\| \ge a)\ ,$$

and starting the sum at an integer k_0 as in (7.97) rather than at 1 we obtain

$$\mathsf{P}\Big(\sup_{k} \|S_k - S_{k_0}\| \ge a\Big) \le 2 \sup_n \mathsf{P}(\|S_n - S_{k_0}\| \ge a)\ .$$

For $a = \delta$ the right hand side above is $\le \delta$ and this proves that

$$\mathsf{P}(\sup_{k_0\le k\le n} \|S_n - S_k\| \ge 2\delta) \le 4\delta\ ,$$

and in turn that a.s. the sequence $(S_k(\omega))_{k\ge 1}$ is a Cauchy sequence. $\square$

Corollary 7.6.3. *If the r.v.s W_i are independent symmetric real-valued then the series $\sum_{i\geq 1} W_i$ converges a.s. provided for some $a > 0$ (or, equivalently, all $a > 0$) we have*

$$\sum_{i\geq 1} \mathsf{E}(W_i^2 \wedge a^2) < \infty . \tag{7.98}$$

Proof. Since $a^2\mathsf{P}(|W_i| \geq a) \leq \mathsf{E}(W_i^2 \wedge a^2)$, the series $\sum_{i\geq 1} \mathsf{P}(|W_i| \geq a)$ converges, and so does the series $\sum_{i\geq 1} W_i \mathbf{1}_{\{|W_i|>a\}}$ because a.s. it has only finitely many non-zero terms. Thus it suffices to prove the convergence of the series $\sum_{i\geq 1} W_i \mathbf{1}_{\{|W_i|\leq a\}}$, but symmetry and (7.98) imply that this series converges in L^2 and hence in measure. The conclusion then follows from Lemma 7.6.2. □

Exercise 7.6.4. Prove the converse of Corollary 7.6.3.

In the next lemma, we assume again that only finitely many of the r.v.s Z_i are not zero. It will be applied to control partial sums.

Lemma 7.6.5. *Consider numbers $(j_n)_{n\geq 0}$, consider $v \geq 1$ and assume (7.46) and (7.47). Consider a number $w \geq 2^{-j_0}$. Then the event*

$$\Omega = \bigcap_{i\geq 1} \{|Z_i(0)| \leq w\} \tag{7.99}$$

satisfies

$$\mathsf{P}(\Omega^c) \leq \sum_{i\geq 1} \mathsf{P}(|Z_i(0)| \geq w) \tag{7.100}$$

and

$$\mathsf{E}\mathbf{1}_\Omega \Big\|\sum_{i\geq 1} \varepsilon_i Z_i\Big\| \leq L\sqrt{v}\sum_{n\geq 0} 2^{n-j_n} + L\sum_{i\geq 1} \mathsf{E}|Z_i(0)|\mathbf{1}_{\{2^{-j_0}\leq |Z_i(0)|\leq w\}} . \tag{7.101}$$

Proof. It is obvious that Ω satisfies (7.100). To prove (7.101) we apply Theorem 7.5.1 with $p = 1$ to obtain

$$\mathsf{E}_\varepsilon \sup_{s\in T}\Big|\sum_{i\geq 1} \varepsilon_i (Z_i(s) - Z_i(0))\Big| \leq Y_1 + Y_2 , \tag{7.102}$$

where

$$\mathsf{E}Y_1 \leq L\sqrt{v}\sum_{n\geq 0} 2^{n-j_n} , \tag{7.103}$$

and

$$Y_2 \leq L\sum_{i\geq 1} |Z_i(0)|\mathbf{1}_{\{|Z_i(0)|\geq 2^{-j_0}\}} . \tag{7.104}$$

Since Ω is independent of the randomness of the sequence (ε_i),

$$\mathsf{E}_\varepsilon \mathbf{1}_\Omega \sup_{s\in T}\Big|\sum_{i\geq 1}\varepsilon_i(Z_i(s)-Z_i(0))\Big| = \mathbf{1}_\Omega \mathsf{E}_\varepsilon \sup_{s\in T}\Big|\sum_{i\geq 1}\varepsilon_i(Z_i(s)-Z_i(0))\Big|$$
$$\leq \mathbf{1}_\Omega Y_1 + \mathbf{1}_\Omega Y_2 \,. \tag{7.105}$$

Now

$$\mathbf{1}_\Omega Y_2 \leq Y_3 := \sum_{i\geq 1}|Z_i(0)|\mathbf{1}_{\{2^{-j_0}\leq |Z_i(0)|\leq w\}} \,.$$

Taking expectation in (7.105) yields

$$\mathsf{E}\mathbf{1}_\Omega \sup_{s\in T}\Big|\sum_{i\geq 1}\varepsilon_i(Z_i(s)-Z_i(0))\Big| \leq L\sqrt{v}\sum_{n\geq 0}2^{n-j_n} + \mathsf{E}Y_3 \,. \tag{7.106}$$

Also,

$$\Big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\Big| \leq \Big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\mathbf{1}_{\{|Z_i(0)|\leq 2^{-j_0}\}}\Big| + Y_2 \,,$$

so that

$$\mathbf{1}_\Omega\Big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\Big| \leq \Big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\mathbf{1}_{\{|Z_i(0)|\leq 2^{-j_0}\}}\Big| + Y_3 \,.$$

Now, using the Cauchy-Schwarz inequality in the second line,

$$\begin{aligned}\mathsf{E}\mathsf{E}_\varepsilon\Big|\sum_{i\geq 1}\varepsilon_i Z_i(0)\mathbf{1}_{\{|Z_i(0)|\leq 2^{-j_0}\}}\Big| &\leq \mathsf{E}\Big(\sum_{i\geq 1}|Z_i(0)|^2\mathbf{1}_{\{|Z_i(0)|\leq 2^{-j_0}\}}\Big)^{1/2} \\ &\leq \Big(\mathsf{E}\sum_{i\geq 1}|Z_i(0)|^2\mathbf{1}_{\{|Z_i(0)|\leq 2^{-j_0}\}}\Big)^{1/2} \\ &\leq 2^{-j_0}\Big(\sum_{i\geq 1}\mathsf{E}|2^{j_0}Z_i(0)|^2\wedge 1\Big)^{1/2} \\ &\leq L\sqrt{v}2^{-j_0} \,,\end{aligned}$$

where the last equality follows from (7.46) and Lemma 7.5.4. Combining with (7.106) we obtain

$$\mathsf{E}\mathbf{1}_\Omega \sup_{s\in T}\Big|\sum_{i\geq 1}\varepsilon_i Z_i(s)\Big| \leq L\sqrt{v}\sum_{n\geq 0}2^{n-j_n} + \mathsf{E}Y_3 \,. \qquad \square$$

Only the case $v=1$ will be used in the present chapter. The case $v\geq 1$ will be used for later purposes.

Proof of Theorem 7.3.4. Assume first the convergence a.s. of the series $\sum_{i\geq 1}\varepsilon_i Z_i$. Let $S_k=\sum_{i\leq k}\varepsilon_i Z_i$, so that the sequence (S_k) converges a.s. Consequently, there exists M such that for each k we have that $\mathsf{P}(\|S_k\|\geq M)\leq \alpha_0$, where α_0 is as in Theorem 7.3.1, and Lemma 7.6.1 implies (7.25) to (7.27).

Let us now turn to the proof of the converse, so we assume (7.25) to (7.27) and our goal is to prove that the sequence (S_k) is a Cauchy sequence for the convergence in measure. For $k \ge 1$ let us define

$$\varphi_{k,j}(s,t) = \sum_{i \ge k} \mathsf{E}(|2^j(Z_i(s) - Z_i(t))|^2 \wedge 1) \ .$$

(Warning: the subscript k in $\varphi_{k,j}$ now means that the summation *starts* at k.) Since (7.25) implies that $\varphi_{1,j_0}(s,0) < \infty$ for each s, and hence $\varphi_{1,j}(s,0) < \infty$ for each s, for each j one has $\lim_{k\to\infty} \varphi_{k,j}(s,0) = 0$. Consequently, given j, there exists k for which $\mu(A) \ge 3/4$, where $A = \{s \in T, \varphi_{k,j}(s,0) < 1/16\}$. Given s and t in T, we have $(A+s)\cap(A+t) \ne \emptyset$, so that if $u \in (A+s)\cap(A+t)$ then $u - s \in A$ and $u - t \in A$. Using the inequality $\varphi_{k,j}(s,t) \le 2\varphi_{k,j}(s,u) + 2\varphi_{k,j}(u,t)$ we then obtain

$$\forall\ s,t \in T\ ,\ \varphi_{k,j}(s,t) \le \frac{1}{4}\ . \tag{7.107}$$

Thus, given j we have proved the existence of k for which (7.107) holds. Consider $\delta > 0$ and n_0 large enough that $\sum_{n \ge n_0} 2^{n-j_n} \le \delta$. Consider $j^* \ge j_0$ large enough that $2^{n_0 - j^*} < \delta/n_0$, and set $j_n^* = \max(j^*, j_n)$. Then

$$\sum_{n\ge 0} 2^{n-j_n^*} \le n_0 2^{n_0-j^*} + \sum_{n \ge n_0} 2^{n-j_n} \le 2\delta\ .$$

Let us then find k for which

$$\forall\ s,t \in T\ ,\ \varphi_{k,j^*}(s,t) \le \frac{1}{4}\ . \tag{7.108}$$

Next we observe that for each $n \ge 1$,

$$\mu\big(\{s \in T\ ;\ \varphi_{k,j_n^*} \le 2^n\}\big) \ge \frac{1}{N_n}\ . \tag{7.109}$$

Indeed this follows from (7.108) if $j_n^* = j^*$ and from (7.26) if $j_n^* = j_n$. Next, we appeal to (7.65), but starting the sums at $i = k$ rather than at $i = 1$. It then follows from (7.108) that $\sum_{i \ge k} \mathsf{E}(|2^{j^*} Z_i(0)|^2 \wedge 1) \le 1/2$, and, consequently $\sum_{i\ge 1} \mathsf{P}(|Z_i(0)| \ge 2^{-j^*}) < \infty$.

The next step is to use Lemma 7.6.5 for $v = 1$, j_n^* rather than j_n and $w = 2^{-j_0^*}$, starting the sums at k rather than at 1. We observe that (7.46) follows from (7.108) while (7.47) follows from (7.109). The last term in the right-hand side (7.101) is 0 by the choice of w. Consequently this inequality shows that the set $\Omega = \cap_{i\ge 1}\{|Z_i(0)| \le 2^{-j_0^*}\}$ satisfies, for any $k' \ge k$

$$\mathsf{E}\mathbf{1}_\Omega \|S_{k'} - S_k\| = \mathsf{E}\mathbf{1}_\Omega \Big\| \sum_{k \le i \le k'} \varepsilon_i Z_i \Big\| \le L \sum_{n \ge 0} 2^{n-j_n^*} \le L\delta\ , \tag{7.110}$$

while

$$\mathsf{P}(\Omega^c) \le \sum_{i \ge k} \mathsf{P}(|Z_i(0)| \ge 2^{-j_0^*}) \;. \tag{7.111}$$

From (7.110) we infer that

$$\mathsf{P}(\|S_{k'} - S_k\| \ge \sqrt{\delta}) \le \mathsf{P}(\mathbf{1}_\Omega \|S_{k'} - S_k\| \ge \delta) + \mathsf{P}(\Omega^c) \le L\sqrt{\delta} + \mathsf{P}(\Omega^c) \;,$$

and (7.111) shows that for k large enough the right-hand side is $\le L\sqrt{\delta}$. This proves that the sequence (S_k) is a Cauchy sequence for the convergence in measure, and Lemma 7.6.2 completes the proof of the convergence of the series $\sum_{i \ge 1} \varepsilon_i Z_i$.

The last statement of the theorem then follows from Theorem 7.3.2. □

7.7 Explicit Computations

In this section we give some examples of concrete results that follow from the abstract theorems that we stated. The link between the abstract theorems and the classical results of Paley and Zygmund and Salem and Zygmund has been thoroughly investigated by Marcus and Pisier [4], and there is no point to reproduce it here. Rather, we develop a specific direction that definitely goes beyond these results. It was initiated in [11] and generalized in [1]. There is actually a seemingly infinite number of variations on the theme we present. We shall consider only questions of convergence. We use the notation of Exercise 3.2.5, so that T is the group of complex numbers of modulus 1, and for $t \in T$, $\chi_i(t) = t^i$ is the i-the power of t. We consider independent r.v.s $(X_i)_{i \ge 1}$ and complex numbers $(a_i)_{i \ge 1}$, and we are interested in the case where

$$Z_i(t) = a_i X_i \chi_i(t) = a_i X_i t^i \;. \tag{7.112}$$

We make the following assumption

$$\sum_{i \ge 1} \mathsf{E}(|a_i X_i|^2 \wedge 1) < \infty \;. \tag{7.113}$$

To study the convergence of the series, without loss of generality, we assume that $a_i \ne 0$ for each i.

Theorem 7.7.1. *Under the previous conditions, for $n \ge 0$ there exists a number λ_n such that*

$$\sum_{i \ge N_n} \mathsf{E}\Big(\frac{|a_i X_i|^2}{\lambda_n^2} \wedge 1\Big) = 2^n \;, \tag{7.114}$$

and the series $\sum_{i \ge i} a_i \varepsilon_i X_i \chi_i$ converges uniformly a.s. whenever

$$\sum_{n \ge 0} 2^n \lambda_n < \infty \;. \tag{7.115}$$

As a consequence we obtain the following (which we leave as an exercise to compare with (3.19)).

Corollary 7.7.2. *If*

$$\sum_{n\geq 0} 2^{n/2}\Big(\sum_{i\geq N_n} |a_i|^2\Big)^{1/2} < \infty \ ,$$

then the series $\sum_{i\geq 1} a_i\varepsilon_i\chi_i$ *converges uniformly a.s.*

Proof. Since $\lambda_n^2 \leq 2^{-n}\sum_{i\geq N_n}|a_i|^2$ because $|X_i| = |\varepsilon_i| = 1$, (7.115) holds. □

Proof of Theorem 7.7.1. First we observe from (7.113) that for any N the function $\Psi(y) := \sum_{i\geq N} \mathsf{E}(|ya_iX_i|^2\wedge 1)$ is continuous and satisfies $\lim_{y\to 0}\Psi(y) = 0$ and $\lim_{y\to\infty}\Psi(y) = \infty$, and this proves the existence of λ_n. The proof will then rely on Theorem 7.3.4. Let us consider $s\in T$ and let us assume that for some integer $n\geq 1$ we have

$$|s-1| \leq \frac{1}{N_{n+1}} \ . \tag{7.116}$$

Let us observe the following inequality, for $i\geq 1$,

$$|s^i - 1| \leq i|s-1| \ . \tag{7.117}$$

We then write, for any integer $j\in\mathbb{Z}$, using also that $|s^i-1|\leq 2$ in the last line,

$$\begin{aligned}\sum_{i\geq 1}\mathsf{E}(|2^j(Z_i(s)-Z_i(0))|^2\wedge 1) &= \sum_{i\geq 1}\mathsf{E}(|2^ja_iX_i(s^i-1)|^2\wedge 1)\\ &\leq \sum_{0\leq m<n}\ \sum_{N_m\leq i<N_{m+1}}\mathsf{E}(|2^j ia_iX_i(s-1)|^2\wedge 1)\\ &\quad+\sum_{i\geq N_n}\mathsf{E}(|2^{j+1}a_iX_i|^2\wedge 1)\ . \end{aligned}\tag{7.118}$$

From (7.114) we observe that

$$\lambda_n 2^{j+1}\leq 1 \Rightarrow \sum_{i\geq N_n}\mathsf{E}(|2^{j+1}a_iX_i|^2\wedge 1) \leq \sum_{i\geq N_n}\mathsf{E}\Big(\frac{|a_iX_i|^2}{\lambda_n^2}\wedge 1\Big) \leq 2^n\ . \tag{7.119}$$

Also, for $i\leq N_{m+1}$ and $m<n$, (7.116) implies $i|s-1|\leq N_{m+1}/N_{n+1}\leq N_n/N_{n+1} = 1/N_n$. Consequently, it follows from (7.114) again that

$$\lambda_m 2^j \leq N_n \Rightarrow \sum_{N_m\leq i<N_{m+1}}\mathsf{E}(|2^j ia_iX_i(s-1)|^2\wedge 1) \leq \sum_{i\geq N_m}\mathsf{E}\Big(\frac{|a_iX_i|^2}{\lambda_m^2}\wedge 1\Big)\leq 2^m\ . \tag{7.120}$$

Consider the largest integer j_n which satisfies both $\lambda_n 2^{j_n+1} \le 1$ and $\lambda_m 2^{j_n} \le N_n$ for each $m < n$. Using (7.118), (7.119) and (7.120) we then get

$$\sum_{i\ge 1} \mathsf{E}(|2^{j_n}(Z_i(s) - Z_i(0))|^2 \wedge 1) \le \sum_{0\le m<n} 2^m + 2^n = 2^{n+1} . \tag{7.121}$$

Moreover the definition of j_n shows that either $\lambda_n 2^{j_n+2} \ge 1$ or $\lambda_m 2^{j_n+1} \ge N_n$ for some $m \le n$, and thus

$$2^{-j_n} \le 4\lambda_n + 2 \sum_{0\le m<n} \frac{\lambda_m}{N_n} . \tag{7.122}$$

Let us denote by U_n the set of points s that satisfy (7.116). Then $\mu(U_n) \ge 1/\pi N_{n+1}$ (where μ is the Haar measure of T), so that for $n \ge 1$ we certainly have $\mu(U_n) \ge 1/N_{n+2}$. In particular we have proved that

$$\mu\Big(\Big\{s \in T\ ;\ \sum_{i\ge 1} \mathsf{E}(|2^{j_n}(Z_i(s) - Z_i(0))|^2 \wedge 1) \le 2^{n+1}\Big\}\Big) \ge \frac{1}{N_{n+2}} ,$$

while (7.115) and (7.122) imply that $\sum_{n\ge 0} 2^{n-j_n} < \infty$. Using Theorem 7.3.4 this completes the proof. □

The following provides a converse of Theorem 7.7.1 under a mild regularity condition.

Theorem 7.7.3. *Assume moreover that the sequence* (X_i) *is i.i.d. and that for a certain number* $C > 0$*, one has*

$$k \le m \le 2k \Rightarrow |a_k| \le C|a_m| . \tag{7.123}$$

Then (7.115) holds whenever the series $\sum_{i\ge i} a_i \varepsilon_i X_i \chi_i$ *converges uniformly a.s.*

Proof. We use Theorem 7.3.4 to obtain a sequence (j_n) with $\sum_{n\ge 0} 2^{n-j_n} < \infty$ and

$$\forall\, n \ge 1\ ,\ \mu\Big(\Big\{s \in T\ ;\ \sum_{i\ge 1} \mathsf{E}(|2^{j_n}(Z_i(s) - Z_i(0))|^2 \wedge 1) \le 2^n\Big\}\Big) \ge \frac{1}{N_n} . \tag{7.124}$$

We will prove that (7.124) implies that

$$\lambda_{n+3} \le LC^2 2^{-j_n} , \tag{7.125}$$

completing the proof. Since $Z_i(s) = a_i X_i s^i$, we deduce from (7.124) that we can find $s \in T$ with

$$|s - 1| \ge \frac{1}{2N_n} \tag{7.126}$$

and

$$\sum_{i\geq 1} \mathsf{E}(|2^{j_n} a_i X_i (s^i - 1)|^2 \wedge 1) \leq 2^n \ . \tag{7.127}$$

The idea is then to show that there are many values of $i \geq 2^{2^n+3}$ for which $|s^i - 1| \geq 1/4$. Indeed we have

$$\sum_{2^p \leq i < 2^{p+1}} s^i = s^{2^p} \frac{s^{2^p} - 1}{s - 1} \ ,$$

so that using (7.126)

$$\Big| \sum_{2^p \leq i < 2^{p+1}} s^i \Big| \leq 4N_n \ ,$$

and hence if $p \geq 2^n + 3$

$$\Big| \sum_{2^p \leq i < 2^{p+1}} (s^i - 1) \Big| \geq 2^p - 4N_n \geq 2^{p-1} \ ,$$

so that

$$\sum_{2^p \leq i < 2^{p+1}} |s^i - 1| \geq 2^{p-1} \ .$$

Since there are 2^p terms on the left-hand side, each of which is ≤ 2, it follows that

$$\operatorname{card} I_p \geq 2^{p-4} \ , \tag{7.128}$$

where

$$I_p = \{ i \ ; \ 2^p \leq i < 2^{p+1} \ , \ |s^i - 1| \geq 1/4 \} \ . \tag{7.129}$$

Now, using (7.123), for $i \in I_p$ we have $|a_i(s^i - 1)| \geq |a_{2^p}|/4C$ and then

$$\mathsf{E}(|2^{j_n} a_i X_i (s^i - 1)|^2 \wedge 1) \geq \mathsf{E}\Big(\Big| \frac{2^{j_n - 2}}{C} a_{2^p} X_{2^p} \Big|^2 \wedge 1 \Big) \ , \tag{7.130}$$

and combining with (7.129),

$$\sum_{2^p \leq i < 2^{p+1}} \mathsf{E}(|2^{j_n} a_i X_i (s^i - 1)|^2 \wedge 1) \geq 2^{p-4} \mathsf{E}\Big(\Big| \frac{2^{j_n - 2}}{C} a_{2^p} X_{2^p} \Big|^2 \wedge 1 \Big) \ . \tag{7.131}$$

Using (7.123) again, for $2^{p-1} \leq i \leq 2^p$ we have $|a_{2^p}| \geq |a_i|/C$ and thus

$$2^{p-4} \mathsf{E}\Big(\Big| \frac{2^{j_n - 2}}{C} a_{2^p} X_{2^p} \Big|^2 \wedge 1 \Big) \geq 2^{-3} \sum_{2^{p-1} \leq i < 2^p} \mathsf{E}\Big(\Big| \frac{2^{j_n - 2}}{C^2} a_i X_i \Big|^2 \wedge 1 \Big) \ . \tag{7.132}$$

Combining with (7.131), summing over $p \geq 2^n + 3$ and combining with (7.127) yields

$$2^{-3} \sum_{i \geq 2^{2^n+2}} \mathsf{E}\Big(\Big| \frac{2^{j_n - 2}}{C^2} a_i X_i \Big|^2 \wedge 1 \Big) \leq 2^n \ , \tag{7.133}$$

and, in particular,

$$\sum_{i \geq N_{n+3}} \mathsf{E}\Big(\Big|\frac{2^{j_n-2}}{C^2} a_i X_i\Big|^2 \wedge 1\Big) \leq 2^{n+3} .$$

By definition of λ_n this implies

$$\frac{2^{j_n-2}}{C^2} \leq \frac{1}{\lambda_{n+3}} .$$

This proves (7.125). □

To give a still more explicit example, we mention the following.

Theorem 7.7.4. *If* (X_i) *denotes an i.i.d. sequence distributed like* X*, the series* $\sum_{i \geq 1} \frac{1}{i} \varepsilon_i X_i \chi_i$ *converges uniformly a.s. if and only if*

$$\mathsf{E}|X| \log\log(|X| + 3) < \infty . \tag{7.134}$$

Proof. Since the sequence $a_k = 1/k$ satisfies (7.123), it suffices from Theorems 7.7.1 and 7.7.3 to prove that (7.134) is equivalent to (7.115). The proof uses standard methods, that are not related to the ideas of this work. It can be found in Lemma 2.1 of [11]. □

7.8 Notes and Comments

The work of Marcus and Pisier on random Fourier series was extended by Marcus [3] to more general situations (that involve the infinitely divisible processes that we will study in Chapter 11). Marcus fails however to obtain necessary and sufficient conditions. Obtaining these intrinsically requires the ideas of "families of distances" as we used in Section 7.2. This is largely done in the paper [10]. The arguments of this paper still require some weak but unnecessary tail conditions, because the chaining is not organized in an optimal way. We finally succeeded to remove then here.

In retrospect it might be hard to understand why the topic of random Fourier series was so popular at one point. Nevertheless, this topic was historically important. The author was lucky to investigate it. It was the ideal setting to invent the concept of "families of distances" because this concept was the most important missing ingredient between the Marcus-Pisier work and the rather complete solution we present here. As the proof of the Bednorz-Latała theorem demonstrates, this concept of families of distances may have some lasting value.

References

1. Fernique, X.: Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens. Université de Montréal, Centre de Recherches Mathématiques, Montreal (1977). iv+217 pp
2. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
3. Marcus, M.B.: ξ-Radial Processes and Random Fourier Series. Mem. Amer. Math. Soc., vol. 68(368) (1987). viii+181 pp
4. Marcus, M.B., Pisier, G.: Random Fourier Series with Applications to Harmonic Analysis. Annals of Mathematics Studies, vol. 101. Princeton University Press, Princeton; University of Tokyo Press, Tokyo (1981). v+151 pp. ISBN: 0-691-08289-8; 0-691-08292-8
5. Marcus, M.B., Pisier, G.: Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. Acta Math. **152**(3–4), 245–301 (1984)
6. Paley, R., Zygmund, A.: On some series of functions (1). Proc. Camb. Philos. Soc. **26**, 337–357 (1930)
7. Paley, R., Zygmund, A.: On some series of functions (2). Proc. Camb. Philos. Soc. **26**, 458–474 (1930)
8. Paley, R., Zygmund, A.: On some series of functions (3). Proc. Camb. Philos. Soc. **28**, 190–205 (1932)
9. Salem, R., Zygmund, A.: Some properties of trigonometric series whose terms have random signs. Acta Math. **91**, 245–301 (1954)
10. Talagrand, M.: Necessary and sufficient conditions for sample continuity of random Fourier series and of harmonic infinitely divisible processes. Ann. Probab. **20**, 1–28 (1992)
11. Talagrand, M.: A borderline random Fourier series. Ann. Probab. **23**, 776–785 (1995)

8. Processes Related to Gaussian Processes

8.1 p-Stable Processes

Consider a number $0 < p \le 2$. A r.v. X is called (real, symmetric) p-stable if for each $\lambda \in \mathbb{R}$ we have

$$\mathsf{E} \exp i\lambda X = \exp\Bigl(-\frac{\sigma^p |\lambda|^p}{2}\Bigr)\,, \tag{8.1}$$

where $\sigma = \sigma_p(X)$ is called the parameter of X. The name "p-stable" comes from the fact that if $X_1, \ldots, X_m$ are independent and p-stable, then $\sum_{j \le m} a_j X_j$ is p-stable, and

$$\sigma_p\Bigl(\sum_{j\le m} a_j X_j\Bigr) = \Bigl(\sum_{j\le m} |a_j|^p \sigma_p(X_j)^p\Bigr)^{1/p}\,.$$

This is obvious from (8.1).

The reason for the restriction $p \le 2$ is that for $p > 2$ no r.v. satisfies (8.1). The case $p = 2$ is the Gaussian case. Despite the formal similarity, the case $p < 2$ is very different. It can be shown that

$$\lim_{s\to\infty} s^p \mathsf{P}(|X| \ge s) = c_p \sigma^p \tag{8.2}$$

where $c_p > 0$ depends on p only. Thus X does not have moments of order p, but it has moments of order q for $q < p$. We refer the reader to [5] for a proof of this and for general background on p-stable processes.

A process $(X_t)_{t\in T}$ is called p-stable if, for every family $(\alpha_t)_{t\in T}$ for which only finitely many of the numbers α_t are not 0, the r.v. $\sum_t \alpha_t X_t$ is p-stable. We can then define a (quasi) distance d on T by

$$d(s,t) = \sigma(X_s - X_t)\,. \tag{8.3}$$

One can also define an equivalent distance by $d(s,t) = (\mathsf{E}|X_s - X_t|^q)^{1/q}$, where $q < p$.

A typical example of p-stable process is given by $X_t = \sum_{i\le n} t_i Y_i$ where $t = (t_i)_{i\le n}$ and $(Y_i)_{i\le n}$ are independent p-stable r.v.s. It can in fact be shown that this example is generic in the sense that "each p-stable process

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*,
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of
Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_8,

(with a finite index set) can be arbitrarily well approximated by a process of this type." It is very instructive to consider the case where $\sigma(Y_i) = 1$ for each i, in which case the distance induced by the process is the ℓ^p distance, $d(X_s, X_t) = \|s - t\|_p$.

Consider first the situation where $T = \{e_1, e_2, \ldots, e_n\}$, the canonical basis of ℓ^p. Then, using (8.2), we observe that for each $\ell \le n$ there exists a set Ω_ℓ of probability $1/n$ on which $X_{e_\ell} \ge n^{1/p}/C'(p)$, so that since the sets Ω_ℓ are independent, $\max_{\ell \le n} X_{e_\ell}$ is at least $n^{1/p}/C'(p)$ on a set of probability about $1/2$. It is then a simple matter to see that

$$\mathsf{E} \sup_{t \in T} X_t \ge \frac{n^{1/p}}{C(p)} , \tag{8.4}$$

where $C(p)$ does not depend on n. The metric space (T, d) consists on n points within distance at most 2 of each other, so (8.4) is dramatically different from the Gaussian case, where in that situation one has $\mathsf{E} \sup_t X_t \le L\sqrt{\log n}$. Consider now the situation where $T = \{(\pm 1, \pm 1, \ldots, \pm 1)\}$. Then

$$\mathsf{E} \sup_{t \in T} X_t = \mathsf{E} \sum_{i \le n} |Y_i| = n\mathsf{E}|Y_1| . \tag{8.5}$$

We observe that since $\operatorname{card} T = 2^n$ then for each $\epsilon > 0$ we have

$$\epsilon(\log N(T, d, \epsilon))^{1-1/p} \le \epsilon n^{1-1/p} .$$

Since the diameter of T is $2n^{1/p}$, for $\epsilon > 2n^{1/p}$ we have $N(T, d, \epsilon) = 1$ and the left-hand side above is 0. Consequently for each $\epsilon > 0$ we have $\epsilon(\log N(T, d, \epsilon))^{1-1/p} \le 2n$. In particular,

$$\epsilon(\log N(T, d, \epsilon))^{1-1/p} \le K(p)\mathsf{E} \sup_t X_t , \tag{8.6}$$

where $K(p)$ depends on p only.

The previous two examples show that in contrast with the Gaussian case, it seems unrealistic to hope to compute $\mathsf{E} \sup_{t \in T} X_t$ as a function of the geometry of (T, d) only. Yet, it turns out not only that (8.6) is true in general, but also that one can extend the lower bound of the majorizing measure Theorem 2.4.1 as follows.

Theorem 8.1.1. *For $1 < p < 2$, there is a number $K(p)$ such that for any p-stable process $(X_t)_{t \in T}$ we have*

$$\gamma_q(T, d) \le K(p)\mathsf{E} \sup_{t \in T} X_t ,$$

where q is the conjugate exponent of p, i.e. $1/q + 1/p = 1$, and where d is as in (8.3).

At the heart of Theorem 8.1.1 is the fact that a p-stable process (X_t) can be represented as a conditionally Gaussian process. That is, we can find two probability spaces $(\Omega\,,\mathsf{P})\,,\,(\Omega'\,,\mathsf{P}')$ and a family $(Y_t)_{t\in T}$ of r.v.s on $\Omega\times\Omega'$ (provided with the product probability), such that

$$\text{Given any finite subset } U \text{ of } T,\ \text{the joint laws of } (Y_t)_{t\in U} \text{ and } (X_t)_{t\in U} \text{ are identical} \tag{8.7}$$

$$\text{Given } \omega\in\Omega\,,\ \text{the process } \omega'\mapsto Y_t(\omega,\omega') \text{ is a centered Gaussian process.} \tag{8.8}$$

This result holds for any value of p with $1\le p<2$. A complete proof is given in Section 11.3, in the more general setting of infinitely divisible processes. A remarkable fact is that to prove Theorem 8.1.1 we do not need to know precisely how the previous representation arises.

We denote by E' integration in P' only. Given ω, we consider the random distance d_ω on T given by

$$d_\omega(s,t)=\big(\mathsf{E}'(Y_s(\omega,\omega')-Y_t(\omega,\omega'))^2\big)^{1/2}\,. \tag{8.9}$$

We define α by

$$\frac{1}{\alpha}:=\frac{1}{p}-\frac{1}{2}\,. \tag{8.10}$$

Lemma 8.1.2. *For all $s,t\in T$ and $\epsilon>0$, we have*

$$\mathsf{P}(d_\omega(s,t)\le\epsilon d(s,t))\le\exp\Big(-\frac{b_p}{\epsilon^\alpha}\Big) \tag{8.11}$$

where $b_p>0$ depends on p only.

Proof. Since the process $Y_t(\omega,\cdot)$ is Gaussian, we have

$$\mathsf{E}'\exp i\lambda(Y_s-Y_t)=\exp\Big(-\frac{\lambda^2}{2}d_\omega^2(s,t)\Big)\,.$$

Taking expectation, using (8.1), and since the pair (Y_s,Y_t) has the same law as the pair (X_s,X_t), we get

$$\exp\Big(-\frac{|\lambda|^p}{2}d^p(s,t)\Big)=\mathsf{E}\exp\Big(-\frac{\lambda^2}{2}d_\omega^2(s,t)\Big)\,. \tag{8.12}$$

Any r.v. Z satisfies

$$\mathsf{P}(Z\le u)\le\exp\Big(\frac{\lambda^2u}{2}\Big)\mathsf{E}\exp\Big(-\frac{\lambda^2}{2}Z\Big)\,.$$

Using this for $Z=d_\omega^2(s,t)$ and $u=\epsilon^2d^2(s,t)$, we get, using (8.12),

$$\mathsf{P}(d_\omega(s,t)\le\epsilon d(s,t))\le\exp\Big(\frac{1}{2}\big(\lambda^2\epsilon^2d^2(s,t)-|\lambda|^pd^p(s,t)\big)\Big)\,,$$

and the result by optimization over λ. □

The content of (8.11) is that, given a pair (s,t), it is rare that $d_\omega(s,t)$ is much smaller than $d(s,t)$. Given two pairs (s,t) and (s',t') we however know nothing about the joint distribution of the r.v.s $d_\omega(s,t)$ and $d_\omega(s',t')$. It is therefore quite surprising that the information contained in this lemma suffices to deduce Theorem 8.1.1 from the majorizing measure Theorem 2.4.1. This will done through the following abstract result about metric spaces.

Theorem 8.1.3. *Consider a (finite) metric space (T,d) and a random distance d_ω on T. Assume that for some $b>0$ we have*

$$\forall s,t\in T\,,\ \forall\epsilon>0\,,\ \mathsf{P}(d_\omega(s,t)\le\epsilon d(s,t))\le\exp\Big(-\frac{b}{\epsilon^\alpha}\Big)\,, \tag{8.13}$$

where $\alpha>2$. Then

$$\mathsf{P}\Big(\gamma_2(T,d_\omega)\ge\frac{1}{K}\gamma_q(T,d)\Big)\ge\frac{3}{4}\,, \tag{8.14}$$

where

$$\frac{1}{q}=\frac{1}{2}-\frac{1}{\alpha}\,,$$

and where K depends on α and b only.

Of course the number $3/4$ plays no special role.

Proof of Theorem 8.1.1. Using Theorem 2.7.5 (c), we may assume that T is finite. Consider the r.v. $Z=\sup_{t\in T}Y_t$. Then Theorem 2.4.1 implies

$$\mathsf{E}'Z\ge\frac{1}{L}\gamma_2(T,d_\omega)\ ,$$

and since $\mathsf{E}'Z\ge0$, taking expectation in this inequality and using (8.14) proves that $\mathsf{E}Z\ge\gamma_q(T,d)/K(p)$. □

Let us now prepare for the proof of Theorem 8.1.3. Replacing d by $b^{1/\alpha}d$, we can and do assume that $b=1$. The following lemma explains how one may use (8.13). It will not be directly used, as we shall need a more elaborate version of the same idea.

Lemma 8.1.4. *Under the hypotheses of Theorem 8.1.3, with probability $\ge 1-\exp(-2^{n+1})$ we have*

$$e_n(T,d_\omega)\ge\frac{1}{K}2^{-n/\alpha}e_n(T,d)\ .$$

Proof. Consider $a<e_n(T,d)$. Consider a subset U of T maximal with respect to the property that $d(s,t)\ge a$ for $s,t\in U, s\ne t$. Then the balls of radius a centered at the points of U cover T. Thus $\operatorname{card}U>N_n$ by definition of $e_n(T,d)$. Consider a subset T_n of U with $\operatorname{card}T_n=N_n$. It follows from (8.13) that the event

$$s, t \in T_n \Rightarrow d_\omega(s,t) \ge u 2^{-n/\alpha} a \tag{8.15}$$

has a probability $\ge 1 - N_n^2 \exp(-2^n/u^\alpha)$, so that one can find u depending only on α so that this probability is $\ge 1 - \exp(-2^{n+1})$, and when (8.15) occurs, we have $e_n(T, d_\omega) \ge u 2^{-n/\alpha} a/2$. □

Recalling the value of α this yields the following:

Corollary 8.1.5. *With probability $\ge 1/2$ we have*

$$\sum_{n \ge 0} 2^{n/2} e_n(T, d_\omega) \ge \frac{1}{K} \sum_{n \ge 0} 2^{n/q} e_n(T, d) \,.$$

This resembles (8.14), except that the γ functionals have been replaced by the corresponding "entropy integrals". But we will have to work quite harder to capture the γ functionals themselves.

Exercise 8.1.6. The goal of the present exercise is to use Corollary 8.1.5 to prove (7.86) (which is the original approach of [7]). Consider i.i.d. r.v.s (θ_i) and assume that for simplicity that for $u \ge 1$ we have $\mathsf{P}(|\theta_i| \ge u) \ge u^{-p}$. Consider numbers b_i. Prove that for $\epsilon > 0$ we have

$$\mathsf{P}\Big(\Big(\sum_i \theta_i^2 b_i^2\Big)^{1/2} \le \epsilon \Big(\sum_i b_i^p\Big)^{1/p}\Big) \le \exp\Big(-\frac{L}{\epsilon^\alpha}\Big) \,,$$

where $1/\alpha = 1/p - 1/2$. (Hint: find a bound for $\mathsf{E}\exp(-\lambda\theta_i^2)$ and proceed as usual when proving an exponential inequality.) Conclude using (3.30).

Suppose now that we have pieces $(H_\ell)_{\ell \le m}$ (where, say, $m = N_{n+3}$) that are well separated for d, say the distance of any two of them is $\ge a$. If we choose a point t_ℓ in H_ℓ, we can as in Lemma 8.1.4 ensure that with large probability $d_\omega(t_\ell, t_{\ell'}) \ge 2^{-n/\alpha} a/K$ whenever $\ell \ne \ell'$, but there is apparently no way to bound $d_\omega(H_\ell, H_{\ell'})$ from below. Still, it must be true in some sense that "most of the points of H_ℓ and $H_{\ell'}$ are at least at distance $2^{-n/\alpha}a/L$ from each other". One way to give a meaning to this idea is to bring in a probability measure, and the next result is actually the crucial point of the proof of Theorem 8.1.3.

Lemma 8.1.7. *Under the hypothesis of Theorem 8.1.3, consider a probability measure μ on T, and assume that for a certain number a and some $n \ge 0$ we have*

$$\mu^{\otimes 2}(\{(x,y) \in T^2 \; ; \; d(x,y) < a\}) \le \frac{1}{N_{n+3}} \,. \tag{8.16}$$

Then with probability $\ge 1 - 2/N_{n+2}$, for each partition $\mathcal{A}_n$ of T with $\operatorname{card} \mathcal{A}_n \le N_n$ *we have*

$$\int \Delta(A_n(t), d_\omega) \mathrm{d}\mu(t) \ge \frac{2^{-n/\alpha} a}{K} \,. \tag{8.17}$$

Proof. The proof has two distinct parts. In the first part, we prove that for some number u depending on α only, if we consider the set

$$B_\omega = \{(x,y) \in T^2 \; ; \; d_\omega(x,y) \le ua2^{-n/\alpha}\} \; , \tag{8.18}$$

then the event Ω defined by

$$\mu^{\otimes 2}(B_\omega) \le \frac{1}{N_{n+2}} \tag{8.19}$$

has probability $\ge 1-2/N_{n+2}$. For this consider $B = \{(x,y) \in T^2; d(x,y) \le a\}$ and $C = T^2 \setminus B$, and observe that by hypothesis we have $\mu^{\otimes 2}(B) \le 1/N_{n+3}$. Then

$$\begin{aligned} \mathsf{E}\mu^{\otimes 2}(B_\omega) &\le \mu^{\otimes 2}(B) + \mathsf{E}\int_C \mathbf{1}_{\{d_\omega(x,y) \le ua2^{-n/\alpha}\}} \mathrm{d}\mu(x)\mathrm{d}\mu(y) \\ &\le \mu^{\otimes 2}(B) + \int_C \mathsf{E}\mathbf{1}_{\{d_\omega(x,y)\le ud(x,y)2^{-n/\alpha}\}} \mathrm{d}\mu(x)\mathrm{d}\mu(y) \\ &\le \frac{1}{N_{n+3}} + \exp\Big(-\frac{2^n}{u^\alpha}\Big) \; , \end{aligned} \tag{8.20}$$

so that by an appropriate choice of u depending on α only, we have $\mathsf{E}\mu^{\otimes 2}(B_\omega) \le 2/N_{n+3} = 2/N_{n+2}^2$ and therefore $\mathsf{P}(\mu^{\otimes 2}(B_\omega) \ge N_{n+2}^{-1}) \le 2/N_{n+2}$.

The second part of the proof is to show that (8.19) implies (8.17). Let us consider a partition $\mathcal{A}_n$ of T with $\operatorname{card}\mathcal{A}_n \le N_n$ and the set

$$D = \bigcup\{A \in \mathcal{A}_n \; ; \; \Delta(A, d_\omega) \le ua2^{-n/\alpha}\} \; .$$

Thus, if $A \in \mathcal{A}_n$ and $A \subset D$, then $A^2 \subset B_\omega$ and thus $\mu(A) \le \sqrt{\mu^{\otimes 2}(B_\omega)} \le 1/N_{n+1}$. Since $\operatorname{card}\mathcal{A}_n \le N_n$ we have $\mu(D) \le 1/2$. Moreover, for $t \notin D$ we have $\Delta(A_n(t), d_\omega) \ge u2^{-n/\alpha}a$, and this completes the proof of (8.17). □

Proof of Theorem 8.1.3. To prove (8.14) we will prove that if a set $U \subset \Omega$ satisfies $\mathsf{P}(U) \ge 1/4$, then

$$\mathsf{E}(\mathbf{1}_U \gamma_2(T, d_\omega)) \ge \frac{1}{K}\gamma_q(T,d) \; . \tag{8.21}$$

Since $U = \{\gamma_2(T,d_\omega) < \gamma_q(T,d)/K\}$ violates (8.21), we must have $\mathsf{P}(U) < 1/4$, and this proves (8.14).

We fix U once and for all with $\mathsf{P}(U) \ge 1/4$. Given a probability measure μ on T and $n \ge 0$ we set

$$F_n(\mu) = \mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int_T \sum_{k\ge n} 2^{k/2}\Delta(A_k(t), d_\omega)\mathrm{d}\mu(t)\Big) \; ,$$

where $\inf_{\mathcal{A}}$ means that the infimum is taken over all admissible sequences $(\mathcal{A}_n)_{n\geq 0}$ of T. Given $A \subset T$, we set

$$F_n(A) = \sup_{\mu} F_n(\mu) \,,$$

where the supremum is over all probability measures μ supported by A. (The reader may like to read Exercise 6.2.6 to motivate this definition.) Since $\int f \mathrm{d}\mu \leq \sup f$, it holds that

$$\inf_{\mathcal{A}} \int_T \sum_{k\geq 0} 2^{k/2} \Delta(A_k(t), d_\omega) \mathrm{d}\mu(t) \leq \gamma_2(T, d_\omega) \;,$$

and therefore

$$F_0(T) \leq \mathsf{E}(\mathbf{1}_U \gamma_2(T, d_\omega)) \;.$$

Next, we claim that

$$\Delta(T, d) \leq K F_0(T) \;. \tag{8.22}$$

(Here and in the rest of the proof, K denotes a number depending on α only, that need not be the same at each occurrence.) To see this, we simply note that since $\mathcal{A}_0 = \{T\}$, we have $A_0(t) = T$ for each t, so that for each probability μ on T we have $F_0(T) \geq F_0(\mu) \geq \mathsf{E}(\mathbf{1}_U \Delta(T, d_\omega))$. Using the bound $\mathsf{E}(f) \geq a\mathsf{P}(f \geq a)$ when $f \geq 0$, we obtain that for any $\epsilon > 0$,

$$F_0(T) \geq \mathsf{E}(\mathbf{1}_U \Delta(T, d_\omega)) \geq \epsilon \Delta(T, d) \mathsf{P}(U \cap \{\Delta(T, d_\omega) \geq \epsilon \Delta(T, d)\}) \;. \tag{8.23}$$

Consider now $s, t \in T$ with $d(s, t) \geq \Delta(T, d)/2$, and ϵ depending on α only with $\exp(-1/(2\epsilon)^\alpha) = 1/8$. Then using (8.13),

$$\mathsf{P}(\Delta(T, d_\omega) \geq \epsilon \Delta(T, d)) \geq \mathsf{P}(d_\omega(s, t) \geq 2\epsilon d(s, t)) \geq 7/8$$

and therefore $\mathsf{P}(U \cap \{\Delta(T, d_\omega) \geq \epsilon \Delta(T, d)\}) \geq 1/8$ and finally $F_0(T) \geq \Delta(T, d)/K$.

Thus (8.21), and hence Theorem 8.1.3 will follow from Theorem 2.7.2 (used for $r = 4, \beta = 1, \theta(n) = 2^{n/q}/K, \xi = 2^{1/q}$ and $\tau = 3$) and Lemma 2.3.5 provided we prove that the functionals F_n satisfy the growth condition of Definition 2.7.1. The purpose of taking $\tau = 3$ is simply that this greatly helps to check this condition, as will become apparent later. To prove the growth condition, we consider a, $n \geq 0$, $m = N_{n+3}$, and points $(t_\ell)_{\ell \leq m}$ in T, with

$$\ell \neq \ell' \Rightarrow d(t_\ell, t_{\ell'}) \geq a > 0 \;. \tag{8.24}$$

We consider sets $H_\ell \subset B(t_\ell, a/4)$, and we shall show that

$$F_n\Big(\bigcup_{\ell \leq m} H_\ell\Big) \geq \frac{2^{n/q} a}{K} + \min_{\ell \leq m} F_{n+1}(H_\ell) \;. \tag{8.25}$$

Consider $c < \min_{\ell \le m} F_{n+1}(H_\ell)$, and consider for each ℓ a probability μ_ℓ supported by H_ℓ, and such that $F_{n+1}(\mu_\ell) > c$. Consider

$$\mu = \frac{1}{m} \sum_{\ell \le m} \mu_\ell \,. \tag{8.26}$$

This is a probability, which is supported by $H := \bigcup_{\ell \le m} H_\ell$. To prove (8.25), it suffices to prove that

$$F_n(\mu) \ge \frac{2^{n/q} a}{K} + c \,. \tag{8.27}$$

Since $\inf(f(x) + g(x)) \ge \inf f(x) + \inf g(x)$, we have

$$F_n(\mu) \ge \mathrm{I} + \mathrm{II}$$

where

$$\mathrm{I} = F_{n+1}(\mu) = \mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int \sum_{k \ge n+1} 2^{k/2} \Delta(A_k(t), d_\omega) \mathrm{d}\mu(t)\Big)$$

$$\mathrm{II} = \mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int 2^{n/2} \Delta(A_n(t), d_\omega) \mathrm{d}\mu(t)\Big) \,,$$

where both infima are over all admissible sequences $(\mathcal{A}_n)$ of T. Using (8.26), we have

$$\mathrm{I} \ge \frac{1}{m} \sum_{\ell \le m} F_{n+1}(\mu_\ell) \ge c$$

so all what remains to prove is that

$$\mathrm{II} \ge \frac{2^{n/q} a}{K} \,. \tag{8.28}$$

We observe that $d(H_\ell, H_{\ell'}) \ge a/2$ for $\ell \ne \ell'$, and thus

$$\{(x,y) \in H^2 \,;\, d(x,y) < a/2\} \subset \bigcup_{\ell \le m} H_\ell^2 \,. \tag{8.29}$$

Since $\mu(H_\ell) = 1/m = 1/N_{n+3}$ for each ℓ this proves (8.16) (for $a/2$ rather than a). Lemma 8.1.7 then implies that with probability $\ge 7/8$ one has $\inf_{\mathcal{A}} \int 2^{n/2} \Delta(A_n(t), d_\omega) \mathrm{d}\mu(t) \ge 2^{n(1/2 - 1/\alpha)} a/K$, and this concludes the proof of (8.28) and of the theorem. □

Our next result extends Theorem 8.1.3 to the case $\alpha = 2$. This will in turn have implications about 1-stable processes. We set $M_0 = 1\,,\ M_n = 2^{N_n}$ for $n \ge 1$. The sequence $M_n = 2^{2^{2^n}}$ grows quite fast. Given a metric space (T,d) we define

$$\gamma_\infty(T,d) = \inf_{\mathcal{B}} \sup_{t \in T} \sum_{n \ge 0} 2^n \Delta(B_n(t)) \,, \tag{8.30}$$

where the infimum is taken over all increasing families of partitions $(\mathcal{B}_n)$ of T with $\operatorname{card} \mathcal{B}_n \le M_n$. This new quantity is a kind of limit of the quantities $\gamma_\alpha(T,d)$ as $\alpha \to \infty$.

Exercise 8.1.8. Consider the quantity $\gamma^*(T,d)$ defined as

$$\gamma^*(T,d) = \inf \sup_{t\in T} \sum_{n\geq 0} \Delta(A_n(t)) \;, \tag{8.31}$$

where the infimum is computed over all admissible sequences of partitions $(\mathcal{A}_n)$. Prove that

$$\frac{1}{L}\gamma^*(T,d) \leq \gamma_\infty(T,d) \leq L\gamma^*(T,d) \;. \tag{8.32}$$

(Hint: given an increasing sequence of partitions $(\mathcal{B}_n)$ with $\operatorname{card}\mathcal{B}_n \leq M_n$ consider the increasing sequence of partitions $(\mathcal{A}_m)$ given by $\mathcal{A}_m = \mathcal{B}_n$ for $2^n \leq m < 2^{n+1}$.)

The formulation (8.31) is more natural than the formulation (8.30). We do not use because of the following technical difficulty: the corresponding function $\theta(n) = 1$ does not satisfy (2.149).

Theorem 8.1.9. *Consider a finite metric space* (T,d) *and a random distance* d_ω *on* T*. Assume that*

$$\forall s,t\in T\,,\; \forall\epsilon > 0\,,\; \mathsf{P}(d_\omega(s,t) < \epsilon d(s,t)) \leq \exp\Big(-\frac{1}{\epsilon^2}\Big)\;.$$

Then

$$\mathsf{P}\Big(\gamma_2(T,d_\omega) \geq \frac{1}{L}\gamma_\infty(T,d)\Big) \geq \frac{3}{4}\;.$$

Proof. The proof of Theorem 8.1.9 closely follows that of Theorem 8.1.3, so we indicate only the necessary modifications. It should be obvious that Theorem 2.7.2 still holds when we replace N_n by M_n. We will use it in that case for $\theta(n) = 2^n/L, r = 4$ and $\tau = 2$. We define

$$F_n(\mu) = \mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int \sum_{k\geq 2^n-1} 2^{k/2}\Delta(A_k(t),d_\omega)d\mu(t)\Big)\;.$$

Here, and everywhere in this proof, the infimum is over all admissible sequences $(\mathcal{A}_n)_{n\geq 0}$ of T. (Thus, as usual, $\operatorname{card}\mathcal{A}_n \leq N_n$.) It suffices to prove that under the condition (8.24) (with now $m = M_{n+2}$) the condition that corresponds to (8.28) holds:

$$\mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int \sum_{2^n-1\leq k<2^{n+1}-1} 2^{k/2}\Delta(A_k(t),d_\omega)\mathrm{d}\mu(t)\Big) \geq \frac{2^n a}{L}\;.$$

For this purpose it suffices to prove that for each $2^n - 1 \leq k < 2^{n+1} - 1$ we have

$$\mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int 2^{k/2}\Delta(A_k(t),d_\omega)\mathrm{d}\mu(t)\Big) \geq \frac{a}{L}\,. \tag{8.33}$$

As in the case of Theorem 8.1.3 we have

$$\mu^{\otimes 2}(\{(x,y) \in T^2 \; ; \; d(x,y) < a/2\}) \le \frac{1}{M_{n+2}} \,. \tag{8.34}$$

For $k < 2^{n+1} - 1$, we have $k \le 2^{n+2} - 3$, so $M_{n+2} \ge N_{k+3}$ and then (8.33) follows from (8.34) and Lemma 8.1.7. □

As promised, we now apply Theorem 8.1.9 to 1-stable processes.

Theorem 8.1.10. *For every* 1*-stable process* $(X_t)_{t\in T}$ *and* $t_0 \in T$ *we have*

$$\mathsf{P}\Big(\sup_{t\in T}(X_t - X_{t_0}) \ge \frac{1}{L}\gamma_\infty(T,d)\Big) \ge \frac{1}{L} \,.$$

To understand the formulation of this theorem, we note that we cannot use expectation to measure the size of $\sup_{t\in T} X_t$, as is shown by (8.2). Also, we observe that when T consists of two points t_0 and t_1, then

$$\sup_{t\in T}(X_t - X_{t_0}) = \max(X_{t_1} - X_{t_0}, 0)$$

is 0 with probability 1/2.

Lemma 8.1.11. *If* $(Y_t)_{t\in T}$ *is a Gaussian process then*

$$\mathsf{P}\Big(\sup_{t\in T}(Y_t - Y_{t_0}) \ge \frac{1}{2}\mathsf{E}\sup_{t\in T}(Y_t - Y_{t_0})\Big) \ge \frac{1}{L} \,.$$

Proof. This is a consequence of the Paley-Zygmund inequality (7.30) and the fact that the r.v. $Z = \sup_{t\in T}(Y_t - Y_{t_0})$ satisfies $\mathsf{E}Z^2 \le L(\mathsf{E}Z)^2$ (a weak consequence of (2.84)). □

Remark 8.1.12. Since $\mathsf{E}Z^2 \le L(\mathsf{E}Z)^2$, Lemma 8.1.11 shows that, assuming $Y_{t_0} = 0$ for some $t_0 \in T$

$$\mathsf{P}\Big(\sup_{t\in T} Y_t \ge \frac{1}{L}\big(\mathsf{E}(\sup_{t\in T} Y_t)^2\big)^{1/2}\Big) \ge \frac{1}{L} \,. \tag{8.35}$$

Proof of Theorem 8.1.10. Combining Theorems 8.1.9 and 2.4.1, we get

$$\mathsf{P}\Big(\mathsf{E}'\sup_{t\in T}(Y_t(\omega,\omega') - Y_{t_0}(\omega,\omega')) \ge \frac{1}{L}\gamma_\infty(T,d)\Big) \ge \frac{1}{L} \,. \tag{8.36}$$

Using Lemma 8.1.11 given ω, we obtain

$$\mathsf{P}'\Big(\sup_{t\in T}(Y_t(\omega,\omega') - Y_{t_0}(\omega,\omega')) \ge \frac{1}{L}\mathsf{E}'\sup_{t\in T}(Y_t(\omega,\omega') - Y_{t_0}(\omega,\omega')\Big) \ge \frac{1}{L} \,.$$

Combining with (8.36) and using Fubini theorem we finally obtain

$$\mathsf{P}\otimes\mathsf{P}'\Big(\sup_{t\in T}(Y_t(\omega,\omega') - Y_{t_0}(\omega,\omega')) \ge \frac{1}{L}\gamma_\infty(T,d)\Big) \ge \frac{1}{L} \,.$$ □

8.2 Order 2 Gaussian Chaos

Consider independent standard Gaussian sequences $(g_i), (g'_j), i, j \geq 1$. Given a double sequence $t = (t_{i,j})_{i,j\geq 1}$ we consider the r.v.

$$X_t = \sum_{i,j\geq 1} t_{i,j} g_i g'_j \ . \tag{8.37}$$

The series converges in L^2 as soon as $\sum_{i,j\geq 1} t_{i,j}^2 < \infty$, but for the present purpose of proving inequalities, we may as well assume than only finitely many coefficients $t_{i,j}$ are not 0. This random variable is called a (decoupled) order 2 Gaussian chaos. There is also a theory of non-decoupled chaos, $\sum_{i>j\geq 1} t_{i,j} g_i g_j$. For the present purposes, this theory reduces to the decoupled case using well understood arguments. For example, it is proved in [1] that

$$\mathsf{E} \sup_{t\in T} \Big| \sum_{i\neq j} t_{i,j} g_i g_j + \sum_{i\geq 1} t_{i,i}(g_i^2 - 1) \Big| \leq L \mathsf{E} \sup_{t\in T} \Big| \sum_{i,j\geq 1} t_{i,j} g_i g'_j \Big| \ . \tag{8.38}$$

Given a finite family T of double sequences $t = (t_{i,j})$, we would like to find upper and lower bounds for the quantity

$$S(T) = \mathsf{E} \sup_{t\in T} X_t \ . \tag{8.39}$$

We will first study the tails of the r.v.s (8.37) (a result which will be considerably extended in Section 8.3), and we will then use chaining to bound $S(T)$. One fundamental feature of the present situation is that there is a method radically different from chaining to bound this quantity. This method is revealed just after Theorem 8.2.2. Thus there is a sharp contrast between chaos and random series. For random series (as argued in Chapter 12) it seems difficult to imagine bounds which do not rely on either chaining or simple comparison properties.

The existence of the alternative method to bound chaos implies that the bounds obtained from chaining are not in general optimal, although, as we shall see, there are optimal under certain rather restrictive conditions.

We find it convenient to assume that the underlying probability space is a product $(\Omega \times \Omega' \ , \ \mathsf{P} = \mathsf{P}_0 \otimes \mathsf{P}')$, so that

$$X_t(\omega, \omega') = \sum_{i,j} t_{i,j} g_i(\omega) g'_j(\omega') \ .$$

We denote by E' integration in ω' only (i.e. conditional expectation given ω). Our first goal is the estimate (8.49) below on the tails of the r.v. X_t.

Conditionally on ω , X_t is a Gaussian r.v. and

$$\mathsf{E}' X_t^2 = \sum_{j\geq 1} \Big(\sum_{i\geq 1} t_{i,j} g_i(\omega) \Big)^2 \ . \tag{8.40}$$

Consider the r.v.

$$\sigma_t = \sigma_t(\omega) = (\mathsf{E}' X_t^2)^{1/2} ,$$

and note that $\mathsf{E}\sigma_t^2 = \mathsf{E}X_t^2$. Then

$$\begin{aligned}\sigma_t &= \sup_\alpha \sum_{j\ge 1} \alpha_j \Big(\sum_{i\ge 1} t_{i,j} g_i(\omega)\Big) \\ &= \sup_\alpha \sum_{i\ge 1} g_i(\omega) \Big(\sum_{j\ge 1} \alpha_j t_{i,j}\Big) := \sup_\alpha g_{t,\alpha} ,\end{aligned} \tag{8.41}$$

where the supremum is over the sequences $\alpha = (\alpha_j)$ with $\sum_{j\ge 1} \alpha_j^2 \le 1$.

Let us define

$$\begin{aligned}\|t\| &= \sup_\alpha \Big(\sum_{i\ge 1}\Big(\sum_{j\ge 1} \alpha_j t_{i,j}\Big)^2\Big)^{1/2} \\ &= \sup\Big\{\sum_{i,j\ge 1} \alpha_j \beta_i t_{i,j} \; ; \; \sum_{j\ge 1} \alpha_j^2 \le 1 \, , \, \sum_{i\ge 1} \beta_i^2 \le 1\Big\} .\end{aligned}$$

If we think of t as a matrix, $\|t\|$ is the operator norm of t from ℓ^2 to ℓ^2. We will also need the Hilbert-Schmidt norm of this matrix, given by

$$\|t\|_{HS} = \Big(\sum_{i,j\ge 1} t_{i,j}^2\Big)^{1/2} .$$

We note that $\|t\| \le \|t\|_{HS}$ by the Cauchy-Schwarz inequality. Also, recalling the Gaussian r.v. $g_{t,\alpha}$ of (8.41),

$$(\mathsf{E}g_{t,\alpha}^2)^{1/2} = \Big(\sum_{i\ge 1}\Big(\sum_{j\ge 1} \alpha_j t_{i,j}\Big)^2\Big)^{1/2} \le \|t\| ,$$

and since $\sigma_t = \sup_\alpha g_{t,\alpha}$, (2.84) implies that for $v > 0$,

$$\mathsf{P}(|\sigma_t - \mathsf{E}\sigma_t| \ge v) \le 2\exp\Big(-\frac{v^2}{2\|t\|^2}\Big) \tag{8.42}$$

so that in particular

$$\mathsf{E}(\sigma_t - \mathsf{E}\sigma_t)^2 \le L\|t\|^2 .$$

Denoting by $\|\cdot\|_2$ the norm in $L^2(\Omega)$, we thus have $\|\sigma_t - \mathsf{E}\sigma_t\|_2 \le L\|t\|$, so that $|\|\sigma_t\|_2 - \mathsf{E}\sigma_t| = |\|\sigma_t\|_2 - |\mathsf{E}\sigma_t|| \le L\|t\|$. Now

$$\|\sigma_t\|_2 = (\mathsf{E}\sigma_t^2)^{1/2} = (\mathsf{E}X_t^2)^{1/2} = \|t\|_{HS} , \tag{8.43}$$

so that $|\mathsf{E}\sigma_t - \|t\|_{HS}| \le L\|t\|$ and (8.42) implies

$$\mathsf{P}(|\sigma_t - \|t\|_{HS}| \ge v + L\|t\|) \le 2\exp\Big(-\frac{v^2}{2\|t\|^2}\Big) . \tag{8.44}$$

Taking $v = \|t\|_{HS}/4$, and distinguishing the cases whether $L\|t\| \le \|t\|_{HS}/4$ or not, we get

$$\mathsf{P}\Big(\sigma_t \le \frac{\|t\|_{HS}}{2}\Big) \le L\exp\Big(-\frac{\|t\|_{HS}^2}{L\|t\|^2}\Big) . \tag{8.45}$$

The random distance d_ω associated to the Gaussian process X_t (at given ω) is

$$d_\omega(s,t) = \sigma_{s-t}(\omega) . \tag{8.46}$$

Considering the two distances on T defined by

$$d_\infty(s,t) = \|t-s\| \, , \, d_2(s,t) = \|t-s\|_{HS} \tag{8.47}$$

we then have shown that

$$\mathsf{P}\Big(d_\omega(s,t) \le \frac{1}{2}d_2(s,t)\Big) \le L\exp\Big(-\frac{d_2^2(s,t)}{Ld_\infty^2(s,t)}\Big) . \tag{8.48}$$

Let us prove another simple classical fact (proved first in [2]).

Lemma 8.2.1. *For $v \ge 0$ we have*

$$\mathsf{P}(|X_t| \ge v) \le L\exp\Big(-\frac{1}{L}\min\Big(\frac{v^2}{\|t\|_{HS}^2}\, , \, \frac{v}{\|t\|}\Big)\Big) . \tag{8.49}$$

Proof. Given ω, the r.v. X_t is Gaussian so that

$$\mathsf{P}'(|X_t| \ge v) \le 2\exp\Big(-\frac{v^2}{2\sigma_t^2}\Big) ,$$

and, given $a > 0$

$$\begin{aligned}\mathsf{P}(|X_t| \ge v) = \mathsf{E}\mathsf{P}'(|X_t| \ge v) &\le 2\mathsf{E}\exp\Big(-\frac{v^2}{2\sigma_t^2}\Big) \\ &\le 2\exp\Big(-\frac{v^2}{2a^2}\Big) + 2\mathsf{P}(\sigma_t \ge a) .\end{aligned}$$

Since $\|t\| \le \|t\|_{HS}$, it follows from (8.44) that $\mathsf{P}(\sigma_t \ge v + L\|t\|_{HS}) \le L\exp(-v^2/2\|t\|^2)$ and thus when $a \ge L\|t\|_{HS}$,

$$\mathsf{P}(\sigma_t \ge a) \le L\exp\Big(-\frac{a^2}{L\|t\|^2}\Big) .$$

Consequently,

$$\mathsf{P}(|X_t| \ge v) \le 2\exp\Big(-\frac{v^2}{2a^2}\Big) + L\exp\Big(-\frac{a^2}{L\|t\|^2}\Big) . \tag{8.50}$$

To finish the proof we take $a = \max\big(L\|t\|_{HS}\, , \, \sqrt{v\|t\|}\big)$ and we observe that the last term in (8.50) is always at most $L\exp(-v/(L\|t\|))$. □

As a consequence of (8.49), we have

$$\mathsf{P}(|X_s - X_t| \geq v) \leq L \exp\Big(-\frac{1}{L}\min\Big(\frac{v^2}{d_2^2(s,t)}, \frac{v}{d_\infty(s,t)}\Big)\Big) \tag{8.51}$$

and Theorem 2.2.23 implies the following.

Theorem 8.2.2. *For a set T of sequences $(t_{i,j})$, we have*

$$S(T) = \mathsf{E}\sup_{t\in T} X_t \leq L\big(\gamma_1(T,d_\infty) + \gamma_2(T,d_2)\big) . \tag{8.52}$$

We analyze now a very interesting example of set T. Given an integer n, we consider

$$T = \{t \; ; \; \|t\| \leq 1 \, , \, t_{i,j} \neq 0 \Rightarrow i \, , \, j \leq n\} . \tag{8.53}$$

Since

$$\sum_{i,j} t_{ij} g_i g'_j \leq \Big(\sum_{i\leq n} g_i^2\Big)^{1/2} \Big(\sum_{j\leq n} g_j'^2\Big)^{1/2} \|t\| ,$$

the Cauchy-Schwarz inequality implies that $S(T) \leq n$. On the other hand, volume arguments show that $\log N(T, d_\infty, 1/4) \geq n^2/L$, so that $\gamma_1(T, d_\infty) \geq n^2/L$. It is also simple to prove that (see [5])

$$\log N(T, d_2, \sqrt{n}/L) \geq n^2/L ,$$

and that $S(T)$ is about n, $\gamma_1(T, d_\infty)$ is about n^2 and $\gamma_2(T, d_2)$ is about $n^{3/2}$. In this case (8.52) is not sharp, which means that there is no hope of reversing this inequality in general. This is so despite the fact that we have used a competent chaining method and that the bounds (8.51) are essentially optimal (as follows e.g. from the left-hand side of (8.97) below). It can also be shown that in the case where the elements t of T satisfy $t_{i,j} = 0$ for $i \neq j$ the bound (8.52) can be reversed. This is essentially proved in Theorem 10.2.8 below.

We now turn to a result involving a very special class of chaos, which we will bound by a method which is different from both the method of Theorem 8.2.2 and of the method used for the set (8.53). To lighten notation we denote by tg the sequence $(\sum_{j\geq 1} t_{i,j} g_j)_{i\geq 1}$, by $\langle\cdot,\cdot\rangle$ the dot product in ℓ^2 and by $\|\cdot\|_2$ the corresponding norm. For $t = (t_{i,j})$ let us write

$$Y_t^* := \sum_{i\geq 1}\Big(\sum_{j\geq 1} t_{i,j} g_j\Big)^2 = \|tg\|_2^2 = \langle tg, tg\rangle = \sum_{i\geq 1}\sum_{j,k\geq 1} t_{i,j} t_{i,k} g_j g_k \tag{8.54}$$

and

$$Y_t := Y_t^* - \mathsf{E}Y_t^* = \sum_{i\geq 1}\sum_{j\neq k} t_{i,j} t_{i,k} g_j g_k + \sum_{i\geq 1}\sum_{j\geq 1} t_{i,j}^2 (g_j^2 - 1) . \tag{8.55}$$

Theorem 8.2.3 ([3]). *For any set T with $0 \in T$ we have*

$$\mathsf{E}\sup_{t\in T}|Y_t| \le L\gamma_2(T,d_\infty)\Big(\gamma_2(T,d_\infty)+\sup_{t\in T}\|t\|_{HS}\Big)\ . \tag{8.56}$$

This theorem belongs to the present section because Y_t is an order 2 chaos, as shown by (8.55). Let us define, with obvious notation,

$$Z_t = \sum_{i,j,k\ge 1} t_{i,j}t_{i,k}g_j g'_k = \langle tg, tg'\rangle\ .$$

The main step of the proof of Theorem 8.2.3 is as follows.

Proposition 8.2.4. *Let $U^2 := \mathsf{E}\sup_{t\in T}\|tg\|_2^2$. Then*

$$\mathsf{E}\sup_{t\in T}|Z_t| \le LU\gamma_2(T,d_\infty)\ . \tag{8.57}$$

Proof of Theorem 8.2.3. Let us define $V = \sup_{t\in T}\|t\|_{HS}$ so that $V^2 = \sup_{t\in T}\|t\|_{HS}^2 = \sup_{t\in T}\sum_{i,j\ge 1}t_{i,j}^2 = \sup_{t\in T}\mathsf{E}Y_t^*$. For $t\in T$ we have $\|tg\|_2^2 = Y_t^* = Y_t + \mathsf{E}Y_t^* \le Y_t + V^2$ and thus

$$U^2 \le V^2 + \mathsf{E}\sup_{t\in T}|Y_t|\ . \tag{8.58}$$

Now, combining (8.38) and (8.55) we have

$$\mathsf{E}\sup_{t\in T}|Y_t| \le L\mathsf{E}\sup_{t\in T}|Z_t|\ , \tag{8.59}$$

so that, combining with (8.58) and (8.57) we obtain

$$U^2 \le V^2 + LU\gamma_2(T,d_\infty)\ ,$$

and thus $U \le L(V+\gamma_2(T,d_\infty))$. Plugging in (8.57) proves the result. □

Proof of Proposition 8.2.4. Without loss of generality we assume that T is finite. Consider an admissible sequence $(\mathcal{A}_n)$ with

$$\sup_{t\in T}\sum_{n\ge 0}2^{n/2}\Delta(A_n(t)) \le 2\gamma_2(T,d_\infty)\ ,$$

where the diameter Δ is for the distance d_∞. For $A\in A_n$ consider an element $t_{A,n}\in A$ and define as usual a chaining by $\pi_n(t) = t_{A_n(t),n}$. Since $0\in T$, without loss of generality we may assume that $\pi_0(t)=0$. We observe that

$$\begin{aligned} Z_{\pi_n(t)} - Z_{\pi_{n-1}(t)} &= \langle(\pi_n(t)-\pi_{n-1}(t))g, \pi_n(t)g'\rangle \\ &\quad + \langle \pi_{n-1}(t)g, (\pi_n(t)-\pi_{n-1}(t))g'\rangle\ . \end{aligned} \tag{8.60}$$

Recalling that we think of each t as an operator on ℓ^2 let us denote by t^* its adjoint. Thus

$$\langle(\pi_n(t)-\pi_{n-1}(t))g,\pi_n(t)g'\rangle=\langle g,(\pi_n(t)-\pi_{n-1}(t))^*\pi_n(t)g'\rangle\ . \tag{8.61}$$

Here of course $(\pi_n(t)-\pi_{n-1}(t))^*\pi_n(t)g'$ is the element of ℓ^2 obtained by applying the operator $(\pi_n(t)-\pi_{n-1}(t))^*$ to the vector $\pi_n(t)g'$. Let us now consider the r.v.s $W=\sup_{t\in T}\|tg\|_2$ and $W'=\sup_{t\in T}\|tg'\|_2$. Then

$$\|(\pi_n(t)-\pi_{n-1}(t))^*\pi_n(t)g'\|_2\le\|(\pi_n(t)-\pi_{n-1}(t))^*\|\|\pi_n(t)g'\|_2\le\Delta(A_n(t))W'.$$

It then follows from (8.61) that, conditionally on g', the quantity $\langle(\pi_n(t)-\pi_{n-1}(t))g,\pi_n(t)g'\rangle$ is simply a Gaussian r.v. G with $(EG^2)^{1/2}\le\Delta(A_n(t))W'$. Thus we obtain that for $u\ge 1$

$$\mathsf{P}\big(|\langle(\pi_n(t)-\pi_{n-1}(t))g,\pi_n(t)g'\rangle|\ge 2^{n/2}u\Delta(A_n(t))W'\big)\le\exp(-u^2 2^n/2)\ .$$

Proceeding in a similar fashion for the second term in (8.60) we get

$$\mathsf{P}\big(|Z_{\pi_n(t)}-Z_{\pi_{n-1}(t)}|\ge 2u2^{n/2}\Delta(A_n(t))(W+W')\big)\le 2\exp(-u^2 2^n/2)\ .$$

Using that $Z_{\pi_0(t)}=0$, and proceeding just as in the proof of the generic chaining bound (2.31), we obtain that for $u\ge L$,

$$\mathsf{P}\Big(\sup_{t\in T}|Z_t|\ge Lu\gamma_2(T,d_\infty)(W+W')\Big)\le L\exp(-u^2)\ .$$

In particular the function $R=\sup_{t\in T}|Z_t|/(W+W')$ satisfies $\mathsf{E}R^2\le L\gamma_2(T,d_\infty)^2$. Since $\mathsf{E}W^2=\mathsf{E}W'^2=U^2$ the Cauchy-Schwarz inequality yields (8.57). □

We return to the study of general chaos processes. When T is "small for the distance d_∞" it follows from (8.51) that the process $(X_t)_{t\in T}$ resembles a Gaussian process, so that there should be a close relationship between $S(T)=\mathsf{E}\sup_{t\in T}X_t$ and $\gamma_2(T,d_2)$. The next result, where we recall the notation (8.39), is a step in this direction. It should be compared with Theorem 5.4.1.

Theorem 8.2.5. *We have*

$$\gamma_2(T,d_2)\le L\big(S(T)+\sqrt{S(T)\gamma_1(T,d_\infty)}\big)\ . \tag{8.62}$$

The example (8.53) provides a situation where this inequality is sharp, since then both the left-hand and the right-hand sides are of order $n^{3/2}$. Combining with Theorem 8.2.2, this implies the following.

Corollary 8.2.6. *If we define*

$$R=\frac{\gamma_1(T,d_\infty)}{\gamma_2(T,d_2)}\ ,$$

then

$$\frac{1}{L(1+R)}\gamma_2(T,d_2)\le S(T)\le L(1+R)\gamma_2(T,d_2)\ . \tag{8.63}$$

In particular, $S(T)$ is of order $\gamma_2(T, d_2)$ when R is of order 1 or smaller.

Proof. The right-hand side is obvious from (8.52). To obtain the left-hand side, we simply write in (8.62) that, since $\sqrt{ab} \le (a+b)/2$,

$$\begin{aligned}\sqrt{S(T)\gamma_1(T, d_\infty)} &= \sqrt{S(T)R\gamma_2(T, d_2)}\\ &\le \frac{1}{2}\Big(\frac{1}{L}\gamma_2(T, d_2) + LS(T)R\Big)\end{aligned}$$

where L is as in (8.62), and together with (8.62) this yields

$$\gamma_2(T, d_2) \le LS(T) + \frac{1}{2}\gamma_2(T, d_2) + LS(T)R\,. \qquad \square$$

Theorem 8.2.5 relies on the following abstract statement.

Theorem 8.2.7. *Consider a finite set T, provided with two distances d_∞ and d_2. Consider a random distance d_ω on T, and a number $\alpha > 0$. Assume that*

$$\forall s, t \in T\,,\ \mathsf{P}\big(d_\omega(s,t) \ge \alpha d_2(s,t)\big) \ge \alpha \tag{8.64}$$

$$\forall s, t \in T\,,\ \mathsf{P}\big(d_\omega(s,t) \le \alpha d_2(s,t)\big) \le \frac{1}{\alpha}\exp\Big(-\alpha\frac{d_2^2(s,t)}{d_\infty^2(s,t)}\Big)\,. \tag{8.65}$$

Consider a number M such that

$$\mathsf{P}(\gamma_2(T, d_\omega) \le M) \ge 1 - \alpha/2\,. \tag{8.66}$$

Then

$$\gamma_2(T, d_2) \le K(\alpha)\big(M + \sqrt{M\gamma_1(T, d_\infty)}\big)\,, \tag{8.67}$$

where $K(\alpha)$ depends on α only.

Proof of Theorem 8.2.5. By (8.48), the pair of distances d_∞ and d_2 of (8.47) satisfies (8.65) whenever α is small enough. The formula (8.41) makes σ_t, and hence σ_{s-t}, appear as the supremum of a Gaussian process. Applying (8.35) to this process yields $\mathsf{P}(\sigma_{s-t} \ge (\mathsf{E}\sigma_{s-t}^2)^{1/2}/L) \ge 1/L$. Consider now $d_\omega(s,t) = \sigma_{s-t}(\omega)$ as in (8.46). Then $\mathsf{E}\sigma_{s-t}^2 = \|s-t\|_2^2 = d_2(s,t)^2$, so that (8.43) implies that (8.64) holds whenever α is small enough.

Next we prove that (8.66) holds for $M = LS(T)/\alpha$. Since $\mathsf{E}\mathsf{E}' \sup_{t\in T} X_t = S(T)$, and since $\mathsf{E}' \sup_{t\in T} X_t \ge 0$, Markov inequality implies

$$\mathsf{P}\Big(\mathsf{E}' \sup_{t\in T} X_t \le 2S(T)/\alpha\Big) \ge 1 - \alpha/2\,.$$

Since $L\mathsf{E}' \sup_{t\in T} X_t \ge \gamma_2(T, d_\omega)$ by Theorem 2.4.1, this proves that (8.66) holds for $M = LS(T)/\alpha$. Thus (8.62) is a consequence of (8.67). $\square$

Proof of Theorem 8.2.7. We consider the subset U of Ω given by $U = \{\gamma_2(T, d_\omega) \le M\}$, so that $\mathsf{P}(U) \ge 1 - \alpha/2$ by hypothesis. Let us fix once and for all an admissible sequence $(\mathcal{C}_n)_{n\ge 0}$ of partitions of T such that

$$\forall t \in T\,, \sum_{n\geq 0} 2^n \Delta(C_n(t), d_\infty) \leq 2\gamma_1(T, d_\infty)\;.$$

We consider an integer $\tau \geq 0$, that will be chosen later. Given a probability measure μ on T, we define

$$F_n(\mu) = \mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int \Big(\sum_{k\geq n} 2^{k/2}\Delta(A_k(t), d_\omega) + \sum_{\ell\geq n} 2^\ell \Delta(C_{\ell+\tau}(t), d_\infty)\Big)\mathrm{d}\mu(t)\Big),$$

where the infimum is over all choices of the admissible sequence $(\mathcal{A}_k)$. Given $A \subset T$, we define

$$F_n(A) = \sup\{F_n(\mu)\;;\; \exists C \in \mathcal{C}_{n+\tau}\,,\, \mu(C \cap A) = 1\}\;.$$

Thus, since $\int f(t)\mathrm{d}\mu(t) \leq \sup_{t\in T} f(t)$, we get

$$\begin{aligned} F_0(T) &\leq \mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}}\Big(\sup_{t\in T}\sum_{k\geq 0} 2^{k/2}\Delta(A_k(t), d_\omega) + \sup_{t\in T}\sum_{\ell\geq 0} 2^\ell\Delta(C_{\ell+\tau}(t), d_\infty)\Big)\Big) \\ &\leq \mathsf{E}\big(\mathbf{1}_U(\gamma_2(T, d_\omega) + 2^{-\tau+1}\gamma_1(T, d_\infty))\big) \\ &\leq M + 2^{-\tau+1}\gamma_1(T, d_\infty)\;, \end{aligned} \tag{8.68}$$

where in the second inequality we have used that

$$\sup_{t\in T}\sum_{\ell\geq 0} 2^{\ell+\tau}\Delta(C_{\ell+\tau}(t), d_\infty) \leq \sup_{t\in T}\sum_{k\geq 0} 2^k \Delta(C_k(t), d_\infty) \leq 2\gamma_1(T, d_\infty)\;.$$

Consider $n \geq 0$, and set $m = N_{n+\tau+3}$. Consider points $(t_\ell)_{\ell\leq m}$ of T, with $d_2(t_\ell, t_{\ell'}) \geq 4a$ when $\ell \neq \ell'$ and sets $H_\ell \subset B_2(t_\ell, a)$. We will prove later that if $\tau \geq \tau_0$, where τ_0 depends only on the value of the constant α, then

$$F_n\Big(\bigcup_{\ell\leq m} H_\ell\Big) \geq \frac{2^{n/2}}{K}a + \min_{\ell\leq m} F_{n+1}(H_\ell)\,. \tag{8.69}$$

Here, as well as in the rest of this proof, K denotes a number depending on α only, not necessarily the same at each occurrence. Using Theorem 2.7.2 with $r = 4$, $\theta(n) = 2^{n/2}/L$, and $\tau + 3$ rather than τ, we then get

$$\gamma_2(T, d_2) \leq K 2^{\tau/2}(F_0(T) + \Delta(T, d_2))\;. \tag{8.70}$$

To bound $\Delta(T, d_2)$, considering $s, t \in T$ with $d_2(t, s) = \Delta(T, d_2)$, we obtain from (8.64) that

$$\mathsf{P}\big(d_\omega(t, s) \geq \alpha\Delta(T, d_2)\big) \geq \alpha\;.$$

Since $\gamma_2(T, d_\omega) \geq d_\omega(t, s)$, and since $1 - \alpha/2 + \alpha > 1$, it follows from (8.66) that $\Delta(T, d_2) \leq LM$. Thus (8.68) and (8.70) imply

$$\gamma_2(T, d_2) \leq K 2^{\tau/2}(M + 2^{-\tau}\gamma_1(T, d_\infty))\;.$$

Optimization over $\tau \geq \tau_0$ then gives (8.67).

We turn to the proof of (8.69). It closely resembles the proof of (8.25). Consider $c < \inf_\ell F_{n+1}(H_\ell)$, and for $\ell \le m$ consider a set $C_\ell \in \mathcal{C}_{n+\tau+1}$ and a probability measure μ_ℓ on $H_\ell \cap C_\ell$ such that $F_{n+1}(\mu_\ell) > c$. There are only at most $N_{n+\tau+1}$ possible values for the set C_ℓ. Since $m = N_{n+\tau+3} \ge N_{n+\tau+2} N_{n+\tau+1}$, we can find a subset I of $\{1, \dots, m\}$ with $\operatorname{card} I \ge N_{n+\tau+2}$ such that for all $\ell \in I$ the set C_ℓ is the same element of $\mathcal{C}_{n+\tau+1}$. In particular there exists an element C_0 of $\mathcal{C}_{n+\tau}$ such that for $C_\ell \subset C_0$ for $\ell \in I$. We define

$$\mu = \frac{1}{\operatorname{card} I} \sum_{\ell \in I} \mu_\ell \ ,$$

so that $\mu(\bigcup_{\ell \le m} H_\ell \cap C_0) = 1$. Thus $F_n(A) \ge F_n(\mu)$ and it suffices to prove that

$$F_n(\mu) \ge \frac{2^{n/2}}{K} a + c \ .$$

When t belongs to the support of μ, t belongs to C_0 and therefore $C_{n+\tau}(t) = C_0$. Proceeding as in the proof of Theorem 8.1.3, it suffices to prove that

$$2^n \Delta(C_0, d_\infty) + \mathsf{E}\Big(\mathbf{1}_U \inf_{\mathcal{A}} \int 2^{n/2} \Delta(A_n(t), d_\omega) \mathrm{d}\mu(t) \Big) \ge \frac{a 2^{n/2}}{K} \ . \tag{8.71}$$

Consider a number K_1 large enough that $\alpha K_1^2 \ge 16$ and $2 \exp(-\alpha K_1^2/8) \le \alpha^2$. Then for $n \ge 0$ one has $\alpha K_1^2 2^{n-3} \ge 2^{n+1}$ and thus

$$\frac{1}{\alpha} \exp(-\alpha K_1^2 2^{n-2}) \le \frac{1}{\alpha} \exp(-\alpha K_1^2/8) \exp(-\alpha K_1^2 2^{n-3}) \le \frac{\alpha}{2N_{n+1}} \ . \tag{8.72}$$

If $\Delta(C_0, d_\infty) > a 2^{-n/2}/K_1$, then (8.71) holds true, so that we may assume that $\Delta(C_0, d_\infty) \le a 2^{-n/2}/K_1$.

First, we prove that the event

$$\mu^{\otimes 2}\big(\{ (x,y) \in T^2 \ ; \ d_2(x,y) \ge a/2 \ , \ d_\omega(x,y) \le a/2 \} \big) \le \frac{1}{N_{n+1}} \tag{8.73}$$

has a probability $\ge 1 - \alpha/2$. For this, let us denote by Y the left-hand side of (8.73). Then, using (8.65) in the second line (and since μ is supported by C_0 so that $d_\infty(x,y) \le a 2^{-n/2}/K_1$ for almost every pair (x,y)) and (8.72) in the last inequality, we obtain

$$\begin{aligned} \mathsf{E} Y &= \int_{\{d_2(x,y) \ge a/2\}} \mathsf{P}(d_\omega(x,y) \le a/2) \mathrm{d}\mu(x) \mathrm{d}\mu(y) \\ &\le \frac{1}{\alpha} \exp\Big(-\alpha \frac{(a/2)^2}{(a 2^{-n/2}/K_1)^2} \Big) = \frac{1}{\alpha} \exp(-\alpha K_1^2 2^{n-2}) \le \frac{\alpha}{2N_{n+1}} \ , \end{aligned}$$

and thus as claimed $\mathsf{P}(Y \le 1/N_{n+1}) \ge 1 - \alpha/2$. Next, as in the proof of Lemma 8.1.7, since $d_2(x,y) \ge a/2$ when x and y belong to two different sets H_ℓ, when (8.73) occurs we have

$$\mu^{\otimes 2}(\{(x,y) \in T^2 \; ; \; d_\omega(x,y) \leq a/2\}) \leq \frac{2}{N_{n+1}} \; , \tag{8.74}$$

and as in the second part of the proof of Lemma 8.1.7 we show that when (8.74) occurs then for each admissible sequence $\mathcal{A}$ one has

$$\int \Delta(A_n(t), d_\omega) \mathrm{d}\mu(t) \geq \frac{a}{L} \; ,$$

and this finishes the proof. □

Let us give a simple consequence of Theorem 8.2.5. We recall the covering numbers $N(T, d, \epsilon)$ of Section 1.2. We recall that $S(T) = \mathsf{E} \sup_{t \in T} X_t$.

Proposition 8.2.8. *There exists a constant L with the following property:*

$$\epsilon \geq L\sqrt{\Delta(T, d_\infty) S(T)} \Rightarrow \epsilon \sqrt{\log N(T, d_2, \epsilon)} \leq L S(T) \; . \tag{8.75}$$

A remarkable feature of (8.75) is that as, we shall now prove, the right-hand side need not hold if $\epsilon \leq \sqrt{\Delta(T, d_\infty) S(T)}/L$ (see however (8.79) below). To see this, let us consider the example (8.53). For $\epsilon = \sqrt{n}/L$ we have $\epsilon\sqrt{\log(N(T, d_2, \epsilon))} \geq n^{3/2}/L$, while $S(T) \leq Ln$, so that the right-hand side of (8.75) does not hold. Moreover, since $\Delta(T, d_\infty) = 2$, ϵ is of the order of $\sqrt{\Delta(T, d_\infty) S(T)}$. This shows that the condition $\epsilon \geq L\sqrt{\Delta(T, d_\infty) S(T)}$ in (8.75) is rather precise.

Proof of Proposition 8.2.8. Assume first that T is finite, $\operatorname{card} T = m$, and consider a number α with

$$\alpha \geq \Delta(T, d_\infty) \, . \tag{8.76}$$

Assume that for certain a number ϵ, we have

$$\forall s, t \in T \, , \; s \neq t \, , \; d_2(s,t) = \|t - s\|_{HS} \geq \epsilon \; . \tag{8.77}$$

Lemma 2.4.2 and Theorem 2.4.1 imply that $\gamma_2(T, d_2) \geq \epsilon\sqrt{\log m}/L$ (see also Exercise 2.2.21). Moreover, $\gamma_1(T, d_\infty) \leq L\alpha \log m$, as is witnessed by an admissible sequence $(\mathcal{A}_n)$ such that if $N_n \geq m$, then each set $A \in \mathcal{A}_n$ contains exactly one point (see Exercise 2.2.21 (b)). Now (8.62) implies

$$\begin{aligned}\frac{\epsilon}{L}\sqrt{\log m} \leq \gamma_2(T, d_2) &\leq L\big(S(T) + \sqrt{S(T)\gamma_1(T, d_\infty)}\big) \\ &\leq L\big(S(T) + \sqrt{S(T)\alpha \log m}\big) \; .\end{aligned} \tag{8.78}$$

Let us denote by L_2 the constant in the previous inequality. Now, if $\epsilon \geq L_3\sqrt{\alpha S(T)}$ where $L_3 = 2(L_2)^2$, we have $\sqrt{S(T)\alpha \log m} \leq \varepsilon\sqrt{\log m}/L_3$, so that (8.78) implies

$$\frac{\epsilon}{L_2}\sqrt{\log m} \leq L_2 S(T) + \frac{1}{2L_2}\epsilon\sqrt{\log m}$$

and therefore $\epsilon\sqrt{\log m} \leq L S(T)$.

If now T is given satisfying (8.76), consider $T' \subset T$ that satisfies (8.77) and has a cardinality m as large as possible. Then we have shown that if $\epsilon \geq L_3\sqrt{\Delta(T,d_\infty)S(T)}$ we must have $\epsilon\sqrt{\log m} \leq LS(T') \leq LS(T)$. Since the cardinality of T' is as large as possible, the balls centered at the points of T' of radius ϵ cover T, so that $N(T,d_2,\epsilon) \leq m$. □

The proof of Proposition 8.2.8 does not use the full strength of Theorem 8.2.7, and we propose the following as a very challenging exercise.

Exercise 8.2.9. Find a direct proof that under the conditions of Theorem 8.2.7 one has

$$\epsilon \geq L\sqrt{M\Delta(T,d_\infty)} \Rightarrow \epsilon\sqrt{\log N(T,d_2,\epsilon)} \leq LM \ ,$$

and use this result to find a more direct proof of Proposition 8.2.8. (Hint: it helps to prove first that $\Delta(T,d_\infty) \leq LM$.)

For completeness let us mention the following, which should of course be compared with (8.75).

Proposition 8.2.10. *For each $\epsilon > 0$, we have*

$$\epsilon(\log N(T,d_2,\epsilon))^{1/4} \leq LS(T) \ . \tag{8.79}$$

In the previous example (8.53), both sides are of order n for $\epsilon = \sqrt{n}/L$.

Research problem 8.2.11. Is it true that

$$\epsilon\sqrt{\log N(T,d_\infty,\epsilon)} \leq LS(T) \ ? \tag{8.80}$$

For a partial result, and a proof of Proposition 8.2.10, see [8].

It is interesting to observe that (8.80) would provide another proof of (8.79). Indeed by (8.80) we would have

$$\log N(T,d_\infty,\alpha) \leq L\frac{S(T)^2}{\alpha^2} \ .$$

Now, if B is a ball $B_\infty(t,\alpha)$ of T for $\alpha = \epsilon^2/L'S(T)$, since $\Delta(B_\infty(t,\alpha),d_\infty) \leq 2\alpha$, for L' large enough the right hand side of (8.75) holds and this inequality implies

$$\log N(B,d_2,\epsilon) \leq \frac{S(T)^2}{\epsilon^2} \ .$$

Since $N(T,d_2,\epsilon) \leq N(T,d_\infty,\alpha)\max_B N(B,d_2,\epsilon)$, combining these yields

$$\log N(T,d_2,\epsilon) \leq L\Big(\frac{S(T)^4}{\epsilon^4} + \frac{S(T)^2}{\epsilon^2}\Big)$$

and this would prove (8.79).

To conclude this section, we describe a way to control $S(T)$ from above, which is really different from both the method of Theorem 8.2.2 and of Theorem 8.2.3.

Given a convex balanced subset U of ℓ^2 (that is, $\lambda U \subset U$ for $|\lambda| \le 1$, or, equivalently, $U = -U$), we define

$$g(U) = \mathsf{E} \sup_{(u_i)\in U} \sum_{i\ge1} u_i g_i$$

$$\sigma(U) = \sup_{(u_i)\in U} \Big(\sum_{i\ge1} u_i^2\Big)^{1/2} .$$

Given convex balanced subsets U and V of ℓ^2, we define

$$T_{U,V} = \Big\{ t = (t_{i,j}) \; ; \; \forall (x_i)_{i\ge1} \, , \, \forall (y_j)_{j\ge1} \, , \\ \sum t_{i,j} x_i y_j \le \sup_{(u_i)\in U} \sum_{i\ge1} x_i u_i \sup_{(v_j)\in V} \sum_{j\ge1} y_j v_j \Big\} .$$

This is a in a sense a generalization of the example (8.53) to other norms than the Euclidean norm. It follows from (2.84) that, if $w > 0$,

$$\mathsf{P}\Big(\sup_{(u_i)\in U} \sum_{i\ge1} g_i u_i \ge g(U) + w\sigma(U) \Big) \le 2\exp\Big(-\frac{w^2}{2}\Big) ,$$

so that (using that for positive numbers, when $ab > cd$ we have either $a > c$ or $b > d$)

$$\mathsf{P}\Big(\sup_{(u_i)\in U} \sum_{i\ge1} g_i u_i \sup_{(v_j)\in V} \sum g'_j v_j \ge g(U)g(V) \\ + \; w(\sigma(U)g(V) + \sigma(V)g(U)) + w^2\sigma(U)\sigma(V) \Big) \\ \le \; 4\exp\Big(-\frac{w^2}{2}\Big) . \tag{8.81}$$

Now

$$\sup_{t\in T_{U,V}} X_t \le \sup_{(u_i)\in U} \sum_{i\ge1} u_i g_i \sup_{(v_j)\in V} \sum_{j\ge1} v_j g'_j ,$$

so that, whenever $g(U), g(V) \le 1$ and $\sigma(U), \sigma(V) \le 2^{-n/2}$ (8.81) yields

$$\mathsf{P}\Big(\sup_{t\in T_{U,V}} X_t \ge (1 + 2^{-n/2}w)^2 \Big) \le 4\exp\Big(-\frac{w^2}{2}\Big) .$$

Changing w into $2^{n/2}w$, (8.81) yields

$$\mathsf{P}\Big(\sup_{t\in T_{U,V}} X_t \ge (1 + w)^2 \Big) \le 4\exp(-2^{n-1}w^2) . \tag{8.82}$$

Proposition 8.2.12. *Consider for* $n \geq 0$ *a family* $\mathcal{C}_n$ *of pairs of convex balanced subsets of* ℓ^2. *Assume that* $\operatorname{card} \mathcal{C}_n \leq N_n$ *and that*

$$\forall (U,V) \in \mathcal{C}_n \,,\, g(U)\,,\, g(V) \leq 1 \,;\; \sigma(U)\,,\, \sigma(V) \leq 2^{-n/2} \,.$$

Then, the set

$$T = \operatorname{conv}\Bigl\{ \bigcup_n \bigcup_{(U,V)\in\mathcal{C}_n} T_{U,V} \Bigr\}$$

satisfies $S(T) \leq L$.

Proof. This should be obvious from (8.82) since

$$\mathsf{P}\Bigl(\sup_T X_t \geq w\Bigr) \leq \sum_n \sum_{(U,V)\in\mathcal{C}_n} \mathsf{P}\Bigl(\sup_{t\in T_{U,V}} X_t \geq w\Bigr) \,. \qquad \square$$

Having found three distinct way of controlling $S(T)$, one should certainly ask whether there are more. It simply seems too early to even make a sensible conjecture about what might be the "most general way to bound a chaos process".

8.3 Tails of Multiple Order Gaussian Chaos

In this section we consider a single order d (decoupled) Gaussian chaos, that is a r.v. X of the type

$$X = \sum_{i_1,\ldots,i_d} a_{i_1,\ldots,i_d} g^1_{i_1} \cdots g^d_{i_d} \,, \tag{8.83}$$

where $a_{i_1,\ldots,i_d}$ are numbers and g^j_i are independent standard Gaussian r.v.s. The sum is finite, each index i_ℓ runs from 1 to m. Our purpose is to estimate the higher moments of the r.v. X as a function of certain characteristics of

$$A := (a_{i_1,\ldots,i_d})_{i_1,\ldots,i_d \leq m} \,. \tag{8.84}$$

Estimating the higher moments of the r.v. X amounts to estimate its tails, and it is self evident that this is a natural question. This topic runs into genuine notational difficulties. One may choose to avoid considering tensors, in which case one faces heavy multi-index notation. Or one may entirely avoid multi-index notation using tensors, but one gets dizzy from the height of the abstraction. We shall not try for elegance in the presentation, but rather to minimize the amount of notation the reader has to assimilate. Our approach will use a dash of tensor vocabulary, but does not require any knowledge of what these are. In any case for the really difficult arguments we shall focus on the case $d = 3$.

Let us start with the case $d = 2$ that we considered at length in the previous section. In that case one may think of A as a linear functional on $\mathbb{R}^{m^2}$ by the formula

$$A(x) = \sum_{i,j} a_{i,j} x_{i,j} \;, \tag{8.85}$$

where $x = (x_{i,j})_{i,j \le m}$ is the generic element of $\mathbb{R}^{m^2}$. It is of course understood that in (8.85) the sum runs over $i, j \le m$. When we provide $\mathbb{R}^{m^2}$ with the canonical Euclidean structure, the norm of A viewed as a linear functional on $\mathbb{R}^{m^2}$ is simply

$$\|A\|_{\{1,2\}} := \Big(\sum_{i,j} a_{i,j}^2\Big)^{1/2} \,. \tag{8.86}$$

This quantity was denoted $\|A\|_{HS}$ in the previous section, but here we need new notation. We may also think of A as a bilinear functional on $\mathbb{R}^m \times \mathbb{R}^m$ by the formula

$$A(x,y) = \sum_{i,j} a_{i,j} x_i y_j \;, \tag{8.87}$$

where $x = (x_i)_{i \le m}$ and $y = (y_i)_{i \le m}$. In that case, if we provide both copies of $\mathbb{R}^m$ with the canonical Euclidean structure, the corresponding norm of A is

$$\|A\|_{\{1\}\{2\}} := \sup\Big\{\Big|\sum_{i,j} a_{i,j} x_i y_j\Big| \;;\; \sum x_i^2 \le 1, \sum y_j^2 \le 1\Big\} \;, \tag{8.88}$$

which is also the operator norm when one see A as a matrix, i.e. an operator from $\mathbb{R}^m$ to $\mathbb{R}^m$. One observes the inequality $\|A\|_{\{1\}\{2\}} \le \|A\|_{\{1,2\}}$.

Let us now turn to the case $d = 3$. One may think of A as a linear functional on $\mathbb{R}^{m^3}$, obtaining the norm

$$\|A\|_{\{1,2,3\}} := \Big(\sum_{i,j,k} a_{i,j,k}^2\Big)^{1/2} \,, \tag{8.89}$$

or think of A as a trilinear functional on $(\mathbb{R}^m)^3$, obtaining the norm

$$\|A\|_{\{1\}\{2\}\{3\}} := \sup\Big\{\Big|\sum_{i,j,k} a_{i,j,k} x_i y_j z_k\Big| \;;\; \sum x_i^2 \le 1, \sum y_j^2 \le 1, \sum z_k^2 \le 1\Big\} \,. \tag{8.90}$$

One may also view A as a bilinear function on $\mathbb{R}^{m^2} \times \mathbb{R}^m$ by the formula

$$A(x,y) = \sum_{i,j,k} a_{i,j,k} x_{i,j} y_k \;, \tag{8.91}$$

for $x = (x_{i,j})_{i,j} \in \mathbb{R}^{m^2}$ and $(y_k) \in \mathbb{R}^m$. One then obtains the norm

$$\|A\|_{\{1,2\}\{3\}} := \sup\Big\{\Big|\sum_{i,j,k} a_{i,j,k} x_{i,j} y_k\Big| \;;\; \sum x_{i,j}^2 \le 1, \sum y_k^2 \le 1\Big\} \,. \tag{8.92}$$

We observe the inequality

$$\|A\|_{\{1\}\{2\}\{3\}} \le \|A\|_{\{1,2\}\{3\}} \le \|A\|_{\{1,2,3\}} \; . \tag{8.93}$$

More generally, given a partition $\mathcal{P} = \{I_1, \dots, I_k\}$ of $\{1, \dots, d\}$ we may define the norm

$$\|A\|_{\mathcal{P}} = \|A\|_{I_1,\dots,I_k} \tag{8.94}$$

by viewing A as a k-linear form C on $F_1 \times \cdots \times F_k$ where $F_\ell = \mathbb{R}^{m^{\operatorname{card} I_\ell}}$ and defining

$$\|A\|_{I_1,\dots,I_k} = \|C\|_{\{1\}\{2\}\cdots\{k\}} \; , \tag{8.95}$$

where of course the right-hand side is defined as in (8.90). When the partition $\mathcal{P}'$ is finer than the partition $\mathcal{P}$, then

$$\|A\|_{\mathcal{P}'} \le \|A\|_{\mathcal{P}} \; . \tag{8.96}$$

The moments of the r.v. X of (8.94) are then evaluated by the following formula.

Theorem 8.3.1 (R. Latała [4]). *For $p \ge 1$ we have*

$$\frac{1}{K(d)} \sum_{\mathcal{P}} p^{\operatorname{card}\mathcal{P}/2} \|A\|_{\mathcal{P}} \le \|X\|_p \le K(d) \sum_{\mathcal{P}} p^{\operatorname{card}\mathcal{P}/2} \|A\|_{\mathcal{P}} \; , \tag{8.97}$$

where $\mathcal{P}$ runs over all partitions of $\{1, \dots, d\}$.

A multidimensional array as in (8.84) will be called a *tensor of order d* (the value of m may depend on the context). Let us denote by $E_1, \dots, E_d$ copies of $\mathbb{R}^m$. The idea is that E_k is the copy that corresponds to the k-th index of A. Given a vector $x \in E_d$ we may then define the contraction $\langle A, x\rangle$ as the tensor $(b_{i_1,\dots,i_{d-1}})$ of order $d-1$ given by

$$b_{i_1,\dots,i_{d-1}} = \sum_{i \le m} a_{i_1,\dots,i_{d-1},i} x_i \; .$$

The summation here is on the d-th index, as is indicated by the fact that $x \in E_d$.

Exercise 8.3.2. Use the upper bound of (8.97) to generalize Theorem 8.2.2 to a set T of tensors of order d. (Hint: this assumes that you know how to transform (8.97) in a tail estimate.)

If G is a standard Gaussian random vector valued in E_d, i.e. $G = (g_i)_{i \le m}$ where g_i are independent standard r.v.s, then $\langle A, G\rangle$ is a random tensor of order $d-1$. We shall deduce Theorem 8.3.1 from the following fact, of independent interest.

Theorem 8.3.3. *For all $\tau \geq 1$ we have*

$$\mathsf{E}\|\langle A, G\rangle\|_{\{1\}\cdots\{d-1\}} \leq K \sum_{\mathcal{P}} \tau^{\operatorname{card}\mathcal{P} - d + 1} \|A\|_{\mathcal{P}} , \tag{8.98}$$

where $\mathcal{P}$ runs over all the partitions of $\{1, \ldots, d\}$.

Here, as well as in the rest of this section, K denotes a number that depends only on the order of the tensor considered and certainly not on τ.

If we think of A as a d-linear form on $E_1 \times \cdots \times E_d$ the left-hand side of (8.98) is

$$\mathsf{E} \sup A(x^1, \ldots, x^{d-1}, G) ,$$

where the supremum is over all choices of x^ℓ with $\|x^\ell\| \leq 1$. Therefore the issue to prove (8.98) is to bound the supremum of a certain complicated Gaussian process.

The bound (8.98) has the mind-boggling feature that the powers of τ in the right-hand side may have different signs. This feature will actually appear very naturally in the course of the proof.

Corollary 8.3.4. *For all $p \geq 1$ on has*

$$\big(\mathsf{E}\|\langle A, G\rangle\|^p_{\{1\}\cdots\{d-1\}}\big)^{1/p} \leq K \sum_{\mathcal{P}} p^{(\operatorname{card}\mathcal{P} - d + 1)/2} \|A\|_{\mathcal{P}} . \tag{8.99}$$

Proof. As we just explained the r.v. $Y = \|\langle A, G\rangle\|_{\{1\}\cdots\{d-1\}}$ is the supremum of Gaussian r.v.s of the type

$$Z = A(x^1, \ldots, x^{d-1}, G) ,$$

where in this formula we view A as a d-linear map on $E_1 \times \cdots \times E_d$ and where x^ℓ is a vector of length ≤ 1. Now, the formula

$$\Big(\mathsf{E}\Big(\sum_i a_i g_i\Big)^2\Big)^{1/2} = \Big(\sum_i a_i^2\Big)^{1/2} = \sup\Big\{\sum_i a_i x_i \; ; \; \sum_i x_i^2 \leq 1\Big\}$$

implies

$$(\mathsf{E}Z^2)^{1/2} = \sup_{\|x\| \leq 1} |A(x^1, \ldots, x^{d-1}, x)| \leq \sigma := \|A\|_{\{1\}\cdots\{d-1\}\{d\}} .$$

It then follows from (2.84) that for $u > 0$ the r.v. Y satisfies

$$\mathsf{P}(|Y - \mathsf{E}Y| \geq u) \leq 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big) .$$

Then (2.22) implies

$$(\mathsf{E}|Y - \mathsf{E}Y|^p)^{1/p} \leq L\sqrt{p}\sigma ,$$

and since $(\mathsf{E}|Y|^p)^{1/p} \leq \mathsf{E}|Y| + (\mathsf{E}|Y - \mathsf{E}Y|^p)^{1/p}$ the result follows from (8.98) used for $\tau = p^{1/2}$. □

Proof of Theorem 8.3.1. First we deduce the upper bound of (8.97) from (8.99) and induction over d. For $d = 1$, (8.97) reflects the growth of the moments of a single Gaussian r.v. as captured by (2.22). Assuming that the result has been proved for $d-1$ we prove it for d. We consider the Gaussian random vector $G = (g_i^d)$, and the order $d-1$ random tensor

$$B = \langle A, G\rangle = (b_{i_1,\dots,i_{d-1}}) \;,$$

where

$$b_{i_1,\dots,i_{d-1}} = \sum_{i \le m} a_{i_1,\dots,i_{d-1},i} g_i^d \;.$$

Thus

$$X = \sum_{i_1,\dots,i_d} a_{i_1,\dots,i_d} g_{i_1}^1 \cdots g_{i_d}^d = \sum_{i_1,\dots,i_{d-1}} b_{i_1,\dots,i_{d-1}} g_{i_1}^1 \cdots g_{i_{d-1}}^{d-1} \;.$$

Let us denote by E' expectation given G. Then the induction hypothesis applied to B implies

$$(\mathsf{E}'|X|^p)^{1/p} \le K \sum_{\mathcal{Q}} p^{\operatorname{card}\mathcal{Q}/2} \|B\|_{\mathcal{Q}} \;, \tag{8.100}$$

where the sum runs over all partitions $\mathcal{Q}$ of $\{1,\dots,d-1\}$. We now compute the p-th moment of both sides, using the triangle inequality in L^p to obtain

$$(\mathsf{E}|X|^p)^{1/p} \le K \sum_{\mathcal{Q}} p^{\operatorname{card}\mathcal{Q}/2} (\mathsf{E}\|B\|_{\mathcal{Q}}^p)^{1/p} \;. \tag{8.101}$$

Let us fix $\mathcal{Q}$ and denote by $I_1,\dots,I_k$ its elements. We claim that

$$\big(\mathsf{E}\|B\|_{\mathcal{Q}}^p\big)^{1/p} = \big(\mathsf{E}\|\langle A, G\rangle\|_{I_1,\dots,I_k}^p\big)^{1/p} \le K p^{-k/2} \sum_{\mathcal{P}} p^{\operatorname{card}\mathcal{P}/2} \|A\|_{\mathcal{P}} \;. \tag{8.102}$$

Since $k = \operatorname{card}\mathcal{Q}$, substitution of this equation in (8.101) finishes the proof of the upper bound of (8.97). Let us now prove (8.102). For this for $\ell \le k$ we define $F_\ell = \mathbb{R}^{m^{\operatorname{card} I_\ell}}$ and we define $F_{k+1} = R^m$. Let us view A as a $(k+1)$-linear form C on the space $F_1 \times \cdots \times F_{k+1}$. Recalling (8.95) let us then apply (8.99) to C (with $d = k+1$). We then obtain the stronger form of (8.102), where the summation in the right-hand side is restricted to the partitions $\mathcal{P}$ whose restriction to $\{1,\dots,d-1\}$ is coarser than $\mathcal{Q}$.

We have proved the upper bound of (8.97) and we turn to the proof of the lower bound, which we shall prove only for $p \ge 2$. First we observe that for $d = 1$, this simply reflects the fact that for a standard Gaussian r.v. g one has $(\mathsf{E}|g|^p)^{1/p} \ge \sqrt{p}/L$. (No, this has not been proved anywhere in this book, but see Exercise 2.2.9.) Next we prove by induction on d that for each d one has

$$(\mathsf{E}|X|^p)^{1/p} \ge \frac{\sqrt{p}}{K}\|A\|_{\{1,2,\dots,d\}} = \frac{\sqrt{p}}{K}\Big(\sum_{i_1,\dots,i_d} a^2_{i_1,\dots,i_d}\Big)^{1/2} . \tag{8.103}$$

For this we consider the random tensor B of order $d-1$ given by

$$b_{i_1,\dots,i_{d-1}} = \sum_{i\le m} a_{i_1,\dots,i_{d-1},i} g_i^d .$$

Applying the induction hypothesis to B given the r.v.s g_i^d, and denoting by E' expectation given these variables, we obtain

$$(\mathsf{E}'|X|^p)^{1/p} \ge \frac{\sqrt{p}}{K}\Big(\sum_{i_1,\dots,i_{d-1}} b^2_{i_1,\dots,i_{d-1}}\Big)^{1/2} .$$

We compute the norm in L^p of both sides, using that for $p\ge 2$ one has $(\mathsf{E}|Y|^p)^{1/p} \ge (\mathsf{E}Y^2)^{1/2}$ to obtain (8.103) for d. (It is only at this place that a tiny extra effort is required if $p\le 2$.)

Let us now prove by induction over k that

$$(\mathsf{E}|X|^p)^{1/p} \ge \frac{p^{k/2}}{K}\|A\|_{I_1,\dots,I_k} . \tag{8.104}$$

The case $k=1$ is (8.103). For the induction from $k-1$ to k let us assume without loss of generality that $I_k = \{r+1,\dots,d\}$ and let us define a random order $d-r$ random tensor C by

$$c_{i_{r+1},\dots,i_d} = \sum_{i_1,\dots,i_r} a_{i_1,\dots,i_d} g^1_{i_1}\cdots g^r_{i_r} ,$$

so that

$$X = \sum_{i_{r+1},\dots,i_d} c_{i_{r+1},\dots,i_d} g^{r+1}_{i_{r+1}}\cdots g^d_{i_d} .$$

Denoting now by $\mathsf{E}^\sim$ expectation only in the r.v.s g_i^ℓ for $r+1\le \ell\le d$, we use (8.103) to obtain

$$(\mathsf{E}^\sim|X|^p)^{1/p} \ge \frac{\sqrt{p}}{K}\Big(\sum_{i_{r+1},\dots,i_d} c^2_{i_{r+1},\dots,i_d}\Big)^{1/2} .$$

Consequently, if $x_{i_{r+1},\dots,i_d}$ are numbers with $\sum_{i_{r+1},\dots,i_d} x^2_{i_{r+1},\dots,i_d} \le 1$, one gets

$$\begin{aligned}(\mathsf{E}^\sim|X|^p)^{1/p} &\ge \frac{\sqrt{p}}{K}\Big|\sum_{i_{r+1},\dots,i_d} c_{i_{r+1},\dots,i_d} x_{i_{r+1},\dots,i_d}\Big| \\ &= \frac{\sqrt{p}}{K}\Big|\sum_{i_1,\dots,i_r} d_{i_1,\dots,i_r} g^1_{i_1}\cdots g^r_{i_r}\Big| ,\end{aligned} \tag{8.105}$$

where

$$d_{i_1,\ldots,i_r} = \sum_{i_{r+1},\ldots,i_d} a_{i_1,\ldots,i_d} x_{i_{r+1},\ldots,i_d} .$$

We now compute the L^p norm of both sides of (8.105), using the induction hypothesis to obtain

$$(\mathsf{E}|X|^p)^{1/p} \geq \frac{p^{k/2}}{K} \|D\|_{I_1,\ldots,I_{k-1}} ,$$

where D is the tensor $(d_{i_1,\ldots,i_r})$. The supremum of the norms in right hand side over the choices of $(x_{i_{r+1},\ldots,i_d})$ with $\sum_{i_{r+1},\ldots,i_d} x^2_{i_{r+1},\ldots,i_d} \leq 1$ is $\|A\|_{I_1,\ldots,I_k}$. (A formal definition of these norms by induction over k would be based exactly on this property.) □

We now prepare for the proof of Theorem 8.3.3. We consider copies $E_1, \ldots, E_k$ of $\mathbb{R}^m$ and for vectors $y^\ell \in E_\ell$, $y^\ell = (y^\ell_i)_{i\leq m}$ we define their tensor product

$$y^1 \otimes \cdots \otimes y^k = \Bigl(\prod_{\ell=1}^{\ell=k} y^\ell_{i_\ell}\Bigr) ,$$

which is simply the vector $(z_{i_1,\ldots,i_k})$ in $\mathbb{R}^{m^k}$ given by $z_{i_1,\ldots,i_k} = y^1_{i_1} \cdots y^k_{i_k}$. Let us consider for $\ell \leq m$ independent standard Gaussian vectors G^ℓ valued in E_ℓ and let us fix vectors $x^\ell \in E_\ell$. For $I \subset \{1, \ldots, k\}$, we use the notation

$$U_I = y^1 \otimes \cdots \otimes y^k ,$$

where $y^\ell = G^\ell$ if $\ell \in I$ and $y^\ell = x^\ell$ otherwise. Thus $U_{\{1,\ldots,k\}} = G^1 \otimes \cdots \otimes G^k$ and $U_\emptyset = x^1 \otimes \cdots \otimes x^k$. We denote by $\|x\|$ the Euclidean norm of a vector x of E_ℓ.

Lemma 8.3.5. *Consider a semi-norm α on $\mathbb{R}^{m^k}$, and denote by $\mathcal{I}_k$ the collection of* non-empty *subsets of $\{1, \ldots, k\}$. Then*

$$\mathsf{P}\Bigl(\alpha(U_{\{1,\ldots,k\}} - U_\emptyset) \leq \sum_{I\in\mathcal{I}_k} 4^{\operatorname{card} I} \mathsf{E}\alpha(U_I)\Bigr) \geq 2^{-k} \exp\Bigl(-\frac{1}{2}\sum_{\ell\leq k} \|x^\ell\|^2\Bigr) . \tag{8.106}$$

Proof. We start with the following observation. If μ denotes the canonical Gaussian measure on $\mathbb{R}^m$ then for each compact symmetric body V of $\mathbb{R}^m$ on has

$$\mu(V + x) \geq \mu(V) \exp\Bigl(-\frac{\|x\|^2}{2}\Bigr) . \tag{8.107}$$

Indeed, if λ denotes Lebesgue's measure on $\mathbb{R}^m$, then, using symmetry in the second line, the parallelogram identity and convexity of the exponential in the third line,

$$\begin{aligned}\mu(x+V) &= \int_{x+V} \exp\Big(-\frac{\|y\|^2}{2}\Big)\mathrm{d}\lambda(y) \\ &= \int_V \exp\Big(-\frac{\|x+y\|^2}{2}\Big)\mathrm{d}\lambda(y) \\ &= \int_V \frac{1}{2}\Big(\exp\Big(-\frac{\|x+y\|^2}{2}\Big) + \exp\Big(-\frac{\|x-y\|^2}{2}\Big)\Big)\mathrm{d}\lambda(y) \\ &\geq \int_V \exp\Big(-\frac{\|x\|^2+\|y\|^2}{2}\Big)\mathrm{d}\lambda(y) \\ &= \exp\Big(-\frac{\|x\|^2}{2}\Big)\mu(V) \; . \end{aligned} \tag{8.108}$$

To prove (8.106) for $k = 1$ we consider the set $V = \{y \in E_1 \; ; \; \alpha(y) \leq 4\mathsf{E}\alpha(G^1)\}$, so that by Markov's inequality, $\mu(V^c) \leq 1/4$ and consequently $\mu(V) \geq 3/4$. Then (8.107) implies

$$\mathsf{P}(\alpha(G^1 - x^1) \leq 4\mathsf{E}\alpha(G^1)) = \mu(V + x_1) \geq \frac{3}{4}\exp\Big(-\frac{\|x^1\|^2}{2}\Big) \; ,$$

which implies (8.106) for $k = 1$. For the induction proof from $k-1$ to k we consider the quantities

$$S = \sum_{I \in \mathcal{I}_{k-1}} 4^{\operatorname{card} I} \alpha(U_{I \cup \{k\}})$$

and

$$T = \sum_{I \in \mathcal{I}_{k-1}} 4^{\operatorname{card} I} \alpha(U_I) \; .$$

In the case $k = 2$, the only case that we shall use, this is simply $S = 4\alpha(G^1 \otimes G^2)$ and $T = 4\alpha(G^1 \otimes x^2)$. We denote by E^k conditional expectation given G^k and we consider the events

$$\Omega_1 = \{\alpha(U_{\{k\}} - U_\emptyset) \leq 4\mathsf{E}\alpha(U_{\{k\}})\} \; ,$$

$$\Omega_2 = \{\alpha(U_{\{1,\ldots,k\}} - U_{\{k\}}) \leq \mathsf{E}^k S\} \; ,$$

and

$$\Omega_3 = \{\mathsf{E}^k S \leq 4\mathsf{E}S + \mathsf{E}T\} \; .$$

When these three events occur simultaneously, we have

$$\begin{aligned}\alpha(U_{\{1,\cdots,k\}} - U_\emptyset) &\leq \alpha(U_{\{1,\ldots,k\}} - U_{\{k\}}) + \alpha(U_{\{k\}} - U_\emptyset) \\ &\leq \mathsf{E}^k S + 4\mathsf{E}\alpha(U_{\{k\}}) \\ &\leq 4\mathsf{E}S + \mathsf{E}T + 4\mathsf{E}\alpha(U_{\{k\}}) \\ &= \sum_{I \in \mathcal{I}_k} 4^{\operatorname{card} I}\mathsf{E}\alpha(U_I) \; . \end{aligned} \tag{8.109}$$

Next, we prove that

$$\mathsf{P}(\Omega_1 \cap \Omega_3) \geq \frac{1}{2} \exp\Big(-\frac{\|x^k\|^2}{2}\Big) . \tag{8.110}$$

For this we consider on E_k the semi-norms

$$\alpha_1(y) = \alpha(x^1 \otimes \cdots \otimes x^{k-1} \otimes y) ,$$

and

$$\alpha_2(y) = \sum_{I \in \mathcal{I}_{k-1}} 4^{\operatorname{card} I} \mathsf{E}\alpha(W_I) ,$$

where $W_I = y^1 \otimes \cdots y^{k-1} \otimes y$ and $y^\ell = G^\ell$ if $\ell \in I$ and $y^\ell = x^\ell$ otherwise. Thus $\mathsf{E}^k S = \alpha_2(G^k)$ and $\mathsf{E}T = \alpha_2(x^k)$. Since $U_{\{k\}} - U_\emptyset = x^1 \otimes \cdots \otimes x^{k-1} \otimes (G^k - x^k)$ we have

$$\Omega_1 = \{\alpha_1(G^k - x^k) \leq 4\mathsf{E}\alpha_1(G^k)\} ,$$
$$\Omega_3 = \{\alpha_2(G^k) \leq 4\mathsf{E}\alpha_2(G^k) + \alpha_2(x^k)\} .$$

Consider the convex symmetric set

$$V = \big\{y \in E_k \ ; \ \alpha_1(y) \leq 4\mathsf{E}\alpha_1(G^k) \ , \ \alpha_2(y) \leq 4\mathsf{E}\alpha_2(G^k)\big\} .$$

Then Markov's inequality implies that $\mathsf{P}(G^k \in V) \geq 1/2$, so that (8.107) yields

$$\mathsf{P}(G^k \in V + x^k) \geq \frac{1}{2} \exp\Big(-\frac{\|x^k\|^2}{2}\Big) . \tag{8.111}$$

The triangle inequality implies

$$\{G^k \in V + x^k\} \subset \Omega_1 \cap \Omega_3 ,$$

so (8.111) implies (8.110).

Finally we prove that if P^k denotes probability given G^k then

$$\mathsf{P}^k(\Omega_2) \geq 2^{-k+1} \exp\Big(-\frac{1}{2} \sum_{\ell \leq k-1} \|x^\ell\|^2\Big) . \tag{8.112}$$

For this we may think of G^k as a given deterministic vector of E_k. We then consider on $\mathbb{R}^{m^{k-1}}$ the norm α' given by $\alpha'(y) = \alpha(y \otimes G^k)$, where if $y = (y_{i_1,\ldots,i_{k-1}})$ and $G^k = (g_i)$ we define $y \otimes G^k = (w_{i_1,\ldots,i_k})$ for $w_{i_1,\ldots,i_k} = y_{i_1\ldots i_{k-1}} g_{i_k}$. We then observe that

$$\Omega_2 = \big\{\alpha'(G^1 \otimes \cdots \otimes G^{k-1} - x^1 \otimes \cdots \otimes x^{k-1}) \leq \mathsf{E}^k S\big\} ,$$

so that (8.112) follows from the induction hypothesis.

Since Ω_1 and Ω_3 depend on G^k only, combining (8.112) and (8.110) proves that the probability that $\Omega_1, \Omega_2, \Omega_3$ occur simultaneously is at least $2^{-k} \exp(-\sum_{\ell \leq k} \|x^\ell\|^2/2)$. Combining with (8.109) completes the proof. □

Having had all this fun with multi-indices and high-order tensors we turn to the proof of Theorem 8.3.3. This will occupy the remainder of this section. In order to make the ideas of this very deep result accessible we shall assume that $d = 3$, and refer to Latała's original paper for the general case. To cover the case $d = 3$ we need only the cases $k = 1$ and $k = 2$ of Lemma 8.3.5. We now draw consequences of this lemma. We recall the entropy numbers $e_n(T, d)$ of (2.34). The next result is classical, and is called "the dual Sudakov inequality". It is *extremely useful.*

Lemma 8.3.6. *Consider a semi-norm α on $\mathbb{R}^m$ and a standard Gaussian r.v. G valued in $\mathbb{R}^m$. Then if d_α is the distance associated with α, the unit ball B of $\mathbb{R}^m$ satisfies*

$$e_n(B, d_\alpha) \le L 2^{-n/2} \mathsf{E}\alpha(G) . \tag{8.113}$$

Proof. The case $k = 1$ of (8.106) implies

$$\mathsf{P}(\alpha(G - x) \le 4\mathsf{E}(G)) \ge \frac{1}{2} \exp\Big(-\frac{\|x\|^2}{2}\Big) ,$$

and, by homogeneity, for $x \in B$ and $\tau > 0$,

$$\mathsf{P}(\alpha(\tau G - x) \le 4\tau \mathsf{E}\alpha(G)) \ge \frac{1}{2} \exp\Big(-\frac{1}{2\tau^2}\Big) . \tag{8.114}$$

The proof then really follows the argument of Exercise 2.2.14. We repeat this argument for the convenience of the reader. If $\epsilon = 4\tau\mathsf{E}\alpha(G)$ and U is a subset of B such that any two points of U are at mutual distances $\ge 3\epsilon$ then the balls for d_α of radius ϵ centered at the points of U are disjoint and consequently (8.114) implies that $\operatorname{card} U \le 2\exp(1/2\tau^2)$. But taking U as large as possible the balls centered at U of radius 3ϵ cover B. Taking τ such that $2\exp(1/2\tau^2) = 2^{2^n}$ concludes the proof. □

Through the remainder of the section, we write

$$B = \big\{x = (x^1, x^2) \in E_1 \times E_2 \ ; \ \|x^1\| \le 1, \|x^2\| \le 1\big\} . \tag{8.115}$$

Lemma 8.3.7. *Consider a subset T of $2B$ and a semi-norm α on $\mathbb{R}^{m^2}$. Consider the distance d_α on T defined for $x = (x^1, x^2)$ and $y = (y^1, y^2)$ by*

$$d_\alpha(x, y) = \alpha(x^1 \otimes x^2 - y^1 \otimes y^2) . \tag{8.116}$$

Let us define

$$\alpha^*(T) = \sup_{x \in T} \big(\mathsf{E}\alpha(x^1 \otimes G^2) + \mathsf{E}\alpha(G^1 \otimes x^2)\big) . \tag{8.117}$$

Then

$$e_n(T, d_\alpha) \le L(2^{-n/2}\alpha^*(T) + 2^{-n}\mathsf{E}\alpha(G^1 \otimes G^2)) . \tag{8.118}$$

Proof. We deduce from (8.106) and homogeneity that for any $\tau > 0$ one has

$$\mathsf{P}\Big(\alpha(\tau^2 G^1 \otimes G^2 - x^1 \otimes x^2) \le W\Big) \ge \frac{1}{4}\exp\Big(-\frac{1}{2\tau^2}\sum_{\ell\le 2}\|x^\ell\|^2\Big) .$$

where

$$W = 4\tau(\mathsf{E}\alpha(x^1\otimes G^2) + \mathsf{E}\alpha(G^1\otimes x^2)) + 16\tau^2\mathsf{E}\alpha(G^1\otimes G^2) .$$

In particular when $x = (x^1, x^2) \in T \subset 2B$ one has $\|x^1\|^2 + \|x^2\|^2 \le 8$ and thus

$$\mathsf{P}\Big(\alpha(\tau^2 G^1\otimes G^2 - x^1\otimes x^2) \le 4\tau\alpha^*(T) + 16\tau^2\mathsf{E}\alpha(G^1\otimes G^2)\Big) \ge \frac{1}{4}\exp\Big(-\frac{4}{\tau^2}\Big) . \tag{8.119}$$

Let

$$\epsilon = 4\tau\alpha^*(T) + 16\tau^2\mathsf{E}\alpha(G^1\otimes G^2) ,$$

and consider a subset U of T such that any two points of U are at mutual distances $\ge 3\epsilon$ for d_α. Then the sets $\{z \in \mathbb{R}^{m^2} ; \alpha(z - x^1\otimes x^2) \le \epsilon\}$ for $x \in U$ are disjoint, so that (8.119) implies

$$\operatorname{card} U \le 4\exp(4\tau^{-2}) , \tag{8.120}$$

and if one takes U maximal for the inclusion this proves that the covering number $N(T, d_\alpha, 3\epsilon)$ is $\le 4\exp(4\tau^{-2})$. Choosing τ so that this quantity is 2^{2^n} finishes the proof. □

Now we prove Theorem 8.3.3 in the case $d-1 = 1$. The argument has already been given at the beginning of the proof of Corollary 8.3.4 but we repeat it for clarity. We simply write

$$\|\langle A, G\rangle\|_{\{1\}} = \sup_{\|x\|\le 1} A(x, G) = \Big(\sum_i\Big(\sum_j a_{i,j}g_j\Big)^2\Big)^{1/2} ,$$

and use of the Cauchy-Schwarz inequality proves that

$$\mathsf{E}\|\langle A, G\rangle\|_{\{1\}} \le \Big(\sum_{i,j} a_{i,j}^2\Big)^{1/2} = \|A\|_{\{1,2\}} . \tag{8.121}$$

Now we start the proof in the case $d-1 = 2$. For a subset T of $E_1\times E_2$ we define

$$F(T) = \mathsf{E}\sup_{x\in T} A(x^1, x^2, G) . \tag{8.122}$$

Since all our spaces are finite dimensional, this quantity is finite whenever T is bounded. The goal is to bound

$$F(B) = \mathsf{E}\|\langle A, G\rangle\|_{\{1\}\{2\}} . \tag{8.123}$$

We consider the semi norm α on $\mathbb{R}^{m^2}$ given for $z = (z_{i,j})_{i,j\le m}$ by

$$\alpha(z) = \Big(\sum_k \Big(\sum_{i,j} a_{i,j,k} z_{i,j}\Big)^2\Big)^{1/2} . \tag{8.124}$$

Then the corresponding distance d_α on $E_1 \times E_2$ given by (8.116) is the canonical distance associated to the Gaussian process $X_x = A(x^1, x^2, G)$. This semi-norm will be used until the end of the proof.

Using the Cauchy-Schwarz inequality, one obtains the relations

$$\mathsf{E}\alpha(G^1 \otimes x^2) \le \|\langle A, x^2\rangle\|_{\{1,3\}} \tag{8.125}$$

$$\mathsf{E}\alpha(x^1 \otimes G^2) \le \|\langle A, x^1\rangle\|_{\{2,3\}} \tag{8.126}$$

$$\mathsf{E}\alpha(G^1 \otimes G^2) \le \|A\|_{\{1,2,3\}} . \tag{8.127}$$

Here, if $A = (a_{i,j,k})$ and $x^2 = (x^2_j)$, then $\langle A, x^2\rangle$ is the matrix $(b_{i,k})$ where $b_{i,k} = \sum_j a_{i,j,k} x^2_j$, and $\|\langle A, x^2\rangle\|_{\{1,3\}} = (\sum_{i,k} b^2_{i,k})^{1/2}$, and to prove (8.125) we simply observe that $\alpha(G^1 \otimes x^2) = (\sum_k (\sum_i b_{i,k} g^1_i)^2)^{1/2}$, so that $\mathsf{E}\alpha(G^1 \otimes x^2) \le (\sum_{i,k} b^2_{i,k})^{1/2} = \|\langle A, x^2\rangle\|_{\{1,3\}}$, etc.

Lemma 8.3.8. *For $u = (u^1, u^2) \in E_1 \times E_2$ and $T \subset 2B$ one has*

$$F(u + T) \le F(T) + 2\|\langle A, u^1\rangle\|_{\{2,3\}} + 2\|\langle A, u^2\rangle\|_{\{1,3\}} . \tag{8.128}$$

Proof. We observe the identity

$$A(x^1 + u^1, x^2 + u^2, G) = A(x^1, x^2, G) + A(u^1, x^2, G) + A(x^1, u^2, G) + A(u^1, u^2, G) . \tag{8.129}$$

We take the supremum over $x \in T$ and then expectation to obtain (using that $\mathsf{E}A(u^1, u^2, G) = 0$)

$$F(T + u) \le F(T) + C_1 + C_2 ,$$

where

$$C_1 = \mathsf{E} \sup_{\|x^2\|\le 2} A(u^1, x^2, G) \; ; \; C_2 = \mathsf{E} \sup_{\|x^1\|\le 2} A(x^1, u^2, G) .$$

We then apply (8.121) to the tensor $\langle A, u^1\rangle$ to obtain $C_1 \le 2\|\langle A, u^1\rangle\|_{\{2,3\}}$ and similarly for C_2. □

This results motivates the introduction on $E_1 \times E_2$ of the norm

$$\alpha^*(x) = \|\langle A, x^1\rangle\|_{\{2,3\}} + \|\langle A, x^2\rangle\|_{\{1,3\}} . \tag{8.130}$$

Then (8.128) reads

$$F(u + T) \le F(T) + 2\alpha^*(u) . \tag{8.131}$$

We denote by d_{α^*} the distance on $E_1 \times E_2$ associated to the norm α^*.

The semi-norm α^* has another use: combining (8.117) with (8.125) and (8.126) we observe the relation

$$\alpha^*(T) \le \sup\{\alpha^*(x)\ ;\ x \in T\}\ . \tag{8.132}$$

Lemma 8.3.9. *We have*

$$e_n(2B, d_{\alpha^*}) \le L2^{-n/2}\|A\|_{\{1,2,3\}}\ . \tag{8.133}$$

Proof. This is a consequence of Lemma 8.3.6 applied to the space $E_1 \times E_2$. A standard Gaussian random vector valued in this space is of the type (G^1, G^2) where G^1 and G^2 are standard Gaussian random vectors. Now, proceeding as in (8.121) we get

$$\mathsf{E}\|\langle A, G^1\rangle\|_{\{2,3\}} \le \|A\|_{\{1,2,3\}}\ ,$$

and similarly $\mathsf{E}\|\langle A, G^1\rangle\|_{\{1,3\}} \le \|A\|_{\{1,2,3\}}$, so that

$$\mathsf{E}\alpha^*(G^1, G^2) \le 2\|A\|_{\{1,2,3\}}\ .$$ □

We lighten notation by setting

$$S_1 = \|A\|_{\{1,2,3\}}\ . \tag{8.134}$$

Given a point $y \in B$ and $a, b > 0$ we define

$$C(y,a,b) = \big\{x \in B - y\ ;\ d_\alpha(0,x) \le a\ ,\ d_{\alpha^*}(0,x) \le b\big\}\ . \tag{8.135}$$

We further define

$$W(a,b) = \sup\{F(C(y,a,b))\ ;\ y \in B\}\ . \tag{8.136}$$

The center of the argument is as follows:

Lemma 8.3.10. *For all values of $a, b > 0$ and $n \ge 0$ we have*

$$W(a,b) \le L2^{n/2}a + Lb + W(L2^{-n/2}b + L2^{-n}S_1, L2^{-n/2}S_1)\ . \tag{8.137}$$

Proof. Consider $y \in B$ so that $B - y \subset 2B$ and $T = C(y,a,b) \subset 2B$. It follows from (8.132) that $\alpha^*(T) \le b$. Combining (8.118) and (8.127) we obtain that $e_n(T, d_\alpha) \le \delta := L(2^{-n/2}b + 2^{-n}S_1)$. Using also (8.133) we find a partition of $T = C(y,a,b)$ into $N_{n+1} = 2^{2^{n+1}}$ sets which are of diameter $\le \delta$ for d_α and of diameter $\le \delta^* := L2^{-n/2}S_1$ for d_{α^*}. Thus we can find points $y_i \in C(y,a,b)$ for $i \le N_{n+1}$ such that

$$C(y,a,b) \subset \bigcup_{i \le N_{n+1}} T_i\ , \tag{8.138}$$

where

$$T_i = \left\{x \in E_1 \times E_2 \; ; \; x \in C(y,a,b) \; , \; d_\alpha(y_i,x) \le \delta \; , \; d_{\alpha^*}(y_i,x) \le \delta^*\right\} .$$

For $x \in B - y$ we have $x - y_i \in B - (y + y_i)$, so that

$$T_i - y_i \subset C(y + y_i, \delta, \delta^*) \tag{8.139}$$

and

$$T_i \subset y_i + C(y + y_i, \delta, \delta^*) . \tag{8.140}$$

Also, since $y_i \in B - y$ we have $y + y_i \in B$, so that

$$F(C(y + y_i, \delta, \delta^*)) \le W(\delta, \delta^*) ,$$

and combining with (8.140) and (8.131), and since $\alpha^*(y_i) = d_{\alpha^*}(y_i, 0) \le b$ because $y_i \in C(y,a,b)$ we obtain

$$F(T_i) \le W(\delta, \delta^*) + 2b . \tag{8.141}$$

The conclusion then follows from (8.138) and the bound

$$F\Big(\bigcup_{i \le M} T_i\Big) \le La\sqrt{\log M} + \max_{i \le M} F(T_i) ,$$

see (2.88). □

Proposition 8.3.11. *For $n \ge 0$ we have*

$$W(a,b) \le L(2^{n/2}a + b + 2^{-n/2}S_1) . \tag{8.142}$$

Proof of Theorem 8.3.3 for $d = 3$. We set

$$S_3 = \|A\|_{\{1\}\{2\}\{3\}}$$
$$S_2 = \|A\|_{\{1\}\{2,3\}} + \|A\|_{\{2\}\{1,3\}} + \|A\|_{\{3\}\{1,2\}} .$$

Since $\alpha(x^1 \otimes x^2) = \sup\{A(x^1,x^2,x^3); \|x^3\| \le 1\}$, we have $d_\alpha(x,0) \le S_3$ for $x \in B$. Therefore $B \subset C(0, S_3, S_2)$ so that

$$F(B) \le W(S_3, S_2) \le L(2^{n/2}S_3 + S_2 + 2^{-n/2}S_1) .$$

Recalling (8.123) and choosing n so that $2^{n/2}$ is about τ proves (8.98). □

Proof of Proposition 8.3.11. Changing n into $n + n_0$ where n_0 is a universal constant, (8.137) implies that for $n \ge n_0$ one has

$$W(a,b) \le L2^{n/2}a + Lb + W(2^{-n/2-2}b + 2^{-n-2}S_1, 2^{-(n+1)/2}S_1) . \tag{8.143}$$

Using this for $a = 2^{-n}S_1$ and $b = 2^{-n/2}S_1$, we obtain

$$W(2^{-n}S_1, 2^{-n/2}S_1) \le L2^{-n/2}S_1 + W(2^{-n-1}S_1, 2^{-(n+1)/2}S_1) .$$

Summation of these relations for $n \ge r$ implies that for $r \ge n_0$

$$W(2^{-r}S_1, 2^{-r/2}S_1) \le L2^{-r/2}S_1 . \tag{8.144}$$

Using this relation we then deduce from (8.143) that

$$W(a, 2^{-n/2}S_1) \le L2^{n/2}a + L2^{-n/2}S_1 ,$$

and bounding the last term of (8.143) using this inequality yields (8.142). □

The proof of Theorem 8.3.3 for the general value of d does not require any essentially new idea. It is more complicated to write because there are more terms when witting the relation corresponding to (8.129). We strongly encourage the reader to carry out this proof in the case $d = 4$, using (8.142) and the induction hypothesis.

8.4 Notes and Comments

Our exposition of Latała's result in Section 8.3 brings no new idea whatsoever compared to his original paper [4]. (Improving the mathematics of Rafał Latała seems extremely challenging.) Whatever part of the exposition might be better than in the original paper draws heavily on J. Lehec's paper [6]. This author found [4] very difficult to read, and included Section 8.3 in an effort to make these beautiful ideas more accessible. It seems most probable that Latała started his work with the case $d = 3$, but one has to do significant reverse engineering to get this less technical case out of his paper.

References

1. Arcones, M., Giné, E.: On decoupling, series expansions, and tail behavior of chaos processes. J. Theor. Probab. **6**, 101–122 (1993)
2. Hanson, D.L., Wright, F.T.: A bound on tail probabilities for quadratic forms of independent random variables. Ann. Math. Stat. **42**(6), 1079–1083 (1971)
3. Kramher, F., Mendelson, S., Rauhut, H.: Suprema of chaos processes and the restricted isometry property. Manuscript (2012)
4. Latała, R.: Estimates of moments and tails of Gaussian chaoses. Ann. Probab. **34**(6), 2315–2331 (2006)
5. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
6. Lehec, J.: Moments of the Gaussian chaos. In: Séminaire de Probabilités XLIII. Lecture Notes in Math., vol. 2006, pp. 327–340. Springer, Berlin (2011)
7. Marcus, M.B., Pisier, G.: Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. Acta Math. **152**(3–4), 245–301 (1984)
8. Talagrand, M.: Sudakov-type minoration for Gaussian chaos. Isr. J. Math **79**, 207–224 (1992)

9. Theory and Practice of Empirical Processes

9.1 Discrepancy Bounds

Throughout this section we consider a probability space (Ω, μ), and (to avoid well understood measurability problems) a countable bounded subset of $L^2(\mu)$, which, following the standard notation in empirical processes theory, we will denote by $\mathcal{F}$ rather than T. (Since $\mathcal{F}$ is countable, there is no need to really distinguish between $\mathcal{L}^2(\mu)$ and $L^2(\mu)$.) To lighten notation we set

$$\mu(f) = \int f \mathrm{d}\mu \, .$$

Consider i.i.d. r.v.s $(X_i)_{i \geq 1}$ valued in Ω, distributed like μ and

$$S_N(\mathcal{F}) := \mathsf{E} \sup_{f \in \mathcal{F}} \Big| \sum_{i \leq N} (f(X_i) - \mu(f)) \Big| \, . \tag{9.1}$$

We have already seen in Chapter 4 the importance of evaluating such quantities. In the present section we consider this question from a theoretical perspective, and we try to focus on the fundamental problems this raises. Let us first mention a basic result (which we already used in Chapter 5).

Proposition 9.1.1. *If $0 \in \mathcal{F}$ we have*

$$S_N(\mathcal{F}) \leq L\big(\sqrt{N}\gamma_2(\mathcal{F}, d_2) + \gamma_1(\mathcal{F}, d_\infty)\big) \, , \tag{9.2}$$

where d_2 and d_∞ are the distances on $\mathcal{F}$ induced by the norms of L^2 and L^∞ respectively.

Proof. This follows from Bernstein's inequality (4.59) and Theorem 2.2.23 just as in the case of Theorem 4.3.6. The requirement that $0 \in \mathcal{F}$ is made necessary by the absolute values in (9.1). □

There is however a very different bound, namely the inequality

$$S_N(\mathcal{F}) \leq 2\mathsf{E} \sup_{f \in \mathcal{F}} \sum_{i \leq N} |f(X_i)| \, . \tag{9.3}$$

To see this we simply write

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_9,

$$\begin{aligned} S_N(\mathcal{F}) &\le \mathsf{E} \sup_{f\in\mathcal{F}} \sum_{i\le N} |f(X_i) - \mu(f)| \\ &\le \mathsf{E} \sup_{f\in\mathcal{F}} \sum_{i\le N} |f(X_i)| + N \sup_{f\in\mathcal{F}} |\mu(f)| \,, \end{aligned}$$

and we observe that the first term in the last line is $\ge$ than the second term through Jensen's inequality. The bound (9.3) does not involve cancellation, and is of a really different nature than (9.2), which involves cancellation in an essential way through Bernstein's inequality.

Having two completely different methods (9.2) and (9.3) to control $S_N(\mathcal{F})$, we can interpolate between them in the spirit of Proposition 5.1.4 as follows.

Proposition 9.1.2. *Consider classes $\mathcal{F}, \mathcal{F}_1$ and $\mathcal{F}_2$ of functions in $L^2(\mu)$, and assume that $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$. Assume that $0 \in \mathcal{F}_1$. Then*

$$\begin{aligned} S_N(\mathcal{F}) = \mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\le N} (f(X_i) - \mu(f)) \Big| &\le L\Big(\sqrt{N}\gamma_2(\mathcal{F}_1, d_2) + \gamma_1(\mathcal{F}_1, d_\infty)\Big) \\ &\quad + 2\mathsf{E} \sup_{f\in\mathcal{F}_2} \sum_{i\le N} |f(X_i)| \,. \end{aligned}$$

Proof. Since $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$, it is clear that $S_N(\mathcal{F}) \le S_N(\mathcal{F}_1) + S_N(\mathcal{F}_2)$. We then use the bound (9.2) for the first term and the bound (9.3) for the second term. □

Is there any other way to control $S_N(\mathcal{F})$ than the method of Proposition 9.1.2? This fundamental problem is related to the situation of Theorem 5.1.5. We formalize it as follows.

Research problem 9.1.3. Consider a class $\mathcal{F}$ of functions in $L^2(\mu)$ with $\mu(f) = 0$ for $f \in \mathcal{F}$. *Given an integer* N, can we find a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ with $0 \in \mathcal{F}_1$ such that the following properties hold:

$$\gamma_2(\mathcal{F}_1, d_2) \le \frac{L}{\sqrt{N}} S_N(\mathcal{F})$$

$$\gamma_1(\mathcal{F}_1, d_\infty) \le L S_N(\mathcal{F})$$

$$\mathsf{E} \sup_{f\in\mathcal{F}_2} \sum_{i\le N} |f(X_i)| \le L S_N(\mathcal{F})\,?$$

In Chapter 12 we shall investigate further generalizations of this question.

Exercise 9.1.4. We say that a countable class $\mathcal{F}$ of functions is a Glivenko-Cantelli class if

$$\lim_{N\to\infty} \mathsf{E} \sup_{f\in\mathcal{F}} \Big| \frac{1}{N} \sum_{i\le N} (f(X_i) - \mu(f)) \Big| = \lim_{N\to\infty} \frac{S_N(\mathcal{F})}{N} = 0 \,.$$

Assuming that $\mathcal{F}$ is uniformly bounded, prove that $\mathcal{F}$ is a Glivenko-Cantelli class if and only if for each $\epsilon > 0$ one can find a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ and an integer N_0 such that $\mathcal{F}_1$ is finite and

$$N \geq N_0 \Rightarrow \mathsf{E} \sup_{f \in \mathcal{F}_2} \frac{1}{N} \sum_{i \leq N} |f(X_i)| \leq \epsilon \,.$$

(Hint: use Theorem 6 of [12] for the "only if" part.)

The main result of this section is a kind of partial answer to Research Problem 9.1.3, where we write $\gamma_2(\mathcal{F})$ rather than $\gamma_2(\mathcal{F}, d_2)$.

Theorem 9.1.5. *Consider a countable class $\mathcal{F}$ of functions in $L^2(\mu)$, with $\mu(f) = 0$ for $f \in \mathcal{F}$. Then we can find a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ where $0 \in \mathcal{F}_1$,*

$$\gamma_2(\mathcal{F}_1, d_2) \leq L\gamma_2(\mathcal{F}) \tag{9.4}$$

$$\gamma_1(\mathcal{F}_1, d_\infty) \leq L\sqrt{N}\gamma_2(\mathcal{F}) \tag{9.5}$$

$$\mathsf{E} \sup_{f \in \mathcal{F}_2} \sum_{i \leq N} |f(X_i)| \leq L(S_N(\mathcal{F}) + \sqrt{N}\gamma_2(\mathcal{F})) \,. \tag{9.6}$$

It is of course a very mild restriction to assume that $\mu(f) = 0$ for $f \in \mathcal{F}$, since we have $S_N(\mathcal{F}) = S_N(\mathcal{F}^*)$ where $\mathcal{F}^* = \{f - \mu(f); f \in \mathcal{F}\}$. This restriction allows for a slightly easier presentation.

The decomposition of Theorem 9.1.5 resembles that of Problem 9.1.3, but uses the quantity $\gamma_2(\mathcal{F})$ rather than $S_N(\mathcal{F})/\sqrt{N}$. Unfortunately, as the following exercise shows, it need not be true that $\gamma_2(\mathcal{F}) \leq LS_N(\mathcal{F})/\sqrt{N}$, so that Theorem 9.1.5 does not solve Research problem 9.1.3. On the other hand, Lemma 9.1.7 below shows that $\gamma_2(\mathcal{F}) \leq L \sup_M S_M(\mathcal{F})/\sqrt{M}$.

Exercise 9.1.6. Given an integer N and a number A construct a finite class $\mathcal{F}$ of functions such that $AS_N(\mathcal{F}) \leq \gamma_2(\mathcal{F})$. (Hint: observe that the fact that for all $f \in \mathcal{F}$ one has $0 \leq f \leq g$ for some g with $\int |g| \mathrm{d}\mu < \infty$ limits the size of $S_N(\mathcal{F})$ but does not limit the size of $\gamma_2(\mathcal{F})$.)

A good setting to illustrate the use of Theorem 9.1.5 is that of Donsker classes, which are classes of functions on which the central limit theorem holds uniformly. The precise definition of Donsker classes includes a number of technicalities that are not related to the topic of this book and we refer e.g. to [5] for this. Here we will concentrate on the essential issue of this theory, the study of classes $\mathcal{F}$ for which

$$\mathcal{C}(\mathcal{F}) := \sup_N \frac{1}{\sqrt{N}} S_N(\mathcal{F}_N) < \infty \,. \tag{9.7}$$

The following easy fact demonstrates how Theorem 9.1.5 is related to Research Problem 9.1.3.

Lemma 9.1.7. *If $\mu(f) = 0$ for each f in $\mathcal{F}$, we have $\gamma_2(\mathcal{F}) \le L\mathcal{C}(\mathcal{F})$.*

Proof. Consider a finite subset T of $\mathcal{F}$. By the ordinary central limit theorem, the joint law of $(N^{-1/2}\sum_{i\le N} f(X_i))_{f\in T}$ converges to the law of a Gaussian process $(g_f)_{f\in T}$ and thus

$$\mathsf{E}\sup_{f\in T} g_f \le \mathcal{C}(\mathcal{F})\ . \tag{9.8}$$

The construction of the process $(g_f)_{f\in T}$ shows that for $f_1, f_2 \in T$ we have $\mathsf{E}g_{f_1}g_{f_2} = \int f_1 f_2 \mathrm{d}\mu$. Now we identify $L^2(\mu)$ with $\ell^2(\mathbb{N}^*)$ through the choice of an arbitrary orthonormal basis. Since the law of a Gaussian process is determined by its covariance, the left-hand side of (9.8) is exactly $g(T)$. This shows that $g(T) \le \mathcal{C}(\mathcal{F})$, and the result follows by Theorem 2.4.1 and (2.163). □

We then have the following characterization of classes for which $\mathcal{C}(\mathcal{F}) < \infty$.

Theorem 9.1.8. *Consider a class of functions $\mathcal{F}$ of $L^2(\mu)$ and assume that $\mu(f) = 0$ for each $f \in \mathcal{F}$. Then we have $\mathcal{C}(\mathcal{F}) < \infty$ if and only if there exists a number A and for each N there exists a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ (depending on N) where $0 \in \mathcal{F}_1$ such that*

$$\gamma_2(\mathcal{F}_1, d_2) \le A$$

$$\gamma_1(\mathcal{F}_1, d_\infty) \le \sqrt{N}A$$

$$\mathsf{E}\sup_{f\in\mathcal{F}_2}\sum_{i\le N}|f(X_i)| \le \sqrt{N}A\ .$$

Proof. The necessity follows from Lemma 9.1.7 and Theorem 9.1.5, while sufficiency follows from Proposition 9.1.1. □

The proof of Theorem 9.1.5 has two essential ingredients, the first of which is the following general principle.

Theorem 9.1.9. *Consider a countable set $T \subset L^2(\mu)$, and a number $u > 0$. Assume that $S = \gamma_2(T, d_2) < \infty$. Then there is a decomposition $T \subset T_1 + T_2$ where*

$$\gamma_2(T_1, d_2) \le LS\ ;\ \gamma_1(T_1, d_\infty) \le LSu \tag{9.9}$$

$$\gamma_2(T_2, d_2) \le LS\ ;\ T_2 \subset \frac{LS}{u}B_1\ . \tag{9.10}$$

Here of course

$$T_1 + T_2 = \big\{t_1 + t_2\ ;\ t_1 \in T_1\,,\, t_2 \in T_2\big\}\ .$$

In words, we can reconstruct T from the two sets T_1 and T_2. These two sets are not really larger than T with respect to γ_2. Moreover, for each of them we have some extra information: we control $\gamma_1(T_1, d_\infty)$, and we control the L^1 norm of the elements of T_2. In some sense Theorem 9.1.9 is an extension of Lemma 5.2.9, which deals with the case where T consists of a single function.

We will give two proofs of Theorem 9.1.9. The first proof relies on Theorem 5.2.7. The second proof avoids the use of Theorem 5.2.7, but is essentially based on the same argument, in a somewhat simpler setting. It is provided only to help the reader form an intuition about what is going on.

First proof of Theorem 9.1.9. We denote by $\Delta_2(A)$ the diameter of a set A for the L^2 norm. We consider an admissible sequence $(\mathcal{A}_n)$ of T such that

$$\forall t \in T\ ,\ \sum_{n \geq 0} 2^{n/2} \Delta_2(A_n(t)) \leq 2S\ . \tag{9.11}$$

We are going to apply Theorem 5.2.7 with $r = 2$ and μ the counting measure. Consider a parameter $u > 0$. We define $j_n(A)$ as the largest integer for which

$$2^{n/2} 2^{-j_n(A)} \geq u \Delta_2(A)\,, \tag{9.12}$$

so that

$$2^{n/2} 2^{-j_n(A)} \leq 2u \Delta_2(A)\ . \tag{9.13}$$

Thus (5.30) is obvious and (9.12) shows that (5.31) holds for $u' = 1/u^2$ instead of u. Thus the hypotheses of Theorem 5.2.7 are satisfied (for $u' = 1/u^2$ instead of u), and we consider the decomposition $T \subset T_1 + T_2 + T_3$ provided by this theorem (using u' instead of u). We observe that (9.11) implies

$$\sup_{t \in T} \sum_{n \geq 0} 2^n 2^{-j_n(A_n(t))} \leq LuS\ . \tag{9.14}$$

From (5.26) and (9.14) we obtain that $\gamma_2(T_1, d_2) \leq LS$, and (5.27) and (9.14) yield $\gamma_1(T_1, d_\infty) \leq LuS$, while (5.28) and (9.14) imply $\|t\|_1 \leq LS/u$ for $t \in T_2$.

Next we prove that

$$t \in T_3 \Rightarrow \|t\|_1 \leq LS/u\ . \tag{9.15}$$

Since $0 \in T$, we have $\|s\|_2 \leq \Delta_2(T)$ for $s \in T$. Using that $|s|\mathbf{1}_{\{|s| \geq v\}} \leq s^2/v$ for $v = 2^{-j_0(T)-1}$, and since $v \geq u\Delta_2(T)/2$ using (9.12) for $n = 0$, we get

$$\|s \mathbf{1}_{\{2|s| \geq 2^{-j_0(T)}\}}\|_1 \leq \|s\|_2^2/v \leq L\Delta_2(T)/u\ .$$

Considering only the term $n = 0$ in (9.11) yields $\Delta_2(T) \leq 2S$, so that for any $s \in T$ we have $\|s \mathbf{1}_{\{2|s| \geq 2^{-j_0(T)}\}}\|_1 \leq LS/u$ and (5.29) implies (9.15).

Setting $T_2' = T_2 + T_3$, we have shown that $\|t\|_1 \leq LS/u$ for $t \in T_2'$. Now since $T \subset T_1 + T_2 + T_3$, we have $T \subset T_1 + T_2'$ so that $T \subset T_1 + T_2''$

where $T_2'' = T_2' \cap (T - T_1)$. We prove that the decomposition $T \subset T_1 + T_2''$ satisfies the required properties. Since $\gamma_2(T, d_2) \le LS$ and $\gamma_2(T_1, d_2) \le LS$, (2.97) implies that $\gamma_2(T - T_1) \le LS$. Consequently $\gamma_2(T_2'', d_2) \le LS$, and also $T_2'' \subset T_2' \subset LSB_1/u$. □

Second proof of Theorem 9.1.9. The idea is simply to write an element of T as the sum of the increments along a chain, and to apply Lemma 5.2.9 to each of these increments. We will also take advantage of the fact that T is countable to write each element of T as the sum of the increments along a chain of finite length, but this is not an essential part of the argument.

As usual, $\Delta_2(A)$ denotes the diameter of A for the distance d_2. We consider an admissible sequence of partitions $(\mathcal{A}_n)_{n \ge 0}$ as in (9.11).

Let us enumerate T as $(t_n)_{n \ge 0}$. By induction over n we pick points $t_{n,A} \in A$ for $A \in \mathcal{A}_n$. We choose any point we want unless $A = A_n(t_n)$, in which case we choose $t_{n,A} = t_n$. Thus each point t of T is of the type $t_{n,A}$ for some n and $A = A_n(t)$. Let us define $\pi_n(t) = t_{n,A}$ where $A = A_n(t)$. We observe that $\pi_n(t) = t$ when $t = t_n$. For $n \ge 1$, let $f_{t,n} = \pi_n(t) - \pi_{n-1}(t)$, so that $f_{t,n}$ depends only on $A_n(t)$ and

$$\|f_{t,n}\|_2 \le \Delta_2(A_{n-1}(t)) \ . \tag{9.16}$$

Using Lemma 5.2.9 with $2^{-n/2}u\|f_{t,n}\|_2$ instead of u we can decompose $f_{t,n} = f_{t,n}^1 + f_{t,n}^2$ where

$$\|f_{t,n}^1\|_2 \le \|f_{t,n}\|_2 \ , \ \|f_{t,n}^1\|_\infty \le 2^{-n/2}u\|f_{t,n}\|_2 \tag{9.17}$$

$$\|f_{t,n}^2\|_2 \le \|f_{t,n}\|_2 \ , \ \|f_{t,n}^2\|_1 \le \frac{2^{n/2}}{u}\|f_{t,n}\|_2 \ . \tag{9.18}$$

Given $t \in T$ we set $g_{t,0}^1 = t_T$ and $g_{t,0}^2 = 0$, while if $n \ge 1$ we set

$$g_{t,n}^1 = t_T + \sum_{1 \le k \le n} f_{t,k}^1 \ , \ g_{t,n}^2 = \sum_{1 \le k \le n} f_{t,k}^2 \ .$$

We set

$$T_n^1 = \{g_{t,m}^1 \ ; \ m \le n \ , \ t \in T\} \ ; \ T_n^2 = \{g_{t,m}^2 \ ; \ m \le n \ , \ t \in T\}$$

$$T_1 = \bigcup_{n \ge 0} T_n^1 \ ; \ T_2 = \bigcup_{n \ge 0} T_n^2 \ .$$

We have $T \subset T_1 + T_2$. Indeed, if $t \in T$, then $t = t_n$ for some n and we have arranged that then $\pi_n(t) = t$. Since $\pi_0(t) = t_T$ we have

$$t - t_T = \pi_n(t) - \pi_0(t) = \sum_{1 \le k \le n} \pi_k(t) - \pi_{k-1}(t) = \sum_{1 \le k \le n} f_{t,k} \ ,$$

so that $t = g_{t,n}^1 + g_{t,n}^2 \in T_1 + T_2$.

Since for $j = 1, 2$ the element $g^j_{t,n}$ depends only on $A_n(t)$, we have $\operatorname{card} T^j_n \le N_0 + \cdots + N_n$, so that $\operatorname{card} T^j_0 = 1$ and $\operatorname{card} T^j_n \le N_{n+1}$. Consider $t^1 \in T_1$, so that $t^1 = g^1_{t,m}$ for some m and some $t \in T$. If $m \le n$ we have $t^1 = g^1_{t,m} \in T^1_n$ so that $d(t^1, T^1_n) = 0$. If $m > n$ we have $g^1_{t,n} \in T^1_n$, so that, using (9.16) and (9.17) in the third inequality we get

$$d_2(t^1, T^1_n) \le d_2(g^1_{t,m}, g^1_{t,n}) \le \sum_{k>n} \|f^1_{t,k}\|_2 \le \sum_{k>n} \Delta_2(A_{k-1}(t)) \ . \tag{9.19}$$

Hence

$$\begin{aligned}\sum_{n\ge 0} 2^{n/2} d_2(t^1, T^1_n) &\le \sum_{n\ge 0, k>n} 2^{n/2} \Delta_2(A_{k-1}(t)) \\ &\le L \sum_{k\ge 1} 2^{k/2} \Delta_2(A_{k-1}(t)) \le LS \ .\end{aligned}$$

It then follows from Theorem 2.3.1 (used for $\tau' = 1$) that $\gamma_2(T_1, d_2) \le LS$. The proof that $\gamma_2(T_2, d_2) \le LS$ is identical.

To control $\gamma_1(T_1, d_\infty)$ we use the same approach. We replace (9.19) by

$$d_\infty(t^1, T^1_n) \le d_\infty(g^1_{t,m}, g^1_{t,n}) \le \sum_{k>n} \|f^1_{t,k}\|_\infty \le \sum_{k>n} 2^{-k/2} u \Delta_2(A_{k-1}(t)) \ .$$

Hence

$$\begin{aligned}\sum_{n\ge 0} 2^n d_\infty(t^1, T^1_n) &\le u \sum_{n\ge 0, k>n} 2^{n-k/2} \Delta_2(A_{k-1}(t)) \\ &\le Lu \sum_{k\ge 1} 2^{k/2} \Delta_2(A_{k-1}(t)) \le LuS \ ,\end{aligned}$$

and it follows again from Theorem 2.3.1 that $\gamma_1(T_1, d_\infty) \le LS$. Finally, (9.18) and (9.17) yield

$$\|g^2_{t,n}\|_1 \le \sum_{k\ge 1} \|f^2_{t,k}\|_1 \le \sum_{k\ge 1} \frac{2^{k/2}}{u} \Delta_2(A_{k-1}(t)) \le \frac{LS}{u} \ ,$$

so that $T_2 \subset LB_1/u$. This completes the proof. □

Proof of Theorem 9.1.5. We use the decomposition of Theorem 9.1.9 with $u = \sqrt{N}$. This produces a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$, where $\mathcal{F}_1$ satisfies (9.4) and (9.5), while $\mathcal{F}_2 \subset L\gamma_2(\mathcal{F}) B_1/\sqrt{N}$. Moreover the construction is such that $\mathcal{F}_2 \subset \mathcal{F} - \mathcal{F}_1$, so that $S_N(\mathcal{F}_2) \le S_N(\mathcal{F}) + S_N(\mathcal{F}_1)$. Combining (9.4), (9.5) and (9.2) yields $S_N(\mathcal{F}_1) \le L\sqrt{N}\gamma_2(\mathcal{F})$. Consequently

$$\mathsf{E} \sup_{f\in\mathcal{F}_2} \Big| \sum_{i\le N} f(X_i) \Big| = S_N(\mathcal{F}_2) \le S_N(\mathcal{F}) + L\sqrt{N}\gamma_2(\mathcal{F}) \ .$$

Since $\int |f| \mathrm{d}\mu \le L\gamma_2(\mathcal{F})/\sqrt{N}$ for $f \in \mathcal{F}_2$, equation (9.6) follows now from the next result, which is the second major ingredient of the proof. □

Theorem 9.1.10 (The Giné-Zinn Theorem [6]). *For a class $\mathcal{F}$ of functions with $\mu(f) = 0$ for f in $\mathcal{F}$ we have*

$$\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\le N}|f(X_i)| \le N\sup_{f\in\mathcal{F}}\int|f|\mathrm{d}\mu + 4\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}f(X_i)\Big| . \tag{9.20}$$

While simple, this is very useful. Let us observe that using Jensen's inequality for the first term we also have the following (proving that (9.20) can essentially be reversed)

$$N\sup_{f\in\mathcal{F}}\int|f|\mathrm{d}\mu + 4\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}f(X_i)\Big| \le 5\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\le N}|f(X_i)| .$$

In order to avoid repetition, we will prove a more general fact. We consider pairs (R_i, X_i) of r.v.s, with $X_i \in \Omega$, $R_i \ge 0$, and we assume that these pairs are independent. We consider a Bernoulli sequence $(\varepsilon_i)_{i\ge1}$, that is an i.i.d. sequence with $\mathsf{P}(\varepsilon_i = \pm1) = 1/2$. We assume that these sequences are independent of the r.v.s (R_i, X_i). We assume that for each ω, only finitely many of the r.v.s $R_i(\omega)$ are not zero.

Lemma 9.1.11. *For a countable class of functions $\mathcal{F}$ we have*

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge1}\big(R_if(X_i) - \mathsf{E}(R_if(X_i))\big)\Big| \le 2\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge1}\varepsilon_iR_if(X_i)\Big| \tag{9.21}$$

and

$$\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\ge1}R_i|f(X_i)| \le \sup_{f\in\mathcal{F}}\sum_{i\ge1}\mathsf{E}(R_i|f(X_i)|)+2\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge1}\varepsilon_iR_if(X_i)\Big| . \tag{9.22}$$

Moreover, if $\mathsf{E}(R_if(X_i)) = 0$ for each $i \ge 1$, then

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge1}\varepsilon_iR_if(X_i)\Big| \le 2\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge1}R_if(X_i)\Big| . \tag{9.23}$$

Proof of Theorem 9.1.10. We take $R_i = 1$ if $i \le N$ and $R_i = 0$ if $i \ge N$, and we combine (9.22) and (9.23). □

Proof of Lemma 9.1.11. Consider an independent copy $(S_i, Y_i)_{i\ge1}$ of the sequence $(R_i, X_i)_{i\ge1}$, which is independent of the sequence $(\varepsilon_i)_{i\ge1}$. Jensen's inequality implies

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge1}\big(R_if(X_i) - \mathsf{E}(R_if(X_i))\big)\Big| \le \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\ge1}(R_if(X_i) - S_if(Y_i))\Big| .$$

Since the sequences $(R_if(X_i) - S_if(Y_i))$ and $(\varepsilon_i(R_if(X_i) - S_if(Y_i)))$ of r.v.s have the same law, we have

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}(R_i f(X_i) - S_i f(Y_i))\Big| = \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\varepsilon_i(R_i f(X_i) - S_i f(Y_i))\Big|$$
$$\leq 2\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\varepsilon_i R_i f(X_i)\Big|$$

and we have proved (9.21). To prove (9.22), we write

$$\sum_{i\geq 1} R_i|f(X_i)| \leq \sum_{i\geq 1}\mathsf{E}(R_i|f(X_i)|) + \sum_{i\geq 1}\big(R_i|f(X_i)| - \mathsf{E}(R_i|f(X_i)|)\big)\,,$$

we take the supremum over f and expectation, and we use (9.21) to get

$$\mathsf{E}\sup_{f\in\mathcal{F}}\sum_{i\geq 1} R_i|f(X_i)| \leq \sup_{f\in\mathcal{F}}\sum_{i\geq 1}\mathsf{E}(R_i|f(X_i)|) + 2\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\varepsilon_i R_i|f(X_i)|\Big|\,.$$

We then conclude with the comparison theorem for Bernoulli processes ([13], Theorem 2.1), which implies

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\varepsilon_i R_i|f(X_i)|\Big| \leq \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1}\varepsilon_i R_i f(X_i)\Big|\,.$$

To prove (9.23), we work conditionally on the sequence $(\varepsilon_i)_{i\geq 1}$. Setting $I = \{i\geq 1; \varepsilon_i = 1\}$ and $J = \{i \geq 1; \varepsilon_i = -1\}$, we obtain

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\leq N}\varepsilon_i R_i f(X_i)\Big| \leq \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in I} R_i f(X_i)\Big| + \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in J} R_i f(X_i)\Big|\,.$$

Now, since $\mathsf{E}R_i f(X_i) = 0$, Jensen's inequality implies

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in I} R_i f(X_i)\Big| \leq \mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\geq 1} R_i f(X_i)\Big|\,. \qquad \square$$

The following is a very powerful practical method to control $S_N(\mathcal{F})$.

Theorem 9.1.12. *Consider a countable class $\mathcal{F}$ of functions in $L^2(\mu)$ with $0\in\mathcal{F}$. Consider an admissible sequence $(\mathcal{A}_n)$ of partitions of $\mathcal{F}$. For $A\in\mathcal{A}_n$, define the function h_A by*

$$h_A(\omega) = \sup_{f,f'\in A}|f(\omega) - f'(\omega)|\,. \tag{9.24}$$

Assume that for a certain $j_0 = j_0(\mathcal{F})$ we have

$$\|h_{\mathcal{F}}\|_2 \leq \frac{2^{-j_0}}{\sqrt{N}}\,. \tag{9.25}$$

Assume that for each $n\geq 1$ and each $A\in\mathcal{A}_n$ we are given a number $j_n(A)\in\mathbb{Z}$ with

$$\int (2^{2j_n(A)} h_A^2) \wedge 1 \mathrm{d}\mu \le \frac{2^n}{N} \tag{9.26}$$

and

$$A \in \mathcal{A}_{n+1} \; , \; B \in \mathcal{A}_n \Rightarrow j_{n+1}(A) \ge j_n(B) \; , \tag{9.27}$$

and let

$$S = \sup_{f \in \mathcal{F}} \sum_{n \ge 0} 2^{n - j_n(A_n(f))} \; . \tag{9.28}$$

Then

$$S_N(\mathcal{F}) = \mathsf{E} \sup_{f \in \mathcal{F}} \Big| \sum_{i \le N} (f(X_i) - \mu(f)) \Big| \le LS \; . \tag{9.29}$$

It is instructive to rewrite (9.25) as $\int 2^{2j_0(\mathcal{F})} h_{\mathcal{F}}^2 \mathrm{d}\mu \le 1/N$ in order to compare it with (9.26).

Corollary 9.1.13. *With the notation of Theorem 9.1.12 define now*

$$S^* = \sup_{f \in \mathcal{F}} \sum_{n \ge 0} 2^{n/2} \| h_{A_n(f)} \|_2 \; . \tag{9.30}$$

Then

$$S_N(\mathcal{F}) \le L\sqrt{N} S^* \; . \tag{9.31}$$

This provide a bound for $S_N(\mathcal{F})/\sqrt{N}$, and is useful in the study of Donsker classes. The reader will observe that $\Delta(A) \le \|h_A\|_2$ for all A, so that (9.30) implies that $\gamma_2(\mathcal{F}, d_2) \le S^*$. This alone is however not sufficient to prove (9.31).

Exercise 9.1.14. Given two (measurable) functions $f_1 \le f_2$ define the bracket $[f_1, f_2]$ as the set of functions $\{f ; f_1 \le f \le f_2\}$. Given a class $\mathcal{F}$ of functions and $\epsilon > 0$ define $N_{[\,]}(\mathcal{F}, \epsilon)$ as the smallest number of brackets $[f_1, f_2]$ with $\|f_2 - f_1\|_2 \le \epsilon$ which can cover $\mathcal{F}$. Use Corollary 9.1.13 to prove that

$$S_N(\mathcal{F}) \le L\sqrt{N} \int_0^\infty \sqrt{\log N_{[\,]}(\mathcal{F}, \epsilon)} \mathrm{d}\epsilon \; . \tag{9.32}$$

Inequality (9.32) is known as Ossiander's bracketing theorem, and (9.31) is simply the "generic chaining version" of it. The proof of Ossiander's bracketing theorem requires a tricky idea beyond the ideas of Dudley's bound. In our approach, we deduce Ossiander's bracketing theorem in a straightforward manner from Theorem 9.1.12, and Theorem 9.1.12 itself is a straightforward consequence of Theorem 5.2.1. None of these simple arguments involves chaining. All the work involving chaining has already been performed in Theorem 5.2.1. This is why we suggested in Section 5.8 that in some sense one might say that Theorem 5.2.1 succeeds in extending the specific chaining argument of Ossiander's bracketing theorem from a specific situation involving brackets to a considerably more general setting.

Proof of Corollary 9.1.13. We have already given similar arguments many times. Define $j_n(A)$ as the largest integer j for which $\|h_A\|_2 \le 2^{n/2-j}/\sqrt{N}$, so that $2^{n/2-j_n(A)} \le 2\sqrt{N}\|h_A\|_2$, and consequently

$$\sum_{n\ge 0} 2^{n-j_n(A_n(f))} \le 2\sqrt{N}\sum_{n\ge 0} 2^{n/2}\|h_{A_n(f)}\|_2 \; .$$

By definition of $j_n(A)$ we have $\|h_A\|_2 \le 2^{n/2-j_n(A)}/\sqrt{N}$. This implies (9.26) and the result then follows from Theorem 9.1.12. □

Proof of Theorem 9.1.12. Let us fix $A \in \mathcal{A}_n$ and consider the r.v.s $W_i = (2^{2j_n(A)}h_A(X_i)^2) \wedge 1$, so that by (9.26) we have $\sum_{i\le N} \mathsf{E}W_i \le 2^n$. Consider a parameter $u \ge 1$. Then Lemma 7.4.3 (b) yields

$$\mathsf{P}\Big(\sum_{i\le N} W_i \ge u2^{n+2}\Big) \le \exp(-u2^{n+1}) \; . \tag{9.33}$$

Consider the event $\Omega(u)$ defined by

$$\forall\, n\ge 0\;,\;\forall\, A\in\mathcal{A}_n\;,\;\sum_{i\le N} 2^{2j_n(A)}h_A(X_i)^2\wedge 1 \le u2^{n+2}\;, \tag{9.34}$$

so that (9.33) and the union bound yield $\mathsf{P}(\Omega(u)) \ge 1 - L\exp(-u)$. Let us consider independent Bernoulli r.v.s ε_i, that are independent of the X_i, and let us recall that E_ε denotes expectation in the r.v.s ε_i only. Given the r.v.s X_i we consider the set T of all sequences of the type $(t_i)_{1\le i\le N} = (f(X_i))_{1\le i\le N}$ for $f \in \mathcal{F}$. To bound $\mathsf{E}\sup_{t\in T}|\sum_{i\le N}\varepsilon_i t_i|$ we appeal to Theorem 5.2.1. We observe that (5.9) is identical to (9.27), and that (5.10) (with $4u$ rather than u) follows from (9.34) since $|f(X_i) - f'(X_i)| \le h_A(X_i)$ for $f, f' \in A$. Also, for $f \in \mathcal{F}$ we have $|f(X_i)| \le h_{\mathcal{F}}(X_i)$, so that

$$|f(X_i)|\mathbf{1}_{\{2|f(X_i)|\ge 2^{-j_0(\mathcal{F})}\}} \le h_{\mathcal{F}}(X_i)\mathbf{1}_{\{2h_{\mathcal{F}}(X_i)\ge 2^{-j_0(\mathcal{F})}\}} \; .$$

We then use (5.12) with $p = 1$ to obtain

$$\begin{aligned}\mathsf{E}_\varepsilon \sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}\varepsilon_i f(X_i)\Big| \le Lu\sup_{f\in\mathcal{F}}\sum_{n\ge 0} 2^{n-j_n(A_n(f))} \\ + L\sum_{i\le N} h_{\mathcal{F}}(X_i)\mathbf{1}_{\{2h_{\mathcal{F}}(X_i)\ge 2^{-j_0(\mathcal{F})}\}} \; .\end{aligned} \tag{9.35}$$

The expectation of the last term is $LN\int h_{\mathcal{F}}\mathbf{1}_{\{2h_{\mathcal{F}}\ge 2^{-j_0(\mathcal{F})}\}}$. Now, since $h\mathbf{1}_{\{h\ge v\}} \le h^2/v$, and using (9.25) in the last inequality,

$$N\int h_{\mathcal{F}}\mathbf{1}_{\{2h_{\mathcal{F}}\ge 2^{-j_0(\mathcal{F})}\}}\mathrm{d}\mu \le N2^{j_0(\mathcal{F})+1}\int h_{\mathcal{F}}^2\mathrm{d}\mu \le 2^{-j_0(\mathcal{F})+1} \; .$$

Consequently, taking expectation in (9.35) and using that $\mathsf{P}(\Omega(u)) \ge 1 - L\exp(-u)$ we obtain

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}\varepsilon_i f(X_i)\Big| \le L\sup_{f\in\mathcal{F}}\sum_{n\ge 0}2^{n-j_n(A_n(f))} = LS \ ,$$

and we conclude the proof using (9.21) for $R_i = 1$ when $i \le N$ and $R_i = 0$ otherwise. □

9.2 How to Approach Practical Problems

A practical problem is of the following type. You are given a class $\mathcal{F}$ with certain properties and you try to bound

$$\mathcal{S} = \sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}(f(X_i) - \mathsf{E}f(X_i))\Big| \ .$$

(Of course, the reader may not agree with this definition of a practical problem. The point here is that one has to deal with a specific class rather than trying to understand in general how to bound $S_N(\mathcal{F})$.) A first point is that if we are only interested in bounding $\mathsf{E}\mathcal{S} = S_N(\mathcal{F})$ it is apparently always good to use (9.21), which yields here

$$\mathsf{E}\mathcal{S} \le 2\mathsf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}\varepsilon_i f(X_i)\Big| \ . \tag{9.36}$$

There are of course situations where this is not sufficient, e.g. when we are interested, say, in bounding the probability that $\mathcal{S}$ is large. In that case, it typically seems that one can proceed as well without (9.21). One just has to work a little harder. A rather typical example of this situation will be given when we present two different proofs of Proposition 14.5.1 below, one that proceeds directly from Bernstein's inequality, and one that uses a device similar to (9.21). As we try here to avoid secondary complications, we will only consider situations where we study $S_N(\mathcal{F})$ and we use (9.36).

To bound the right-hand side of (9.36) we think of the process as a Bernoulli process conditionally on the r.v.s X_i and we use chaining. That is, we have to bound along the chain the quantities

$$\Big|\sum_{i\le N}\varepsilon_i(\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i))\Big| \ .$$

For this we find a decomposition

$$\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i) = v_i + w_i \ , \tag{9.37}$$

and we write that by the subgaussian inequality, with probability $\mathsf{P}_\varepsilon \ge 1 - \exp(-L2^n)$ we have

$$\Big|\sum_{i\le N}\varepsilon_i(\pi_n(f)(X_i)-\pi_{n-1}(f)(X_i))\Big|\le\sum_{i\le N}|v_i|+L2^{n/2}\Big(\sum_{i\le N}w_i^2\Big)^{1/2}\,. \quad (9.38)$$

A specific way to find such a decomposition is to write $v_i=(\pi_n(f)(X_i)-\pi_{n-1}(f)(X_i))\mathbf{1}_{\{i\in I\}}$ for a cleverly chosen subset I of $\{1,\ldots,N\}$, in which case (9.38) becomes

$$\Big|\sum_{i\le N}\varepsilon_i(\pi_n(f)(X_i)-\pi_{n-1}(f)(X_i))\Big|\le\sum_{i\in I}|\pi_n(f)(X_i)-\pi_{n-1}(f)(X_i)|$$
$$+L2^{n/2}\Big(\sum_{i\notin I}|\pi_n(f)(X_i)-\pi_{n-1}(f)(X_i)|^2\Big)^{1/2}\,. \quad (9.39)$$

This is of course what we have been doing since Chapter 5, and the reader is probably disappointed to hear such an obvious advice. The reason we repeat it is to insist that if Problem 9.1.3 has a positive answer there is *never* any other way to proceed that the method we just described. And if Problem 9.1.3 has “a positive answer for all practical purposes” (i.e. that the only counterexamples are far-fetched and unnatural), then in natural situations there will be no other way to proceed. This certitude might be of great help in finding proofs. To demonstrate this, there seems to be no better way (as we do in the next section) than a choosing a few deep results from recent literature, and presenting the best proofs we can. Of course, there are many other aspects to the theory of empirical processes, (even restricted to its applications to Analysis and Banach Space theory), which it is beyond the scope of this book to cover.

In the remainder of this chapter, we consider two facets of the following problem. Consider independent r.v.s X_i valued in $\mathbb{R}^m$. Denoting by $\langle\cdot,\cdot\rangle$ the canonical duality of $\mathbb{R}^m$ with itself, and T a subset of $\mathbb{R}^m$, we are interested in bounding the quantity

$$\sup_{t\in T}\Big|\sum_{i\le N}(\langle X_i,t\rangle^2-\mathsf{E}\langle X_i,t\rangle^2)\Big|\,. \quad (9.40)$$

9.3 The Class of Squares of a Given Class

The present section will culminate (after the proof of a simpler result in the same spirit) in the proof of the following deep fact, due to S. Mendelson and G. Paouris. The firm belief that there is no possible approach other than the method explained in the previous section (and hence that matters just *can't* be very complicated) greatly helped the author to discover in a short time the proof he presents. There is no claim whatsoever that this proof contains any important idea which is not in the original proof. The point, which is of course difficult to convey, is that it was so helpful to know beforehand which route to take. (It certainly also helped to know that the result was true!) Given what we understand now, it simply seems very unlikely that we can find any other really different method to prove this result, and moreover, if the author is right with his conjectures, no such method exists at all.

As in the previous section we consider a probability space (Ω, μ), and denote by $(X_i)_{i \le N}$ r.v.s valued in Ω of law μ.

Theorem 9.3.1 ([11]). *Consider a (countable) class of functions $\mathcal{F}$ with $0 \in \mathcal{F}$. Assume that for a certain number $q > 4$ and a number C we have*

$$\forall f \in \mathcal{F} \ , \ \forall u > 0 \ , \ \mu(\{|f| \ge u\}) \le \Big(\frac{C}{u}\Big)^q \ . \tag{9.41}$$

Moreover consider two distances d_1 and d_2 on $\mathcal{F}$. Assume that given $f, f' \in \mathcal{F}$, then

$$\forall u > 0 \ , \ \mu(\{|f - f'| \ge u\}) \le 2 \exp\Big(-\min\Big(\frac{u^2}{d_2(f,f')^2}, \frac{u}{d_1(f,f')}\Big)\Big) \ . \tag{9.42}$$

Let $S = \gamma_2(\mathcal{F}, d_2) + \gamma_1(\mathcal{F}, d_1)$. Then

$$\mathsf{E} \sup_{f \in \mathcal{F}} \Big| \sum_{i \le N} (f(X_i)^2 - \mathsf{E} f^2) \Big| \le K\sqrt{N} C S + K S^2 \ . \tag{9.43}$$

Here and below, the number K depends on q only. The point of the theorem is that we use information on the class $\mathcal{F}$ to bound the empirical process on the class $\mathcal{F}^2 = \{f^2; f \in \mathcal{F}\}$.

To better understand the statement, let us observe from (9.42) that since $0 \in \mathcal{F}$, for any f in $\mathcal{F}$ we have

$$\mu(\{|f| \ge u\}) \le 2 \exp\Big(-\min\Big(\frac{u^2}{\Delta(T,d_2)^2}, \frac{u}{\Delta(T,d_1)}\Big)\Big) \ , \tag{9.44}$$

and that for large enough values of u this is much stronger than (9.41). The point of (9.41) however is that we may have $C \ll \Delta(T, d_2)$ and $C \ll \Delta(T, d_1)$, in which case for small values of u the inequality (9.41) is better than (9.44).

As an example of relevant situation, let us consider the case where $\Omega = \mathbb{R}^m$ and where μ is the canonical Gaussian measure on $\mathbb{R}^m$, i.e. the law of an independent sequence $(g_i)_{i \le m}$ of standard Gaussian r.v.s. Let us recall that we denote $\langle \cdot, \cdot \rangle$ the canonical duality between $\mathbb{R}^m$ and itself. We observe that μ is isotropic, that is

$$\int \langle t, x \rangle^2 \mathrm{d}\mu(x) = \|t\|_2^2 \tag{9.45}$$

for any $t \in \mathbb{R}^m$, where of course $\|t\|_2^2$ denotes the Euclidean norm of t. Thus if X_i has law μ, then $\mathsf{E}\langle X_i, t \rangle^2 = \|t\|_2^2$. Consider a subset T of $\mathbb{R}^m$, which is seen as a set $\mathcal{F}$ of functions on Ω through the canonical duality $\langle \cdot, \cdot \rangle$. The left-hand side of (9.43) is then simply

$$\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}(\langle X_i,t\rangle^2-\|t\|_2^2)\Big|\,. \tag{9.46}$$

A bound for this quantity is relevant in particular to the problem of signal reconstruction, i.e. of (approximately) finding the transmitted signal $t\in T$ when observing only the data $(\langle X_i,t\rangle)_{i\le N}$, see [10] for details. In these applications one does not like to have $0\in\mathcal{F}$, but one assumes instead that $\mathcal{F}$ is symmetric (i.e. $-f\in\mathcal{F}$ if $f\in\mathcal{F}$). It is simple to show that (9.43) still holds in this case. (Let us also observe that (9.43) does not hold when $\mathcal{F}$ is reduced to a single non-zero function.)

Now (1.4) implies that (9.42) holds when d_2 is (twice) the Euclidean distance and $d_1=0$, so that we can bound the quantity (9.46) as in (9.43) for $S=\gamma_2(T,d_2)$. The reader can find in Theorem 9.3.7 below a self-contained proof of a result sufficient to cover this case, yet significantly simpler than Theorem 9.3.1.

Let us now briefly mention a generalization of the precedent example to measures which are more general than the canonical Gaussian measure, and for which one needs to consider a condition as (9.42) with $d_1\neq 0$. Let us say that a probability μ on $\mathbb{R}^m$ is log-concave if $\mu(\lambda A+(1-\lambda)B)\ge\mu(A)^\lambda\mu(B)^{1-\lambda}$ for any measurable sets A,B and any $0\le\lambda\le 1$. Let us say that μ is unconditional if it is invariant under any change of signs of the coordinates. Consider then an isotropic (as in (9.45)) unconditional, sign invariant probability measure μ on $\mathbb{R}^m$. Viewing elements t of $\mathbb{R}^m$ as functions on the space $(\mathbb{R}^m,\mu)$ through the canonical duality, a result of R. Latała [8] proves (when combined with Bernstein's inequality in the form (9.63) below) that (9.42) holds where d_2 is a constant multiple of the Euclidean distance, and d_1 a multiple of the ℓ_∞ distance. Therefore Theorem 9.3.1 is also applicable to such measures.

Now, what is a possible strategy to prove Theorem 9.3.1? First, rather than the left-hand side of (9.43) we shall bound $\mathsf{E}\sup_{f\in\mathcal{F}}|\sum_{i\le N}\varepsilon_i f(X_i)^2|$, where (ε_i) are independent Bernoulli r.v.s, independent of the r.v.s (X_i). Setting $\mathcal{F}^2=\{f^2;f\in\mathcal{F}\}$ we have to bound the empirical process on the class $\mathcal{F}^2$. There is a natural chaining $(\pi_n(f))$ on $\mathcal{F}$, witnessing the value of $S=\gamma_2(\mathcal{F},d_2)+\gamma_1(\mathcal{F},d_1)$. There simply seems to be no other way than to use the chaining $(\pi_n(f)^2)$ on $\mathcal{F}^2$, and to use the strategy (9.37), which we shall use in the form (9.39). That is, to control the "increments along the chain"

$$\sum_{i\le N}\varepsilon_i(\pi_n(f)(X_i)^2-\pi_{n-1}(f)(X_i)^2)\,,$$

let us think that the r.v.s (X_i) have already been chosen. We will then find a set $I\subset\{1,\dots,N\}$ (depending on n and on f) for which we control both quantities

$$\sum_{i\in I}|\pi_n(f)(X_i)^2-\pi_{n-1}(f)(X_i)^2| \tag{9.47}$$

and

$$\Big(\sum_{i\notin I}|\pi_n(f)(X_i)^2-\pi_{n-1}(f)(X_i)^2|^2\Big)^{1/2} . \tag{9.48}$$

Now, we may expect to use the deviation inequality (9.42) to sufficiently control the sequences $(|\pi_n(f)(X_i)-\pi_{n-1}(f)(X_i)|)_{i\le N}$, but since

$$\begin{aligned}&\pi_n(f)(X_i)^2-\pi_{n-1}(f)(X_i)^2\\ &=(\pi_n(f)(X_i)-\pi_{n-1}(f)(X_i))(\pi_n(f)(X_i)+\pi_{n-1}(f)(X_i)) ,\end{aligned}$$

it seems impossible to achieve anything unless we have some control of the sequence $(\pi_n(f)(X_i)+\pi_{n-1}(f)(X_i))_{i\le N}$, which most likely means that we must gain some control of the sequence $(f(X_i))_{i\le N}$ for all $f\in\mathcal{F}$. In fact we shall prove in particular that

$$\mathsf{E}\sup_{f\in\mathcal{F}}\Big(\sum_{i\le N}f(X_i)^2\Big)^{1/2}\le K(C\sqrt{N}+S) ,$$

a formula which greatly contributes to explain the strange right-hand side in (9.43).

Before we start the proof we must understand the tail behavior of sums $\sum_{i\ge1}a_iY_i$ where a_i are numbers and where the independent r.v.s Y_i satisfy the tail condition (9.42). The methods are elementary and standard. The results are of constant use.

Lemma 9.3.2. *Consider a r.v. Y (not necessarily centered) and assume that for two numbers $A>0$ and $B>0$ we have*

$$\forall\, u>0 ,\ \mathsf{P}(|Y|\ge u)\le 2\exp\Big(-\min\Big(\frac{u^2}{A^2},\frac{u}{B}\Big)\Big) . \tag{9.49}$$

Then for each λ we have

$$0\le\lambda\le 1/(2B)\Rightarrow \mathsf{E}\exp\lambda|Y|\le L\exp(L\lambda^2A^2) . \tag{9.50}$$

Moreover when Y is centered and $B\le A$ we also have

$$0\le\lambda\le 1/(2B)\Rightarrow \mathsf{E}\exp\lambda Y\le \exp(L\lambda^2A^2) . \tag{9.51}$$

Let us observe that (9.51) cannot hold without restriction on A as the case $A=0$ shows. The condition $B\le A$ is always satisfied in practice.

Proof. We write

$$\mathsf{E}\exp\lambda|Y|=\int_0^\infty\mathsf{P}(\exp\lambda|Y|\ge v)\mathrm{d}v=1+\lambda\int_0^\infty\exp(\lambda u)\mathsf{P}(|Y|\ge u)\mathrm{d}u \tag{9.52}$$

and, since $u^2/A^2\le u/B$ for $u\le A^2/B$, using (9.49) we get

$$\lambda \int_0^{A^2/B} \exp(\lambda u) \mathsf{P}(|Y| \geq u) \mathrm{d}u \leq 2\lambda \int_0^{\infty} \exp(\lambda u) \exp\Big(-\frac{u^2}{A^2}\Big) \mathrm{d}u$$
$$\leq L \exp(L\lambda^2 A^2) \ . \tag{9.53}$$

Since $u/B \leq u^2/A^2$ for $u \geq A^2/B$ and since $\lambda \leq 1/(2B)$, using (9.49) again we get

$$\lambda \int_{A^2/B}^{\infty} \exp(\lambda u) \mathsf{P}(|Y| \geq u) \mathrm{d}u \leq 2\lambda \int_0^{\infty} \exp(\lambda u) \exp\Big(-\frac{u}{B}\Big) \mathrm{d}u$$
$$\leq 2\lambda \int_0^{\infty} \exp\Big(-\frac{u}{2B}\Big) \mathrm{d}u = 4\lambda B \leq 2 \ .$$

Combining with (9.52) and (9.53) proves (9.50).

To prove (9.51) it suffices to consider the case $|\lambda| A \leq 1/4$ for otherwise the result follows from (9.50). Then $2|\lambda| B \leq 1/2$ and (9.50) used for 2λ rather than λ implies $\mathsf{E} \exp 2|\lambda Y| \leq L$. Since $\exp x \leq 1 + x + x^2 \exp|x|$ we have $\mathsf{E} \exp \lambda Y \leq 1 + \lambda^2 \mathsf{E}(Y^2 \exp|\lambda Y|)$. Using the Cauchy-Schwarz inequality on the last term proves that $\mathsf{E} \exp \lambda Y \leq 1 + L\lambda^2 A^2$ since $\mathsf{E} \exp 2|\lambda Y| \leq L$ and since $\mathsf{E} Y^4 \leq L A^4$ from (9.49) because $B \leq A$. □

Lemma 9.3.3. *Consider i.i.d. copies* $(Y_i)_{i \leq k}$ *of a r.v.* Y *which satisfies the condition (9.49). Then for numbers* $(a_i)_{i \leq k}$ *and any* $u > 0$ *we have*

$$\mathsf{P}\Big(\Big|\sum_{i \leq k} a_i Y_i\Big| \geq u\Big) \leq L^k \exp\Big(-\frac{1}{L} \min\Big(\frac{u^2}{A^2 \sum_{i \leq k} a_i^2}, \frac{u}{B \max_{i \leq k} |a_i|}\Big)\Big) \ . \tag{9.54}$$

Proof. For any $\lambda > 0$ the left-hand side of (9.54) is at most

$$\exp(-\lambda u) \mathsf{E} \exp\Big(\lambda \sum_{i \leq k} |a_i||Y_i|\Big) = \exp(-\lambda u) \prod_{i \leq k} \mathsf{E} \exp(\lambda |a_i||Y_i|) \ ,$$

so that if $2B\lambda \max_{i \leq k} |a_i| \leq 1$ by Lemma 9.3.2 this is bounded by

$$\exp(-\lambda u) L^k \exp\Big(L_1 \lambda^2 A^2 \sum_{i \leq k} a_i^2\Big) \ .$$

As in the proof of Bernstein's inequality (4.59) we conclude by taking

$$\lambda = \min\Big(\frac{1}{2B \max_{i \leq k} |a_i|}, \frac{u}{2L_1 A^2 \sum_{i \leq k} a_i^2}\Big) \ . \qquad \square$$

A convenient way to use (9.54) is the following, which is now obvious:

Lemma 9.3.4. *Consider i.i.d. copies* $(Y_i)_{i \leq k}$ *of a r.v.* Y *which satisfies the condition (9.49). If* $w > 0$ *and*

$$v = LA\sqrt{w\sum_{i\le k} a_i^2} + LBw\max_{i\le k}|a_i|\ , \tag{9.55}$$

then

$$\mathsf{P}\Big(\Big|\sum_{i\le k} a_i Y_i\Big| \ge v\Big) \le L^k \exp(-w)\ . \tag{9.56}$$

Exercise 9.3.5. Consider a centered r.v. Y which satisfies (9.49) with $B \le A$. Using (9.51) improve (9.54) into

$$\mathsf{P}\Big(\Big|\sum_{i\le k} a_i Y_i\Big| \ge u\Big) \le L\exp\Big(-\frac{1}{L}\min\Big(\frac{u^2}{A^2\sum_{i\le k} a_k^2}, \frac{u}{B\max_{i\le k}|a_i|}\Big)\Big)\ . \tag{9.57}$$

We recommend that the reader studies the following exercise. The results there are often needed.

Exercise 9.3.6. Given a probability μ, for a measurable function f we define the following two norms (Orlicz norms).

$$\|f\|_{\psi_1} = \inf\Big\{A > 0\ ;\ \int \exp\Big(\frac{|f|}{A}\Big)\mathrm{d}\mu \le 2\Big\} \tag{9.58}$$

and

$$\|f\|_{\psi_2} = \inf\Big\{A > 0\ ;\ \int \exp\Big(\frac{f^2}{A^2}\Big)\mathrm{d}\mu \le 2\Big\}\ . \tag{9.59}$$

(a) Prove that

$$\int \exp|f|\mathrm{d}\mu \le 2^k \Rightarrow \|f\|_{\psi_1} \le k\ . \tag{9.60}$$

(Hint: Use Hölder's inequality.)
(b) Prove that

$$\forall\, u > 0\ ,\ \mathsf{P}(|f| \ge u) \le 2\exp(-u) \Rightarrow \|f\|_{\psi_1} \le L$$

and

$$\forall\, u > 0\ ,\ \mathsf{P}(|f| \ge u) \le 2\exp(-u^2) \Rightarrow \|f\|_{\psi_2} \le L\ .$$

(c) Prove that

$$\|f\|_{\psi_1} \le L\|f\|_{\psi_2} \tag{9.61}$$

and

$$\|f_1 f_2\|_{\psi_1} \le \|f_1\|_{\psi_2}\|f_2\|_{\psi_2}\ . \tag{9.62}$$

(d) On a rainy day, obtain a completely uninteresting and useless result by computing the exact value of $\|g\|_{\psi_2}$ where g is a standard Gaussian r.v.
(e) If (ε_i) denote independent Bernoulli r.v.s and (a_i) denote real numbers prove that

$$\Big\|\sum_i a_i\varepsilon_i\Big\|_{\psi_2} \le L\Big(\sum_i a_i^2\Big)^{1/2}\ .$$

(Hint: Use the subgaussian inequality (3.2.2).)
(f) Observe that any r.v. Y satisfies (9.49) with $A = B = \|Y\|_{\psi_1}$ and deduce from (9.57) that if the r.v.s Y_i are independent and centered then for $v > 0$ it holds

$$\mathsf{P}\Big(\sum_{i\ge 1} Y_i \ge v\Big) \le \exp\Big(-\frac{1}{L}\min\Big(\frac{v^2}{\sum_{i\le N}\|Y_i\|_{\psi_1}^2}, \frac{v}{\max_{i\le N}\|Y_i\|_{\psi_1}}\Big)\Big)\ . \tag{9.63}$$

Then rewrite a self-contained proof of this inequality.
(g) Deduce from (9.57) that if the r.v.s Y_i are independent and centered then for $v > 0$ it holds

$$\mathsf{P}\Big(\sum_{i\ge 1} Y_i \ge v\Big) \le \exp\Big(-\frac{v^2}{L\sum_{i\le N}\|Y_i\|_{\psi_2}^2}\Big)\ . \tag{9.64}$$

We recall the norm ψ_2 of (9.59). We denote by d_{ψ_2} the associated distance. Before proving Theorem 9.3.1 we prove the following simpler (and earlier) result, corresponding to a case where $d_2 = d_{\psi_2}$ and $d_1 = 0$, and where (9.41) is replaced by a much stronger condition.

Theorem 9.3.7 ([7], [10]). *Consider a (countable) class of functions $\mathcal{F}$ with $0 \in \mathcal{F}$. Assume that*

$$\forall f \in \mathcal{F}\ ,\ \|f\|_{\psi_2} \le \Delta^*\ . \tag{9.65}$$

Then

$$\mathsf{E}\sup_{\mathcal{F}}\Big|\sum_{i\le N}(f(X_i)^2 - \mathsf{E}f^2)\Big| \le L\sqrt{N}\Delta^*\gamma_2(\mathcal{F}, d_{\psi_2}) + L\gamma_2(\mathcal{F}, d_{\psi_2})^2\ . \tag{9.66}$$

We need one more simple fact.

Lemma 9.3.8. *If $u \ge 1$ then*

$$\mathsf{P}\Big(\sum_{i\le N}|f(X_i)| \ge 2uN\|f\|_{\psi_1}\Big) \le \exp(-uN)\ . \tag{9.67}$$

Proof. By homogeneity we may assume $\|f\|_{\psi_1} = 1$. Then

$$\mathsf{E}\exp\sum_{i\le N}|f(X_i)| \le 2^N \le e^N\ ,$$

so that $\mathsf{P}(\sum_{i\le N}|f(X_i)| \ge N(u+1)) \le \exp(-uN)$. □

We now prepare the proof of Theorem 9.3.7. We consider an admissible sequence $(\mathcal{A}_n)$ of partitions of $\mathcal{F}$ such that

$$\forall f \in \mathcal{F} \ , \ \sum_{n\geq 0} 2^{n/2}\Delta(A_n(f), d_{\psi_2}) \leq 2\gamma_2(\mathcal{F}, d_{\psi_2}) \ . \tag{9.68}$$

For each $A \in A_n$ we choose a point $f_A \in A$. (This is an unimportant detail, but we can arrange that this point does not depend on n.) For $A \in \mathcal{A}_n$ with $n \geq 1$ we denote by A' the unique element of $\mathcal{A}_{n-1}$ that contains A. This defines as usual a chaining in $\mathcal{F}$, by choosing $\pi_n(f) = f_A$ where $A = A_n(f)$.

We consider Bernoulli r.v.s (ε_i) independent of the r.v.s (X_i). We denote by n_1 the largest integer with $2^{n_1} \leq N$, so that $N \leq 2^{n_1+1}$.

Lemma 9.3.9. *Consider a parameter $u \geq 1$ and the event $\Omega(u)$ defined by the following conditions:*

$$\begin{aligned} &\forall n \ , \ 1 \leq n \leq n_1 \ , \ \forall A \in \mathcal{A}_n \ , \\ &\Big|\sum_{i\leq N} \varepsilon_i (f_A(X_i)^2 - f_{A'}(X_i)^2)\Big| \leq Lu2^{n/2}\sqrt{N}\Delta^*\Delta(A', d_{\psi_2}) \end{aligned} \tag{9.69}$$

$$\forall n > n_1 \ , \ \forall A \in \mathcal{A}_n \ , \ \sum_{i\leq N}(f_A(X_i) - f_{A'}(X_i))^2 \leq Lu2^n\Delta(A', d_{\psi_2})^2 \ . \tag{9.70}$$

$$\forall A \in \mathcal{A}_{n_1} \ , \ \sum_{i\leq N} f_A(X_i)^2 \leq LuN\Delta^{*2} \ . \tag{9.71}$$

Then

$$\mathsf{P}(\Omega(u)) \geq 1 - L\exp(-u) \ . \tag{9.72}$$

Proof. We first prove that, given $1 \leq n \leq n_1$, and given $A \in \mathcal{A}_n$ the inequality of (9.69) occurs with probability $\geq 1 - L\exp(-2u2^n)$. We observe that

$$f_A^2 - f_{A'}^2 = (f_A - f_{A'})(f_A + f_{A'}) \ ,$$

and that the first factor $f_A - f_{A'}$ has $\|\cdot\|_{\psi_2}$ norm $\leq \Delta(A', d_{\psi_2})$ while the second factor $(f_A + f_{A'})$ has $\|\cdot\|_{\psi_2}$ norm $\leq 2\Delta^*$ since $f_A, f_{A'} \in \mathcal{F}$. Consequently $\|f_A^2 - f_{A'}^2\|_{\psi_1} \leq 2\Delta^*\Delta(A', d_{\psi_2})$ by (9.62) and the r.v. $Y_i = \varepsilon_i(f_A(X_i)^2 - f_{A'}(X_i)^2)$ is centered and $\|Y_i\|_{\psi_1} \leq 2\Delta^*\Delta(A', d_{\psi_2})$. We use (9.63) to obtain that for any $v > 0$,

$$\mathsf{P}\Big(\Big|\sum_{i\leq N} Y_i\Big| \geq 2v\Delta^*\Delta(A', d_{\psi_2})\Big) \leq 2\exp\Big(-\frac{1}{L}\min\Big(\frac{v^2}{N}, v\Big)\Big) \ . \tag{9.73}$$

Since $\sqrt{N} \geq 2^{n/2}$ and $u \geq 1$, for $v = Lu2^{n/2}\sqrt{N}$, then $v^2/N \geq Lu2^n$ and $v \geq Lu2^n$. This implies as we claimed that the inequality in (9.69) occurs with probability $\geq 1 - L\exp(-2u2^n)$.

Next we prove that given $n > n_1$ and $A \in \mathcal{A}_n$ the inequality in (9.70) occurs with probability $\geq 1 - L\exp(-2u2^n)$. Since it is obvious from the definition that $\|f^2\|_{\psi_1} \leq \|f\|_{\psi_2}^2$, the function $f = (f_A - f_{A'})^2$ satisfies $\|f\|_{\psi_1} \leq \Delta(A', d_{\psi_2})^2$. Also $u2^n/N \geq 1$ since $n > n_1$. Using Lemma 9.3.8 for $u2^n/N$ rather than u we obtain indeed that the right-hand side of (9.70) occurs with probability $\geq 1 - L\exp(-2u2^n)$.

Using again Lemma 9.3.8, and since $\|f_A^2\|_{\psi_1} \leq \|f_A\|_{\psi_2}^2 \leq \Delta^{*2}$, we obtain that for any $A \in \mathcal{A}_{n_1}$ inequality (9.71) holds with probability $\geq 1 - L\exp(-2Nu)$.

Finally we use the union bound. Since $\operatorname{card} \mathcal{A}_n \leq N_n = 2^{2^n}$ and in particular $\operatorname{card} \mathcal{A}_{n_1} \leq N_{n_1} \leq 2^N$, and since $\sum_{n \geq 0} 2^{2^n} \exp(-2u2^n) \leq L\exp(-u)$ the result follows. □

We consider the random norm $W(f)$ given by

$$W(f) = \Big(\sum_{i \leq N} f(X_i)^2\Big)^{1/2} . \tag{9.74}$$

Lemma 9.3.10. *On the event $\Omega(u)$ we have*

$$\forall f \in \mathcal{F} ,\ W(f) \leq L\sqrt{u}(\gamma_2(\mathcal{F}, d_{\psi_2}) + \sqrt{N}\Delta^*) . \tag{9.75}$$

Proof. Given $f \in \mathcal{F}$ we denote by $\pi_n(f)$ the element f_A where $A = A_n(f)$. We also observe that $A_{n-1}(f)$ is the unique element A' in $\mathcal{A}_{n-1}$ which contains A.

First, (9.70) implies that for $n > n_1$ one has

$$W(\pi_n(f) - \pi_{n-1}(f)) \leq L2^{n/2}\sqrt{u}\Delta(A_{n-1}(f), d_{\psi_2}) . \tag{9.76}$$

Moreover (9.71) implies $W(\pi_{n_1}(f)) \leq L\sqrt{Nu}\Delta^*$. Writing $f = \pi_{n_1}(f) + \sum_{n > n_1}(\pi_n(f) - \pi_{n-1}(f))$, using the triangle inequality for W and (9.68) concludes the proof. □

Proof of Theorem 9.3.7. Let us recall the event $\Omega(u)$ of Lemma 9.3.9. First we prove that when this event occurs then

$$\sup_{f \in \mathcal{F}} \Big|\sum_{i \leq N} \varepsilon_i f(X_i)^2\Big| \leq Lu\gamma_2(\mathcal{F}, d_{\psi_2})(\gamma_2(\mathcal{F}, d_{\psi_2}) + \sqrt{N}\Delta^*) , \tag{9.77}$$

which by taking expectation and using (9.72) implies

$$\mathsf{E} \sup_{f \in \mathcal{F}} \Big|\sum_{i \leq N} \varepsilon_i f(X_i)^2\Big| \leq L\big(\sqrt{N}\Delta^*\gamma_2(\mathcal{F}, d_{\psi_2}) + \gamma_2(\mathcal{F}, d_{\psi_2})^2\big) ,$$

and the conclusion by (9.21). Since $0 \in \mathcal{F}$ we may assume that $\pi_0(f) = 0$. We deduce from (9.69) that for each n with $1 \leq n \leq n_1$, one has

$$\Big|\sum_{i\le N}\varepsilon_i(\pi_n(f)(X_i)^2-\pi_{n-1}(f)(X_i)^2)\Big|\le Lu2^{n/2}\sqrt{N}\Delta^*\Delta(A_{n-1}(f),d_{\psi_2})\,. \tag{9.78}$$

For $n>n_1$ we write

$$\Big|\sum_{i\le N}\varepsilon_i(\pi_n(f)(X_i)^2-\pi_{n-1}(f)(X_i)^2)\Big|\le\sum_{i\le N}\big|\pi_n(f)(X_i)^2-\pi_{n-1}(f)(X_i)^2\big|\,. \tag{9.79}$$

Recalling the random norm $W(f)$ of (9.74), and since $(a-b)^2=(a-b)(a+b)$, using the Cauchy-Schwarz inequality the right-hand side of (9.79) is at most

$$\begin{aligned}&W(\pi_n(f)-\pi_{n-1}(f))W(\pi_n(f)+\pi_{n-1}(f))\\&\le W(\pi_n(f)-\pi_{n-1}(f))(W(\pi_n(f))+W(\pi_{n-1}(f))\,,\end{aligned} \tag{9.80}$$

and from (9.76) and (9.75) this is at most

$$Lu2^{n/2}\Delta(A_{n-1}(f),d_{\psi_2})(\gamma_2(\mathcal{F},d_{\psi_2})+\sqrt{N}\Delta^*)\,.$$

Combining with (9.78) and summation over n using (9.68) proves (9.77) and concludes the proof of Theorem 9.3.7. □

Exercise 9.3.11. When considering (9.58) one does not really follow the strategy outlined in Section 9.2. Write an alternate (and slightly longer) proof of Theorem 9.3.7 that would strictly follow this strategy.

We now turn to the proof of Theorem 9.3.1 itself. The highlight of this theorem is that it uses only the weak condition (9.41) to control the "diameter" of $\mathcal{F}$. This makes the result difficult, so the first time reader should probably jump now to Section 9.4. During the course of this proof, we shall have to prove that certain inequalities hold whatever the choice of a certain set $I\subset\{1,\dots,N\}$. The number $\binom{N}{k}$ of such sets will be very relevant. To control this number, we use the well known inequality

$$\binom{N}{k}\le\Big(\frac{eN}{k}\Big)^k=\exp(k\log(eN/k))\,. \tag{9.81}$$

The value of k for which this number is about 2^{2^n} is important, because then the number of sets of cardinality k is about the cardinality of a partition $\mathcal{A}_n$ that we use during the chaining. We gather properties of this number through the following elementary lemma. We define again n_1 as the largest integer for which $2^{n_1}\le N$.

Lemma 9.3.12. *For an integer $n\le n_1$, let us define $k(n)$ as the smallest integer $k(n)\ge 1$ such that*

$$2^n\le k(n)\log(eN/k(n))\,. \tag{9.82}$$

Then for $n<n_1$ we have

$$k(n) \geq 2 \Rightarrow 2(k(n)-1) < k(n+1) \Rightarrow k(n+1) \geq \frac{3}{2}k(n) \tag{9.83}$$

and

$$k(n+1) \leq 8k(n) \; . \tag{9.84}$$

Moreover, for $k(n) > 1$ *and* $n < n_1$ *we have*

$$k(n)\log(eN/k(n)) \leq 2^{n+1} \; . \tag{9.85}$$

Proof. The basic idea is that the function $\log(eN/x)$ varies slowly, so that we expect that $k(n) = 2^n U(n)$ where $U(n)$ varies slowly. The details however require some work. Consider the function $\varphi(x) = x\log(eN/x)$ for $x \geq 0$, so that for $x \leq N$ we have $\varphi(x) \geq x$. Then $k(n)$ is defined as the smallest integer such that $2^n \leq \varphi(k(n))$, and $k(n) \leq N$ since $n \leq n_1$. For $a > 0$ we have

$$\varphi(ax) = a\varphi(x) - ax\log a \; ,$$

so that

$$\varphi(2x) < 2\varphi(x) \; . \tag{9.86}$$

To prove (9.83) we observe that by definition of $k(n)$,

$$\varphi(k(n)-1) \leq 2^n \; ,$$

so that since $k(n)-1 \leq N$, (9.86) implies $\varphi(2(k(n)-1)) < 2^{n+1} \leq \varphi(k(n+1))$ and (9.83). Moreover, when $k(n) > 1$, we have $k(n) \leq 2(k(n)-1)$, so that $\varphi(k(n)) \leq \varphi(2(k(n)-1)) \leq 2^{n+1}$ and this proves (9.85).

To prove (9.84) we may assume $k(n) \leq N/8$ for otherwise (9.84) holds since $k(n+1) \leq N$. We then note that $\varphi(x) \geq x\log(8e)$ for $x \leq N/8$. Also, since $\log 8 < 3$ we have $6\log(8e) = 6(1+\log 8) \geq 8\log 8$ and hence for $x \leq N/8$ we have $6\varphi(x) \geq 8x\log 8$. Therefore $\varphi(8x) = 8\varphi(x) - 8x\log 8 \geq 2\varphi(x)$, so that since $\varphi(k(n)) \geq 2^n$ we have $\varphi(8k(n)) \geq 2^{n+1} \geq \varphi(k(n+1))$ and this proves (9.84). □

We will keep using the notation

$$\varphi(k) = k\log(eN/k) \; ,$$

so that (9.81) becomes

$$\binom{N}{k} \leq \left(\frac{eN}{k}\right)^k = \exp\varphi(k) \; . \tag{9.87}$$

We observe that $\varphi(k) \geq k$, that the sequence $(\varphi(k))$ increases, and thus $\varphi(k) \geq \varphi(1) = 1 + \log N$.

We continue the preparations for the proof of Theorem 9.3.1. In the beginning of the proof of Theorem 2.2.23 we have learned how to construct an admissible sequence $(\mathcal{A}_n)$ of partitions of $\mathcal{F}$ which, for $j = 1, 2$ satisfies

$$\forall f \in \mathcal{F} , \sum_{n \geq 0} 2^{n/j} \Delta(A_n(f), d_j) \leq 2\gamma_j(\mathcal{F}, d_j) . \tag{9.88}$$

For each $A \in A_n$ we choose a point $f_A \in A$. This defines as usual a chaining in $\mathcal{F}$, by choosing $\pi_n(f) = f_A$ where $A = A_n(f)$. (Again we can arrange that this point does not depend on n.) For $A \in \mathcal{A}_n$ with $n \geq 1$ we denote by A' the unique element of $\mathcal{A}_{n-1}$ that contains A. To lighten notation, for $n \geq 1$ and $A \in \mathcal{A}_n$ we write

$$\delta_A = f_A - f_{A'} , \tag{9.89}$$

and for $j = 1, 2$ we define

$$\Delta_A^j = \Delta(A', d_j) . \tag{9.90}$$

For a number $u \geq 0$ we define

$$d(u, k) = 2C\Big(\frac{eN}{k}\Big)^{2/q} \exp\Big(\frac{u}{qk}\Big) , \tag{9.91}$$

so that (9.41) implies

$$\forall f \in \mathcal{F} , \ \mu(\{|f| \geq d(u, k)\}) \leq 2^{-q}\Big(\frac{k}{eN}\Big)^2 \exp(-u/k) . \tag{9.92}$$

In the next proposition we start to gather the information we need. Of course, the usefulness of these conditions will only become clear gradually. We recall that we denote by n_1 the largest integer with $2^{n_1} \leq N$, and we recall the number $k(n)$ of Lemma 9.3.12, defined for $0 \leq n \leq n_1$.

Proposition 9.3.13. *Consider a number $u \geq L$. Then with probability $\geq 1 - L\exp(-u)$ the following occurs. Consider any integer $0 \leq n \leq n_1$, any integer $k(n) \leq k \leq N$ and any subset I of $\{1, \dots, N\}$ with* $\operatorname{card} I = k$*. Then*

$$A \in \mathcal{A}_n \Rightarrow \exists i \in I , \ |f_A(X_i)| < d(u, k) , \tag{9.93}$$

$$n \geq 1 , \ A \in \mathcal{A}_n \Rightarrow \sum_{i \in I} |\delta_A(X_i)| \leq L(\sqrt{u}\Delta_A^2\sqrt{k\varphi(k)} + u\Delta_A^1\varphi(k)) . \tag{9.94}$$

Proof. First we claim that (9.93) occurs with probability $\geq 1 - L\exp(-u)$ for $u \geq L$. Given $A \in \mathcal{A}_n$ we observe from (9.92) that for any $i \in I$,

$$\mathsf{P}(|f_A(X_i)| \geq d(u, k)) \leq 2^{-q}\Big(\frac{k}{eN}\Big)^2 \exp(-u/k) ,$$

so that given a set I with cardinality k,

$$\begin{aligned}\mathsf{P}(\forall\, i \in I, |f_A(X_i)| \geq d(u, k)) &\leq 2^{-qk}\Big(\frac{k}{eN}\Big)^{2k} \exp(-u) \\ &= 2^{-qk} \exp(-u) \exp(-2\varphi(k)) .\end{aligned}$$

Moreover, using (9.87), since $\varphi(k) \geq \varphi(k(n)) \geq 2^n$ and since

$$\sum_{k(n)\leq k\leq N} 2^{-qk}\exp(-\varphi(k)) \leq \exp(-2^n)\sum_{k\geq 1} 2^{-qk} \leq L\exp(-2^n) ,$$

so that

$$\begin{aligned}\sum_{n\geq 0} 2^{2^n}\sum_{k(n)\leq k\leq N}\binom{N}{k}2^{-qk}\exp(-2\varphi(k)) &\leq \sum_{n\geq 0}2^{2^n}\sum_{k(n)\leq k\leq N}2^{-qk}\exp(-\varphi(k))\\ &\leq L\sum_{n\geq 0}2^{2^n}\exp(-2^n)\leq L , \qquad (9.95)\end{aligned}$$

and this proves the claim.

Next, we claim that (9.94) occurs with probability $\geq 1-L\exp(-u)$. From the hypothesis (9.42) the independent r.v.s $\delta_A(X_i)$ satisfy (9.49) for $A = \Delta_A^2$ and $B = \Delta_A^1$, so by Lemma 9.3.4 for any coefficients (a_i) we then have

$$\begin{aligned}&\mathsf{P}\Big(\Big|\sum_{i\leq k} a_i\delta_A(X_i)\Big| \geq L\Delta_A^2\sqrt{u\varphi(k)\sum_{i\leq k}a_i^2} + L\Delta_A^1 u\varphi(k)\max_{i\leq k}|a_i|\Big)\\ &\leq L^k\exp(-u\varphi(k)) . \qquad (9.96)\end{aligned}$$

Consequently, the event Ω_u on which

$$\Big|\sum_{i\in I} a_i\delta_A(X_i)\Big| \leq L\Delta_A^2\sqrt{uk\varphi(k)} + L\Delta_A^1 u\varphi(k) \qquad (9.97)$$

for every choice of $n \geq 1$, $A \in \mathcal{A}_n$, every integer k with $k(n) \leq k \leq N$, every set $I \subset \{1,\dots,N\}$ with $\operatorname{card} I = k$ and every coefficients $(a_i)_{i\in I}$ with $a_i = \pm 1$ satisfies (using in the third line that $L^k \leq \exp(L'\varphi(k))$ since $\varphi(k) \geq k$)

$$\begin{aligned}\mathsf{P}(\Omega_u^c) &\leq \sum_{n\geq 1}\operatorname{card}\mathcal{A}_n\sum_{k(n)\leq k\leq N}\binom{N}{k}L^k\exp(-u\varphi(k))\\ &\leq \sum_{n\geq 1}\operatorname{card}\mathcal{A}_n\sum_{k(n)\leq k\leq N}L^k\exp(-(u-1)\varphi(k))\\ &\leq \sum_{n\geq 1}2^{2^n}\exp(-(u-L)2^n) , \qquad (9.98)\end{aligned}$$

so that indeed $\mathsf{P}(\Omega_u^c) \leq L\exp(-u)$ for $u \geq L$. Now, when (9.97) occurs for each choice of $a_i = \pm 1$ we have

$$\sum_{i\in I}|\delta_A(X_i)| \leq L\Delta_A^2\sqrt{uk\varphi(k)} + L\Delta_A^1 u\varphi(k) . \qquad \square$$

We need the following elementary fact.

Lemma 9.3.14. *In* $\mathbb{R}^k$ *there is a set* U *with* $\operatorname{card} U \le 5^k$ *consisting of vectors of norm* ≤ 1*, with the property that* $x \in 2\operatorname{conv} U$ *whenever* $\|x\| \le 1$*. Consequently,*

$$\forall\, x \in \mathbb{R}^k \ , \ \exists a \in U \ , \ \sum_{i \le k} a_i x_i \ge \frac{1}{2}\Big(\sum_{i \le k} x_i^2\Big)^{1/2} \ . \tag{9.99}$$

Proof. It follows from (2.41) that there exists a subset U of the unit ball of $\mathbb{R}^k$ with $\operatorname{card} U \le 5^k$ of the unit sphere such that every point of this ball is within distance $\le 1/2$ of a point of U. Given a point x of the unit ball we can inductively pick points u_ℓ in U such that $\|x - \sum_{1 \le \ell \le n} 2^{\ell-1} u_\ell\| \le 2^{-n}$ and this proves that $x \in 2\operatorname{conv} U$. Let us denote by $\langle \cdot, \cdot \rangle$ the canonical dot product on $\mathbb{R}^k$. Given $x \in \mathbb{R}^k$ and using that $x/\|x\| \in 2\operatorname{conv} U$ we obtain that $\|x\|^2 = \langle x, x \rangle \le 2\|x\| \sup_{a \in U} \langle x, a \rangle$ which proves (9.99). □

Using Lemma 9.3.14, for each $1 \le k \le N$, and each subset I of $\{1, \dots, N\}$ of cardinality k, we construct a subset $\mathcal{S}_{k,I}$ of the unit ball of $\mathbb{R}^I$ with $\operatorname{card} \mathcal{S}_{k,I} \le 5^k$, such that $2\operatorname{conv} \mathcal{S}_{k,I}$ contains this unit ball. Consequently

$$x \in \mathbb{R}^I \Rightarrow \sup_{a \in \mathcal{S}_{k,I}} \sum_{i \in I} a_i x_i \ge \frac{1}{2}\Big(\sum_{i \in I} x_i^2\Big)^{1/2} \ . \tag{9.100}$$

Proposition 9.3.15. *Consider a number* $u \ge L$*. Then with probability* $\ge 1 - L\exp(-u)$ *the following occurs. Consider any integer* n*. Then, if* $n \le n_1$*, for any subset* I *of* $\{1, \dots, N\}$ *with* $\operatorname{card} I = k(n)$*,*

$$A \in \mathcal{A}_n \Rightarrow \Big(\sum_{i \in I} \delta_A(X_i)^2\Big)^{1/2} \le L(\sqrt{u} 2^{n/2} \Delta_A^2 + u 2^n \Delta_A^1) \ , \tag{9.101}$$

while if $n > n_1$*,*

$$A \in \mathcal{A}_n \Rightarrow \Big(\sum_{i \le N} \delta_A(X_i)^2\Big)^{1/2} \le L(\sqrt{u} 2^{n/2} \Delta_A^2 + u 2^n \Delta_A^1) \ . \tag{9.102}$$

Proof. We observe that (9.87) implies

$$\log \operatorname{card} \bigcup_{\operatorname{card} I = k} \mathcal{S}_{k,I} \le k \log 5 + \varphi(k) \le 3\varphi(k) \ . \tag{9.103}$$

We proceed exactly as in the case of (9.94) to obtain that for $u \ge L$, with probability $\ge 1 - \exp(-u)$, for each set I with $\operatorname{card} I = k(n)$, each sequence (a_i) in $\mathcal{S}_{k(n),I}$ and each A in $\mathcal{A}_n$ we have, using in the second line that $\varphi(k(n)) \le L 2^n$ by (9.85),

$$\Big|\sum_{i\in I} a_i\delta_A(X_i)\Big| \le L\Delta_A^2\sqrt{u\varphi(k(n))} + L\Delta_A^1 u\varphi(k(n))$$
$$\le L\Delta_A^2\sqrt{u2^n} + L\Delta_A^1 u2^n \,. \tag{9.104}$$

Namely, (9.96) proves that for each set I with $\operatorname{card} I = k(n)$, each sequence (a_i) in $\mathcal{S}_{k(n)}$ and each A in $\mathcal{A}_n$, the probability that (9.104) fails is at most $L^{k(n)}\exp(-u\varphi(k(n))) \le L^{2^n}\exp(-u2^n)$, and using (9.103) the sum of these quantities over all choices of I, of the sequence (a_i) and of A is $\le L\exp(-u)$ for $u \ge L$.

Using (9.100) this implies (9.101). In a similar manner for $n > n_1$, (9.104) occurs for $k = N$ with probability $\ge 1 - \exp(-u2^n)$. This completes the proof. □

In the remainder of this section we denote by $\Omega(u)$ the event that (9.93), (9.94), (9.101) and (9.102) occur. Thus Propositions 9.3.13 and 9.3.15 imply that $\mathsf{P}(\Omega(u)) \ge 1 - L\exp(-u)$.

We are now ready for the crucial step of gaining control over

$$\sup_{f\in\mathcal{F}}\sum_{i\le N} f(X_i)^2 \,.$$

Without loss of generality we shall assume throughout the proof that $\mathcal{F}$ is finite. We recall that $S = \gamma_1(T,d_1) + \gamma_2(T,d_2)$ and the definition (9.91) of $d(u,k)$. We abandon all pretense to get a decent dependence in u by using the trivial bound

$$d(u,k) \le 2C\Big(\frac{eN}{k}\Big)^{2/q}\exp\Big(\frac{u}{q}\Big) \,. \tag{9.105}$$

(To get a decent dependence on u one simply has to use (9.44) rather than (9.41) for the small values of k, a direction which we do not pursue.)

Proposition 9.3.16. *Assume that $\Omega(u)$ occurs. Then for each set $I \subset \{1,\ldots,N\}$ and each $f \in \mathcal{F}$ we have, setting $k = \operatorname{card} I$,*

$$\Big(\sum_{i\in I} f(X_i)^2\Big)^{1/2} \le L\sqrt{u}\gamma_2(\mathcal{F},d_2) + Lu\gamma_1(\mathcal{F},d_1) + KC\exp(u/q)N^{2/q}k^{1/2-2/q}$$
$$\le LuS + KC\exp(u/q)N^{2/q}k^{1/2-2/q} \,. \tag{9.106}$$

Proof. The proof we present is identical to the proof of the similar statement in [11]. We shall use this statement only when $k = N$. This case follows from the case $k = k(n_1)$ since $k(n_1) \ge N/L$, so that $\{1,\ldots,N\}$ can be covered by L sets of cardinality n_1. We shall prove the statement only in the case $k = k(n_0)$ for some integer $1 \le n_0 \le n_1$, and we let the reader convince herself using (9.84) and the same covering argument as above that this implies the general case. We consider any subset I of $\{1,\ldots,N\}$, any numbers $(v_i)_{i\in I}$ such that $\sum_{i\in I} v_i^2 \le 1$, and we aim to prove that

$$\sum_{i\in I} v_i f(X_i) \le L\sqrt{u}\gamma_2(\mathcal{F},d_2) + Lu\gamma_1(\mathcal{F},d_1) + KC\exp(u/q)N^{2/q}k^{1/2-2/q} \;, \tag{9.107}$$

where $k = \operatorname{card} I$. The proof is a chaining argument. Since we assume that $\mathcal{F}$ is finite, to prove (9.107) it suffices to consider the case where $f = f_A$ for some $A \in \mathcal{A}_{n_2}$ where $n_2 \ge n_0$ is large enough so that $\pi_{n_2}(f) = f$ for each f. Consider any integer $m \ge 0$. Then we have the chaining identity

$$\begin{aligned} v_i f(X_i) &= v_i \pi_{n_2}(f)(X_i) \\ &= \sum_{m<n\le n_2} v_i(\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i)) + v_i\pi_m(f)(X_i) \;. \end{aligned} \tag{9.108}$$

For each $i \in I$ let us now choose any way we like an integer $m = m(i)$ with $0 \le m(i) \le n_2$. (The idea will be to choose $m = m(i)$ so that the term $\pi_m(f)(X_i)$ is suitably small.) For $0 \le m \le n_2$ and $0 \le n \le n_2$ let us define

$$J(m) = \{i \in I \;;\; m(i) = m\} \;;\; I(n) = I \setminus \bigcup_{m\ge n} J(m) \;, \tag{9.109}$$

so that $i \in I(n) \Leftrightarrow m(i) < n$, and summation of the identities (9.108) with $m = m(i)$ over $i \in I$ yields the relation

$$\begin{aligned} \sum_{i\in I} v_i f(X_i) = &\sum_{1\le n\le n_2} \sum_{i\in I(n)} v_i(\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i)) \\ &+ \sum_{0\le m\le n_2} \sum_{i\in J(m)} v_i\pi_m(f)(X_i) \;. \end{aligned} \tag{9.110}$$

Let us now turn things around: the previous relation holds whatever the sets $J(m)$, provided they define a partition of I, and provided $I(n)$ is defined by (9.109).

Let us define these sets in a way that suits our program. Define first n^* as the smallest integer for which $k(n^*+1) \ge 2$, so that $k(n^*) = 1$. For $n < n^*$ let us set $I^*(n) = \emptyset$. For $n^* \le n \le n_0$ let us consider a subset $I^*(n)$ of I with $\operatorname{card} I^*(n) = k(n) - 1$ such that

$$i \notin I^*(n) \Rightarrow |\pi_n(f)(X_i)| \le d(u, k(n)) \;. \tag{9.111}$$

This is possible by (9.93) and since $\pi_n(f) = f_A$ for $A = A_n(f)$. We observe that $I^*(n^*) = \emptyset$ since $\operatorname{card} I^*(n^*) = k(n^*) - 1 = 0$. For $n \ge n_0$ we define $I^*(n) = I$. For $n^* \le n < n_0$ let us define $J(n) = I^*(n+1) \setminus I^*(n)$, so that $J(n^*) = I(n^*+1)$ and $J(n) = \emptyset$ for $n \ge n_0$. We observe that the sets $J(n)$ are disjoint, and form a partition of I. We define $I(n) = \bigcup_{m\le n} I^*(m) = I \setminus \bigcup_{m\ge n} J(m)$, so that the second part of (9.109) holds, (9.110) holds and becomes

$$\begin{aligned} \sum_{i\in I} v_i f(X_i) = &\sum_{n^*<n\le n_2} \sum_{i\in I(n)} v_i(\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i)) \\ &+ \sum_{n^*\le n<n_0} \sum_{i\in J(n)} v_i\pi_n(f)(X_i) \;. \end{aligned} \tag{9.112}$$

We shall bound separately each of the double sums in the right-hand side of (9.112). We first bound the first sum. We observe that

$$\delta_{A_n(f)} = \pi_n(f) - \pi_{n-1}(f) , \tag{9.113}$$

and that by the Cauchy-Schwarz inequality, and since $\sum_{i \le N} v_i^2 \le 1$,

$$\Big| \sum_{i \in I(n)} v_i(\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i)) \Big| \le \Big(\sum_{i \in I(n)} \delta^2_{A_n(f)} \Big)^{1/2} .$$

Using (9.83), we obtain that $k(m) \le (2/3)^{-(n-m)} k(n)$ when $k(m)-1 > 0$ and $m \le n$. Thus $\operatorname{card} I(n) \le \sum_{m \le n} \operatorname{card} I^*(m) \le \sum_{m \le n} (k(m) - 1) \le Lk(n)$ for all n. We deduce from (9.101) for $n \le n_1$ and from (9.102) for $n > n_1$ that

$$\Big| \sum_{i \in I(n)} v_i(\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i)) \Big| \le L(\sqrt{u} 2^{n/2} \Delta^2_{A_n(f)} + u 2^n \Delta^1_{A_n(f)}) .$$

Since for $j = 1, 2$ we have $\sum_{n \ge 1} 2^{n/j} \Delta^j_{A_n(f)} \le L\gamma_j(\mathcal{F}, d_j)$ by (9.88), we obtain

$$\Big| \sum_{n^* < n \le n_2} \sum_{i \in I(n)} v_i(\pi_n(f)(X_i) - \pi_{n-1}(f)(X_i)) \Big| \le LuS . \tag{9.114}$$

Next we bound the second double sum in the right-hand side of (9.112). We write

$$\Big| \sum_{i \in J(n)} v_i \pi_{n-1}(f)(X_i) \Big| \le \max_{i \in J(n)} |\pi_{n-1}(f)(X_i)| \sum_{i \in J(n)} |v_i| . \tag{9.115}$$

We use (9.111) for $n \ge n^*$ and since $J(n) \cap I^*(n) = \emptyset$,

$$\max_{i \in J(n)} |\pi_n(f)(X_i)| \le d(u, k(n)) . \tag{9.116}$$

Now, by the Cauchy-Schwarz inequality

$$\sum_{i \in J(n)} |v_i| \le \sqrt{\operatorname{card} J(n)} \Big(\sum_{i \in J(n)} v_i^2 \Big)^{1/2}$$

and moreover $\operatorname{card} J(n) \le \operatorname{card} I(n+1) \le Lk(n+1) \le Lk(n)$ by (9.84), so that

$$\Big| \sum_{i \in J(n)} v_i \pi_{n-1}(f)(X_i) \Big| \le Ld(u, k(n)) \sqrt{k(n)} \Big(\sum_{i \in J(n)} v_i^2 \Big)^{1/2} . \tag{9.117}$$

Let us also observe from (9.83), and since $k(n^*+1) = 2$, $k(n_0) = k$ and $1 - 4/q > 0$,

$$\sum_{n^* \le n \le n_0} k(n)^{1-4/q} \le K k^{1-4/q} \; . \tag{9.118}$$

Since the sets $J(n)$ are disjoint and $k(n_0) = k$ we obtain by the Cauchy-Schwarz inequality, using also (9.105) and (9.118), and since $\sum_i v_i^2 \le 1$,

$$\begin{aligned} \Big| \sum_{n^* \le n \le n_0} \sum_{i \in J(n)} v_i \pi_n(f)(X_i) \Big| &\le L \sum_{n^* \le n \le n_0} d(u, k(n)) \sqrt{k(n)} \Big(\sum_{i \in J(n)} v_i^2 \Big)^{1/2} \\ &\le L \Big(\sum_{n^* \le n \le n_0} k(n) d(u, k(n))^2 \Big)^{1/2} \\ &\le LCN^{2/q} \exp(u/q) \Big(\sum_{n^* \le n \le n_0} k(n)^{1-4/q} \Big)^{1/2} \\ &\le KCN^{2/q} k^{1/2-2/q} \exp(u/q) \; . \end{aligned} \tag{9.119}$$

Combining with (9.112) and (9.114) we have proved (9.107). □

Before proving Theorem 9.3.1 we need a last observation.

Lemma 9.3.17. *Consider numbers $(c_i)_{i \le N}$, $c_i \ge 0$, and assume that for a certain $n \le n_1$, for each subset I of $\{1, \ldots, N\}$ with* $\operatorname{card} I = k \ge k(n)$ *we have*

$$\sum_{i \in I} c_i \le \varphi(k) \; . \tag{9.120}$$

Then we can find a subset I of $\{1, \ldots, N\}$ with $\operatorname{card} I = k(n) - 1$ *such that for any $r > 0$,*

$$\sum_{i \notin I} c_i^r \le KN \; . \tag{9.121}$$

Proof. Without loss of generality we may assume that the sequence (c_i) is non-increasing. Then for $k \ge k(n)$ we have

$$k c_k \le \sum_{i \le k} c_i \le \varphi(k) = k \log(eN/k) \; ,$$

so that $c_k \le \log(eN/k)$ and $\sum_{k \ge k(n)} c_k^r \le \sum_{1 \le k \le N} (\log(eN/k))^r \le KN$ with huge room to spare. □

Here is another, simpler, version of the same principle.

Lemma 9.3.18. *Consider numbers $(x_i)_{i \le N}$ and assume that for a certain $n \le n_1$, for each subset I of $\{1, \ldots, N\}$ with* $\operatorname{card} I = k \ge k(n)$ *there is $i \in I$ with $|x_i| \le d(u, k)$. Then we can find a subset I of $\{1, \ldots, N\}$ with* $\operatorname{card} I = k(n) - 1$ *such that for any $r > 0$,*

$$\sum_{i \notin I} |x_i|^r \le \sum_{j \ge k(n)} d(u, j)^r \; . \tag{9.122}$$

Proof. Assuming, as we may, that the sequence $(|x_i|)_{i\le N}$ is non-increasing, it is obvious that $|x_j| \le d(u,j)$ for $j \ge k(n)$. □

Let us fix $1 < r < q/4$, so that (9.105) implies

$$\Big(\sum_{1\le j\le N} d(u,j)^{2r}\Big)^{1/2r} \le KC\exp(u/q)N^{1/2r}\ . \tag{9.123}$$

Using again (9.21), to conclude the proof of Theorem 9.3.1 it suffices to prove the following, where we recall that E_ε denotes expectation in the r.v.s (ε_i) only.

Proposition 9.3.19. *On $\Omega(u)$ we have*

$$\mathsf{E}_\varepsilon \sup_{f\in\mathcal{F}}\Big|\sum_{i\le N}\varepsilon_i f(X_i)^2\Big| \le Lu^2\exp(u/q)S(\sqrt{N}C+S)\ . \tag{9.124}$$

As explained before, we shall control "along the chain" the r.v.s

$$\sum_{i\le N}\varepsilon_i(f_A(X_i)^2 - f_{A'}(X_i)^2)\ . \tag{9.125}$$

The basic principle is as follows.

Lemma 9.3.20. *When the event $\Omega(u)$ occurs, then for each $n\ge 1$ and each $A\in\mathcal{A}_n$ we can write*

$$f_A(X_i)^2 - f_{A'}(X_i)^2 = v_i + w_i\ , \tag{9.126}$$

where

$$\sum_{i\le N}|v_i| \le Ku(2^{n/2}\Delta_A^2 + 2^n\Delta_A^1)\big(Su + \exp(u/q)C\sqrt{N}\big)\ , \tag{9.127}$$

and where $w_i = 0$ for $n > n_1$ while for $n\le n_1$

$$\Big(\sum_{i\le N} w_i^2\Big)^{1/2} \le Ku\exp(u/q)(\Delta_A^2+\Delta_A^1)\sqrt{N}C\ . \tag{9.128}$$

Proof. It implements the strategy of (9.39): We shall find a suitable set I for which

$$v_i = (f_A(X_i)^2 - f_{A'}(X_i)^2)\mathbf{1}_I$$

is the "peaky part", and

$$w_i = (f_A(X_i)^2 - f_{A'}(X_i)^2)\mathbf{1}_{I^c}$$

is the "spread part". We define $s_i = f_A(X_i) + f_{A'}(X_i)$, so that

$$f_A(X_i)^2 - f_{A'}(X_i)^2 = s_i\delta_A(X_i)$$

and

$$\Big(\sum_{i\le N} w_i^2\Big)^{1/2} = \Big(\sum_{i\in I} s_i^2\delta_A(X_i)^2\Big)^{1/2} \ ; \ \sum_{i\le N}|v_i| = \sum_{i\in I}|s_i\delta_A(X_i)| \ . \tag{9.129}$$

Let us assume first that $n \le n_1$. We then deduce from (9.93), (9.123) and Lemma 9.3.18 (used once for f_A and once for $f_{A'}$) that we can find a subset I_1 of $\{1,\ldots,N\}$ with $\operatorname{card} I_1 = 2(k(n)-1)$ and

$$\Big(\sum_{i\notin I_1} s_i^{2r}\Big)^{1/2r} \le K\exp(u/q)N^{1/2r}C \ . \tag{9.130}$$

Next, since $\sqrt{k\varphi(k)} \le \varphi(k)$, we deduce from (9.94) that for any subset J of $\{1,\ldots,N\}$ of cardinality k, we have

$$A\in\mathcal{A}_n \Rightarrow \sum_{i\in J}|\delta_A(X_i)| \le L(\sqrt{u}\Delta_A^2 + u\Delta_A^1)\varphi(k) \le Lu(\Delta_A^2+\Delta_A^1)\varphi(k) \ .$$

Let r' be the conjugate exponent of r. We then deduce from Lemma 9.3.17 that we can find a subset I_2 of $\{1,\ldots,N\}$ with $\operatorname{card} I_2 = k(n)-1$ and

$$\Big(\sum_{i\notin I_2}\delta_A(X_i)^{2r'}\Big)^{1/2r'} \le Ku(\Delta_A^2+\Delta_A^1)N^{1/2r'} \ . \tag{9.131}$$

We define $I = I_1\cup I_2$. Then (9.128) follows from (9.129), (9.130) and (9.131), using Hölder's inequality:

$$\Big(\sum_{i\notin I} s_i^2\delta_A(X_i)^2\Big)^{1/2} \le \Big(\sum_{i\notin I} s_i^{2r}\Big)^{1/2r}\Big(\sum_{i\notin I}\delta_A(X_i)^{2r'}\Big)^{1/2r'} \ .$$

Next, since $I=\emptyset$ when $k(n)=1$ and $\operatorname{card} I \le 2k(n)$ when $k(n)>1$, (9.101) implies

$$\Big(\sum_{i\in I}\delta_A(X_i)^2\Big)^{1/2} \le L(\sqrt{u}\Delta_A^2 2^{n/2} + u\Delta_A^1 2^n) \le Lu(\Delta_A^2 2^{n/2} + \Delta_A^1 2^n) \ , \tag{9.132}$$

while, using (9.106) for $k=N$,

$$\Big(\sum_{i\in I} s_i^2\Big)^{1/2} \le \Big(\sum_{i\le N} s_i^2\Big)^{1/2} \le LuS + K(\exp u/q)C\sqrt{N} \ , \tag{9.133}$$

so that (9.127) follows from (9.129) and the Cauchy-Schwarz inequality in the case $n\le n_1$.

Assume now that $n\ge n_1$. We then define $I=\{1,\cdots,N\}$ and (9.127) follows similarly, since then (9.132) and (9.133) now hold for $I=\{1,\ldots,N\}$.
$\square$

Proof of Proposition 9.3.19. For $f \in \mathcal{F}$ consider the sequences $(v_{i,f,n})$ and $(w_{i,f,n})$ obtained from Lemma 9.3.20 in the case $A = A_n(f)$ so that

$$\pi_n(f)(X_i)^2 - \pi_{n-1}(f)(X_i)^2 = v_{i,f,n} + w_{i,f,n} \;, \tag{9.134}$$

with

$$z_{f,n} := \sum_{i \le N} |v_{i,f,n}| \le Lu^2 \exp(u/q)(2^{n/2}\Delta^2_{A_n(f)} + 2^n \Delta^1_{A_n(f)})(S + C\sqrt{N}) \tag{9.135}$$

and

$$Z_{f,n} := \Big(\sum_{i \le N} w^2_{i,f,n}\Big)^{1/2} \le Lu^2 \exp(u/q)(\Delta^2_{A_n(f)} + \Delta^1_{A_n(f)})C\sqrt{N} \;. \tag{9.136}$$

Thus, given a parameter $v \ge 1$ the subgaussian inequality (3.2.2) implies

$$\mathsf{P}_\varepsilon\bigg(\Big|\sum_{i \le N} \varepsilon_i(\pi_n(f)(X_i)^2 - \pi_{n-1}(f)(X_i)^2)\Big| \ge 2v2^{n/2}Z_{f,n} + z_{f,n}\bigg) \le \exp(-v2^n),$$

and proceeding as usual it suffices to show that

$$\sum_{n \ge 1} 2^{n/2} Z_{f,n} + z_{f,n} \le Lu^2 S \exp(u/q)(S + \sqrt{N}C) \;.$$

But this follows from (9.88), (9.135) and (9.136). □

9.4 When Not to Use Chaining

In this section we work in the space $\mathbb{R}^n$ provided with the Euclidean norm $\|\cdot\|$. We denote by $\langle\cdot,\cdot\rangle$ the canonical duality of $\mathbb{R}^n$ with itself. We consider a sequence $(X_i)_{i \le N}$ of independent $\mathbb{R}^n$-valued random vectors and we assume that

$$\|x\| \le 1 \Rightarrow \mathsf{E} \exp|\langle X_i, x\rangle| \le 2 \tag{9.137}$$

and

$$\max_{i \le N} \|X_i\| \le (Nn)^{1/4} \;. \tag{9.138}$$

Theorem 9.4.1 ([1], [2]). *Assume $N \ge n$. Then with probability $\ge 1 - L\exp(-(Nn)^{1/4}) - L\exp(-n)$ we have*

$$\sup_{\|x\| \le 1} \Big|\sum_{i \le N} (\langle x, X_i\rangle^2 - \mathsf{E}\langle x, X_i\rangle^2)\Big| \le L\sqrt{nN} \;. \tag{9.139}$$

We refer the reader to [1] for various interpretations of this result.

Generally speaking, chaining is rarely the way to go when dealing with the unit ball B of $\mathbb{R}^n$. This is because (recalling the entropy numbers (2.34)) in the series $\sum_k 2^{k/2} e_k(B)$ (and in many others) the sum is basically equal to the largest term (as seen e.g from Exercise 2.2.13 (d)). Trying to use chaining is then an overkill, and simpler arguments yield shorter proofs. Being certain of this greatly helped to organize the proofs of [1] the way we present now. (And, again, knowing that the result is true greatly helped too.) This requires no new idea whatsoever compared to [1], and similar arguments were actually already given in [3]. We shall use Lemma 9.3.14 instead of chaining. The difficult part of the proof of Theorem 9.4.1 is the control of the random quantities

$$A_k := \sup_{\|x\|\le 1} \sup_{\operatorname{card} I \le k} \Big(\sum_{i\in I} \langle x, X_i\rangle^2\Big)^{1/2} . \tag{9.140}$$

Proposition 9.4.2. *For $u > 0$, with probability $\ge 1 - L\exp(-u)$ we have*

$$\forall\, k \ge 1\ ,\ A_k \le L\Big(u + \sqrt{k}\log\Big(\frac{eN}{k}\Big) + \max_{i\le N} \|X_i\|\Big) . \tag{9.141}$$

Corollary 9.4.3. *If $N \le n$ and $\max \|X_i\| \le u$, then with probability $\ge 1 - L\exp(-u)$ we have*

$$\sup_{\|x\|\le 1} \sum_{i\le N} \langle x, X_i\rangle^2 \le L(u^2 + N) .$$

This shows that the condition $N \ge n$ in Theorem 9.4.1 is not a restriction, since for $N \le n$ and $u = (Nn)^{1/4}$ we have $L(u^2 + N) \le L\sqrt{Nn}$.

Proof. Use (9.141) for $k = N$. □

We prepare for the proof of Proposition 9.4.2. We first observe the fundamental identity

$$A_k = \sup_{\|x\|\le 1} \sup_{\operatorname{card} I\le k} \sup_{\sum_{i\in I} a_i^2 \le 1} \sum_{i\in I} a_i \langle x, X_i\rangle = \sup_{\operatorname{card} I\le k} \sup_{\sum_{i\in I} a_i^2\le 1} \Big\|\sum_{i\in I} a_i X_i\Big\| . \tag{9.142}$$

We recall the notation $\varphi(k) = k\log(eN/k)$, so that the quantity $\sqrt{k}\log(eN/k)$ occurring in (9.141) is $\varphi(k)/\sqrt{k}$.

Lemma 9.4.4. *Consider $x \in \mathbb{R}^n$ with $\|x\| \le 1$ and an integer $1 \le k \le N$. Then for $u > 0$, with probability $\ge 1 - L(k/eN)^{3k}\exp(-u)$ the following occurs. For each set $I \subset \{1, \ldots, N\}$ with $\operatorname{card} I = m \ge k$ we have*

$$\sum_{i\in I} |\langle X_i, x\rangle| \le 6\varphi(m) + u . \tag{9.143}$$

Proof. Given a set I with card $I = m$, (9.137) implies $\mathsf{E}\exp\sum_{i\in I}|\langle X_i, x\rangle| \leq 2^m \leq \exp\varphi(m)$ and thus

$$\mathsf{P}\Big(\sum_{i\in I}|\langle X_i, x\rangle| \geq 6\varphi(m)+u\Big) \leq \exp(-5\varphi(m))\exp(-u) .$$

Since there at most $\exp\varphi(m)$ choices for I by (9.87) the union bound implies

$$\sum_{\operatorname{card} I\geq k}\mathsf{P}\Big(\sum_{i\in I}|\langle X_i, x\rangle| \geq 6\varphi(m)+u\Big) \leq \sum_{m\geq k}\exp(-4\varphi(m))\exp(-u) .$$

Now we observe that $\varphi(m) \geq \varphi(k)$ for $m \geq k$ and that $\varphi(m) \geq m$, so that

$$\sum_{m\geq k}\exp(-4\varphi(m)) \leq \exp(-3\varphi(k))\sum_{m\geq 1}\exp(-\varphi(m)) \leq L\exp(-3\varphi(k)) . \quad \square$$

We recall the sets $\mathcal{S}_{k,I}$ of (9.100).

Lemma 9.4.5. *With probability* $\geq 1 - L\exp(-u)$ *the following occurs. Consider disjoint subsets* I, J *of* $\{1,\ldots,N\}$ *with* card $I = m \geq$ card $J = k$, *and consider any* $a \in \mathcal{S}_{k,J}$. *Then*

$$\sum_{i\in I}\Big|\Big\langle X_i, \sum_{j\in J}a_jX_j\Big\rangle\Big| \leq (6\varphi(m)+u)\Big\|\sum_{j\in J}a_jX_j\Big\| . \tag{9.144}$$

Proof. Given J and $a \in \mathcal{S}_{k,J}$, the probability that (9.144) occurs for each choice of I of cardinality m and disjoint of J is at least $1-L(k/eN)^{3k}\exp(-u)$, as is shown by using Lemma 9.4.4, used given the r.v.s X_j for $j \in J$ and for $x = y/\|y\|$, $y = \sum_{j\in J}a_jX_j$. There are at most $\exp\varphi(k)$ choices of J of cardinality k and for each such J there are at most 5^k choices for a. Moreover

$$\sum_{k\leq N}(k/eN)^{3k}5^k\exp\varphi(k) = \sum_{k\leq N}(k/eN)^{2k}5^k \leq \sum_{k\geq 1}e^{-2k}5^k \leq L .$$

The result then follows from the union bound. $\square$

Corollary 9.4.6. *For* $u > 0$, *with probability* $\geq 1 - L\exp(-u)$ *the following occurs. Consider disjoint subsets* I, J *of* $\{1,\ldots,N\}$ *with* card $I = m \geq$ card $J = k$, *and consider any sequence* $(a_i)_{i\in J}$ *with* $\sum_{i\in J}a_i^2 \leq 1$. *Then*

$$\sum_{i\in I}\Big|\Big\langle X_i, \sum_{j\in J}a_jX_j\Big\rangle\Big| \leq L(\varphi(m)+u)A_k . \tag{9.145}$$

Proof. With probability $\geq 1 - L\exp(-u)$, (9.144) occurs for every choice of $a \in \mathcal{S}_{k,J}$. We prove that then (9.145) holds. Since $\sum_{j\in J}a_j^2 \leq 1$ for $a \in \mathcal{S}_{k,J}$, for each such sequence we then have

$$\sum_{i\in I}\Big|\Big\langle X_i, \sum_{j\in J}a_jX_j\Big\rangle\Big| \leq (6\varphi(m)+u)A_k .$$

Now, each sequence $(b_j)_{j\in J}$ with $\sum_{j\in J}b_j^2 \leq 1$ is in the convex hull of $2\mathcal{S}_{k,J}$, and this proves (9.145). $\square$

Proposition 9.4.7. *When the event of Corollary 9.4.6 occurs, we have*

$$\forall\, k \geq 1\ ,\ A_k^2 \leq \max_{i\leq N} \|X_i\|^2 + L(u+\varphi(k)/\sqrt{k})A_k\ . \tag{9.146}$$

Proof of Proposition 9.4.2. We use that $BA_k \leq (B^2+A_k^2)/2$ with $B = L(u+\varphi(k)/\sqrt{k})$ to deduce (9.141) from (9.146). □

Proof of Proposition 9.4.7. We fix once and for all an integer k. Consider a subset W of $\{1,\dots,N\}$ with $\operatorname{card} W = k$. Consider $(a_i)_{i\in W}$ with $\sum_{i\in W} a_i^2 \leq 1$. The plan is to bound

$$\Big\|\sum_{i\in W} a_i X_i\Big\|^2 = \sum_{i\in W} a_i^2\|X_i\|^2 + \sum_{i,j\in W, i\neq j} \langle a_iX_i, a_jX_j\rangle\ . \tag{9.147}$$

First, we use the obvious bound for the first term:

$$\sum_{i\in W} a_i^2\|X_i\|^2 \leq \max_{i\leq N}\|X_i\|^2\ . \tag{9.148}$$

For the second term, we use a standard "decoupling device". Consider independent Bernoulli r.v.s ε_i and observe that for $i\neq j$ we have $\mathsf{E}(1-\varepsilon_i)(1+\varepsilon_j) = 1$, so that by linearity of expectation, and denoting by E_ε expectation in the r.v.s ε_i only,

$$\sum_{i,j\in W, i\neq j}\langle a_iX_i, a_jX_j\rangle = \mathsf{E}_\varepsilon \sum_{i,j\in W, i\neq j}(1+\varepsilon_i)(1-\varepsilon_j)\langle a_iX_i, a_jX_j\rangle\ .$$

Given (ε_i) observe that if $I = \{i\in I; \varepsilon_i = 1\}$ and $J = W\setminus I$,

$$\begin{aligned}\frac{1}{4}\sum_{i,j\in W, i\neq j}(1+\varepsilon_i)(1-\varepsilon_j)\langle a_iX_i, a_jX_j\rangle &= \sum_{i\in I, j\in J}\langle a_iX_i, a_jX_j\rangle\\ &= \Big\langle \sum_{i\in I} a_iX_i, \sum_{j\in J} a_jX_j\Big\rangle\ . \end{aligned}\tag{9.149}$$

We now think of (a_i), I and J as fixed and we proceed to bound the right-hand side of (9.149), by suitably grouping the terms depending on the values of the coefficients (a_i). Let $\kappa = \operatorname{card} I$, and consider the largest integer ℓ_1 with $2^{\ell_1} \leq 2\operatorname{card} I = 2\kappa$. Let us enumerate $I = \{i_1,\dots,i_\kappa\}$ in such a way that the sequence $(|a_{i_s}|)_{1\leq s\leq\kappa}$ is non-increasing. For $0\leq \ell < \ell_1$, let $I_\ell = \{i_1,\dots,i_{2^\ell}\}$ and $\alpha_\ell = |a_{i_{2^\ell}}|$, so that

$$\alpha_\ell^2 2^\ell = \alpha_\ell^2 \operatorname{card} I_\ell \leq \sum_{i\in I_\ell} a_i^2 \leq 1\ ,$$

and thus

$$\alpha_\ell \leq 2^{-\ell/2}\ . \tag{9.150}$$

Let us set $I_{\ell_1} = I$ and for $1 \le \ell \le \ell_1$ let us set $I'_\ell = I_\ell \setminus I_{\ell-1}$, so that since $|a_i| \le \alpha_{\ell-1} \le 2^{-\ell/2+1}$ we have

$$i \in I'_\ell \Rightarrow |a_i| \le 2^{-\ell/2+1} . \tag{9.151}$$

We set $I'_0 = I_0$ so that (9.151) still holds for $\ell = 0$. The sets I'_ℓ for $0 \le \ell \le \ell_1$ form a partition of I.

For $0 \le \ell \le \ell_1$ let us set $y_\ell = \sum_{i \in I'_\ell} a_i X_i$, so that $\sum_{i \in I} a_i X_i = \sum_{0 \le \ell \le \ell_1} y_\ell$. Then for each vector x and each $0 \le \ell \le \ell_1$ we have

$$|\langle y_\ell, x\rangle| = \Big|\Big\langle \sum_{i \in I'_\ell} a_i X_i, x \Big\rangle\Big| \le \sum_{i \in I'_\ell} |a_i||\langle X_i, x\rangle| \le 2^{-\ell/2+1} \sum_{i \in I'_\ell} |\langle X_i, x\rangle| . \tag{9.152}$$

Let us then define similarly for $0 \le \ell \le \ell_2$ sets $J_\ell \subset J$ with $\operatorname{card} J_\ell = 2^\ell$ for $\ell < \ell_2$, sets J'_ℓ and elements $z_\ell = \sum_{j \in J'_\ell} a_j X_j$ so that $\sum_{j \in J} a_j X_j = \sum_{0 \le \ell \le \ell_2} z_\ell$. Without loss of generality we assume $\operatorname{card} I \ge \operatorname{card} J$. We write

$$\Big\langle \sum_{i \in I} a_i X_i, \sum_{j \in J} a_j X_j \Big\rangle = \Big\langle \sum_{0 \le \ell \le \ell_1} y_\ell, \sum_{0 \le \ell' \le \ell_2} z_{\ell'} \Big\rangle = \mathrm{I} + \mathrm{II} , \tag{9.153}$$

where

$$\mathrm{I} = \sum_{0 \le \ell \le \ell_1} \Big\langle y_\ell, \sum_{0 \le \ell' \le \ell} z_{\ell'} \Big\rangle \ ; \ \mathrm{II} = \sum_{0 \le \ell' \le \ell_2} \Big\langle \sum_{0 < \ell < \ell'} y_\ell, z_{\ell'} \Big\rangle .$$

This identity is obvious if we observe that $\mathrm{I} = \sum_{\ell' \le \ell} \langle y_\ell, z_{\ell'}\rangle$ and $\mathrm{II} = \sum_{\ell' > \ell} \langle y_\ell, z_{\ell'}\rangle$. We bound I. First we use (9.152) to obtain

$$\Big|\Big\langle y_\ell, \sum_{0 \le \ell' \le \ell} z_{\ell'} \Big\rangle\Big| \le 2^{-\ell/2+1} \sum_{i \in I_\ell} \Big|\Big\langle X_i, \sum_{0 \le \ell' \le \ell} z_{\ell'} \Big\rangle\Big| .$$

The key point is that $\sum_{0 \le \ell' \le \ell} z_{\ell'} = \sum_{i \in J_\ell} a_i X_i$, so that we may use (9.145) for I_ℓ and J_ℓ, and $\operatorname{card} I_\ell = \min(2^\ell, \operatorname{card} I) \ge \min(2^\ell, \operatorname{card} J) = \operatorname{card} J_\ell$ to obtain (recalling that $k = \operatorname{card} W \ge \operatorname{card} J$ so that $A_{\operatorname{card} J} \le A_k$)

$$\sum_{i \in I_\ell} \Big|\Big\langle X_i, \sum_{0 \le \ell' \le \ell} z_{\ell'} \Big\rangle\Big| \le L(u + \varphi(2^\ell)) A_k ,$$

so that

$$\mathrm{I} \le L A_k \sum_{0 \le \ell \le \ell_1} 2^{-\ell/2} (u + \varphi(2^\ell)) .$$

It is elementary to prove that the function $x \mapsto x^{1/4} \log(eN/x)$ increases for $1 \le x \le Ne^{-3}$. This implies that for $1 \le x \le y \le N$ we have

$$\sqrt{x} \log(eN/x) \le L(xy)^{1/4} \log(eN/y) , \tag{9.154}$$

since indeed this holds true with $L = 1$ when $y \le Ne^{-3}$. This inequality clearly remains true for $x \le 2y$ rather than $x \le y$. Consequently, using

(9.154) for $y = k$ and $x = 2^\ell \le 2^{\ell_1} \le 2\kappa \le 2k \le 2y$, for $\ell \le \ell_1$ we have $2^{\ell/2}\log(eN/2^\ell) \le L2^{\ell/4}k^{1/4}\log(eN/k)$ and thus

$$\sum_{0\le\ell\le\ell_1} 2^{-\ell/2}\varphi(2^\ell) = \sum_{0\le\ell\le\ell_1} 2^{\ell/2}\log(eN/2^\ell) \le L\sqrt{k}\log(eN/k) = L\varphi(k)/\sqrt{k}\ ,$$

so we finally obtain that $\mathrm{I} \le L(u + \varphi(k)/\sqrt{k})A_k$. The same argument proves that this bound also holds for II (using now that $\operatorname{card} I_{\ell'-1} \le 2^{\ell'-1} \le \operatorname{card} J_{\ell'}$ if $\ell' \le \ell_2$).

So, we have proved that the right-hand side of (9.146) bounds the left-hand side of (9.147), irrelevant of the choice of W with $\operatorname{card} W = k$ and of the $(a_i)_{i\in W}$. Recalling (9.142) this completes the proof. □

We complete the proof of Theorem 9.4.1 by reproducing the arguments of [2] for the convenience of the reader. We consider a subset U of $\mathbb{R}^n$ with $\operatorname{card} U \le 5^n$, consisting of elements of norm ≤ 1 and such that its convex hull contains the ball of $\mathbb{R}^n$ centered at the origin with radius 1/2. Thus

$$\begin{aligned}&\sup_{\|x\|,\|y\|\le 1/2}\Big|\sum_{i\le N}(\langle x,X_i\rangle\langle y,X_i\rangle - \mathsf{E}\langle x,X_i\rangle\langle y,X_i\rangle)\Big| \\ &\le \sup_{x,y\in U}\Big|\sum_{i\le N}(\langle x,X_i\rangle\langle y,X_i\rangle - \mathsf{E}\langle x,X_i\rangle\langle y,X_i\rangle)\Big|\ . \end{aligned} \tag{9.155}$$

The plan is to assume that

$$\forall\, k\ge 1\ ,\ A_k \le L(\sqrt{k}\log(eN/k) + (Nn)^{1/4})\ , \tag{9.156}$$

and to prove that then with probability $\ge 1-\exp(-n)$ the right-hand side of (9.155) is $\le L\sqrt{Nn}$. This complete the proof of Theorem 9.4.1 because Proposition 9.4.2 and (9.138) show that (9.156) occurs with probability $\ge 1 - L\exp(-(Nn)^{1/4})$.

Consider a truncation level $B \ge 0$ and define

$$Z_i(x,y) = \langle x,X_i\rangle\langle y,X_i\rangle\mathbf{1}_{\{|\langle x,X_i\rangle\langle y,X_i\rangle|\le B\}}$$

and

$$Y_i(x,y) = \langle x,X_i\rangle\langle y,X_i\rangle\mathbf{1}_{\{|\langle x,X_i\rangle\langle y,X_i\rangle|> B\}}\ ,$$

so that $\langle x,X_i\rangle\langle y,X_i\rangle = Z_i(x,y) + Y_i(x,y)$. (This argument is yet another instance of a decomposition in a "spread part" and a "picky part". The peaky part will be controlled as usual without using cancellations, i.e. we will control $\sum_i |Y_i(x,y)|$.) We bound the right-hand side of (9.155) by $\mathrm{I}+\mathrm{II}+\mathrm{III}$, where

$$\mathrm{I} = \sup_{x,y\in U}\Big|\sum_{i\le N}(Z_i(x,y) - \mathsf{E}Z_i(x,y))\Big|\ , \tag{9.157}$$

$$\mathrm{II} = \sup_{x,y\in U} \sum_{i\le N} |Y_i(x,y)| \;, \tag{9.158}$$

$$\mathrm{III} = \sup_{x,y\in U} \sum_{i\le N} |\mathsf{E} Y_i(x,y)| \;. \tag{9.159}$$

The fun is to bound II. Let us fix $x, y \in U$ and set

$$I = \{i \le N \; ; \; |Y_i(x,y)| > B\} = \{i \le N \; ; \; |Y_i(x,y)| \ne 0\} \; .$$

Thus, if $m = \operatorname{card} I$ we have, using the Cauchy-Schwarz inequality in the second inequality

$$mB \le \sum_{i\in I} |Y_i(x,y)| \le \Big(\sum_{i\in I} \langle x, X_i\rangle^2\Big)^{1/2} \Big(\sum_{i\in I} \langle y, X_i\rangle^2\Big)^{1/2} \le A_m^2 \;, \tag{9.160}$$

and thus from (9.156)

$$mB \le L_1(m(\log(eN/m))^2 + \sqrt{Nn}) \; . \tag{9.161}$$

We shall use this to bound m. Without loss of generality we assume from Corollary 9.4.3 that $N > n$ and then $N > \sqrt{Nn}$. Thus we may consider the smallest integer $k_0 \le N$ such that $k_0(\log(eN/k_0))^2 > \sqrt{Nn}$. Let us now choose $B = 2L_1(\log(eN/k_0))^2$, so that if $m \ge k_0$ then (9.161) implies

$$\begin{aligned} 2L_1 m(\log(eN/k_0))^2 = mB &\le L_1(m(\log(eN/m))^2 + \sqrt{Nn}) \\ &\le L_1(m(\log(eN/k_0))^2 + \sqrt{Nn}) \;, \end{aligned}$$

and thus

$$m(\log(eN/k_0))^2 \le \sqrt{Nn} \; .$$

Since $m \ge k_0$ this is impossible by definition of k_0, so that $m < k_0$. By definition of k_0, we have $m(\log(eN/m))^2 \le \sqrt{Nn}$ and thus by (9.156) that $A_m \le L(Nn)^{1/4}$, and finally by (9.160) and since $\sum_{i\in I} |Y_i(x,y)| = \sum_{i\le N} |Y_i(x,y)|$ that $\mathrm{II} \le L\sqrt{Nn}$.

Next, since always $\mathrm{II} \le \sum_{i\le N} \|X_i\|^2 \le N\sqrt{Nn}$ and since $\mathrm{II} \le L\sqrt{Nn}$ when (9.156) occurs, i.e. with probability $\ge 1 - \exp(-(Nn)^{1/4})$, we have $\mathrm{III} \le \mathsf{E}\,\mathrm{II} \le L\sqrt{Nn}$.

Finally, since $(\log x)^2 \le L\sqrt{x}$ for $x \ge e$,

$$\sqrt{Nn} < k_0(\log(eN/k_0))^2 \le Lk_0\sqrt{N/k_0}$$

and thus $k_0 \ge n/L$. Therefore, with huge room to spare,

$$B = 2L_1(\log(eN/k_0))^2 \le L\sqrt{N/n} \; .$$

Since $|Z_i(x,y)| \le B$ and $\mathsf{E} Z_i(x,y)^2 \le L$ (using the Cauchy-Schwarz inequality and (9.137)), it follows from Bernstein's inequality (4.59) that

$$\mathsf{P}\Big(\Big|\sum_{i\le N}(Z_i(x,y)-\mathsf{E}Z_i(x,y))\Big|\ge t\Big)\le 2\exp\Big(-\min\Big(\frac{t^2}{LN},\frac{t}{L\sqrt{N/n}}\Big)\Big)\,.$$

The right-hand side is $\le 5^{-3n}$ for $t = L\sqrt{Nn}$. There are at most 5^{2n} choices for the pair $(x,y)\in U^2$, so that by the union bound $\mathrm{I}\le L\sqrt{Nn}$ with probability $\ge 1-\exp(-n)$. This completes the proof of Theorem 9.4.1.

9.5 Notes and Comments

The paper [10] started the study of the left-hand side of (9.43) and proved the bound (9.66). This paper brings out in particular the importance of the control of $\sup_{f\in\mathcal{F}}(\sum_{i\le N} f(X_i)^2)^{1/2}$, and apparently is the first which contains a version of Proposition 9.3.16. A crucial step is performed in [9], which deals again with the case $d_1 = 0$, and proves now (9.43) with $\Delta(T, d_{\psi_1})$ instead of C, where d_{ψ_1} denotes of course the distance associated to the Orlicz norm of (9.58). The point is that this quantity is often significantly smaller than $\Delta(T, d_2)$ (which basically coincides with $\Delta(T, d_{\psi_2})$ under (9.42)). The paper [9] also makes a further step in the direction of Proposition 9.3.16.

Many of the ideas of the proof of Theorem 9.4.1 go back to a seminal paper of J. Bourgain [4]. It would be nice if one could deduce this theorem from a general principle such as Theorem 9.3.1, but unfortunately we do not know how to do this, even when the sequence (X_i) is i.i.d.

References

1. Adamczak, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. J. Am. Math. Soc. **23**(2), 535–561 (2010)
2. Adamczak, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Sharp bounds on the rate of convergence of the empirical covariance matrix. C. R. Math. Acad. Sci. Paris **349**(3–4), 195–200 (2011)
3. Adamczak, R., Latała, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. Ann. Inst. Henri Poincaré Probab. Stat. **48**(4), 1103–1136 (2012)
4. Bourgain, J.: Random points in isotropic convex sets. In: Convex Geometric Analysis, Berkeley, CA, 1996. Math. Sci. Res. Inst. Publ., vol. 34, pp. 53–58. Cambridge Univ. Press, Cambridge (1999)
5. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics, vol. 63. Cambridge University Press, Cambridge (1999)
6. Giné, E., Zinn, J.: Some limit theorems for empirical processes. Ann. Probab. **12**, 929–989 (1984)
7. Klartag, B., Mendelson, S.: Empirical processes and random projections. J. Funct. Anal. **225**(1), 229–245 (2005)

8. Latała, R.: Weak and strong moments of random vectors. In: Marcinkiewicz Centenary Volume. Banach Center Publ., vol. 95, pp. 115–121. Polish Acad. Sci. Inst. Math, Warsaw (2011)
9. Mendelson, S.: Empirical processes with a bounded ψ_1-diameter. Geom. Funct. Anal. **20**(4), 988–1027 (2010)
10. Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.: Reconstruction and subgaussian operators in asymptotic geometric analysis. Geom. Funct. Anal. **17**(4), 1248–1282 (2007)
11. Mendelson, S., Paouris, G.: On generic chaining and the smallest singular value of random matrices with heavy tails. J. Funct. Anal. **262**(9), 3775–3811 (2012)
12. Talagrand, M.: The Glivenko-Cantelli problem. Ann. Probab. **15**, 837–870 (1987)
13. Talagrand, M.: Regularity of infinitely divisible processes. Ann. Probab. **21**, 362–432 (1993)

10. Partition Scheme for Families of Distances

10.1 The Partition Scheme

The goal of the present section is to generalize to the setting of "families of distances" the partitioning scheme of Section 2.7. In Section 10.2 we will apply this tool to the study of "canonical processes".

We consider a family of maps $(\varphi_j)_{j\in\mathbb{Z}}$, with the following properties:

$$\varphi_j : T \times T \to \mathbb{R}^+ \cup \{\infty\}\,,\ \varphi_j \geq 0\,,\ \varphi_j(s,t) = \varphi_j(t,s)\ .$$

These maps play the role of a family of distances (although it probably would be better to think of φ_j as the square of a distance rather than as of a distance).

We recall that a functional F on a set T is a non-increasing map from the subsets of T to $\mathbb{R}$. We consider functionals $F_{n,j}$ on T for $n \geq 0, j \in \mathbb{Z}$. We assume

$$F_{n+1,j} \leq F_{n,j}\ ;\ F_{n,j+1} \leq F_{n,j}\ . \tag{10.1}$$

We define

$$B_j(t,c) = \{s \in T\ ;\ \varphi_j(s,t) \leq c\}\ .$$

We will assume that the functionals satisfy a "growth condition", that is very similar in spirit to Definition 2.7.1. This condition involves as main parameter an integer $\kappa \geq 4$. We set $r = 2^{\kappa-2}$, so that $r \geq 4$. The role of r is as in (2.73), the larger r, the weaker the growth condition. The reason why we take r of the type $r = 2^{\kappa-2}$ for an integer κ is purely technical convenience.

The growth condition, that also involves as secondary parameter an integer $n_0 \geq 1$, is as follows.

Definition 10.1.1. *We say that the functionals $F_{n,j}$ satisfy the* growth condition *(for n_0 and r) if the following occurs. Consider any $j \in \mathbb{Z}$, any $n \geq n_0$ and $m = N_n$. Consider any sets $(H_\ell)_{1\leq\ell\leq m}$ that are separated in the following sense: there exist points $t, t_1, \ldots, t_m$ in T for which*

$$\forall \ell \leq m\,,\ t_\ell \in B_j(t, 2^n)\ , \tag{10.2}$$

$$\forall \ell, \ell' \leq m\,,\ \ell \neq \ell'\,,\ \varphi_{j+1}(t_\ell, t_{\ell'}) > 2^{n+1} \tag{10.3}$$

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_10,

and $H_\ell \subset B_{j+2}(t_\ell, 2^{n+\kappa})$. *Then*

$$F_{n,j}\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge 2^n r^{-j-1} + \min_{\ell \le m} F_{n+1,j+1}(H_\ell)\ . \tag{10.4}$$

We have not made assumptions on how φ_j relates to φ_{j+1}; but we have little chance to prove (10.4) unless $B_{j+2}(t_\ell, 2^{n+\kappa})$ is quite smaller than $B_{j+1}(t_\ell, 2^{n+1})$.

To understand the preceding conditions we will carry out the case where

$$\varphi_j(s,t) = r^{2j} d(s,t)^2 \tag{10.5}$$

for a distance d on T. The reader is encouraged to carry out the more general case where $\varphi_j(s,t) = r^{\alpha j} d(s,t)^\beta$ for $\alpha, \beta > 0$. Denoting by $B(t,b)$ the ball for d of center t and radius b, we thus have

$$B_j(t,c) = B(t, r^{-j}\sqrt{c})\,.$$

Thus in (10.3) we require that

$$\forall \ell, \ell' \le m\,,\ \ell \ne \ell'\,,\ d(t_\ell, t_{\ell'}) \ge 2^{(n+1)/2} r^{-j-1} := a\ . \tag{10.6}$$

On the other hand, the condition $H_\ell \subset B_{j+2}(t_\ell, 2^{n+\kappa})$ means

$$H_\ell \subset B(t_\ell, 2^{(n+\kappa)/2} r^{-j-2}) = B(t_\ell, \eta a)\,,$$

for $\eta = 2^{(\kappa-1)/2}/r = \sqrt{2}/\sqrt{r}$. Thus, as r gets larger, η gets smaller, and the sets H_ℓ become better separated. Also, (10.4) reads as

$$F_{n,j}\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge 2^{(n-1)/2} a + \min_{\ell \le m} F_{n+1,j+1}(H_\ell)\ ,$$

which strongly resembles (2.147) for $\theta(n) = 2^{n/2}$ and $\beta = 1$.

Theorem 10.1.2. *Assume that the functionals* $F_{n,j}$ *are as above, and in particular satisfy the growth condition of Definition 10.1.1, and that, for some* $j_0 \in \mathbb{Z}$ *we have*

$$\forall s,t \in T\ ,\ \varphi_{j_0}(s,t) \le 2^{n_0}\ . \tag{10.7}$$

Then there exists an admissible sequence $(\mathcal{A}_n)$ *and for each* $A \in \mathcal{A}_n$ *an integer* $j_n(A) \in \mathbb{Z}$ *such that*

$$A \in \mathcal{A}_n\,,\ C \in \mathcal{A}_{n-1}\,,\ A \subset C \Rightarrow j_{n-1}(C) \le j_n(A) \le j_{n-1}(C) + 1 \tag{10.8}$$

$$\forall t \in T\ ,\ \sum_{n \ge n_0} 2^n r^{-j_n(A_n(t))} \le L(rF_{n_0,j_0}(T) + 2^{n_0} r^{-j_0}) \tag{10.9}$$

$$\forall n \ge n_0\,,\ \forall A \in \mathcal{A}_n\,,\ \exists t_{n,A} \in T\,,\ A \subset B_{j_n(A)}(t_{n,A}, 2^n)\ . \tag{10.10}$$

To make sense out of this, we again carry out the case (10.5). Then (10.7) means that $\Delta(T,d) \le r^{-j_0}2^{n_0/2}$, while (10.10) implies $A \subset B(r^{-j_n(A)}2^{n/2})$ and hence $\Delta(A,d) \le r^{-j_n(A)}2^{n/2+1}$. Moreover (10.9) implies

$$\forall t \in T\,,\ \sum_{n\ge n_0} 2^{n/2}\Delta(A_n(t),d) \le L(rF_{n_0,j_0}(T) + 2^{n_0}r^{-j_0})\,.$$

Taking for j_0 the largest integer such that $\Delta(T,d) \le r^{-j_0}2^{n_0/2}$, we get

$$\forall t \in T\,,\ \sum_{n\ge n_0} 2^{n/2}\Delta(A_n(t),d) \le Lr\big(F_{n_0,j_0}(T) + 2^{n_0/2}\Delta(T,d)\big)\,.$$

This relation resembles the relation one gets by combining (2.149) with Lemma 2.3.5, and the parameter n_0 plays a role similar to τ.

The proof of Theorem 10.1.2 follows closely the proof of Theorem 2.7.2, and of course the reader should master this latter result before attempting to read it. The proof relies on the following, where again the functionals are as above.

Lemma 10.1.3 (The Decomposition Lemma). *Consider a set $C \subset T$ and assume that for some $t_C \in T$ and some integers $j \in \mathbb{Z}$ and $n \ge n_0$ we have $C \subset B_j(t_C, 2^n)$. Then we can find a partition $(A_\ell)_{\ell \le m'}$ of C, where $m' \le m = N_n$, such that for each $\ell \le m'$ we have* **either**

$$\exists t_\ell \in C\,,\ A_\ell \subset B_{j+1}(t_\ell, 2^{n+1}) \tag{10.11}$$

or else

$$2^{n-1}r^{-j-1} + \sup_{t\in A_\ell} F_{n+1,j+1}(A_\ell \cap B_{j+2}(t, 2^{n+\kappa})) \le F_{n,j}(C)\,. \tag{10.12}$$

Proof. We consider $\epsilon = 2^{n-1}r^{-j-1}$. We set $D_0 = C$. First we choose t_1 in D_0 with

$$F_{n+1,j+1}(D_0 \cap B_{j+2}(t_1, 2^{n+\kappa})) \ge \sup_{t\in D_0} F_{n+1,j+1}(D_0 \cap B_{j+2}(t, 2^{n+\kappa})) - \epsilon\,.$$

We then set $A_1 = D_0 \cap B_{j+1}(t_1, 2^{n+1})$ and $D_1 = D_0 \backslash A_1$. If D_1 is not empty, we choose t_2 in D_1 such that

$$F_{n+1,j+1}(D_1 \cap\ B_{j+2}(t_2, 2^{n+\kappa})) \ge \sup_{t\in D_1} F_{n+1,j+1}(D_1 \cap B_{j+2}(t, 2^{n+\kappa})) - \epsilon\,,$$

and we set $A_2 = D_1 \cap B_{j+1}(t_2, 2^{n+1})$ and $D_2 = D_1 \backslash A_2$. We continue in this manner until either we exhaust C or we construct D_{m-1}. In the latter case we set $A_m = D_{m-1}$ and we stop the construction.

It is obvious by construction that if $\ell < m$ then A_ℓ satisfies (10.11), so that to conclude the proof we show that $A = A_m = D_{m-1}$ satisfies (10.12). Consider $1 \le \ell < m$. By construction of t_ℓ we have, since $A = D_{m-1} \subset D_{\ell-1}$,

$$\begin{aligned}\forall t \in D_{\ell-1},\ & F_{n+1,j+1}(A \cap B_{j+2}(t, 2^{n+\kappa})) \\ & \le F_{n+1,j+1}(D_{\ell-1} \cap B_{j+2}(t, 2^{n+\kappa})) \\ & \le F_{n+1,j+1}(D_{\ell-1} \cap B_{j+2}(t_\ell, 2^{n+\kappa})) + \epsilon\ . \end{aligned} \tag{10.13}$$

Consider $t \in A$ and set $H_\ell = D_{\ell-1} \cap B_{j+2}(t_\ell, 2^{n+\kappa})$ for $1 \le \ell < m$ and $H_m = A \cap B_{j+2}(t, 2^{n+\kappa})$. By (10.13), for $\ell < m$ we have $F_{n+1,j+1}(H_m) \le F_{n+1,j+1}(H_\ell) + \epsilon$ and thus

$$\inf_{\ell \le m} F_{n+1,j+1}(H_\ell) \ge F_{n+1,j+1}(H_m) - \epsilon\ .$$

Define $t_m = t$. We have $\varphi_{j+1}(t_\ell, t_{\ell'}) > 2^{n+1}$ for $\ell \ne \ell'$, and $t_\ell \in C \subset B_j(t_C, 2^n)$, so that (10.4) implies

$$\begin{aligned} F_{n,j}(C) \ge F_{n,j}\Big(\bigcup_{\ell \le m} H_\ell\Big) &\ge 2^n r^{-j-1} + F_{n+1,j+1}(H_m) - \epsilon \\ &= 2^n r^{-j-1} + F_{n+1,j+1}(A \cap B_{j+2}(t, 2^{n+\kappa})) - \epsilon \end{aligned}$$

and recalling the choice of ϵ this proves that (10.12) holds for $A = A_m$. □

Proof of Theorem 10.1.2. Together with $C \in \mathcal{A}_n$, for $n \ge n_0$ we construct an integer $j_n(C) \in \mathbb{Z}$ and a point $t_{n,C} \in T$ satisfying the following condition, where $j = j_n(C)$:

$$C \subset B_j(t_{n,C}\,,\, 2^n)\ . \tag{10.14}$$

To start the construction we pick an arbitrary point $t_{0,T} \in T$ and we define $\mathcal{A}_{n_0} = \{T\}$ and $j_{n_0}(T) = j_0$. Thus (10.14) holds by (10.7).

To construct $\mathcal{A}_{n+1}$ once $\mathcal{A}_n$ has been constructed, to each element C of $\mathcal{A}_n$ we apply the Decomposition Lemma with $j = j_n(C)$ to split C in $m' \le N_n$ pieces $A_1, \ldots, A_{m'}$. (Thus, since $N_n^2 \le N_{n+1}$, $\mathcal{A}_{n+1}$ contains at most N_{n+1} sets.) Let A be one of these sets.

When A satisfies (10.12), we set $j_{n+1}(A) = j = j_n(C)$ and $t_{n+1,A} = t_{n,C}$, so that (10.14) for A follows from the same relation for C. On the other hand, when A satisfies (10.11), we define $j_{n+1}(A) = j + 1$ and $t_{n+1,A} = t_\ell$. Then (10.14) for A follows from (10.11). Thus (10.10) holds, and since (10.8) holds by construction, it remains only to prove (10.9).

Let us first observe that by construction, the following holds:

$$A \in \mathcal{A}_{n+1}\ ,\ C \in \mathcal{A}_n\ ,\ A \subset C\ ,\ j_{n+1}(A) = j_n(C) + 1 \Rightarrow t_{n+1,A} \in C\ . \tag{10.15}$$

Let us fix once and for all a point $u \in T$ and to lighten notation let $j(n) = j_n(A_n(u))$. Let $a_n = 2^n r^{-j(n)}$. Consider the set

$$J = \{n_0\} \cup \big\{m > n_0\ ;\ j(m-1) = j(m)\ ,\ j(m+1) = j(m) + 1\big\}\ .$$

Consider $m \in J$, $m > n_0$ and $n = m - 1$. Then $j(m-1) = j(n)$, i.e. $j(n+1) = j(n)$, and it follows by construction that when we split $C = A_n(u)$

according to the Decomposition Lemma, $A_m(u) = A_{n+1}(u)$ is a piece A_ℓ that satisfies (10.12), i.e.

$$\frac{1}{4r}a_m + \sup_{t\in A_m(u)} F_{m,j(m)+1}(A_m(u)\cap B_{j(m)+2}(t,2^{m+\kappa-1})) \\ \le F_{m-1,j(m)}(A_{m-1}(u)) \ . \tag{10.16}$$

Next, we prove that $\sum_{n\ge n_0} a_n \le 4\sum_{n\in J} a_n$. To see this we simply observe that if we define

$$I_0 = \{n_0\} \cup \{k > n_0 \ ; \ \forall\, n \ge n_0 \ , \ n \ne k \ , \ a_n < a_k 2^{|k-n|}\} \ ,$$

then Lemma 2.6.3 used for $\alpha = 2$ implies $\sum_{n\ge n_0} a_n \le 4\sum_{n\in I_0} a_n$, while obviously (as in (2.140)) $I_0 \subset J$. Next, we apply again Lemma 2.6.3 to the sequence $(a_j)_{j\in J}$. That is, if we enumerate J as a sequence $(n_\ell)_{\ell\ge 0}$, we define

$$I = \{n_0\} \cup \{n_k \ , \ k \ge 1 \ ; \ \forall\, s > 1 \ , \ s \ne k \ , \ a_{n_s} < a_{n_k} 2^{|s-k|}\} \ .$$

Then Lemma 2.6.3 implies $\sum_{n\in J} a_n \le 4\sum_{n\in I} a_n$. The point of this construction is as follows. Consider $m \in I$, with $m > n_0$, and the largest p for which $j(p) = j(m+1)$, so that $p \in J$. Moreover, since $m \in I$ we have $a_p < 2a_m$, which means $2^p r^{-j(p)} < 2^{m+1} r^{-j(m)}$ and since $j(p) = j(m)+1$ and $r = 2^{\kappa-2}$ this means that $p+1 \le m+\kappa-1$. Let $A^* = A_{p+1}(u)$. Since $j(p+1) = j(p)+1$ it follows from (10.15) that $t^* := t_{p+1,A^*} \in A_p(u) \subset A_m(u)$. Now, (10.10) implies

$$A^* = A_{p+1}(u) \subset B_{j(p+1)}(t^*, 2^{p+1}) \\ = B_{j(m)+2}(t^*, 2^{p+1}) \subset B_{j(m)+2}(t^*, 2^{m+\kappa-1}) \ ,$$

so that

$$A_{p+1}(u) = A^* \subset A_m(u) \cap B_{j(m)+2}(t^*, 2^{m+\kappa-1})$$

and (10.16) implies

$$\frac{1}{4r}a_m + F_{m,j(m)+1}(A_{p+1}(u)) \le F_{m-1,j(m)}(A_{m-1}(u)) \ . \tag{10.17}$$

For $m \ge n_0$ let us define $x(m) = F_{m-1,j(m)}(A_{m-1}(u))$. This sequence is decreasing from (10.1). Since $m \le p+1$ and $j(m)+1 \le j(p+2)$, (10.1) implies $F_{m,j(m)+1}(A_{p+1}(u)) \ge x(p+2)$, so that for $m \in I$

$$\frac{1}{4r}a_m \le x(m) - x(p+2) \ . \tag{10.18}$$

If we number the elements of $I \setminus \{n_0\}$ as an sequence $(m(i))_{i\ge 1}$, when $m = m(i)$ then $p \le m(i+1)$ and therefore $p+2 \le m(i+3)$. Thus

$$\frac{1}{4r}a_{m(i)} \le x(m(i)) - x(m(i+3)) \ .$$

Summation of these relations, we obtain

$$\sum_{i\geq 1} a_{m(i)} \leq Lrx(m(1)) \leq LrF_{n_0,j_0} \,.$$

Since $j_{n_0}(T) = j_0$, we have $a_{n_0} = 2^{-n_0} r^{-j_0}$ and therefore

$$\sum_{i\in I} a_n \leq L(rF_{n_0,j_0} + 2^{-n_0} r^{-j_0}) \,.$$

This completes the proof. □

10.2 The Structure of Certain Canonical Processes

In this section we prove a far reaching generalization of Theorem 2.4.1. We consider independent, centered, symmetric r.v. $(Y_i)_{i\geq 1}$. We assume that

$$U_i(x) = -\log P(|Y_i| \geq x) \tag{10.19}$$

is convex. Since it is a matter of normalization, we assume that $U_i(1) = 1$. Since $U_i(0) = 0$ we then have $U_i'(1) \geq 1$.

Given $t = (t_i)_{i\geq 1} \in \ell^2$, we define

$$X_t = \sum_{i\geq 1} t_i Y_i \,.$$

The condition $t \in \ell^2$ is to ensure the convergence of the series. (Very little of the results we will present is lost if one assumes that only finitely many of the coefficients t_i are not 0). The aim of this section is to study collections of such r.v. as t varies over a set T, and in particular "to compute $\mathsf{E}\sup_{t\in T} X_t$ as a function of the geometry of T". The case where $U_i(x) = x^p$ for a certain $p \geq 1$ was obtained by this author in [3] and we owe the present more general setting to a further effort by R. Latała [1]. It is in truth a rather amazing fact that this can be done at all at the present level of generality. Even the question of understanding precisely the size of the tails of one single r.v X_t is far from obvious. The definitions we are going to introduce represent the outcome of many steps of abstraction, and the ideas behind them can be understood only gradually.

A first idea "is to redefine the function U_i as x^2 for $0 \leq x \leq 1$." In order to preserve convexity, we consider the function $\hat{U}_i(x)$ (defined on all $\mathbb{R}$) given by

$$\hat{U}_i(x) = \begin{cases} x^2 \text{ if } 0 \leq |x| \leq 1 \\ 2U_i(|x|) - 1 \text{ if } |x| \geq 1 \,, \end{cases} \tag{10.20}$$

so that this function is convex.

Given $u > 0$, we define

$$\mathcal{N}_u(t) = \sup\Big\{\sum_{i\geq 1} t_i a_i \; ; \; \sum_{i\geq 1} \hat{U}_i(a_i) \leq u\Big\}$$

and

$$B(u) = \{t \; ; \; \mathcal{N}_u(t) \leq u\} \, .$$

Given a number r, we further define

$$\varphi_j(s,t) = \inf\{u > 0 \; ; \; s - t \in r^{-j} B(u)\} \tag{10.21}$$

when the set in the right-hand side is not empty and $\varphi_j(s,t) = \infty$ otherwise.

To get a feeling of what happens let us first carry out the meaning of these definitions in simple cases. The simplest case is when $U_i(x) = x^2$ for all i. It is rather immediate then that

$$x^2 \leq \hat{U}_i(x) \leq 2x^2 \; ; \; \sqrt{\frac{u}{2}}\|t\|_2 \leq \mathcal{N}_u(t) \leq \sqrt{u}\|t\|_2,$$

so that $B_2(0,\sqrt{u}) \subset B(u) \subset B_2(0,\sqrt{2u})$, where B_2 denotes the ball of ℓ^2, and

$$\frac{1}{2} r^{2j}\|s-t\|_2^2 \leq \varphi_j(s,t) \leq r^{2j}\|s-t\|_2^2 \, , \tag{10.22}$$

so we are basically in the situation of (10.5).

The second simplest example is the case where for all i we have $U_i(x) = x$ for $x \geq 0$. In that case we have $|x| \leq \hat{U}_i(x) = 2|x| - 1 \leq x^2$ for $|x| \geq 1$. Thus $\hat{U}_i(x) \leq x^2$ and $\hat{U}_i(x) \leq 2|x|$ for all $x \geq 0$, and hence

$$\sum_{i\geq 1} a_i^2 \leq u \Rightarrow \sum_{i\geq 1} \hat{U}_i(a_i) \leq u$$

and

$$\sum_{i\geq 1} 2|a_i| \leq u \Rightarrow \sum_{i\geq 1} \hat{U}_i(a_i) \leq u \, .$$

Consequently, we have $\mathcal{N}_u(t) \geq \sqrt{u}\|t\|_2$ and $\mathcal{N}_u(t) \geq u\|t\|_\infty/2$. Moreover, if $\sum_{i\geq 1} \hat{U}_i(a_i) \leq u$, writing $b_i = a_i \mathbf{1}_{\{|a_i|\geq 1\}}$ and $c_i = a_i \mathbf{1}_{\{|a_i|<1\}}$ we have $\sum_{i\geq 1} |b_i| \leq u$ (since $\hat{U}_i(x) \geq |x|$ for $|x| \geq 1$) and $\sum_{i\geq 1} c_i^2 \leq u$ (since $\hat{U}_i(x) \geq x^2$ for $|x| \leq 1$). Consequently

$$\sum_{i\geq 1} t_i a_i = \sum_{i\geq 1} t_i b_i + \sum_{i\geq 1} t_i c_i \leq u\|t\|_\infty + \sqrt{u}\|t\|_2 \, ,$$

and we have shown that

$$\frac{1}{2}\max(u\|t\|_\infty, \sqrt{u}\|t\|_2) \leq \mathcal{N}_u(t) \leq (u\|t\|_\infty + \sqrt{u}\|t\|_2) \, ,$$

and thus

$$\frac{1}{2}\{t\,;\,\|t\|_\infty \le 1\,,\,\|t\|_2 \le \sqrt{u}\} \subset B(u) \subset 2\{t\,;\,\|t\|_\infty \le 1\,,\,\|t\|_2 \le \sqrt{u}\}\,. \tag{10.23}$$

In this case the functions φ_j have a genuinely more complicated structure than in the case of (10.22).

The third simplest example is the case where for some $p \ge 1$ and for all i we have $U_i(x) = x^p$ for $x \ge 0$, and the reader who truly wants to understand what really is going on would do well to work out a version of the general result in this special case. (The cases $p > 2$ and $p < 2$ offer significant differences.)

Our first result provides suitable upper bounds for $\mathsf{E}\sup_t X_t$. We recall that the definition (10.21) of the function φ_j involves the parameter r.

Theorem 10.2.1. *Assume that there exists an admissible sequence $(\mathcal{A}_n)$ of $T \subset \ell^2$, and for $A \in \mathcal{A}_n$ an integer $j_n(A) \in \mathbb{Z}$ such that*

$$\forall A \in \mathcal{A}_n\,, \forall s, s' \in A\,,\, \varphi_{j_n(A)}(s, s') \le 2^{n+1}\,. \tag{10.24}$$

Then

$$\mathsf{E}\sup_{t\in T} X_t \le L \sup_{t\in T} \sum_{n\ge 0} 2^n r^{-j_n(A_n(t))}\,. \tag{10.25}$$

Let us first interpret this statement in the case where $U_i(x) = x^2$ for each i. Then (and more generally when $U_i(x) \ge x^2/L$ for $x \ge 1$) we have $\varphi_j(s,t) \le Lr^{2j}\|s-t\|_2^2$, so that (10.24) holds as soon as $r^{2j_n(A)}\Delta(A, d_2)^2 \le 2^n/L$, where of course d_2 denotes the distance induced by the norm of ℓ^2. Taking for $j_n(A)$ the largest integer that satisfies this inequality implies that the right-hand side of (10.25) is bounded by $Lr\sup_{t\in T}\sum_{n\ge 0} 2^{n/2}\Delta(A_n(t), d_2)$. Taking the infimum over the admissible sequences $(\mathcal{A}_n)$ this yields

$$\mathsf{E}\sup_{t\in T} X_t \le Lr\gamma_2(T, d_2)\,.$$

Let us now interpret Theorem 10.2.1 when $U_i(x) = x$ for each i. When $\|s-t\|_\infty \le r^{-j}/L$, (10.23) implies $\varphi_j(s,t) \le Lr^{2j}\|s-t\|_2^2$, so that (10.24) holds whenever $r^{j_n(A)}\Delta(A, d_\infty) \le 1/L$ and $r^{2j_n(A)}\Delta(A, d_2)^2 \le 2^n/L$, where of course d_∞ denotes the distance induced by the norm of ℓ^∞. Taking for $j_n(A)$ the largest integer that satisfies both conditions yields

$$r^{-j_n(A)} \le Lr\big(\Delta(A, d_\infty) + 2^{-n/2}\Delta(A, d_2)\big)\,,$$

so that (10.25) implies

$$\mathsf{E}\sup_{t\in T} X_t \le Lr \sup_{t\in T} \sum_{n\ge 0}\big(2^n\Delta(A_n(t), d_\infty) + 2^{n/2}\Delta(A_n(t), d_2)\big)\,. \tag{10.26}$$

At the beginning of the proof of Theorem 2.2.23, we have explained how, given two admissible sequences that behave well for two different distances, one can construct an admissible sequence that behaves well for both distances. Thus (10.26) implies

$$\mathsf{E}\sup_{t\in T} X_t \le Lr\big(\gamma_2(T,d_2)+\gamma_\infty(T,d_1)\big)\,. \tag{10.27}$$

This resembles Theorem 2.2.23, and could actually be deduced from this theorem and an appropriate version of Bernstein's inequality.

It will be a simple adaptation of the proof of Theorem 2.2.22 to deduce Theorem 10.2.1 from the following, that provides a sharp description of the size of the tails of an individual r.v. X_t.

Proposition 10.2.2. *If $u>0$, $v\ge 1$, we have*

$$\mathsf{P}(X_t \ge Lv\mathcal{N}_u(t)) \le \exp(-uv)\,. \tag{10.28}$$

Proof of Theorem 10.2.1. We consider an arbitrary element t_0 of T and we set $T_0=\{t_0\}$. For $n\ge 1$ we consider a set T_n such that

$$\forall A\in\mathcal{A}_n\,,\ \mathrm{card}(A\cap T_n)=1\,.$$

For $t\in T$ we define $\pi_n(t)$ by $\{\pi_n(t)\}=A_n(t)\cap T_n$. For any integer k and any t in T_k we have

$$X_t - X_{t_0} = \sum_{1\le n\le k} X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\,. \tag{10.29}$$

For $v\ge 1$ consider the event Ω_v defined by

$$\forall n\ge 1\,,\ \forall s\in T_n\,,\ \forall s'\in T_{n-1}\,,\ |X_s - X_{s'}| \le Lv\mathcal{N}_{2^n}(s-s')\,, \tag{10.30}$$

where L is as in (10.28). Then (10.28) and the fact that $\mathrm{card}\,T_n\cdot\mathrm{card}\,T_{n-1}\le N_nN_{n-1}\le 2^{2^{n+1}}$ imply

$$\mathsf{P}(\Omega_v^c)\le p(v):=\sum_{n\ge 1} 2^{2^{n+1}}\exp(-v2^n)\,. \tag{10.31}$$

The definition of φ_j and (10.24) imply

$$\forall s,s'\in A\in\mathcal{A}_n\,,\ s-s'\in r^{-j_n(A)}B(2^{n+1})\,. \tag{10.32}$$

Since $\pi_n(t),\pi_{n-1}(t)\in A_{n-1}(t)$, using (10.32) for $n-1$ rather than n yields

$$\pi_n(t)-\pi_{n-1}(t)\in r^{-j_{n-1}(A_{n-1}(t))}B(2^n)\,,$$

and the definition of $B(u)$ implies

$$\mathcal{N}_{2^n}(\pi_n(t) - \pi_{n-1}(t)) \le 2^n r^{-j_{n-1}(A_{n-1}(t))} .$$

When the event Ω_v occurs, using (10.30) for $s = \pi_n(t)$ and $s' = \pi_{n-1}(t)$ we get

$$|X_{\pi_n(t)} - X_{\pi_{n-1}}(t)| \le Lv 2^n r^{-j_{n-1}(A_{n-1}(t))} .$$

Consequently (10.29) implies that for $t \in T_k$

$$|X_t - X_{t_0}| \le Lv \sum_{1 \le n \le k} 2^n r^{-j_{n-1}(A_{n-1}(t))} ,$$

and thus

$$\sup_{t \in T_k} |X_t - X_{t_0}| \le Lv \sup_{t \in T} \sum_{1 \le n \le k} 2^n r^{-j_{n-1}(A_{n-1}(t))} ,$$

so that

$$\mathsf{P}\Big(\sup_{t \in T_k} |X_t - X_{t_0}| > Lv \sup_{t \in T} \sum_{1 \le n \le k} 2^n r^{-j_{n-1}(A_{n-1}(t))}\Big) \le \mathsf{P}(\Omega_v^c) ,$$

and using (10.31) we get (after a simple computation)

$$\mathsf{E} \sup_{t \in T_k} |X_t - X_{t_0}| \le L \sup_{t \in T} \sum_{1 \le n \le k} 2^n r^{-j_{n-1}(A_{n-1}(t))} ,$$

which implies the conclusion since k is arbitrary. □

Exercise 10.2.3. (a) Use Proposition 10.2.2 to prove that if $p \ge 1$ then $\|X_t\|_p \le Lpr^{-j}\varphi_j(t)$.
(b) Deduce Theorem 10.2.1 from Exercise 2.2.25.

The proof of Proposition 10.2.2 requires several lemmas. For $\lambda \ge 0$ we define $V_i(\lambda) = \sup_x(\lambda x - \hat{U}_i(x))$, so that $V_i(\lambda) < \infty$ for $\lambda < \lambda_i$, where $\lambda_i = \lim_{x \to \infty} \hat{U}_i(x)/x \ge 1 \in]0, \infty]$. (The limit exists since $\hat{U}_i$ is convex). Moreover (taking $x = 0$), we have $V_i \ge 0$, and V_i is convex with $V_i(0) = 0$. Taking $x = \lambda/2$, and since $\hat{U}_i(x) = x^2$ for $x < 1$, we observe that

$$\lambda \le 2 \Rightarrow V_i(\lambda) \ge \frac{\lambda^2}{4} \tag{10.33}$$

and taking $x = 1$ that

$$V_i(\lambda) \ge \lambda - 1 . \tag{10.34}$$

Lemma 10.2.4. *For $\lambda \ge 0$ we have*

$$\mathsf{E} \exp \lambda Y_i \le \exp V_i(L\lambda) . \tag{10.35}$$

Proof. Since $U_i'(1) \geq 1$, for $x \geq 1$ we have $U_i(x) \geq x$, so that by (10.19) we have $\mathsf{P}(|Y_i| \geq x) \leq e^{-x}$ and hence (using e.g. that $x^2 \leq L \exp |x|/6$),

$$\mathsf{E} Y_i^2 \exp \frac{|Y_i|}{2} \leq L \; .$$

The elementary inequality $e^x \leq 1 + x + x^2 e^{|x|}$ yields that, if $\lambda \leq 1/2$,

$$\mathsf{E} \exp \lambda Y_i \leq 1 + \lambda^2 \mathsf{E} Y_i^2 \exp \lambda |Y_i| \leq 1 + L\lambda^2 \leq \exp L\lambda^2 \; . \tag{10.36}$$

Now since $\lambda \leq 1/2$, we have $\lambda^2 \leq 4V_i(\lambda)$, and since V_i is convex, $V_i \geq 0$ and $V_i(0) = 0$ we have $4LV_i(\lambda) \leq V_i(4L\lambda)$, so that $L\lambda^2 \leq V_i(4L\lambda)$. This completes the proof of (10.35) in the case $\lambda \leq 1/2$.

Assume now that $\lambda \geq 1/2$, and observe that

$$\begin{aligned} \mathsf{E} \exp \lambda |Y_i| &= 1 + \lambda \int_0^\infty \exp \lambda x \, \mathsf{P}(|Y_i| \geq x) \mathrm{d}x \\ &\leq 1 + \lambda \int_0^\infty \exp(\lambda x - U_i(x)) \mathrm{d}x \; . \end{aligned} \tag{10.37}$$

We will prove that, for $x \geq 0$,

$$\lambda x - U_i(x) \leq \frac{V_i(6\lambda)}{2} - \lambda x \; . \tag{10.38}$$

Combining with (10.37), this yields

$$\begin{aligned} \mathsf{E} \exp \lambda |Y_i| &\leq 1 + \lambda \int_0^\infty \exp\Big(\frac{V_i(6\lambda)}{2} - \lambda x\Big) \mathrm{d}x \\ &= 1 + \exp \frac{V_i(6\lambda)}{2} \leq 2 \exp \frac{V_i(6\lambda)}{2} \\ &\leq \exp V_i(6\lambda) \end{aligned}$$

because (using (10.34) in the last inequality) $V_i(6\lambda) \geq V_i(3) \geq 2$, completing the proof of (10.35).

To prove (10.38) we first consider the case where $x \leq 1$. Then $4\lambda x \leq 4\lambda$, $4\lambda \leq 6\lambda - 1$ (since $\lambda \geq 1/2$) and $6\lambda - 1 \leq V_i(6\lambda)$ by (10.34), so that $4\lambda x \leq V_i(6\lambda)$. Thus we have

$$\lambda x - U_i(x) \leq \lambda x \leq \frac{V_i(6\lambda)}{2} - \lambda x \, .$$

When $x \geq 1$ we have $U_i(x) \geq \hat{U}_i(x)/2$ and then

$$\lambda x - U_i(x) \leq \lambda x - \frac{\hat{U}_i(x)}{2} \leq \frac{V_i(4\lambda)}{2} - \lambda x$$

by definition of V_i. Since $V_i(4\lambda) \leq V_i(6\lambda)$ the proof is complete. □

Lemma 10.2.5. *We have*

$$\sum_{i\geq 1} V_i\Big(\frac{u|t_i|}{\mathcal{N}_u(t)}\Big) \leq u \;.$$

Proof. It suffices to show that given numbers $x_i \geq 0$, we have

$$\sum_{i\geq 1} \frac{u|t_i|x_i}{\mathcal{N}_u(t)} - \sum_{i\geq 1} \hat{U}_i(x_i) \leq u \;. \tag{10.39}$$

If $\sum_{i\geq 1} \hat{U}_i(x_i) \leq u$, then by definition of $\mathcal{N}_u(t)$ we have $\sum_{i\geq 1} |t_i|x_i \leq \mathcal{N}_u(t)$ so we are done. If $\sum_{i\geq 1} \hat{U}_i(x_i) = \theta u$ with $\theta > 1$, then (since $\hat{U}_i(0) = 0$ and $\hat{U}_i$ is convex) we have $\sum_{i\geq 1} \hat{U}_i(x_i/\theta) \leq u$, so that by definition of $\mathcal{N}_u$, $\sum_{i\geq 1} |t_i|x_i \leq \theta\mathcal{N}_u(t)$ and the left-hand side of (10.39) is in fact ≤ 0. □

Lemma 10.2.6. *If $v \geq 1$ we have*

$$\mathcal{N}_{uv}(t) \leq v\mathcal{N}_u(t) \;. \tag{10.40}$$

Proof. Consider numbers a_i with $\sum_{i\geq 1} \hat{U}_i(a_i) \leq uv$. For $v \geq 1$ we have $\hat{U}_i(a_i/v) \leq \hat{U}_i(a_i)/v$, so that $\sum_{i\geq 1} \hat{U}_i(a_i/v) \leq u$. By definition of $\mathcal{N}_u$ we then have $\sum_{i\geq 1} t_i a_i/v \leq \mathcal{N}_u(t)$ i.e. $\sum_{i\geq 1} t_i a_i \leq v\mathcal{N}_u(t)$. The definition of $\mathcal{N}_{uv}$ then implies (10.40). □

Proof of Proposition 10.2.2. Since by Lemma 10.2.6 we have $v\mathcal{N}_u(t) \geq \mathcal{N}_{vu}(t)$, we can assume $v = 1$. Lemma 10.2.4 implies

$$\begin{aligned}\mathsf{P}(X_t \geq y) &\leq \exp(-\lambda y)\mathsf{E}\exp\lambda X_t \\ &\leq \exp\Big(-\lambda y + \sum_{i\geq 1} V_i(L_0\lambda||t_i||)\Big) \;.\end{aligned}$$

We choose $y = 2L_0\mathcal{N}_u(t)$, $\lambda = 2u/y$, and we apply Lemma 10.2.5 to get

$$-\lambda y + \sum_{i\geq 1} V_i(L_0\lambda t_i) \leq -2u + u = -u \,.$$

□

Let us now turn to the converse of Theorem 10.2.1. We assume the following regularity conditions. For some constant C_0, we have

$$\forall i \geq 1 \,,\, \forall s \geq 1 \,,\, U_i(2s) \leq C_0 U_i(s) \;. \tag{10.41}$$

$$\forall i \geq 1 \,,\, U_i'(0) \geq 1/C_0 \,. \tag{10.42}$$

Here, $U_i'(0)$ is the right derivative at 0 of the function $U_i(x)$. Condition (10.41) is often called "the $\varDelta_2$ condition".

Theorem 10.2.7. *Under conditions (10.41) and (10.42) we can find r_0 (depending on C_0 only) and a number $K = K(C_0)$ such when $r \geq r_0$, for each subset T of ℓ^2 there exists an admissible sequence $(\mathcal{A}_n)$ of T and for $A \in \mathcal{A}_n$ an integer $j_n(A) \in \mathbb{Z}$ such that (10.24) holds together with*

$$\sup_{t \in T} \sum_{n \geq 0} 2^n r^{-j_n(A_n(t))} \leq K(C_0) r \mathsf{E} \sup_{t \in T} X_t \,. \tag{10.43}$$

Together with Theorem 10.2.1, this essentially allows the computation of $\mathsf{E} \sup_{t \in T} X_t$ as a function of the geometry of T. It is not very difficult to prove that Theorem 10.2.7 still holds true without condition (10.42), and this is done in [1]. But it is an entirely different matter to remove condition (10.41). The difficulty is of the same nature as in the study of Bernoulli processes. Now that the Bernoulli conjecture has been solved, on can hope that this will eventually be done.

Let us interpret Theorem 10.2.7 in the case where $U_i(x) = x^2$ for $x \geq 1$. In that case (and more generally when $U_i(x) \leq x^2/L$ for $x \geq 1$), we have

$$\varphi_j(s,t) \geq r^{2j} \|s - t\|_2^2 / L \,, \tag{10.44}$$

so that (10.24) implies that $\Delta(A, d_2) \leq L 2^{n/2} r^{-j_n(A)}$ and (10.43) implies

$$\sup_{t \in T} \sum_{n \geq 0} 2^{n/2} \Delta(A_n(t), d_2) \leq L r \mathsf{E} \sup_{t \in T} X_t \,,$$

and hence

$$\gamma_2(T, d_2) \leq L r \mathsf{E} \sup_{t \in T} X_t \,. \tag{10.45}$$

Thus, we have proved (an extension of) Theorem 2.4.1.

Next consider the case where $U_i(x) = x$ for all x. Then (10.23) implies (10.44), and thus (10.45). It also implies that $\varphi_j(s,t) = \infty$ whenever $\|s - t\|_\infty > 2r^{-j}$, because then $r^j(s-t) \notin B(u)$ whatever the value of u. Consequently, (10.24) implies that $\Delta(A, d_\infty) \leq L r^{-j_n(A)}$, and (10.43) yields

$$\gamma_1(T, d_\infty) \leq L r \mathsf{E} \sup_{t \in T} X_t \,.$$

Recalling (10.27) (and since here r is a universal constant) we thus have proved the following very pretty fact.

Theorem 10.2.8. *Assume that the r.v.s Y_i are independent, symmetric and satisfy $\mathsf{P}(|Y_i| \geq x) = \exp(-x)$. Then*

$$\frac{1}{L}(\gamma_2(T, d_2) + \gamma_1(T, d_\infty)) \leq \mathsf{E} \sup_{t \in T} X_t \leq L(\gamma_2(T, d_2) + \gamma_1(T, d_\infty)) \,.$$

Corollary 10.2.9. *If T is a set of sequences,*

$$\gamma_2(\operatorname{conv} T, d_2) + \gamma_1(\operatorname{conv} T, d_\infty) \leq L(\gamma_2(T, d_2) + \gamma_1(T, d_\infty)) \,. \tag{10.46}$$

Research problem 10.2.10. Given a geometrical proof of (10.46).

A far more general question occurs in Problem 10.2.15 below.

Exercise 10.2.11. Prove that it is not true that for a set T of sequences one has

$$\gamma_1(\operatorname{conv} T, d_\infty) \le L\gamma_1(T, d_\infty) .$$

(Hint: consider the set T of coordinate functions on $\{-1, 1\}^k$.)

We now prepare for the proof of Theorem 10.2.7.

Lemma 10.2.12. *Under (10.41), given $\rho > 0$ we can find r_0, depending on C_0 and ρ only, such that if $r \ge r_0$, for $u \in \mathbb{R}^+$ we have*

$$B(4ru) \subset \rho r B(u) . \tag{10.47}$$

Proof. We claim that for a constant C_1 depending only on C_0 we have

$$\forall u > 0 \,,\, \hat{U}_i(2u) \le C_1 \hat{U}_i(u) . \tag{10.48}$$

Indeed, it suffices to prove this for u large, where this follows from the Δ_2 condition (10.41). Consider an integer k large enough that $2^{-k+2} \le \rho$ and let $r_0 = C_1^k$. Assuming that $r \ge r_0$, we prove (10.47).

Consider $t \in B(4ru)$. Then $\mathcal{N}_{4ru}(t) \le 4ru$ by definition of $B(4ru)$, so that for any numbers $(a_i)_{i \ge 1}$ we have

$$\sum_{i \ge 1} \hat{U}_i(a_i) \le 4ru \Rightarrow \sum_{i \ge 1} a_i t_i \le 4ru . \tag{10.49}$$

Consider numbers b_i with $\sum_{i \ge 1} \hat{U}_i(b_i) \le u$. Then by (10.48) we have $\hat{U}_i(2^k b_i) \le C_1^k \hat{U}_i(b_i) \le r\hat{U}_i(b_i)$, so that $\sum_{i \ge 1} \hat{U}_i(2^k b_i) \le ru \le 4ru$, and (10.49) implies $\sum_{i \ge 1} 2^k b_i t_i \le 4ru$. Since $2^k \ge 4/\rho$ we have shown that

$$\sum_{i \ge 1} \hat{U}_i(b_i) \le u \Rightarrow \sum_{i \ge 1} b_i \frac{t_i}{\rho r} \le u ,$$

so that $\mathcal{N}_u(t/\rho r) \le u$ and thus $t/\rho r \in B(u)$ i.e. $t \in r\rho B(u)$. □

Theorem 10.2.13. *Under Condition (10.42) we can find a number $\rho > 0$ with the following property. Given any points $t_1, \ldots, t_m$ in ℓ^2 such that*

$$\ell \ne \ell' \Rightarrow t_\ell - t_{\ell'} \notin B(u) \tag{10.50}$$

and given any sets $H_\ell \subset t_\ell + \rho B(u)$, we have

$$\mathsf{E} \sup_{t \in \bigcup H_\ell} X_t \ge \frac{1}{L} \min(u, \log m) + \min_{\ell \le m} \mathsf{E} \sup_{t \in H_\ell} X_t . \tag{10.51}$$

The proof of this statement is very similar to the proof of (2.89). The first ingredient is a suitable version of Sudakov minoration, asserting that, under (10.50)

$$\mathsf{E}\sup_{\ell\le m} X_{t_\ell} \ge \frac{1}{L}\min(u,\log m) \tag{10.52}$$

and the second is a "concentration of measure" result that quantifies the deviation of $\sup_{t\in H_\ell} X_t$ from its mean. Condition (10.42) is used there, to assert that the law of Y_i is the image of the probability ν of density $e^{-2|x|}$ with respect to Lebesgue measure by a Lipschitz map. This allows to apply the result of concentration of measure concerning ν first proved in [2]. Since neither of these arguments is closely related to our main topic, we refer the reader to [3] and [1].

Proof of Theorem 10.2.7. Consider ρ as in Theorem 10.2.13. If $r = 2^{\kappa-2}$, where κ is large enough (depending on C_0 only), Lemma 10.2.12 shows that (10.47) holds for each $u > 0$. We fix this value of r, and we prove that the functionals $F_{n,j}(A) = 2L_0\mathsf{E}\sup_{t\in A} X_t$, where L_0 is the constant of (10.51), satisfy the growth condition of Definition 10.1.1 for $n_0 = 1$. Consider $n \ge 1$, and points (t_ℓ) for $\ell \le m = N_n$ as in (10.3). By definition of φ_{j+1} we have

$$\ell \ne \ell' \Rightarrow t_\ell - t_{\ell'} \notin r^{-j-1}B(2^{n+1})\ . \tag{10.53}$$

Consider then sets $H_\ell \subset B_{j+2}(t_\ell, 2^{\kappa+n})$. By definition of φ_{j+2}, we have $B_{j+2}(t_\ell, 2^{\kappa+n}) = t_\ell + r^{-j-2}B(2^{\kappa+n})$. Using (10.47) for $u = 2^n$ (and since $2^\kappa = 4r$) we obtain that $B(2^{\kappa+n}) \subset \rho r B(2^n)$ and therefore $H_\ell \subset t_\ell + \rho r^{-j-1}B(2^n)$. Since $\log m = 2^n \log 2 \ge 2^{n-1}$, we can then appeal to (10.51) to obtain the desired relation

$$F_{n,j}\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge 2^n r^{-j-1} + \min_{\ell\le m} F_{n+1,j+1}(H_\ell)$$

that completes the proof of the growth condition.

Using (10.52) for $n = 2$ and homogeneity yields

$$s,t\in T\ ,\ s-t\notin aB(1) \Rightarrow \frac{a}{L_0} \le \mathsf{E}\max(X_s,X_t) \le \mathsf{E}\sup_{t\in T} X_t\,. \tag{10.54}$$

Let us denote by j_0 the largest integer such that $r^{-j_0} > L_0\mathsf{E}\sup_{t\in T} X_t$, so that

$$r^{-j_0} \le L_0 r\mathsf{E}\sup_{t\in T} X_t\ . \tag{10.55}$$

For $s,t \in T$, using (10.54) for $a = r^{-j_0}$ implies $s - t \in r^{-j_0}B(1)$ and thus $\varphi_{j_0}(s,t) \le 1$, that is (10.7) holds for $n_0 = 1$ and this value of j_0. Thus we are in a position to apply Theorem 10.1.2 to construct an admissible sequence $(\mathcal{A}_n)$. Using (10.55), (10.9) implies

$$\forall t\in T\ ,\ \sum_{n\ge1} 2^n r^{-j_n(A_n(t))} \le Lr\mathsf{E}\sup_{t\in T} X_t\ .$$

Setting $j_0(T) = j_0$, this yields (10.43) since then $r^{-j_0(A_{n_0}(t))} = r^{-j_0}$.

To finish the proof, it remains to prove (10.24). By definition of $B_j(t,u)$ and of φ_j, we have

$$s \in B_j(t,u) \Rightarrow \varphi_j(s,t) \le u \Rightarrow s - t \in r^{-j} B(u)\,.$$

Thus (10.10) implies

$$\forall n \ge 1\,,\ \forall A \in \mathcal{A}_n\,,\ \forall s \in A\,,\ s - t_{n,A} \in r^{-j_n(A)} B(2^n)\,.$$

Since $B(u)$ is a convex symmetric set, we have

$$s - t_{n,A} \in r^{-j_n(A)} B(2^n),\ s' - t_{n,A} \in r^{-j_n(A)} B(2^n) \Rightarrow \frac{s - s'}{2} \in r^{-j_n(A)} B(2^n)$$
$$\Rightarrow \varphi_{j_n(A)}\Big(\frac{s}{2}, \frac{s'}{2}\Big) \le 2^n\,,$$

and finally

$$\forall n \ge 1\,, \forall A \in \mathcal{A}_n\,, \forall s, s' \in A\,,\ \varphi_{j_n(A)}\Big(\frac{s}{2}, \frac{s'}{2}\Big) \le 2^n\,.$$

This is not exactly (10.24), but of course to get rid of the factor 1/2 it would have sufficed to apply the above proof to $2T = \{2t; t \in T\}$ instead of T. □

Exercise 10.2.14. Prove that under conditions (10.41) and (10.42) we can find an admissible increasing sequence $\mathcal{A}_n$ of partitions of T such that $\sup_{t\in T} \Delta_n(A_n(t)) \le K(C_0) \mathsf{E} \sup_{t\in T} X_t$, where Δ_n denotes the diameter for the distance $\|X_s - X_t\|_{2^n}$. That is, the upper bound of Exercise 2.2.25 can be reversed.

As a consequence of Theorems 10.2.1 and 10.2.7, we have the following geometrical result. Consider a set $T \subset \ell^2$, an admissible sequence $(\mathcal{A}_n)$ of T and for $A \in \mathcal{A}_n$ an integer $j_n(A)$ such that (10.24) holds true. Then there is an admissible sequence $(\mathcal{B}_n)$ of $\operatorname{conv} T$ and for $B \in \mathcal{B}_n$ an integer $j_n(B)$ that satisfies (10.24) and

$$\sup_{t\in \operatorname{conv} T} \sum_{n\ge 0} 2^n r^{-j_n(B_n(t))} \le K(C_0) r \sup_{t\in T} \sum_{n\ge 0} 2^n r^{-j_n(A_n(t))}\,. \tag{10.56}$$

Research problem 10.2.15. Give a geometrical proof of this fact.

This is a far-reaching generalization of Research Problem 2.4.22.

The following generalizes Theorem 2.4.18.

Theorem 10.2.16. *Assume (10.41) and (10.42). Consider a countable subset T of ℓ^2, with $0 \in T$. Then we can find a sequence (x_n) of vectors of ℓ^2 such that*

$$T \subset (K(C_0) \mathsf{E} \sup_{t\in T} X_t)\ \operatorname{conv}(\{x_n\ ;\ n \ge 2\} \cup \{0\})$$

and, for each n,

$$\mathcal{N}_{\log n}(x_n) \le 1\,.$$

The point of this result is that, whenever the sequence $(x_n)_{n\geq 2}$ satisfies $\mathcal{N}_{\log n}(x_n) \leq 1$, then $\mathsf{E}\sup_{n\geq 2} X_{x_n} \leq L$. To see this, we simply write, by (10.28), that for $v \geq 1$,

$$\mathsf{P}\Big(\sup_{n\geq 2}|X_{x_n}| \geq Lv\Big) \leq \sum_{n\geq 2}\mathsf{P}(|X_{x_n}| \geq Lv\mathcal{N}_{\log n}(x_n)) \\ \leq \sum_{n\geq 2}\exp(-v\log n) \leq L\exp(-v/2)\ . \quad (10.57)$$

Proof. We consider a sequence of partitions of T as provided by Theorem 10.2.7. We choose $t_{0,T} = 0$, and for $A \in \mathcal{A}_n$, $n \geq 1$ we select $t_{n,A} \in \mathcal{A}_n$, making sure (as in the proof of Theorem 2.4.18) that each point of T is of the form $t_{n,A}$ for a certain A and a certain n . For $A \in \mathcal{A}_n$, $n \geq 1$, we denote by A' the unique element of $\mathcal{A}_{n-1}$ that contains A.

We define
$$u_A = \frac{t_{n,A} - t_{n-1,A'}}{2^{n+1}r^{-j_{n-1}(A')}}$$
and $U = \{u_A; A \in \mathcal{A}_n, n \geq 1\}$. Consider $t \in T$, so that $t = t_{n,A}$ for some n and some $A \in \mathcal{A}_n$, and, since $A_0(t) = T$ and $t_{0,T} = 0$,

$$t = t_{n,A} = \sum_{1\leq k\leq n} t_{k,A_k(t)} - t_{k-1,A_{k-1}(t)} = \sum_{1\leq k\leq n} 2^{k+1}r^{-j_{k-1}(A_{k-1}(t))}u_{A_k(t)}\ .$$

Since $\sum_{k\geq 0} 2^k r^{-j_k(A_k(t))} \leq K(C_0)\mathsf{E}\sup_{t\in T} X_t$ by (10.43), this shows that

$$T \subset (K(C_0)\mathsf{E}\sup_{t\in T} X_t)\ \mathrm{conv}\, U\ .$$

Next, we prove that $\mathcal{N}_{2^{n+1}}(u_A) \leq 1$ whenever $A \in \mathcal{A}_n$. The definition of φ_j and (10.24) imply

$$\forall s\,,\ s' \in A\ , s - s' \in r^{-j_n(A)}B(2^{n+1})\ ,$$

and the homogeneity of $\mathcal{N}_u$ yields

$$\forall s\,,\ s' \in A\ , \mathcal{N}_{2^{n+1}}(s - s') \leq r^{-j_n(A)}2^{n+1}\ .$$

Since $t_{n,A}, t_{n-1,A'} \in A'$, using this for $n-1$ rather than n and A' instead of A we get
$$\mathcal{N}_{2^n}(t_{n,A} - t_{n-1,A'}) \leq 2^n r^{-j_{n-1}(A')}\ ,$$
and thus $\mathcal{N}_{2^n}(u_A) \leq 1/2$, so that $\mathcal{N}_{2^{n+1}}(u_A) \leq 1$ using (10.40).

Let us enumerate $U = (x_n)_{n\geq 2}$ in such a manner that the points of the type u_A for $A \in \mathcal{A}_1$ are enumerated before the points of the type u_A for $A \in \mathcal{A}_2$, etc. Then if $x_n = u_A$ for $A \in \mathcal{A}_k$, we have $n \leq N_0 + N_1 + \cdots + N_k \leq N_k^2$ and therefore $\log n \leq 2^{k+1}$. Thus $\mathcal{N}_{\log n}(x_n) \leq \mathcal{N}_{2^{k+1}}(x_n) = \mathcal{N}_{2^{k+1}}(u_A) \leq 1$. □

References

1. Latała, R.: Sudakov minoration principle and supremum of some processes. Geom. Funct. Anal. **7**, 936–953 (1997)
2. Talagrand, M.: A new isoperimetric inequality for product measure and the concentration of measure phenomenon. In: Israel Seminar (GAFA). Springer Lecture Notes in Math., vol. 1469, pp. 94–124 (1991)
3. Talagrand, M.: The supremum of certain canonical processes. Am. J. Math. **116**, 283–325 (1994)

11. Infinitely Divisible Processes

11.1 A Well-Kept Secret

The reader having never worked with infinitely divisible processes is unlikely to feel comfortable with formulas such as (11.6) below, so we explain here, in an informal but essentially correct way what infinitely divisible processes really are, and why they fit so well with the other objects we study. Consider a σ-finite measure space (Ω, ν). There is a canonical way to construct a countable random subset of Ω, through a Poisson point process of intensity measure ν. This is explained on page 340, but the exact properties of this subset are not very important until one starts to prove theorems. Let us enumerate this subset as a sequence $(Z_i)_{i \geq 1}$. Consider an independent Bernoulli sequence $(\varepsilon_i)_{i \geq 1}$. Consider a function t on (Ω, ν). Then, when $\int \min(t^2, 1) \mathrm{d}\nu < \infty$, one can show that the series $X_t = \sum_i \varepsilon_i t(Z_i)$ converges a.s. An infinitely divisible (symmetric, without Gaussian component) process is then simply a collection of such r.v.s, for $t \in T$, where T a set of functions on Ω. Obviously this absolutely canonical construction is connected with the idea of empirical processes. Let us observe that given the randomness of the (Z_i), an infinitely divisible process is a Bernoulli process, a fact that will be used in a fundamental way. For example, the "harmonic infinitely divisible processes" considered in Theorem 11.2.1 are simply the following. Considering a compact group T, the set G of continuous characters on T, and a σ-finite measure ν on $\mathbb{R}G$. Consider the sequence (Z_i) associated to this measure as above. Then Z_i is the multiple of a character on T and the process $(X_t)_{t \in T}$ is distributed as the process $(\sum_i \varepsilon_i Z_i(t))_{t \in T}$. Given the choice of the r.v.s Z_i this process is then a random Fourier series and it is not surprising that the complete understanding we have of these series bears on the study of such processes.

Not only the previous definition of infinitely divisible processes is rather direct, it also immediately brings forward what is (for the type of results we consider) the fruitful point of view, to think of an infinitely divisible process as a class of functions. Nonetheless, since the topic has a long history, we will follow the traditional method of introducing infinitely divisible r.v.s through the properties of their characteristic function, and we will develop the present point of view only gradually.

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_11,

11.2 Overview of Results

We start by recalling some classical facts. A number of these facts will be proved in Section 11.3. A reader needing more details can refer to her favorite textbook.

A Poisson r.v. X of expectation a is a r.v. such that

$$\forall n \geq 0 \; ; \; \mathsf{P}(X = n) = \frac{a^n}{n!} \exp(-a) \; , \tag{11.1}$$

and thus $\mathsf{E}X = a$. Then, for any $b \in \mathbb{C}$,

$$\mathsf{E} b^X = \exp(-a) \sum_{n \geq 0} b^n \frac{a^n}{n!} = \exp(-a(1-b)) \; , \tag{11.2}$$

and in particular

$$\mathsf{E} \exp(i\alpha X) = \exp\bigl(-a(1 - \exp(i\alpha))\bigr) \; . \tag{11.3}$$

Consequently, the sum of two independent Poisson r.v.s is Poisson. Consider now (finitely many) independent Poisson r.v.s X_k with $\mathsf{E}X_k = a_k$, and numbers $\beta_k \geq 0$. Then, by independence,

$$\begin{aligned} \mathsf{E} \exp\Bigl(i\alpha \sum_k \beta_k X_k\Bigr) &= \exp\Bigl(-\sum_k a_k (1 - \exp(i\alpha\beta_k))\Bigr) \\ &= \exp\Bigl(-\int (1 - \exp(i\alpha\beta)) \mathrm{d}\nu(\beta)\Bigr) \; , \end{aligned} \tag{11.4}$$

where ν is the discrete positive measure on $\mathbb{R}^+$ such that for each $\beta \in \mathbb{R}^+$ we have $\nu(\{\beta\}) = \sum\{a_k ; \beta_k = \beta\}$. Let us observe the formula

$$\mathsf{E} \sum_k \beta_k X_k = \sum_k \beta_k a_k = \int \beta \mathrm{d}\nu(\beta) \; .$$

We recall the notation $x \wedge 1 = \min(x, 1)$. We say that a r.v. X is *positive infinitely divisible* if there exists a positive measure ν on $\mathbb{R}^+$ such that

$$\int (\beta \wedge 1) \mathrm{d}\nu(\beta) < \infty \; , \tag{11.5}$$

$$\forall \alpha \in \mathbb{R} \; , \; \mathsf{E} \exp i\alpha X = \exp\Bigl(-\int (1 - \exp(i\alpha\beta)) \mathrm{d}\nu(\beta)\Bigr) \; . \tag{11.6}$$

It is appropriate to think of X as a (continuous) sum of independent r.v.s of the type βY where Y is a Poisson r.v. and $\beta \geq 0$. This is a sum of quantities that are ≥ 0 and *there is no cancellation* in this sum. The r.v. X need not have an expectation. In fact, it has an expectation if and only $\int \beta d\nu(\beta) < \infty$

(and the value of this expectation is then $\int \beta d\nu(\beta)$). The reader need not be very concerned with the case where X does not have an expectation because our main results require technical conditions which imply that X has an expectation.

Consider again a Poisson r.v. Y of expectation a and an independent copy Y' of Y. Then (11.3) implies

$$\mathsf{E} \exp i\alpha(Y - Y') = \exp(-2a(1 - \cos(\alpha))) \ . \tag{11.7}$$

Thus, when a r.v. X is a sum of independent terms $\beta_k(Y_k - Y_k')$ where Y_k and Y_k' are independent Poisson r.v.s of expectation a_k and $\beta_k \geq 0$, we have

$$\mathsf{E} \exp i\alpha X = \exp\Big(- \int (1 - \cos(\alpha\beta)) \mathrm{d}\nu(\beta) \Big) \ , \tag{11.8}$$

where now ν is the discrete positive measure on $\mathbb{R}^+$ such that $\nu(\{\beta\}) = 2 \sum \{a_k ; \beta_k = \beta\}$ for each $\beta \in \mathbb{R}^+$. Let us observe the inequality

$$\mathsf{P}(X \neq 0) \leq 2 \sum_k a_k = |\nu| \ , \tag{11.9}$$

where $|\nu|$ denotes the total mass of ν.

We say that a r.v. X is *infinitely divisible* (real, symmetric, without Gaussian component) if it satisfies (11.8) for a positive measure ν on $\mathbb{R}^+$ such that

$$\int (\beta^2 \wedge 1) \mathrm{d}\nu(\beta) < \infty \ . \tag{11.10}$$

(We shall prove the existence of X in Section 11.3.) It is appropriate to think of X as a continuous sum of independent r.v.s of the type $\beta(Y - Y')$ where Y and Y' are independent Poisson r.v.s with the same expectation. These r.v.s are symmetric rather than positive, and there is *a lot of cancellation* when one adds them. This is why the formula (11.8) makes sense under the condition (11.10) rather than the much stronger condition (11.5). (The essence of the proof that (11.8) makes sense under (11.10) is simply to bring out cancellation through computation of second moments.) This dichotomy no cancellation versus cancellation considerably influences the rest of this chapter.

If T is a finite set, a stochastic process $(X_t)_{t \in T}$ is called (real, symmetric, without Gaussian component) infinitely divisible if there exists a positive measure ν on $\mathbb{R}^T$ such that $\int_{\mathbb{R}^T} (\beta(t)^2 \wedge 1) \mathrm{d}\nu(\beta) < \infty$ for all t in T, and such that for all families $(\alpha_t)_{t \in T}$ of real numbers we have

$$\mathsf{E} \exp i \sum_{t \in T} \alpha_t X_t = \exp\Big(- \int_{\mathbb{R}^T} \Big(1 - \cos\big(\sum_{t \in T} \alpha_t \beta(t)\big)\Big) \mathrm{d}\nu(\beta) \Big) \ . \tag{11.11}$$

The positive measure ν is called the *Lévy measure* of the process. (It turns out to be unique provided one assumes, which changes nothing, that $\nu(\{0\}) = 0$.)

As a consequence of this formula, each of the linear combinations $\sum_{t \in T} \alpha_t X_t$ is an infinitely divisible r.v. To get a feeling for this formula, consider the case where ν consists of a mass a at a point $\beta \in \mathbb{R}^T$. Then, in distribution, we have $(X_t)_{t \in T} = (\beta(t)(Y - Y'))_{t \in T}$ where Y and Y' are independent Poisson r.v.s of expectation $a/2$. One can then view the formula (11.11) as saying that the general case is obtained by taking a (kind of continuous) sum of independent processes of the previous type. Let us point out right away that a lot of cancellation occurs when taking such sums.

Let us comment on our definition of infinitely divisible processes. It is a very mild restriction to exclude Gaussian components, since these are so well understood, as was seen in Chapter 2. Also, for the purpose of studying the supremum of the process $(X_t)_{t \in T}$, it is essentially not a restriction to consider only the symmetric case, using the symmetrization procedure that we have met several times, i.e. replacing the process $(X_t)_{t \in T}$ by the process $(X_t - X'_t)_{t \in T}$ where $(X'_t)_{t \in T}$ is an independent copy of $(X_t)_{t \in T}$. In summary, our definition of infinitely divisible processes is the most general one for the purpose of studying the supremum of such processes.

For the type of inequalities we wish to prove, it is not a restriction to assume that T is finite, but it is still useful to consider also the case where T is infinite. In that case, we still say that the process $(X_t)_{t \in T}$ is infinitely divisible if (11.11) holds for each family $(\alpha_t)_{t \in T}$ such that only finitely many coefficients are not 0. Now ν is a "cylindrical measure" that is known through its projections on $\mathbb{R}^S$ for S finite subset of T, projections that are positive measures (and satisfy the obvious compatibility conditions, see e.g. [6] on how these projections can be glued together).

An infinitely divisible process indexed by T is thus parameterized by a cylindrical measure on $\mathbb{R}^T$ (with the sole restriction that $\int (\beta(t)^2 \wedge 1) \mathrm{d}\nu(\beta) < \infty$ for each $t \in T$). This is a *huge* class, and only some extremely special subclasses have yet been studied in any detail. The best known such subclass is that of infinitely divisible processes with *stationary increments*. Then $T = \mathbb{R}^+$ and ν is the image of $\mu \otimes \lambda$ under the map $(x, u) \mapsto (x \mathbf{1}_{\{t \ge u\}})_{t \in \mathbb{R}^+}$, where μ is a positive measure on $\mathbb{R}$ such that $\int (x^2 \wedge 1) \mathrm{d}\mu(x) < \infty$ and where λ is Lebesgue measure. More likely than not a probabilist selected at random (!) will think that infinitely divisible processes are intrinsically discontinuous. This is simply because he has this extremely special case as a mental picture. In a certain sense, the processes we study are to infinitely divisible processes with stationary increments what general Gaussian processes are to Brownian motion. As will be apparent later (through Rosinski's representation) discontinuity in the case of processes with stationary increments is created by the fact that ν is supported by the discontinuous functions $t \mapsto x \mathbf{1}_{\{t \ge u\}}$ and is certainly not intrinsic to infinitely divisible processes. In fact, some lesser known classes of infinitely divisible processes studied in the literature, such as moving averages (see e.g. [2]) are often continuous. They are still very much more special than the structures we consider.

Continuity will not be studied here, and was mentioned simply to stress that we deal with hugely general and complicated structures, and it is almost surprising that so much can be said about them.

Next, we show that p-stable processes are a special case of infinitely divisible processes. Consider $1 \le p < 2$ and denote by λ Lebesgue's measure on $\mathbb{R}^+$. Consider a probability measure m on R^T. Assume that

$$\nu \text{ is the image of } \lambda \otimes m \text{ under the map } (x, \gamma) \mapsto x^{-1/p}\gamma\,. \tag{11.12}$$

In that case, if the Lévy measure is given by (11.12), the process $(X_t)_{t \in T}$ is p-stable. To see this, we observe the formula

$$\int_{\mathbb{R}^+} (1 - \cos(a x^{-1/p})) \mathrm{d}\lambda(x) = C(p)|a|^p\,,$$

which is obvious through change of variable. Then, for each real θ we have

$$\begin{aligned}
&\int_{\mathbb{R}^T} \Big(1 - \cos\big(\theta \sum_{t \in T} \alpha_t \beta(t)\big)\Big) \mathrm{d}\nu(\beta) \\
&= \int_{\mathbb{R}^T} \int_{\mathbb{R}^+} \Big(1 - \cos\big(\theta x^{-1/p} \sum_{t \in T} \alpha_t \gamma(t)\big)\Big) \mathrm{d}\lambda(x) \mathrm{d}m(\gamma) \\
&= |\theta|^p \frac{\sigma^p}{2}\,,
\end{aligned} \tag{11.13}$$

where

$$\sigma^p = 2C(p) \int_{\mathbb{R}^T} \Big|\sum_{t \in T} \alpha_t \gamma(t)\Big|^p \mathrm{d}m(\gamma)\,. \tag{11.14}$$

Then (11.11) and (8.1) show that the r.v. $\sum_{t \in T} \alpha_t X(t)$ is p-stable (so that the process (X_t) is stable by definition).

The goal of the present chapter is, following our general philosophy, to try to relate the size of the r.v. $\sup_{t \in T} X_t$ with a proper measure of "the size of T". The functions

$$\varphi(s, t, u) = \int_{\mathbb{R}^T} \big((u^2 |\beta(s) - \beta(t)|^2) \wedge 1\big) \mathrm{d}\nu(\beta) \tag{11.15}$$

will help measure the size of T. Given a number r, we will consider the functions

$$\varphi_j(s, t) = \varphi(s, t, r^j)\ .$$

For example, in the case (11.12) of p-stable processes, using change of variables again, it is quite straightforward to prove that

$$\varphi_j(s, t) = C'(p)(r^j d(s, t))^p\ , \tag{11.16}$$

where the distance d is given by

$$d(s,t)^p = \int |\beta(s) - \beta(t)|^p \mathrm{d}m(\beta) \,. \tag{11.17}$$

Motivated by Section 3.1 we may expect that our goal will be much easier to reach when "there is stationarity", and we will consider this case first. For this we need to consider complex-valued infinitely divisible processes.

Let us assume for simplicity that T is an metrizable compact group. Let us consider the dual G of T, i.e. the set of continuous characters on T. We say that the process $(X_t)_{t\in T}$ is (complex valued, symmetric) *harmonic* infinitely divisible if there exists a positive σ-finite measure ν on $\mathbb{C}G \subset \mathbb{C}^T$ (provided with the σ-algebra generated by the coordinate functions) such that $\int_{\mathbb{C}^T} |\beta(t)|^2 \wedge 1 \mathrm{d}\nu(\beta) < \infty$ for all t in T, and such that for all families $(\alpha_t)_{t\in T}$ of complex numbers, finitely many of which only are not 0, we have

$$\mathsf{E} \exp i\Re \sum_{t\in T} \alpha_t X_t = \exp\Big(- \int_{\mathbb{C}^T} \Big(1 - \cos \Re\big(\sum_{t\in T} \alpha_t \beta(t) \big) \Big) \mathrm{d}\nu(\beta) \Big) \,, \tag{11.18}$$

where $\Re(z)$ denotes the real part of the complex number z. Of course this seems a formidable formula, but in Section 11.1 we have already tried to take the scare out of it. An example of process satisfying this condition is $X_t = \beta(t)(Y - Y')$ where $\beta \in \mathbb{C}G$ and where Y and Y' are independent Poisson r.v.s of expectation a. Then ν has mass $2a$ at β, and as previously independent sums of such r.v.s also satisfy (11.18).

We denote by μ the Haar measure on T, and we assume $r = 2$ for simplicity. We recall that $N_0 = 1$ and that $N_n = 2^{2^n}$ for $n \geq 1$.

Theorem 11.2.1. *There exists a number $\alpha_1 > 0$ with the following property. Assume that the process $(X_t)_{t\in T}$ is harmonic infinitely divisible. Assume that for each finite set $S \subset T$ and a certain number M we have*

$$\mathsf{P}\Big(\sup_{t\in S} |X_t| \geq M \Big) \leq \alpha_1 \,. \tag{11.19}$$

Then there exist integers j_n for $n \geq 0$ such that

$$\forall\, s,t \in T \,,\ \varphi_{j_0}(s,t) \leq 3 \,, \tag{11.20}$$

and, for $n \geq 1$,

$$\mu(\{s \in T \,;\ \varphi_{j_n}(s,0) \leq 2^n\}) \geq N_n^{-1} \,, \tag{11.21}$$

and

$$\sum_{n\geq 0} 2^{n-j_n} \leq LM \,. \tag{11.22}$$

In words, if we control the size of the process $(X_t)_{t\in T}$ then we control the size of T "as measured with the distances φ_j".

We turn to a converse of this theorem. First we must observe that conditions (11.20) and (11.21) are automatically satisfied when ν has a total mass ≤ 1 because then $\varphi_j(s,t) \leq 1$, so that these conditions alone are certainly not sufficient to imply much about the boundedness of the process $(X_t)_{t\in T}$.

Exercise 11.2.2. Given a number $A > 0$ construct examples of harmonic infinitely divisible processes for which (11.22) holds for $M = 1$ but where

$$\mathsf{P}\Big(\sup_{t\in T}|X_t| \geq A\Big) \geq 1/2\ .$$

On the other hand if U is a subset of $\mathbb{C}^T$, it is true in some sense that "the part of ν supported by U influences the values of the process $(X_t)_{t\in T}$ only on an event of probability $\leq 2\nu(U)$". In the case of single r.v. this have been shown in (11.9), and in the case of an entire process this will become clear later. Thus we may expect at best that (11.22) suffices to control the values of the process outside a set of small probability. The next theorem asserts that this is the case.

Theorem 11.2.3. *Consider a harmonic infinitely divisible process and numbers j_n, $n \geq 0$ that satisfy the conditions (11.20) and (11.21). Then for any number $w \geq 2^{-j_0}$ we can find an event Ω with the property that*

$$\mathsf{P}(\Omega^c) \leq \nu(\{\beta\ ;\ |\beta(0)| \geq w\}) \tag{11.23}$$

such that for any finite subset S of T

$$\mathsf{E}\mathbf{1}_\Omega \sup_{s\in S}|X_s| \leq L\sum_{n\geq 0} 2^{n-j_n} + L\int_{2^{-j_0}\leq|\beta(0)|\leq w} |\beta(0)|\mathrm{d}\nu(\beta)\ . \tag{11.24}$$

This of course is of interest only if $\sum_{n\geq 0} 2^{n-j_n} < \infty$. In that case the right-hand side of (11.24) is finite for any value of w, but as w becomes large the right-hand side of (11.23) becomes small (since because $\int |\beta(0)|^2 \wedge 1\mathrm{d}\nu < \infty$ we have $\nu(\{\beta\ ;\ |\beta(0)| \geq 1\}) < \infty$). Theorem 11.2.3 is particularly satisfactory in the situation where

$$\int_{|\beta(0)|\geq 2^{-j_0}} |\beta(0)|\mathrm{d}\nu(\beta) < \infty\,,$$

since in that case one can take $w = \infty$ and $\Omega^c = \emptyset$.

The two previous theorems together provide a complete understanding of "when the process $\sup_t |X_t|$ is bounded". Let us illustrate this in the simpler case of "harmonic p-stable processes". That is, we assume that there exists a probability measure m on G such that if ν is the positive measure on $\mathbb{C}G$ given by (11.12), then (11.18) holds, where the integral is on $\mathbb{C}G$ rather than $\mathbb{R}^T$. Then we have, where d is the distance (11.17), and where $\gamma_q(T,d)$ is as in Definition 2.2.19:

Corollary 11.2.4. *If $1 < p < 2$ then $\mathsf{E}\sup_{t\in T}|X_t| < \infty$ if and only if $\gamma_q(T,d) < \infty$. If $p = 1$, then $\sup_{t\in T}|X_t| < \infty$ a.s. if and only if there exists a sequence (ϵ_n) such that $\sum_n \epsilon_n < \infty$ and*

$$\mu(\{s \in T\ ;\ d(s,t) \leq \epsilon_n\}) \geq N_n^{-1}\ .$$

To avoid ambiguities, let us stress that unless we deal with harmonic infinitely divisible processes, which are complex-valued, we consider only real-valued process. This is the case for all the forthcoming results.

One of our main results will be, in a sense, an extension of Theorem 11.2.1 to the case of infinitely divisible processes that are not necessarily harmonic. Our main tool will be the use of Theorem 10.1.2 where $\varphi_j(s,t) = \varphi(s,t,r^j)$ a certain value of r. Unfortunately, in order to be able to prove a suitable growth condition related to these quantities, we need a technical condition on ν. It is a major open problem to decide whether our result holds without this technical condition. To a large extent, the difficulty seems the same as in the Bernoulli conjecture (so that now that this conjecture has been solved one may hope for eventual progress on this problem). It is part of a larger circle of problems which are discussed in Chapter 12. We introduce this condition now.

Definition 11.2.5. *Consider $\delta > 0$ and $C_0 > 0$. We say that condition $H(C_0, \delta)$ holds if for all $s, t \in T$, and all $u > 0$, $v > 1$ we have*

$$\nu(\{\beta\ ;\ |\beta(s) - \beta(t)| \ge uv\}) \le C_0 v^{-1-\delta} \nu(\{\beta\ ;\ |\beta(s) - \beta(t)| \ge u\})\ .$$

Without loss of generality we assume that $\delta < 1$.

Condition $H(C_0, \delta)$ is certainly annoying, since it rules out important cases, such as when ν charges only one point. A large class of measures ν that satisfy condition $H(C_0, \delta)$ can be constructed as follows. Consider a measure μ on $\mathbb{R}$, and assume that

$$\forall u > 0\ ,\ \forall v > 1\ ,\ \mu(\{x\ ;\ |x| \ge uv\}) \le C_0 v^{-1-\delta} \mu(\{x\ ;\ |x| \ge u\})\ . \quad (11.25)$$

Consider a probability measure m on $\mathbb{R}^T$, and assume that ν is the image of $\mu \otimes m$ under the map $(x, \gamma) \mapsto x\gamma$. Then ν satisfies condition $H(C_0, \delta)$. This follows from (11.25) and the formula

$$\nu(\{\beta\ ;\ |\beta(s) - \beta(t)| \ge u\}) = \int \mu(\{x\ ;\ |x||\gamma(s) - \gamma(t)| \ge u\}) \mathrm{d}m(\gamma)\ .$$

In particular condition $H(C_0, \delta)$ holds when μ has density x^{-p-1} with respect to Lebesgue's measure on $\mathbb{R}^+$ and $1 < p < 2$.

Theorem 11.2.6. *Under condition $H(C_0, \delta)$, there exists a number $r \ge 4$ (depending only on C_0 and δ), an admissible sequence of partitions $\mathcal{A}_n$ and for $A \in \mathcal{A}_n$ a number $j_n(A) \in \mathbb{Z}$ such that (10.8) holds i.e.*

$$A \in \mathcal{A}_n\ ,\ C \in \mathcal{A}_{n-1}\ ,\ A \subset C \Rightarrow j_{n-1}(C) \le j_n(A) \le j_{n-1}(C) + 1 \quad (10.8)$$

together with

$$\forall\, n \ge 0\ ,\ \forall\, A \in \mathcal{A}_n\ ,\ \forall s, s' \in A\ ,\ \varphi_{j_n(A)}(s, s') = \varphi(s, s', r^{j_n(A)}) \le 2^{n+2} \quad (11.26)$$

$$\forall t \in T\ ,\ \sum_{n \ge 0} 2^n r^{-j_n(A_n(t))} \le K \mathsf{E} \sup_{t \in T} X_t\ . \quad (11.27)$$

Here of course K denotes a number that depends on C_0 and δ only and that need not be the same at each occurrence. It is legitimate to use expectation to control the size of $\sup_{t\in T} X_t$ in (11.27) because $H(C_0,\delta)$ implies that all the variables X_t do have an expectation.

Of course the level of abstraction reached here might make it hard for the reader to immediately understand the power of Theorem 11.2.6. As a first illustration, let us prove that in the case where ν is obtained as in (11.12), we recover Theorem 8.1.1. By change of variable, it is obvious that

$$\int_{\mathbb{R}^+} ((ax^{-1/p})^2 \wedge 1)\mathrm{d}\lambda(x) = C_1(p)|a|^p \,,$$

so that

$$\begin{aligned}\varphi(s,t,u) &= \int_{\mathbb{R}^T}\int_{\mathbb{R}^+}\Big(\big(x^{-1/p}u|\gamma(s)-\gamma(t)|\big)^2 \wedge 1\Big)\mathrm{d}\lambda(x)\mathrm{d}m(\gamma)\\ &= C_1(p)u^p\int_{\mathbb{R}^T}|\gamma(s)-\gamma(t)|^p\mathrm{d}m(\gamma)\,.\end{aligned}$$

Using (11.13) and (11.14) when $\sum_{t\in T}\alpha_t\beta(t) = \beta(t)-\beta(s)$ and comparing with (8.1) and (8.3) we obtain

$$\varphi(s,t,u) = C_2(p)u^p d^p(s,t)\,,$$

so that (11.26) implies $\Delta(A,d) \le K2^{n/p}r^{-j_n(A)}$, and (11.27) yields

$$\sum_{n\ge 0} 2^{n/q}\Delta(A_n(t),d) \le K\mathsf{E}\sup_{t\in T} X_t\,,$$

where $1/q = 1-1/p$ and thus $\gamma_q(T,d) \le K\mathsf{E}\sup_{t\in T} X_t$, which is the content of Theorem 8.1.10.

In the present chapter we shall go much beyond the negative fact that the inequality (11.27) of Theorem 11.2.6 cannot be reversed in general. To grab the headline right away, let us state that shall prove in a very precise sense in Theorem 11.2.10 below that

Theorem 11.2.6 **exactly** captures the part of the boundedness
that is due to **cancellation** . (11.28)

To give a slightly more precise statement we shall show that in distribution we can write $X_t = X'_t + X''_t$ where
(a) the infinitely divisible process (X'_t) is of a special type for which the inequality (11.27) of Theorem 11.2.6 can be reversed,
(b) the boundedness of the infinitely divisible process (X''_t) owes nothing to cancellation.

We now start the description of a representation of infinitely divisible processes invented by J. Rosinski ([4]). Not only this representation is an essential technical tool, but it will also help us in formulating our results, and make the link with the point of view of Section 11.1. That is, it allows to think in a natural manner of infinitely divisible processes as *classes of functions on a probability space*, a fundamental shift of point of view which brings a connection with the ideas of Section 9.1.

We denote Lebesgue's measure on $\mathbb{R}^+$ by λ. We consider a Poisson point process Π on $\mathbb{R}^+$ with intensity λ. That is, Π is a random subset of $\mathbb{R}^+$ with the following properties. First, for any bounded Borel subset A of $\mathbb{R}^+$,

$$\mathrm{card}(A \cap \Pi) \text{ is a Poisson r.v. of expectation } \lambda(A) , \tag{11.29}$$

and moreover

$$\begin{gathered}\text{If } A_1, \ldots, A_k \text{ are disjoint Borel sets, the r.v.s} \\ (\mathrm{card}(A_\ell \cap \Pi))_{\ell \le k} \text{ are independent .}\end{gathered} \tag{11.30}$$

We denote by $(\tau_i)_{i \ge i}$ an increasing enumeration of Π. (Equivalently, $(\tau_i)_{i\ge1}$ is the sequence of arrival times of a Poisson process of parameter 1, that is $\tau_i = \Gamma_1 + \cdots + \Gamma_i$, where the sequence $(\Gamma_k)_{k\ge1}$ is i.i.d. and $\mathsf{P}(\Gamma_k \ge u) = e^{-u}$. The equivalence is however non-trivial and will not be used.)

Consider a measurable function $G : \mathbb{R}^+ \times \mathbb{R}^T \to \mathbb{R}^T$. Consider a probability measure m on $\mathbb{R}^T$. We denote by $(Y_i)_{i\ge1}$ an i.i.d. sequence of $\mathbb{R}^T$-valued r.v.s, distributed like m, and by $(\varepsilon_i)_{i\ge1}$ a Bernoulli sequence. We assume that the sequences (τ_i), (ε_i) and (Y_i) are independent of each other, and that T is finite.

Theorem 11.2.7 (Rosinski's representation [4]). *Denote by ν the image measure of $\lambda \otimes m$ under G, and assume that it is a Lévy measure, i.e. that $\int_{\mathbb{R}^T} (|\beta(t)|^2 \wedge 1) \mathrm{d}\nu(\beta) < \infty$ for each t in T. Then the series $\sum_{i\ge1} \varepsilon_i G(\tau_i, Y_i)$ converges a.e. in $\mathbb{R}^T$ and its law is the law of the symmetric infinitely divisible process of Lévy measure ν.*

In practice, we are not given λ and m, but ν. There are many ways to represent ν as the image of a product $\lambda \otimes m$ under a measurable transformation. One particular method is very fruitful (and is also brought to light in [4] among other interesting methods). Consider a probability measure m such that ν is absolutely continuous with respect to m. There are of course many possible choices, but, remarkably enough, the particular choice of m does not seem relevant. Consider a Radon-Nikodym derivative g of ν with respect to m and define $G(u, \beta) = R(u, \beta)\beta$ where

$$R(u, \beta) = \mathbf{1}_{[0, g(\beta)]}(u) . \tag{11.31}$$

For simplicity we write $R_i = R(\tau_i, Y_i)$. Theorem 11.2.7 implies that the sum

$$\sum_{i\ge1} \varepsilon_i R_i Y_i \tag{11.32}$$

is distributed like the infinitely divisible process of Lévy measure ν. This representation of the process will be called *Rosinski's representation*. Let us note that

$$R_i \in \{0,1\} \; ; \; R_i \text{ is a non-increasing function of } \; \tau_i \; . \tag{11.33}$$

Conditionally on the sequence $(\tau_i)_{i\geq 1}$, the sequence $(R_i Y_i)_{i\geq 1}$ is independent. This sequence (τ_i) is a nuisance, but a secondary one. For all practical purposes one can almost think as if $\tau_i = i$. More precisely, it turns out that the influence of the sequence $(\tau_i)_{i\geq 1}$ will be felt only through the following two quantities (that exist from the law of large numbers)

$$\alpha_- = \min_{i\geq 1} \frac{\tau_i}{i} \; ; \; \alpha_+ = \max_{i\geq 1} \frac{\tau_i}{i} \; . \tag{11.34}$$

Conditionally in the sequences $(\tau_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$, the process (11.32) is a Bernoulli process. In particular, an infinitely divisible processes can be represented as a mixture of Bernoulli processes. This is essential for our approach.

Next, we describe a class of infinitely divisible process for which boundedness owes nothing to cancellation Given a finite set T, we say that the process $(X_t)_{t\in T}$ is *positive infinitely divisible* if there exists a positive measure ν on $(\mathbb{R}^+)^T$, such that

$$\forall t \in T \; , \; \int (\beta(t) \wedge 1) \mathrm{d}\nu(\beta) < \infty \; ,$$

and that for each family $(\alpha_t)_{t\in T}$ of real numbers we have

$$\mathsf{E} \exp i \sum_{t\in T} \alpha_t X_t = \exp\Big(- \int \big(1 - \exp(i \sum_{t\in T} \alpha_t \beta(t))\big) \mathrm{d}\nu(\beta) \Big) \; .$$

We will call ν the *Lévy measure* of the process. (it turns out to be unique provided one assumes that $\nu(\{0\}) = 0$.) While by "infinitely divisible process" we understand that the process is symmetric, a *positive* infinitely divisible process is certainly not symmetric. It is positive since a r.v satisfying (11.6) is positive. This also follows from another version of Rosinski's representation, which asserts that the process

$$\sum_{i\geq 1} R_i Y_i(t) \tag{11.35}$$

has the same law as $(X_t)_{t\in T}$. This is also proved in [4] and later here in Section 11.3. The representation (11.35) will also be called the Rosinski representation of the positive infinitely divisible process. The important feature here is that all terms in (11.35) are non-negative. There is no cancellation in this sum, so that the boundedness of a positive infinitely divisible process owes nothing to cancellation.

Next, we show that certain (symmetric) infinitely divisible processes are in a natural way dominated by a positive infinitely divisible process. Consider a (symmetric) infinitely divisible process $(X_t)_{t\in T}$ with Lévy measure ν and assume that

$$\forall t \in T\,,\ \int (|\beta(t)| \wedge 1)\mathrm{d}\nu(\beta) < \infty\,.$$

Consider the positive measure ν' on $(\mathbb{R}^+)^T$ which is the image of ν under the map $\beta \mapsto |\beta|$, where $|\beta|(t) = |\beta(t)|$. Then

$$\forall t \in T\,,\ \int (\beta(t) \wedge 1)\mathrm{d}\nu'(\beta) < \infty$$

so ν' is the Lévy measure of a positive infinitely divisible process that we denote by $(|X|_t)$. If ν is the image of $\lambda\otimes m$ under the map $(x,\beta) \mapsto R(x,\beta)\beta$, then, since $R(x,\beta) \geq 0$, ν' is the image of $\lambda \otimes m$ under the map

$$(x,\beta) \mapsto |R(x,\beta)\beta| = R(x,\beta)|\beta|\,.$$

Thus if $\sum_{i\geq 1} \varepsilon_i R_i Y_i$ is a Rosinski representation of the infinitely divisible process (X_t), then $\sum_{i\geq 1} R_i |Y_i|$ is a Rosinski representation of the positive infinitely divisible process $(|X|_t)$. Hence

$$\mathsf{E}\sup_{t\in T} X_t = \mathsf{E}\sup_{t\in T} \sum_{i\geq 1} \varepsilon_i R_i Y_i(t) \leq \mathsf{E}\sup_{t\in T} \sum_{i\geq 1} R_i |Y_i(t)| = \mathsf{E}\sup_{t\in T} |X|_t\,.$$

It we control $\mathsf{E}\sup_{t\in T} |X|_t$, we can certainly claim that we control $\mathsf{E}\sup_{t\in T} X_t$ in a way that involves no cancellation.

Let us now describe a completely different method to control $\mathsf{E}\sup_{t\in T} X_t$. Consider a Borel subset Ω of $\mathbb{R}^T$ with $m(\Omega^c) = 0$. On T consider the distance $d_\infty(s,t)$ given by $d_\infty(s,t) = \sup_{\beta\in\Omega} |\beta(s) - \beta(t)|$, and the distance $d_2(s,t)$ given by $d_2^2(s,t) = \int_\Omega (\beta(s) - \beta(t))^2 \mathrm{d}\nu(\beta)$.

Theorem 11.2.8. *We have*

$$\mathsf{E}\sup_{t\in T} X_t \leq L\big(\gamma_2(T,d_2) + \gamma_1(T,d_\infty)\big)\,. \tag{11.36}$$

A reader expert about infinitely divisible processes will wonder why she has never seen anything even remotely resembling this result. This is because Theorem 11.2.8 usually does not apply to the entire infinitely divisible process, but only to the part "where the boundedness is explained by the cancellation". This statement will soon be made precise.

In the following definition as usual ν is the image of $\lambda\otimes m$ under the map $(x,\beta) \mapsto R(x,\beta)\beta$.

Definition 11.2.9. *We say that an infinitely divisible process is S-certified if $\gamma_1(T,d_\infty) \leq S$ and $\gamma_2(T,d_2) \leq S$, where, for a certain set $\Omega \subset \mathbb{R}^T$ with $m(\Omega^c) = 0$, we have*

$$d_\infty(s,t) = \sup_{\beta \in \Omega} |\beta(s) - \beta(t)| \,,$$

and

$$d_2(s,t) = \Big(\int_\Omega (\beta(s) - \beta(t))^2 \mathrm{d}\nu(\beta)\Big)^{1/2} \,.$$

Thus, Theorem 11.2.8 asserts that if the process $(X_t)_{t\in T}$ is S-certified, then $\mathsf{E}\sup_{t\in T} X_t \le LS$, so that we "certify that we control the size the process".

We now come to the main result of this chapter, the Decomposition Theorem for infinitely divisible processes.

Theorem 11.2.10 (The Decomposition Theorem). *Consider an infinitely divisible process $(X_t)_{t\in T}$, and assume that condition $H(C_0,\delta)$ of Definition 11.2.5 holds. Let $S = \mathsf{E}\sup_{t\in T} X_t$. Then we can write in distribution*

$$X_t = X'_t + X''_t$$

where both processes $(X'_t)_{t\in T}$ and $(X''_t)_{t\in T}$ are infinitely divisible with the following properties: (X'_t) is KS-certified, and $\mathsf{E}\sup_{t\in T} |X''|_t \le KS$.

In other words, we know two ways to control $\mathsf{E}\sup_{t\in T} X_t$. One way is that the process is S-certified. The other way is that we already control $\mathsf{E}\sup_{t\in T} |X|_t$. Under condition $H(C_0,\delta)$ there is *no* other method: every situation is a combination of these.

To prove this theorem, it will be convenient to adopt a different point of view (that was outlined in Section 11.1). This will also bring to light the fact that the present material is closely connected to the material of Section 9.1. To make this more apparent, rather than considering $\beta \in \Omega \subset \mathbb{R}^T$ as a function of $t \in T$, we will think of $t \in T$ as a function of β, by the formula $t(\beta) = \beta(t)$. Since ν is a Lévy measure, we have

$$\forall t \in T\,, \int_\Omega (t(\beta)^2 \wedge 1)\mathrm{d}\nu(\beta) < \infty \,. \tag{11.37}$$

Conversely, assume that we are given a (σ-finite) positive measure space (Ω,ν) and a (countable) set T of measurable functions on Ω such that (11.37) holds. Consider a probability measure m such that ν is absolutely continuous with respect to m and a function g such that $\nu = gm$. Consider an i.i.d. sequence (Y_i) distributed like m, and set $R_i = \mathbf{1}_{[0,g(Y_i)]}(\tau_i)$. Then Rosinski's representation

$$X_t = \sum_{i\ge 1} \varepsilon_i R_i t(Y_i)$$

defines an infinitely divisible process $(X_t)_{t\in T}$. Its Lévy measure $\bar\nu$ is the image of ν under the map $\omega \mapsto (t(\omega))_{t\in T}$. If, moreover,

$$\forall t \in T\,, \int_\Omega (|t(\beta)| \wedge 1)\mathrm{d}\nu(\beta) < \infty \,, \tag{11.38}$$

we can define a positive infinitely divisible process $(|X|_t)_{t\in T}$ by

$$|X|_t = \sum_{i\geq 1} R_i |t(Y_i)| \,.$$

The distances d_2 and d_∞ of Theorem 11.2.8 are simply the distances on T induced by the norms of $L^2(\nu)$ and $L^\infty(\nu)$ respectively.

Let us repeat: for the purpose of studying boundedness, an infinitely divisible process is essentially *a class of functions on a measure space*. This idea is usually implemented in the literature by defining the process X_t as "the stochastic integral of a function $f(t,.)$ with respect to an independently scattered symmetric infinitely divisible random measure without Gaussian component", but it is identical to what we do here.

We conclude our results on infinitely divisible processes by a "bracketing theorem" in the spirit of Ossiander's Theorem (Theorem 9.1.12). In this theorem, we still think of T as a (countable) set of measurable functions on (Ω, m).

Theorem 11.2.11. *Consider an admissible sequence $(\mathcal{A}_n)$ of T, and for $A\in\mathcal{A}_n$ consider $h_A(\omega)=\sup_{s,t\in A}|t(\omega)-s(\omega)|$. Assume that for $A\in\mathcal{A}_n$ we are given $j_n(A)\in\mathbb{Z}$ satisfying*

$$A\in\mathcal{A}_n\,,\; C\in\mathcal{A}_{n-1}\,,\; A\subset C \Rightarrow j_n(A)\geq j_{n-1}(C)\;.$$

Assume that for some numbers $r\geq 2$ and $S>0$ we have

$$\forall A\in\mathcal{A}_n\,,\; \int \big(r^{2j_n(A)}h_A^2\wedge 1\big)\mathrm{d}\nu \leq 2^n \tag{11.39}$$

$$\int h_T \mathbf{1}_{\{2h_T\geq r^{-j_0(T)}\}}\mathrm{d}\nu \leq S\,, \tag{11.40}$$

and

$$\forall t\in T\,,\; \sum_{n\geq 0} 2^n r^{-j_n(A_n(t))} \leq S\,. \tag{11.41}$$

Then $\mathsf{E}\sup_{t\in T}|X_t|\leq LS$.

Even though the principle of Theorem 11.2.11 goes back at least to [8], the power of this principle does not seem to have been understood. To illustrate this power, we will deduce from Theorem 11.2.10 a sample theorem of the much more recent work by Marcus and Rosinski from [3].

11.3 Rosinski's Representation

We start with some simple observations.

Lemma 11.3.1. *Consider a Borel function* $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$*, and assume that* $\int \varphi(x)\mathrm{d}\lambda(x) < \infty$*. Then*

$$\mathsf{E}\sum_{i\geq 1}\varphi(\tau_i) = \int \varphi(x)\mathrm{d}\lambda(x) \; . \tag{11.42}$$

Proof. When f is the indicator of a Borel set, (11.42) is a consequence of (11.29). When f has a bounded support, (11.42) follows from linearity and approximation, and this yields the general case using Fatou's lemma. □

Corollary 11.3.2. *Consider a countable set* T*, a probability measure* m *on* $\mathbb{R}^T$*, and an i.i.d. sequence* (Y_i) *of law* m*, which is independent of the r.v.s* τ_i*. Consider a Borel function* $\psi : \mathbb{R}^+ \times \mathbb{R}^T \to \mathbb{R}^+$*. Then, whenever the right-hand side is finite we have*

$$\mathsf{E}\sum_{i\geq 1}\psi(\tau_i, Y_i) = \int \psi(x,y)\mathrm{d}\lambda(x)\mathrm{d}m(y) \; . \tag{11.43}$$

Proof. Use (11.42) for the function $\varphi(x) = \mathsf{E}\psi(x, Y_1) = \int \psi(x,y)\mathrm{d}m(y)$. □

Lemma 11.3.3. *Consider a continuous function* $\varphi : \mathbb{R}^+ \to \mathbb{C}$*. Then, for each number* a*,*

$$\mathsf{E}\prod_{\tau_i < a}\varphi(\tau_i) = \exp\Big(-\int_0^a (1-\varphi(x))\mathrm{d}\lambda(x)\Big) \; . \tag{11.44}$$

Proof. Let us first assume that φ takes finitely many values on the interval $[0,a]$, so that there it a Borel partition $(A_1, \ldots, A_k)$ of this interval and numbers $b_1, \ldots, b_k$ such that $\varphi(x) = b_j$ for $x \in A_j$. Then

$$\prod_{\tau_i \leq a}\varphi(\tau_i) = \prod_{j\leq k} b_j^{\mathrm{card}\{i\,;\,\tau_i \in A_j\}} \; ,$$

and using (11.30),

$$\mathsf{E}\prod_{\tau_i \leq a}\varphi(\tau_i) = \prod_{j\leq k} \mathsf{E} b_j^{\mathrm{card}\{i\,;\,\tau_i \in A_j\}} \; .$$

Now $\mathrm{card}\{i; \tau_i \in A_j\}$ is a Poisson r.v. of expectation $\lambda(A_j)$ so that (11.2) implies

$$\mathsf{E} b_j^{\mathrm{card}\{i\,;\,\tau_i \in A_j\}} = \exp(-(1-b_j)\lambda(A_j)) \; ,$$

and (11.44) is proved in this case. The continuous case follows by approximation. □

Theorem 11.3.4. *Consider a finite set T, a probability measure m on $\mathbb{R}^T$, and an i.i.d. sequence (Y_i) distributed like m. Consider also an independent Bernoulli sequence (ε_i), and assume that the sequences (Y_i), (τ_i) and (ε_i) are mutually independent. Consider a Borel function $\psi : \mathbb{R}^+ \times \mathbb{R}^T \to \mathbb{R}^+$. Then if*

$$\int (\psi(x,y)^2 \wedge 1) \mathrm{d}\lambda(x) \mathrm{d}m(y) < \infty \tag{11.45}$$

the series $\sum_{k \ge 1} \varepsilon_k \psi(\tau_k, Y_k)$ converges a.s. and its sum X satisfies

$$\mathsf{E} \exp(iX) = \exp\Big(-\int (1 - \cos \psi(x,y)) \mathrm{d}\lambda(x) \mathrm{d}m(y)\Big) . \tag{11.46}$$

If we assume further that

$$\int (|\psi(x,y)| \wedge 1) \mathrm{d}\lambda(x) \mathrm{d}m(y) < \infty , \tag{11.47}$$

then the series $\sum_{k \ge 1} \psi(\tau_k, Y_k)$ converges a.s. and its sum X satisfies

$$\mathsf{E} \exp(iX) = \exp\Big(-\int (1 - \exp(i\psi(x,y)) \mathrm{d}\lambda(x) \mathrm{d}m(y)\Big) . \tag{11.48}$$

Throughout this chapter we denote by E^τ expectation given the sequence $(\tau_i)_{i \ge 1}$. The reader will carefully distinguish the notation E^τ, which means expectation **given** τ from the notation E_ε, which means expectation **in** ε **only**.

Proof. Let us first assume (11.45), and define

$$\varphi(x) = \int \cos \psi(x,y) \mathrm{d}m(y) . \tag{11.49}$$

Then, given any $a > 0$, using first independence and then (11.44) in the last step,

$$\begin{aligned} \mathsf{E} \exp\Big(i \sum_{\tau_k \le a} \varepsilon_k \psi(\tau_k, Y_k)\Big) &= \mathsf{E}\mathsf{E}_\varepsilon \exp\Big(i \sum_{\tau_k \le a} \varepsilon_k \psi(\tau_k, Y_k)\Big) \\ &= \mathsf{E} \prod_{\tau_k \le a} \cos(\psi(\tau_k, Y_k)) \\ &= \mathsf{E}\mathsf{E}^\tau \prod_{\tau_k \le a} \cos(\psi(\tau_k, Y_k)) = \mathsf{E} \prod_{\tau_k \le a} \varphi(\tau_k) \\ &= \mathsf{E} \exp\Big(-\int_0^a (1 - \varphi(x)) \mathrm{d}\lambda(x)\Big) . \end{aligned} \tag{11.50}$$

Moreover, Corollary 11.3.2 and (11.45) imply

$$\mathsf{E} \sum_{i \ge 1} \psi(\tau_i, Y_i)^2 \wedge 1 = \int (\psi(x,y)^2 \wedge 1) \mathrm{d}\lambda(x) \mathrm{d}m(y) < \infty . \tag{11.51}$$

Consequently for almost every realization of the sequence (τ_i) we have $\mathsf{E}^\tau \sum_{i\ge 1} \psi(\tau_i, Y_i)^2 \wedge 1 < \infty$, and Corollary 7.6.3 proves that the series $\sum_{i\ge 1} \varepsilon_i \psi(\tau_i, Y_i)$ converges a.s. To obtain (11.46) we simply let $a \to \infty$ in (11.50). This is justified in the left-hand side by dominated convergence (since we take expected value of a function of modulus 1) and in the right-hand side by (11.45) and the inequality

$$|1-\varphi(x)| \le \int |1-\cos\psi(x,y)| \mathrm{d}m(y) \le 2\int (\psi(x,y)^2 \wedge 1)\mathrm{d}m(y) \ ,$$

using that $|1-\cos x| \le 2(x^2 \wedge 1)$.

Let us now assume (11.47). As in (11.50) we obtain

$$\mathsf{E}\exp\Big(i\sum_{\tau_k\le a}\psi(\tau_k, Y_k)\Big) = \mathsf{E}\exp\Big(-\int_0^a (1-\varphi(x))\mathrm{d}\lambda(x)\Big) \ , \tag{11.52}$$

where now

$$\varphi(x) = \int \exp i\psi(x,y)\mathrm{d}m(y) \tag{11.53}$$

satisfies

$$|1-\varphi(x)| \le \int (|\psi(x,y)| \wedge 1)\mathrm{d}m(y) \ . \tag{11.54}$$

Corollary 11.3.2 and (11.47) imply that $\mathsf{E}\sum_i |\psi(\tau_i, Y_i)| \wedge 1 < \infty$, and this implies in turn that the series $\sum_{i\ge 1}\psi(\tau_i, Y_i)$ is absolutely convergent a.s. We then conclude as before by letting $a \to \infty$ in (11.52), this being now justified in the right-hand side by (11.54) and (11.47). □

Proof of Theorem 11.2.7. It suffices to prove for each t the convergence a.s. of the series $\sum_{i\ge 1}\varepsilon_i G(\tau_i, Y_i)(t)$, since this implies the convergence a.s. of the series $\sum_{i\ge 1}\varepsilon_i G(\tau_i, Y_i)$. This convergence follows from Theorem 11.3.4, used for the functions $\psi(x,y) = G(x,y)(t)$. Moreover, given numbers $(\alpha_t)_{t\in T}$, the right-hand side of (11.46) for the choice $\psi(x,y) = \sum_{t\in T}\alpha_t G(x,y)(t)$ is

$$\begin{aligned}&\exp\Big(-\int \Big(1-\cos\big(\sum_t \alpha_t G(x,y)(t)\big)\Big)\mathrm{d}\lambda(x)\mathrm{d}m(y)\Big)\\ &= \exp\Big(-\int \Big(1-\cos\big(\sum_t \alpha_t \beta(t)\big)\Big)\mathrm{d}\nu(\beta)\Big) \ .\end{aligned}$$

Consequently the process $(X_t)_{t\in T}$ given by $X_t = \sum_{i\ge 1}\varepsilon_i G(\tau_i, Y_i)(t)$ is infinitely divisible with Lévy measure ν (by definition). □

In a similar manner, using now the case (11.47) one proves that the process (11.35) has the same law as the process (X_t) in the case of positive infinitely divisible processes.

Perhaps we can remove some of the mystery underlying Rosinski's representation if we observe that when $R_i = \mathbf{1}_{\{\tau_i \le g(Y_i)\}}$ the family of points

$\{Y_i; R_i \neq 0\}$ is exactly a Poisson point process of intensity measure ν, a fact that we leave as a teaser to the reader. Rosinski's representation in some sense coincides with the representation $X_t = \sum_i \varepsilon_i Z_i(t)$, where Z_i is an enumeration of a Poisson point process of intensity measure ν, a representation which must have been known for a very long time. The point of Rosinski's representation is that it cleverly visualizes the independence properties underlying the Poisson point process (Z_i). (All of this is made very clear in Rosinski's paper [5].)

Let us now investigate the case where $G(x,y) = yx^{-1/p}$ of (11.12). Combining Theorem 11.2.7 and (11.13) we obtain that the process $X = \sum_{k\geq 1} \varepsilon_k Y_k \tau_k^{-1/p}$ satisfies

$$\exp\Bigl(i\sum_t \alpha_t X_t\Bigr) = \exp\Bigl(-C(p)\int \bigl|\sum_t \alpha_t \gamma(t)\bigr|^p \mathrm{d}m(\gamma)\Bigr) . \tag{11.55}$$

Suppose now that the probability measure m is the image of a product $\theta \otimes \mu$ under that map $(x,y) \to xy$, where $x \in \mathbb{R}$ and $y \in \mathbb{R}^T$, and that $\int |x|^p \mathrm{d}\theta(x) = 1$. Then it is immediate that

$$\int \bigl|\sum_t \alpha_t \gamma(t)\bigr|^p \mathrm{d}m(\gamma) = \int \bigl|\sum_t \alpha_t \gamma(t)\bigr|^p \mathrm{d}\mu(\gamma) .$$

Consequently, consider an i.i.d sequence (η_k) of law θ. Then the process

$$X = \sum_{k\geq 1} \varepsilon_k \eta_k Y_k \tau_k^{-1/p} \tag{11.56}$$

satisfies

$$\exp\Bigl(i\sum_t \alpha_t X_t\Bigr) = \exp\Bigl(-C(p)\int \bigl|\sum_t \alpha_t \gamma(t)\bigr|^p \mathrm{d}\mu(\gamma)\Bigr) . \tag{11.57}$$

It is proved in [1], Theorem 5.2 that for any p-stable process on $\mathbb{R}^T$ there exists a probability measure on $\mathbb{R}^T$ that satisfies (11.57), and we have just proved that such a process admits the representation (11.56). In particular we may choose η_k to be Gaussian, in which case (11.56) represents a p-stable process as a mixture of Gaussian processes, as used in Section 8.1.

Theorem 11.2.7 admits a complex valued version, where now m is a measure on $\mathbb{C}^T$ and $G : \mathbb{R}^+ \times \mathbb{C}^T \to \mathbb{C}^T$. The statement "its law is the law of the symmetric infinitely divisible process of Lévy measure ν" has now to be understood as meaning that (11.18) holds. The necessary modifications of the proof are left to the reader.

11.4 The Harmonic Case

In this section we prove Theorems 11.2.1 and 11.2.3. The basic idea is to use Rosinski's representation, so that given the sequence (τ_i) the process (X_t)

is equal in distribution to the sum of a random Fourier series of the type studied in Section 7.2. Theorems 11.2.1 and 11.2.3 will then be fairly easy consequences of the results of Section 7.2. They do not use Theorem 10.1.2. The rest of this chapter can be read independently of the results of the present section.

We consider a probability measure m on $\mathbb{C}G$ such that ν is absolutely continuous with respect to m, a Radon-Nikodym derivative g of ν with respect to m and we define $R(u,\beta) = \mathbf{1}_{[0,g(\beta)]}(u)$. We consider i.i.d. r.v.s (Y_i) of law m, and $R_i = R(\tau_i, Y_i)$. Then, by Rosinski's Theorem 11.2.7 (or more precisely, the complex version of this theorem explained at the end of the previous section) the following holds:

$$\text{For each } t \text{ in } T, \text{ the series } \sum_{i\geq 1} \varepsilon_i R_i Y_i(t) \text{ converges a.s.} \tag{11.58}$$

Moreover, writing

$$X_t^* = \sum_{i\geq 1} \varepsilon_i R_i Y_i(t) \ , \tag{11.59}$$

then for any countable set $S \subset T$,

$$(X_t)_{t\in S} \text{ has the same distribution as } (X_t^*)_{t\in S} \ . \tag{11.60}$$

We start the arguments with four very simple facts, that are of constant use when using Rosinski's representation. The first one is obvious.

Lemma 11.4.1. *Consider $\alpha > 0$ and a non-increasing function θ on $\mathbb{R}^+$. Then*

$$\alpha \sum_{i\geq 1} \theta(\alpha i) \leq \int_0^\infty \theta(x)\mathrm{d}\lambda(x) \leq \alpha \sum_{i\geq 0} \theta(\alpha i) \ . \tag{11.61}$$

Since $R(x,\beta) \in \{0,1\}$, the following is also obvious.

Lemma 11.4.2. *If $h(0) = 0$ then*

$$h(R(x,\beta)\beta) = R(x,\beta)h(\beta) \ . \tag{11.62}$$

Lemma 11.4.3. *Consider a non-negative measurable function h on $\mathbb{R}^T$, with $h(0) = 0$. Then*

$$\mathsf{E}\sum_{i\geq 1} h(R_i Y_i) = \mathsf{E}\sum_{i\geq 1} R_i h(Y_i) = \int h(\beta)\mathrm{d}\nu(\beta) \ . \tag{11.63}$$

Proof. Since g is a density of ν with respect to m, this follows from (11.43) used for the function $\psi(x,y) = \mathbf{1}_{\{x\leq g(y)\}} h(y)$. □

We recall the quantities α_- and α_+ of (11.34).

Lemma 11.4.4. *Consider a non-negative measurable function h on $\mathbb{R}^T$, with $h(0) = 0$. Then*

$$\frac{1}{\alpha_+}\int_{\mathbb{R}^T} h(\beta)\mathrm{d}\nu(\beta) - \int_{\mathbb{R}^T} h(\beta)\mathrm{d}m(\beta) \le \sum_{i\ge 1} \mathsf{E}^\tau h(R_iY_i) \le \frac{1}{\alpha_-}\int_{\mathbb{R}^T} h(\beta)\mathrm{d}\nu(\beta)\ . \tag{11.64}$$

Proof. Given β, the function $\theta(x) = h(R(x,\beta)\beta)$ is non-increasing since its value is $h(\beta) \ge 0$ for $x \le g(\beta)$ and $h(0) = 0$ for $x > g(\beta)$. Thus, using that $\tau_i \le \alpha_+ i$ in the first inequality, and (11.61) in the second one we get

$$\sum_{i\ge 1} h(R(\tau_i,\beta)\beta) + h(\beta) \ge \sum_{i\ge 0} h(R(\alpha_+ i,\beta)\beta) \ge \frac{1}{\alpha^+}\int_0^\infty h(R(x,\beta)\beta)\mathrm{d}\lambda(x)\ .$$

Since Y_i is distributed like m and since ν is the law of $\lambda\otimes m$ under the map $(x,\beta)\mapsto R(x,\beta)\beta$, integrating both sides in β with respect to m yields the left-hand side of (11.64). The right-hand side is similar. □

Proof of Theorem 11.2.1. There are some technical details involved in this proof, but the overall idea is completely straightforward. Let us set

$$Z_i = R_i(\tau_i, Y_i)Y_i\ ,$$

so that given the sequence (τ_i) these r.v.s are independent. We will show that by choosing a rather generic realization of the sequence (τ_i), and working given this sequence we inherit from the hypothesis (11.19) enough information on the partial sums of the series $\sum_i \varepsilon_i Z_i$ to use Lemma 7.6.1 . The information provided by this lemma is then brought back using (11.64).

First, we explain how to find a suitable realization of the sequence (τ_i). Since $\lim_{i\to\infty}\tau_i/i = 1$ a.s, we have $\mathsf{P}(\alpha^+ \le 2) > 0$. Recalling the constant α_0 of Theorem 7.3.1, we define $\alpha_1 = \alpha_0\mathsf{P}(\alpha^+ \le 2)/4$. Let us assume that the harmonic process $(X_t)_{t\in T}$ satisfies (11.19). Consider a dense subset U of T, so that (11.19) implies $\mathsf{P}(\sup_{t\in U}|X_t| \ge M) \le \alpha_1$. Recalling (11.59) it follows from (11.60) that $\mathsf{P}(\sup_{t\in U}|X_t^*| \ge M) \le \alpha_1$. Denoting by P^τ the conditional probability given τ, the r.v. $Y = \mathsf{P}^\tau(\{\sup_{t\in U}|X_t^*| \ge M\})$ satisfies $\mathsf{E}Y = \mathsf{P}(\{\sup_{t\in U}|X_t^*| \ge M\}) \le \alpha_1$. Consequently, using Markov's inequality and the definition of α_1,

$$\mathsf{P}(\{Y \ge \alpha_0/2\}) \le \frac{\alpha_1}{\alpha_0/2} = \mathsf{P}(\alpha^+ \le 2\})/2\ .$$

Therefore one can find a realization of the sequence (τ_i) for which $\alpha^+ \le 2$, the series $\sum_i \varepsilon_i R_i Y_i$ converges P^τ a.s. for each $t\in U$, and moreover $Y \le \alpha_0/2$ i.e.

$$\mathsf{P}^\tau\Big(\sup_{t\in U}|X_t^*| \ge M\Big) \le \frac{\alpha_0}{2}\ . \tag{11.65}$$

We think now of this sequence (τ_i) as being fixed once and for all, so that to lighten notation we write P rather than P^τ. Next, we prove that we control the partial sums $S_t^k = \sum_{i \le k} \varepsilon_i Z_i(t)$ uniformly over k. We claim that

$$\forall\, k \ge 1\;,\; \mathsf{P}\Big(\sup_{t \in U} |S_t^k| \ge M\Big) \le \alpha_0\;. \tag{11.66}$$

To see this, we set $X_t^{(k)} = \sum_{i > k} \varepsilon_i Z_i(t)$, so that $X_t^* = S_t^k + X_t^{(k)}$ where these two terms are independent. Also, given numbers $(x(t))_{t \in U}$ we have

$$\mathsf{P}\Big(\sup_{t \in U} |x(t) + X_t^{(k)}| \ge \sup_{t \in U} |x(t)|\Big) \ge \frac{1}{2}\;. \tag{11.67}$$

This is simply because for each t we have $\mathsf{P}(|x(t) + X_t^{(k)}| \ge |x(t)|) \ge 1/2$ by symmetry of the r.v. $X_t^{(k)}$. Using (11.67) for $x(t) = S_t^k$ and using independence yields

$$\mathsf{P}\Big(\sup_{t \in U} |X_t^*| \ge M\Big) \ge \frac{1}{2}\mathsf{P}\Big(\sup_{t \in U} |S_t^k| \ge M\Big)\;,$$

and combining with (11.65) proves (11.66). Since each Z_i is a continuous function on T we get

$$\forall\, k \ge 1\;,\; \mathsf{P}\Big(\sup_{t \in T} |S_t^k| \ge M\Big) \le \alpha_0\;. \tag{11.68}$$

This is the information we need to apply Lemma 7.6.1. Consider the sequence $(j_n)_{n \ge 0}$ produced by this lemma. We observe that since $r = 2$ we have $2^{-j_0} \le LM$. In order to avoid a conflict of notation we denote by φ_j^* the quantities (7.89), i.e.

$$\varphi_j^*(s,t) = \sum_{i \ge 1} \mathsf{E}(|2^j(Z_i(s) - Z_i(t))^2| \wedge 1)\;.$$

We use the left-hand side of (11.64) with the function $h(\beta) = |2^j(\beta(s) - \beta(t))|^2 \wedge 1$ to obtain (since $\alpha^+ \le 2$ and $\int h(\beta)\mathrm{d}\nu(\beta) = \varphi_j(s,t)$) that

$$\varphi_j(s,t) \le 2\bigg(\int h(\beta)\mathrm{d}m(\beta) + \varphi_j^*(s,t)\bigg)\;. \tag{11.69}$$

Since $h \le 1$ we observe first that

$$\varphi_j^*(s,0) \le 2^n \Rightarrow \varphi_j(s,0) \le 2^{n+2}\;, \tag{11.70}$$

and also that

$$\varphi_j^*(s,t) \le 1/4 \Rightarrow \varphi_j(s,t) \le 5/2 \le 3\;, \tag{11.71}$$

and then (11.21) and (11.22) are consequences of (7.90) and (7.91) respectively. □

We now prepare for the proof of Theorem 11.2.3.

Lemma 11.4.5. *We have*

$$\mathsf{E}\frac{1}{\sqrt{\alpha^-}} < \infty \ .$$

Proof. We write

$$\mathsf{P}(\alpha^- < u) \le \sum_{i\ge 1} \mathsf{P}(\tau_i \le ui) = \sum_{i \ge 1} \mathsf{P}(\operatorname{card}(\Pi \cap [0, ui]) \ge i) \ ,$$

and we show by elementary estimates that the first term dominates and that the sum is $\le Lu$. □

Proof of Theorem 11.2.3. The idea is straightforward: to use Lemma 7.6.5 given the randomness of the sequence (τ_i). We apply the right-hand side of (11.64) with the function $h(\beta) = |2^j(\beta(s) - \beta(t))|^2 \wedge 1$ to obtain

$$\varphi_j^*(s,0) := \sum_{i\ge 1} \mathsf{E}(|2^j(Z_i(s) - Z_i(0))|^2 \wedge 1) \le \frac{1}{\alpha^-}\varphi_j(s,0) \ .$$

Combining with (11.20) and (11.21) we obtain that for $v = L/\alpha^-$,

$$\forall\, s,t \in T \ , \ \varphi_j^*(s,t) \le \frac{v}{4}$$

and

$$\mu(\{s \ ; \ \varphi_{j_n}^*(s,t) \le v2^n\}) \ge N_n^{-1} \ .$$

That is, (7.46) and (7.47) hold for the value $v = L/\alpha^-$.

Consider the event

$$\Omega = \bigcap_{i \ge 1} \{|Z_i(0)| \le w\} \ .$$

We apply Lemma 7.6.5 given the randomness of the sequence (τ_i) and with $p = 1$ to the sums $S_t^k = \sum_{1\le i\le k} \varepsilon_i Z_i(t)$ and we let $k \to \infty$ to obtain

$$\mathsf{E}^\tau \mathbf{1}_\Omega \sup_{s\in S} \Big|\sum_i \varepsilon_i Z_i(s)\Big| \le \frac{L}{\sqrt{\alpha^-}} \sum_{n\ge 0} 2^{n-j_n} + L \sum_i \mathsf{E}^\tau |Z_i(0)| \mathbf{1}_{\{2^{-j_0} \le |Z_i(0)| < w\}} \ . \tag{11.72}$$

Taking expectation in (11.72) and using Lemma 11.4.5 we obtain

$$\mathsf{E}\mathbf{1}_\Omega \sup_{s\in S} \Big|\sum_i \varepsilon_i Z_i(s)\Big| \le L \sum_{n \ge 0} 2^{n-j_n} + L \sum_i \mathsf{E}|Z_i(0)| \mathbf{1}_{\{2^{-j_0} \le |Z_i(0)| < w\}} \ . \tag{11.73}$$

Since, using (11.63) in the equality,

$$\mathsf{P}(\Omega^c) \le \mathsf{E}\sum_{i \ge 1} \mathsf{P}(|Z_i(0)| \ge w) = \nu(\{\beta \ ; \ |\beta(0)| \ge w\})$$

and since

$$\sum_i \mathsf{E}|Z_i(0)|\mathbf{1}_{\{2^{-j_0}\le|Z_i(0)|<w\}} = \int_{\{2^{-j_0}\le|\beta(0)|\le w\}} |\beta(0)|\mathrm{d}\nu(\beta)$$

by (11.63) again, the proof is complete. □

Proof of Corollary 11.2.4. The formula (11.16) proves that

$$\varphi_j(s,0)\le 2^n \Leftrightarrow d(s,0)\le C''(p)2^{n/p-j}$$

so that the existence of the integers j_n that satisfy (11.21) and $\sum_{n\ge0} 2^{n-j_n} < \infty$ is equivalent to the existence of a sequence (ϵ_n) such that

$$\sum_{n\ge0} 2^{n(1-1/p)}\epsilon_n < \infty \tag{11.74}$$

and

$$\mu(\{s\in T\ ;\ d(s,0)\le\epsilon_n\})\ge N_n^{-1}\ . \tag{11.75}$$

When $p>1$, (3.4) shows that (11.74) and (11.75) are equivalent to $\gamma_q(T,d)<\infty$.

Assuming $\sup_{t\in T} X_t<\infty$ a.s., (11.19) holds for M large enough, so that Theorem 11.2.1 implies that we can find the sequence (ϵ_n) which satisfies (11.74) and (11.75) and this concludes the proof of the "only if" part.

To prove the converse, we assume first $p>1$, so that then $\gamma_q(T,d)<\infty$. Moreover, since ν is given by (11.12) where m is supported by G (so that $\gamma(0)=1$ a.s. for m),

$$\int_{|\beta(0)|>a} |\beta(0)|\mathrm{d}\nu(\beta) = \int_{x^{-1/p}\ge a} x^{-1/p}\mathrm{d}x < \infty\ ,$$

and (11.24) used with $w=\infty$ completes the proof in that case. When $p=1$, and since the set Ω of Theorem 11.2.3 does not depend on the finite set S, denoting by $\mathcal{S}$ the essential supremum of the family $\sup_{t\in S}|X_t|$ over the possible choices of S, (11.24) implies that given $\delta>0$ one can find a set Ω with $\mathsf{P}(\Omega^c)<\delta$ and $\mathsf{E}(\mathbf{1}_\Omega\mathcal{S})<\infty$, so that $\mathcal{S}<\infty$ a.s. (which is what we meant by $\sup_{t\in T}|X_t|<\infty$ a.s.) □

Exercise 11.4.6. Deduce from (11.24) the estimate $\mathsf{P}(\mathcal{S}\ge u)\le C(\log u)/u$ for a number C independent of u.

This is not sharp. In [7] the author proves that in fact $\mathsf{P}(\mathcal{S}\ge u)\le C/u$ for a number C independent of u. One should of course wonder whether this inequality can be obtained as a consequence of a general theorem such as here.

Exercise 11.4.7. In the case $p=1$, prove that the condition $\sum_n \epsilon_n<\infty$ is equivalent to the condition $\gamma_\infty(T,d)<\infty$ where the quantity $\gamma_\infty(T,d)$ is defined in (8.30).

11.5 Proof of the Decomposition Theorem

The proof of Theorem 11.2.6 requires the most work. We delay this proof until the next section, and in the present section we prove all the other statements.

Proof of Theorem 11.2.8. Let us denote by E^τ and P^τ expectation and probability given the sequence $(\tau_i)_{i\geq 1}$. We will prove that

$$\mathsf{E}^\tau \sup_{t\in T} X_t \leq L\Big(\frac{1}{\sqrt{\alpha_-}}\gamma_2(T,d_2)+\gamma_1(T,d_\infty)\Big)\,. \tag{11.76}$$

Taking expectation and using Lemma 11.4.5 then finishes the proof.

To prove (11.76), consider $s,t\in T$, and let $G_i=\varepsilon_i R_i(Y_i(s)-Y_i(t))$. Thus $|G_i|\leq d_\infty(s,t)$, and the right-hand side of (11.64) used for $h(\beta)=(\beta(s)-\beta(t))^2$ yields $\sum_{i\geq 1}\mathsf{E}^\tau G_i^2\leq d_2(s,t)^2/\alpha_-$. Bernstein's inequality (Lemma 4.3.4) implies

$$\mathsf{P}^\tau\Big(\big|\sum_{i\geq 1} G_i\big|\geq v\Big)\leq \exp\Big(-\frac{1}{L}\min\big(\frac{v^2\alpha_-}{d_2(s,t)^2}\,,\,\frac{v}{d_\infty(s,t)}\big)\Big)\,,$$

so that (11.76) follows from Theorem 2.2.23. □

In the rest of this section, we always think of T as a set of functions over Ω. The next result is in the spirit of the Giné-Zinn Theorem (Theorem 9.1.10).

Theorem 11.5.1. *We have*

$$\mathsf{E}\sup_{t\in T}|X|_t\leq L\Big(\mathsf{E}\sup_{t\in T}|X_t|+\sup_{t\in T}\int|t(\beta)|\mathrm{d}\nu(\beta)\Big)\,.$$

Proof. As explained, if a Rosinski representation of X_t is $\sum_{i\geq 1}\varepsilon_i R_i t(Y_i)$, a Rosinski representation of $|X|_t$ is $\sum_{i\geq 1}R_i|t(Y_i)|$. We will need to use (11.64) and a minor technical difficulty arises because $1/\alpha_-$ is not integrable. This is why below we consider the first term separately. We write

$$\mathsf{E}\sup_{t\in T}|X|_t=\mathsf{E}\sup_{t\in T}\sum_{i\geq 1}R_i|t(Y_i)|\leq \mathrm{I}+\mathrm{II} \tag{11.77}$$

$$\mathrm{I}=\mathsf{E}\sup_{t\in T}R_1|t(Y_1)|\quad;\quad \mathrm{II}=\mathsf{E}\sup_{t\in T}\sum_{i\geq 2}R_i|t(Y_i)|\,. \tag{11.78}$$

We control the term I as follows:

$$\mathsf{E}\sup_{t\in T}R_1|t(Y_1)|=\mathsf{E}\sup_{t\in T}|\varepsilon_1 R_1 t(Y_1)|\leq \mathsf{E}\sup_{t\in T}\Big|\sum_{i\geq 1}\varepsilon_i R_i t(Y_i)\Big|=\mathsf{E}\sup_{t\in T}|X_t|\,. \tag{11.79}$$

To control the term II, we denote by E^τ expectation at $(\tau_i)_{i\geq 1}$ given. Given the sequence (τ_i), the pairs of r.v.s (R_i,Y_i) are independent. We appeal to (9.22) to get

$$\begin{aligned}\mathsf{E}^\tau \sup_{t\in T}\sum_{i\geq 2} R_i|t(Y_i)| &\leq \sup_{t\in T}\sum_{i\geq 2}\mathsf{E}^\tau R_i|t(Y_i)| + 2\mathsf{E}^\tau \sup_{t\in T}\Big|\sum_{i\geq 2}\varepsilon_i R_i t(Y_i)\Big| \\ &\leq \sup_{t\in T}\sum_{i\geq 2}\mathsf{E}^\tau R_i|t(Y_i)| + 2\mathsf{E}^\tau \sup_{t\in T}|X_t|\,. \end{aligned} \tag{11.80}$$

An obvious adaptation of the right-hand side inequality of (11.64) gives

$$\sum_{i\geq 2}\mathsf{E}^\tau R_i|t(Y_i)| \leq \frac{1}{\alpha^*}\int |t(\beta)|\mathrm{d}\nu(\beta)\,, \tag{11.81}$$

where $\alpha^* = \inf_{i\geq 2}\tau_i/i$. Writing

$$\mathsf{P}(\alpha^* < \epsilon) \leq \sum_{i\geq 2}\mathsf{P}(\tau_i \leq \epsilon i) = \sum_{i\geq 2}\mathsf{P}(\mathrm{card}\,\Pi\cap[0,\epsilon_i]\geq i)\,,$$

one sees through simple estimates that $\mathsf{E}(1/\alpha^*)^{3/2} < \infty$. We plug (11.81) in (11.80), we take expectation, and we combine with (11.77), (11.78) and (11.79) to conclude the proof. □

Proof of Theorem 11.2.10. We still think of T as a set of functions on (Ω, m). Without loss of generality we may assume that $0\in T$, so that $\mathsf{E}\sup_{t\in T}|X_t| \leq 2S$ by Lemma 2.2.1.

The main argument consists in constructing a decomposition $T\subset T_1+T_4$, where $\gamma_1(T_1,d_\infty)\leq KS$, $\gamma_2(T_1,d_2)\leq KS$, $0\in T_1$ and

$$\sup_{t\in T_4}\int |t(\beta)|d\nu(\beta) \leq KS\,.$$

Once this is done, it follows from Theorem 11.2.8 and Lemma 2.2.1 that $\mathsf{E}\sup_{t\in T_1}|X_t|\leq KS$. We may assume that $T_4\subset T-T_1$, simply by replacing T_4 by $T_4\cap(T-T_1)$. Thus $\mathsf{E}\sup_{t\in T_4}|X_t| \leq \mathsf{E}\sup_{t\in T_1}|X_t| + \mathsf{E}\sup_{t\in T}|X_t| \leq KS$. Theorem 11.5.1 then implies $\mathsf{E}\sup_{t\in T_4}|X|_t\leq KS$. Finally, the decomposition $X_t = X'_t + X''_t$ is obtained by fixing a decomposition $t = t_1+t_2$ for each t in T with $t_1\in T_1$, $t_2\in T_4$, and setting $X'_t = X_{t_1}$, $X''_t = X_{t_2}$.

To decompose T as above we first use Theorem 11.2.6 to find a number r (depending only on C_0 and δ), an admissible sequence $(\mathcal{A}_n)$ of T and for $A\in\mathcal{A}_n$ an integer $j_n(A)\in\mathbb{Z}$ that satisfies (10.8) i.e.

$$A\in\mathcal{A}_n\,,\, C\in\mathcal{A}_{n-1}\,,\, A\subset C \Rightarrow j_{n-1}(C)\leq j_n(A)\leq j_{n-1}(C)+1 \tag{10.8}$$

and

$$\forall s,t\in A\in\mathcal{A}_n\,,\ \varphi(s,t,r^{j_n(A)-1})\leq 2^{n+2} \tag{11.82}$$

$$\forall t\in T\,,\ \sum_{n\geq 0}2^n r^{-j_n(A_n(t))}\leq KS\,. \tag{11.83}$$

We then use Theorem 5.2.7, with $\mu=\nu$ and $u=L$. We consider the decomposition $T\subset T_1+T_2+T_3$ provided by this theorem. We set $T_4 = T_2+T_3$, so that $T \subset T_1 + T_4$. By (5.26), (5.27) and (11.83) we have $\gamma_2(T_1, d_2) \le KS$ and $\gamma_1(T_1, d_\infty) \le KS$. Moreover (5.28), used for $p = 1$ implies $\|t\|_1 \le KS$ for $t \in T_2$, as is obvious since

$$V^{2j_{n+1}(A_{n+1}(t)) - j_n(A_n(t))}\delta_{n+1}(A_{n+1}(t))^2 = 2^{n+3} r^{-j_n(A_n(t))} \ .$$

To conclude the proof it suffices to show that $\|t\|_1 \le KS$ for $t \in T_3$, and, from (5.29), it suffices to show that $\|t\|_1 \le KS$ for $t = |s|\mathbf{1}_{\{2|s| \ge r^{-j_0(T)}\}}$ and $s \in T$. Now, since $0 \in T$, using (11.82) for $n = 0$ and $A = T$ yields

$$\begin{aligned}\nu(\{\beta \ ; \ |s(\beta)| \ge r^{-j_0(T)}/2\}) &\le 4 \int \big((s r^{j_0(T)})^2 \wedge 1\big) \mathrm{d}\nu \\ &\le 4\varphi(s, 0, r^{j_0(T)}) \le 16 \ .\end{aligned}$$

It follows from condition $H(C_0, \delta)$ that for $v \ge 1$ one has

$$\nu(\{\beta \ ; \ |s(\beta)| \ge v r^{-j_0(T)}/2\}) \le 16 C_0 v^{-\delta - 1} \ ,$$

so that as in (2.5) we obtain $\|t\|_1 \le K r^{-j_0(T)}$, and since $r^{-j_0(T)} \le KS$ by (11.83), that $\|t\|_1 \le KS$. The proof is complete. □

Proof of Theorem 11.2.11. The plan is to use Theorem 5.2.1. The proof is nearly identical to that of Theorem 9.1.12. First we deduce from (11.39) and (11.64) that

$$\forall\, A \in \mathcal{A}_n \ , \ \sum_{i \ge 1} \mathsf{E}^\tau R_i (r^{2j_n(A)} h_A(Y_i)^2 \wedge 1) \le \frac{2^n}{\alpha_-} \ . \tag{11.84}$$

Consider a parameter $u \ge 1$. Then since $\alpha_- \le 1$ Lemma 7.4.3 (b) implies

$$\mathsf{P}^\tau\Bigg(\sum_{i \ge 1} R_i (r^{2j_n(A)} h_A(Y_i)^2 \wedge 1) \le \frac{u 2^n}{\alpha_-}\Bigg) \ge 1 - \exp(-u 2^{n-1}) \ .$$

Consequently for $u \ge L$ the event $\Omega(u)$ given by

$$\forall\, n \ge 0 \ , \ \forall\, A \in \mathcal{A}_n \ , \ \sum_{i \ge 1} R_i (r^{2j_n(A)} h_A(Y_i)^2 \wedge 1) \le \frac{u 2^n}{\alpha_-}$$

satisfies $P(\Omega(u)) \ge 1 - L\exp(-u)$. We observe the fundamental fact: if $s, t \in A$ then

$$\sum_{i \ge 1} R_i (r^{2j_n(A)} (t(Y_i) - s(Y_i))^2) \wedge 1 \le \sum_{i \ge 1} R_i (r^{2j_n(A)} h_A(Y_i)^2) \wedge 1 \ ,$$

and therefore using (5.12) with $p = 1$ we obtain that when $\Omega(u)$ occurs

$$\mathsf{E}_\varepsilon \sup_{t \in T} \Big| \sum_{i \ge 1} \varepsilon_i R_i t(Y_i) \Big| \le \frac{Ku}{\sqrt{\alpha_-}} S + K \sum_{i \ge 1} R_i h_T(Y_i) \mathbf{1}_{\{2h_T(Y_i) \ge r^{-j_0(T)}\}} \ , \tag{11.85}$$

where K depends on r only. Since $\mathsf{P}(\Omega(u)) \geq 1 - L\exp(-u)$, integrating in the r.v.s Y_i yields as usual

$$\mathsf{E}^\tau \sup_{t\in T}\Big|\sum_{i\geq 1}\varepsilon_i R_i t(Y_i)\Big| \leq \frac{K}{\sqrt{\alpha_-}}S + K\sum_{i\geq 1}\mathsf{E}^\tau R_i h_T(Y_i)\mathbf{1}_{\{2h_T(Y_i)\geq r^{-j_0(T)}\}} ,$$

and taking expectation and using Lemma 11.4.5,

$$\mathsf{E}\sup_{t\in T}\Big|\sum_{i\geq 1}\varepsilon_i R_i t(Y_i)\Big| \leq KS + K\sum_{i\geq 1}\mathsf{E}R_i h_T(Y_i)\mathbf{1}_{\{2h_T(Y_i)\geq r^{-j_0(T)}\}} . \quad (11.86)$$

Now (11.63) yields

$$\sum_{i\geq 1}\mathsf{E}R_i h_T(Y_i)\mathbf{1}_{\{2h_T(Y_i)\geq r^{-j_0(T)}\}} = \int h_T(\beta)\mathbf{1}_{\{2h_T(\beta)\geq r^{-j_0(T)}\}}\mathrm{d}\nu(\beta) ,$$

and (11.40) proves that this quantity is $\leq KS$. Combining with (11.86) proves that $\mathsf{E}\sup_{t\in T}|X_t| \leq KS$. □

We have imposed condition (11.40) in order to get clean statement. Its use is simply to control the size of the last term in (11.85). This hypothesis is absolutely inessential: this term is a.s. finite because the sum contains only finitely many non-zero terms. Its size can then be controlled in specific cases through specific methods (such as is required in the case $p = 1$ below).

As an application of Theorem 11.2.11, let us discuss (a part of) Theorem 2.1 of [3] (translated in our language). Consider a measured space (Ω, ν), where ν is σ-finite, and a set T of measurable functions on Ω such that $\int t^2 \wedge 1\mathrm{d}\nu < \infty$ for each $t \in T$. Then as explained this class of functions defines an infinitely divisible process $(X_t)_{t\in T}$. Consider a number $q > 2$ and the conjugate exponent p. Assume that there is a distance d on T such that $\gamma_q(T, d) < \infty$. Assume moreover that for a certain function $\psi(\omega)$ the following condition holds:

$$\forall\, s,t \in T\ ,\ |t(\omega) - s(\omega)| \leq \psi(\omega)d(s,t)\ , \quad (11.87)$$

and that, for all $u > 0$ and a number B we have

$$\nu(\{\psi \geq u\}) \leq \frac{B^p}{u^p} . \quad (11.88)$$

Then

$$\mathsf{E}\sup X_t \leq K(p)B\gamma_q(T,d) . \quad (11.89)$$

To prove this consider an admissible sequence of partitions $(\mathcal{A}_n)$ of T such that $\sup_{t\in T}\sum_{n\geq 0} 2^{n/q}\Delta(A_n(t)) \leq 2\gamma_q(T,d)$. It follows from (11.87) that for $A \in \mathcal{A}_n$ the function $h_A(\omega) := \sup_{s,t}|s(\omega) - t(\omega)|$ satisfies

$$h_A(\omega) \leq \Delta(A,d)\psi(\omega) . \quad (11.90)$$

We then observe that for $u > 0$ we have, using (11.88) in the inequality,

$$\int (u\psi)^2 \wedge 1 \mathrm{d}\nu = \int_0^1 \nu(\{(u\psi)^2 \geq w\})\mathrm{d}w \leq \int_0^1 \frac{(Bu)^p}{w^{p/2}}\mathrm{d}w = (C(p)Bu)^p \ . \tag{11.91}$$

For $A \in \mathcal{A}_n$ let us define $j_n(A)$ as the largest integer for which

$$2^{j_n(A)}\Delta(A,d) \leq \frac{2^{n/p}}{BC(p)} \ .$$

We observe that (11.90) implies that $2^{j_n(A)}h_A \leq u\psi$ for $u = 2^{n/p}/BC(p)$, so that (11.91) implies (11.39). Moreover (11.88) implies that

$$\int \psi \mathbf{1}_{\{\psi \geq u\}} \mathrm{d}\nu \leq K(p)B^p u^{1-p} \ ,$$

so that since $h_T \leq \Delta(T,d)\psi$, using the choice $u = 2^{-j_0(T)-1}\Delta(T,d)^{-1}$ we obtain

$$\int h_T \mathbf{1}_{\{2h_T \geq 2^{-j_0(T)}\}} \mathrm{d}\nu \leq K(p)B^p \Delta(T,d)^{p-1} 2^{j_0(T)(p-1)}$$

and since $2^{j_0(T)}\Delta(T,d) \leq K(p)/B$ by definition of $j_0(T)$, the above integral in bounded by

$$K(p)B\Delta(T,d) \leq S := K(p)B\gamma_q(T,d) \ ,$$

and we have shown that (11.40) holds for this value of S. Moreover, as we have done many times, the definition of $j_n(A)$ proves that (11.41) holds with the same value of S, so that Theorem 11.5.1 proves (11.89). Moreover, using the equivalence (8.32) of the quantity $\gamma_\infty(T,d)$ with the quantity $\gamma^*(T,d)$ of (8.31) basically the same proof works in the case $p = 1$ and $q = \infty$ (except that then one cannot use expectation to control the size of the supremum of the process). Since there is nothing intrinsic in the choice of conditions such as (11.87) and (11.88) one can of course imagine many other statements of the same type.

It should also be mentioned that the bracketing theorem of [8] is already sufficient to prove results such as (11.89).

11.6 Proof of the Main Lower Bound

In this section we prove Theorem 11.2.6. Let us recall that in the proof of Theorem 8.1.1 we used in an essential way that p-stable processes can be represented as a mixture of Gaussian processes. Here we will use in an essential way that infinitely divisible processes can be represented as mixtures of Bernoulli processes. Since we do not understand Bernoulli processes as well as we understand Gaussian processes we can expect that the proof of Theorem 11.2.6 will be more difficult than the proof of Theorem 8.1.1. The proof we present is very much simpler than the original proof of [8], but still requires a real effort. Before we start this proof, we collect a number of simple lemmas, aiming in particular at understanding the behavior of the independent sequence $(R_i Y_i)_{i\geq 1}$.

We denote by P^τ conditional probability given the sequence $(\tau_i)_{i\geq 1}$, and we recall the definition (11.15) of $\varphi(s,t,u)$.

Lemma 11.6.1. *Consider $s,t\in T$.*
(a) Assume that $2\alpha_+ \leq \varphi(s,t,u)$. Then

$$\mathsf{P}^\tau\Bigl(\sum_{i\geq 1}\bigl(R_i u^2(Y_i(s)-Y_i(t))^2\bigr)\wedge 1 \leq \frac{\varphi(s,t,u)}{8\alpha_+}\Bigr) \leq \exp\Bigl(-\frac{\varphi(s,t,u)}{8\alpha_+}\Bigr)\,. \quad (11.92)$$

(b) If $A \geq 4\varphi(s,t,u)/\alpha_-$ then

$$\mathsf{P}^\tau\Bigl(\sum_{i\geq 1}\bigl(R_i u^2(Y_i(s)-Y_i(t))^2\bigr)\wedge 1 \geq A\Bigr) \leq \exp\Bigl(-\frac{A}{2}\Bigr)\,. \quad (11.93)$$

Proof. We set $W_i = (R_i u^2(Y_i(s)-Y_i(t))^2)\wedge 1$ and we define $h(\beta) = (u^2(\beta(s)-\beta(t))^2)\wedge 1$, so that $W_i = h(R_i Y_i)$. Also, the definition of φ implies $\int h(\beta)\mathrm{d}\nu(\beta) = \varphi(s,t,u)$. We then use Lemma 11.4.4 to get, using the assumption $2\alpha_+\leq\varphi(s,t,u)$ of (a) in the first inequality,

$$\frac{1}{2\alpha_+}\varphi(s,t,u) \leq \frac{1}{\alpha_+}\varphi(s,t,u) - 1 \leq \sum_{i\geq 1}\mathsf{E}^\tau W_i \leq \frac{1}{\alpha_-}\varphi(s,t,u)$$

and we use Lemma 7.4.3 to conclude. □

We now explore some consequences of condition $H(C_0,\delta)$. We recall that K denotes a quantity depending on C_0 and δ only.

Lemma 11.6.2. *Assume condition $H(C_0,\delta)$. Then for $s,t\in T$ and $u>0$, we have, for any $v\geq 1$,*

$$\varphi(s,t,uv) \geq \frac{v^{1+\delta}}{K}\varphi(s,t,u)\,.$$

Proof. We write $f(\beta) = |\beta(s)-\beta(t)|$, so that

$$\begin{aligned}\varphi(s,t,u) &= \int\bigl((u^2 f^2(\beta))\wedge 1\bigr)\mathrm{d}\nu(\beta)\\ &= \int_0^1 \nu(\{\beta\,;\,u^2f^2(\beta)\geq x\})\mathrm{d}x\\ &= \int_0^{1/v^2}\nu(\{\beta\,;\,u^2f^2(\beta)\geq x\})\mathrm{d}x + \int_{1/v^2}^1 \nu(\{\beta\,;\,u^2f^2(\beta)\geq x\})\mathrm{d}x\,.\end{aligned}$$

Setting $x = y/v^2$ in the first term and $x = y^2/v^2$ in the second term we get

$$\varphi(s,t,u) = \frac{1}{v^2}\varphi(s,t,uv) + \frac{1}{v^2}\int_1^v 2y\nu\big(\{\beta\,;\ f(\beta) \ge \frac{y}{uv}\}\big)\mathrm{d}y\ . \tag{11.94}$$

Now, for $y \ge 1$, condition $H(C_0,\delta)$ implies

$$\nu\big(\{\beta\,;\ f(\beta) \ge \frac{y}{uv}\}\big) \le C_0 y^{-1-\delta}\nu(\{\beta\,;\ uvf(\beta)\ge 1\}) \le C_0 y^{-1-\delta}\varphi(s,t,uv)\ .$$

Since we assume $\delta < 1$, substitution in (11.94) yields

$$\varphi(s,t,u) \le \frac{K}{v^2}\varphi(s,t,uv)(1+\int_1^v y^{-\delta}\mathrm{d}y) \le Kv^{-1-\delta}\varphi(s,t,uv)\ . \qquad \square$$

Lemma 11.6.3. *Assume condition $H(C_0,\delta)$. Consider $s,t \in T$ and $u > 0$. Set*

$$W_i = R_i|Y_i(s) - Y_i(t)|\ .$$

Then

$$\sum_{i\ge 1} \mathsf{E}^\tau (uW_i\mathbf{1}_{\{uW_i\ge 1\}})^{1+\delta/2} \le \frac{K}{\alpha_-}\varphi(s,t,u)\ . \tag{11.95}$$

Proof. We set $f(\beta) = |\beta(s)-\beta(t)|$, so that $W_i = f(R_iY_i)$. Consider the function $h(\beta) = (uf(\beta)\mathbf{1}_{\{uf(\beta)\ge 1\}})^{1+\delta/2}$, so that $h(R_iY_i) = (uW_i\mathbf{1}_{\{uW_i\ge 1\}})^{1+\delta/2}$. We use Lemma 11.4.4 to see that the left-hand side of (11.95) is bounded by

$$\frac{1}{\alpha_-}\int_{\mathbb{R}^T} h(\beta)\mathrm{d}\nu(\beta) = \frac{1}{\alpha_-}\int_{\{uf(\beta)\ge 1\}} (uf(\beta))^{1+\delta/2}\mathrm{d}\nu(\beta) \le \mathrm{I} + \mathrm{II}$$

where

$$\mathrm{I} = \frac{1}{\alpha_-}\nu(\{\beta\,;\ uf(\beta)\ge 1\}) \le \frac{1}{\alpha_-}\varphi(s,t,u)$$

$$\mathrm{II} = \frac{1}{\alpha_-}\int_1^\infty \nu(\{\beta\,;\ (uf(\beta))^{1+\delta/2} \ge x\})\mathrm{d}x\ .$$

Now, for $x \ge 1$, condition $H(C_0,\delta)$ implies

$$\begin{aligned}\nu(\{\beta\,;\ (uf(\beta))^{1+\delta/2} \ge x\}) &= \nu(\{\beta\,;\ uf(\beta) \ge x^{1/(1+\delta/2)}\}) \\ &\le C_0 x^{-\frac{1+\delta}{1+\delta/2}}\nu(\{\beta\,;\, uf(\beta)\ge 1\})\end{aligned}$$

and since $\nu(\{\beta\,;\ uf(\beta)\ge 1\}) \le \varphi(s,t,u)$ the result follows. $\square$

Lemma 11.6.4. *Consider independent r.v.s $(V_i)_{i\ge 1}$ such that $V_i \ge 0$, and consider $0 < \delta < 2$. Assume that for a certain number $S > 0$ we have*

$$\sum_{i\ge 1} \mathsf{E}V_i^{1+\delta/2} \le S\ . \tag{11.96}$$

Then

$$\mathsf{P}\Big(\sum_{i\ge 1} V_i\mathbf{1}_{\{V_i\ge 1\}} \ge 4S\Big) \le LS^{-\delta/2}\ . \tag{11.97}$$

Proof. We may assume $S \geq 1$, for there is nothing to prove otherwise. We set

$$U_i = V_i \mathbf{1}_{\{1 \leq V_i \leq S\}} - \mathsf{E}(V_i \mathbf{1}_{\{1 \leq V_i \leq S\}}) \, .$$

Since $x\mathbf{1}_{\{x \geq 1\}} \leq x^{1+\delta}$, we have

$$\sum_{i \geq 1} \mathsf{E}(V_i \mathbf{1}_{\{1 \leq V_i \leq S\}}) \leq \sum_{i \geq 1} \mathsf{E} V_i^{1+\delta/2} \leq S,$$

so that

$$\mathsf{P}\Big(\sum_{i \geq 1} V_i \mathbf{1}_{\{1 \leq V_i \leq S\}} \geq 4S\Big) \leq \mathsf{P}\Big(\sum_{i \geq 1} U_i \geq 3S\Big) \, .$$

Since

$$\mathsf{P}\Big(\sum_{i \geq 1} V_i \mathbf{1}_{\{V_i \geq 1\}} \geq 4S\Big) \leq \mathsf{P}\Big(\sum_{i \geq 1} V_i \mathbf{1}_{\{1 \leq V_i \leq S\}} \geq 4S\Big) + \sum_{i \geq 1} \mathsf{P}(V_i \geq S) \, ,$$

we obtain

$$\mathsf{P}\Big(\sum_{i \geq 1} V_i \mathbf{1}_{\{V_i \geq 1\}} \geq 4S\Big) \leq \mathsf{P}\Big(\sum_{i \geq 1} U_i \geq 3S\Big) + \sum_{i \geq 1} \mathsf{P}(V_i \geq S) \, .$$

Now Markov's inequality yields $\mathsf{P}(V_i \geq S) \leq S^{-(1+\delta/2)} \mathsf{E} V_i^{1+\delta/2}$, so that

$$\sum_{i \geq 1} \mathsf{P}(V_i \geq S) \leq \frac{1}{S^{1+\delta/2}} \sum_{i \geq 1} \mathsf{E} V_i^{1+\delta/2} \leq S^{-\delta/2} \, .$$

Also, we note that $|U_i| \leq S$, so that since $1 - \delta/2 \geq 0$ we have, using in the second inequality that $(a+b)^c \leq L(a^c + b^c)$ for $c \leq 2$, and since $\mathsf{E}(V_i \mathbf{1}_{\{V_i \geq 1\}})^{1+\delta/2} \leq \mathsf{E} V_i^{1+\delta/2}$,

$$\mathsf{E} U_i^2 \leq S^{1-\delta/2} \mathsf{E}|U_i|^{1+\delta/2} \leq L S^{1-\delta/2} \mathsf{E} V_i^{1+\delta/2} \, . \tag{11.98}$$

Thus using (11.98) and (11.96) we get

$$\mathsf{P}\Big(\sum_{i \geq 1} U_i \geq 3S\Big) \leq \frac{1}{9S^2} \mathsf{E}\Big(\sum_{i \geq 1} U_i\Big)^2 = \frac{1}{9S^2} \sum_{i \geq 1} \mathsf{E} U_i^2 \leq L S^{-\delta/2} \, . \qquad \square$$

Corollary 11.6.5. *Assume condition $H(C_0, \delta)$. Consider $s, t \in T$ and $u > 0$. Then for $S \geq K\varphi(s,t,u)/\alpha_-$ we have*

$$\mathsf{P}^\tau\Big(u \sum_{i \geq 1} R_i |Y_i(s) - Y_i(t)| \mathbf{1}_{\{u|Y_i(s) - Y_i(\tau)| \geq 1\}} \geq 4S\Big) \leq L S^{-\delta/2} \, . \tag{11.99}$$

Proof. Let $W_i = R_i |Y_i(s) - Y_i(t)|$. Then, since $R_i \in \{0, 1\}$, we get

$$V_i := u W_i \mathbf{1}_{\{uW_i \geq 1\}} = u R_i |Y_i(s) - Y_i(t)| \mathbf{1}_{\{u|Y_i(s) - Y_i(\tau)| \geq 1\}} \, .$$

Now (11.95) implies

$$\sum_{i\geq 1} \mathsf{E}^\tau V_i^{1+\delta/2} \leq \frac{K}{\alpha_-}\varphi(s,t,u) ,$$

and the conclusion follows from (11.97) used for P^τ. □

Lemma 11.6.6. *Consider a measure μ on $\mathbb{R}$ such that $\int(x^2\wedge 1)\mathrm{d}\mu(x) < \infty$, and a r.v. X that satisfies*

$$\forall \alpha \in \mathbb{R}\,,\; \mathsf{E}\exp i\alpha X = \exp\Big(-\int_{\mathbb{R}}(1-\cos\alpha x)\mathrm{d}\mu(x)\Big)\ . \tag{11.100}$$

Then

$$\int_{\mathbb{R}}\Big(\Big(\frac{x}{2\mathsf{E}|X|}\Big)^2 \wedge 1\Big)\mathrm{d}\mu(x) \leq L\ . \tag{11.101}$$

Proof. Since $\cos x \geq 1-|x|$, whenever $0\leq\alpha\leq 1/2\mathsf{E}|X|$ we have

$$\mathsf{E}\cos\alpha X \geq 1-\alpha\mathsf{E}|X| \geq \frac{1}{2}\ .$$

Now (11.100) implies that for such α,

$$\frac{1}{2} \leq \mathsf{E}\cos\alpha X = \exp\Big(-\int_{\mathbb{R}}(1-\cos\alpha x)\mathrm{d}\mu(x)\Big)\ ,$$

and hence

$$\int_{\mathbb{R}}(1-\cos\alpha x)\mathrm{d}\mu(x) \leq \log 2\ .$$

Averaging the previous inequality over $0\leq\alpha\leq 1/2\mathsf{E}|X|$, we get

$$\int_{\mathbb{R}}\left(1-\frac{\sin(x/(2\mathsf{E}|X|))}{x/(2\mathsf{E}|X|)}\right)\mathrm{d}\mu(x) \leq \log 2$$

and the result since $y^2\wedge 1 \leq L(1-(\sin y)/y)$. □

The proof of Theorem 11.2.6 will rely upon the application of Theorem 10.1.2 to suitable functionals, and we turn to the task of defining these. Consider an integer $\kappa\geq 6$, to be determined later, and $r=2^{\kappa-4}$. Consider the maps

$$\varphi_j(s,t) = \varphi(s,t,r^j)$$

for $j\in\mathbb{Z}$. From Lemma 11.6.2, we note that

$$\varphi_{j+1}(s,t) \geq \frac{r^{1+\delta}}{K}\varphi_j(s,t)\ . \tag{11.102}$$

Given the sequences $(\tau_i)_{i\geq 1}$, $(Y_i)_{i\geq 1}$, to each $t\in T$ we can associate the sequence $\mathcal{S}(t) = (R_iY_i(t))_{i\geq 1}$. To a subset A of T we can associate $\mathcal{S}(A) =$

$\{\mathcal{S}(t)\,;\, t \in A\}$. This is a random set of sequences. Since the process (11.32) is distributed like (X_t), we have the identity

$$\mathsf{E}\sup_{t\in T} X_t = \mathsf{E}b(\mathcal{S}(T)) \;, \tag{11.103}$$

where the notation b is defined in (5.1). We recall the notation (5.101). For $j \in \mathbb{Z}$ we consider the map $\Psi_j : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}\times\mathbb{N})$ given by $\Psi_j(t) = (\psi_{\ell r^{-j},(\ell+1)r^{-j}}(t_i))_{(i,\ell)}$. For a set of sequences U, we use the notation $b_j(U) = b(\Psi_j(U))$. Obvious changes to the proof of Proposition 5.6.2 show the fundamental fact that

$$b(U) \ge b_j(U) \ge b_{j+1}(U) \;. \tag{11.104}$$

We will perform some constructions with the sets $\mathcal{S}(T)$. These sets depend on the randomness of τ_i and Y_i. We must expect that, given this randomness, at least for some values of t the sequence $\mathcal{S}(t)$ will not be well behaved, and we must introduce a device to remove such sequences. We define a *random subset* Z of T as a subset of T that depends on the sequences $(\tau_i)_{i\ge1}$ and $(Y_i)_{i\ge1}$ and is such that for each $t \in T$ the set $\{t \in Z\}$ is an event (i.e. is measurable). We should think of a random subset of T as a (small) set of badly behaved points.

Let us fix once and for all a number α such that $\mathsf{P}(\Omega_0) \ge 3/4$, where

$$\Omega_0 = \Big\{\frac{1}{\alpha} \le \alpha_- \le \alpha_+ \le \alpha\Big\} \;. \tag{11.105}$$

Consider two decreasing sequences $c(n), d(n) > 0$, tending to 0. These will be determined after we make some calculations, but we should think of them as being fixed once and for all. Given a probability measure μ on a T, we first define the number $F_{n,j}(\mu)$ as the supremum of the numbers c with the following property: there exists a random subset Z of T with $\mathsf{E}\mu(Z) \le d(n)$ for which

$$\forall\, U \subset T \;,\; U \cap Z = \emptyset \;,\; \mu(U) \ge c(n) \Rightarrow b_j(\mathcal{S}(U)) \ge c\mathbf{1}_{\Omega_0} \;. \tag{11.106}$$

In words, this means that the set Z of badly behaved points is small in average, and that, when Ω_0 occurs, any subset U of T which is not too small and contains no badly behaved points satisfies $b_j(\mathcal{S}(U)) \ge c$.

Given a subset A of T, we then define

$$F_{n,j}(A) = \sup\{F_{n,j}(\mu)\;;\; \mu(A) = 1\} \;.$$

This is obviously an increasing function of A. Let us make a simple observation.

Lemma 11.6.7. *If $c < F_{n,j}(A)$ we can find a probability measure μ on A and a random set $Z \subset A$ with $\mathsf{E}\mu(Z) \le d(n)$ for which (11.106) holds.*

The new information here is that $Z \subset A$.

Proof. By definition of the functional $F_{n,j}$ we can find a probability measure μ on A and a random set $Z \subset T$ with $\mathsf{E}\mu(Z) \le d(n)$ for which (11.106) holds (so the point is to prove that we can take $Z \subset A$). We then have

$$\forall U \subset T \,,\; U \cap (Z \cap A) = \emptyset \,,\; \mu(U) \ge c(n) \Rightarrow b_j(\mathcal{S}(U)) \ge c\mathbf{1}_{\Omega_0} \,, \qquad (11.107)$$

as is seen by applying (11.106) to $U \cap A$ rather than U. Consequently, (11.106) still holds if we replace Z by $Z \cap A$. □

Since we assumed $c(n+1) \le c(n)$ and $d(n+1) \le d(n)$, it is obvious that $F_{n+1,j} \le F_{n,j}$; and it follows from (11.104) that $F_{n,j+1} \le F_{n,j}$.

Lemma 11.6.8. *If $d(n) \le 1/8$ and $c(n) \le 1/2$, then $F_{n,j}(T) \le 2\mathsf{E}\sup_{t\in T} X_t$ for all $j \in \mathbb{Z}$.*

Proof. Consider $c < F_{n,j}(T)$, a probability measure μ on T with $F_{n,j}(\mu) > c$ and a random subset Z of T with $\mathsf{E}\mu(Z) \le d(n)$ that satisfy (11.106). Since $d(n) \le 1/8$, then $\mathsf{P}(\mu(Z) \le 1/2) \ge 3/4$ and thus $\mathsf{P}(\Omega_1) \ge 1/2$ where $\Omega_1 = \Omega_0 \cap \{\mu(Z) \le 1/2\}$. Now, when $\mu(Z) \le 1/2$ we have $\mu(T \setminus Z) \ge 1/2 \ge c(n)$ and (11.106) implies $b_j(\mathcal{S}(T\setminus Z)) \ge c\mathbf{1}_{\Omega_1}$ and thus

$$b_j(\mathcal{S}(T)) \ge c\mathbf{1}_{\Omega_1} \,.$$

Since $b_j(\mathcal{S}(T)) \ge 0$, taking expectation and since $\mathsf{P}(\Omega_1) \ge 1/2$ we get $c \le 2\mathsf{E}b_j(\mathcal{S}(T))$, and using (11.104) that $c \le 2\mathsf{E}b(\mathcal{S}(T)) = 2\mathsf{E}\sup_{t\in T} X_t$. □

Lemma 11.6.9. *Assuming condition $H(C_0, \delta)$, and $r \ge K$, there exists $j = j_0 \in \mathbb{Z}$ such that*

$$\forall s,t \in T \,,\; \varphi_{j-1}(s,t) \le 1 \qquad (11.108)$$

$$r^{-j} \le 4r\mathsf{E}\sup_{t\in T} X_t \,. \qquad (11.109)$$

Proof. Consider $s, t \in T$. By Lemma 2.2.1 we have

$$\mathsf{E}|X_s - X_t| \le 2\mathsf{E}\sup_{t\in T} X_t := 2S \,.$$

Defining μ as the image of ν under the map $\beta \mapsto \beta(s) - \beta(t)$, (11.11) implies that the r.v. $X = X_s - X_t$ satisfies (11.100). From (11.101) we get that $\varphi(s,t,1/(4S)) \le L$. Consider the largest integer j such that $r^{-j} \ge 4S$, so that $\varphi_j(s,t) \le L$ and $r^{-j} \le 4rS$. It then follows from (11.102) that if $r \ge K$ we have $\varphi_{j-1}(s,t) \le 1$. □

Let us observe that for any fixed number u, $\delta(s,t) = \varphi(s,t,u)$ is the square of a distance on T, so that for elements s, s' and t of T,

$$\varphi(s,s',u)^{1/2} \le \varphi(s,t,u)^{1/2} + \varphi(s',t,u)^{1/2} \,,$$

and the inequality $(a+b)^2 \le 2(a^2+b^2)$ implies

$$\varphi(s,s',u) \le 2(\varphi(s,t,u) + \varphi(s',t,u)) . \tag{11.110}$$

The following is the crucial step of the proof of Theorem 11.2.6. We recall that $r = 2^{\kappa-2}$.

Proposition 11.6.10. *There exists a number K_1, depending on C_0 and δ only, and sequences $(d(n))_{n\ge0}$ and $(c(n))_{n\ge0}$, also depending on C_0 and δ only, tending to 0 as $n\to\infty$, such that if $n \ge K_1$ and $\kappa \ge K_1$, the functionals $K_1 F_{n,j}$ satisfy the growth condition of Definition 10.1.1.*

Proof of Theorem 11.2.6. We choose $n_0 \ge K_1$ large enough that $d(n_0) \le 1/8$ and $c(n_0) \le 1/2$, and depending only on C_0 and δ. Consider the value of j_0 constructed in Lemma 11.6.9, so that $2^{n_0} r^{-j_0} \le K \mathsf{E} \sup_{t\in T} X_t$ by (11.109) and $F_{n_0,j_0}(T) \le L \mathsf{E} \sup_{t\in T} X_t$ by Lemma 11.6.8. We then apply Theorem 10.1.2 with these values of j_0 and n_0. We observe that for $n \ge n_0$ (11.26) follows from (10.10) and (11.110). Finally for $A \in \mathcal{A}_n$ with $n \le n_0$ we define $j_n(A) = j_0 - 1$, and (11.108) shows that (11.26) remains true for $n \le n_0$. This choice implies that $\sum_{n\le n_0} 2^n r^{-j_n(A_n(t))} \le K 2^{n_0} r^{-j_0+1} \le K 2^{n_0} r^{-j_0}$ and (11.27) follows from (10.9). This completes the proof of Theorem 11.2.6. □

Proof of Proposition 11.6.10. Let us assume that we are given points $(t_\ell)_{\ell\le m}$, as in (10.3), where $m = N_n$, and consider sets $H_\ell \subset B_{j+1}(t_\ell, 2^{n+\kappa})$.

The basic idea is that we want to apply (5.76) to the sets $H'_\ell := \Psi_j(\mathcal{S}(H_\ell))$. After all, the points t_ℓ for $\ell \le m$ are far from each other, so that the points $u_\ell := \Psi_j(\mathcal{S}(t_\ell))$ should also be far from each other. Also, since the points of H_ℓ are close to t_ℓ, the points of H'_ℓ should be close to u_ℓ. Unfortunately, none of this is literally true:

(a) It is not always true that the points $u_\ell := \Psi_j(\mathcal{S}(t_\ell))$ for $\ell \le m$ are far from each other.

(b) It is not true that all the points of H'_ℓ are close to u_ℓ.

Fortunately, sufficiently of it is true:

(c) It is very likely that sufficiently many of the points u_ℓ are far from each other, so that it will suffice to work using only the corresponding values of ℓ.

(d) For most of the points t of H_ℓ it is true that $\Psi_j(\mathcal{S}(t))$ is close to u_ℓ.

The exceptional bad points of H_ℓ will be absorbed into a suitable random set (which is the main purpose of introducing random sets).

We turn to the proof of (c) above. Since $\varphi_j(t_\ell, t_{\ell'}) \ge 2^n$ for $\ell \ne \ell'$, using (11.92) for $u = r^j$ yields

$$\mathsf{P}^\tau\Big(\sum_{i\ge1}\big(R_i(Y_i(t_\ell) - Y_i(t_{\ell'}))^2\big) \wedge r^{-2j} \le \frac{2^n r^{-2j}}{8\alpha_+}\Big) \le \exp\Big(-\frac{2^n}{8\alpha_+}\Big) .$$

Recalling (11.105), we have

$$\mathsf{P}\Big(\Omega_0 \cap \Big\{\sum_{i\geq 1}\big(R_i(Y_i(t_\ell)-Y_i(t_{\ell'}))^2\big)\wedge r^{-2j} \leq \frac{2^n r^{-2j}}{8\alpha}\Big\}\Big) \leq \exp\Big(-\frac{2^n}{8\alpha}\Big)\,. \tag{11.111}$$

It follows from (5.111) that

$$|x-y|^2 \wedge c^2 \leq 4\sum_{\ell\in Z}|\psi_{c\ell,c(\ell+1)}(s_i)-\psi_{c\ell,c(\ell+1)}(t_i)|^2$$

and using this for $c = r^{-j}$ implies

$$\sum_{i\geq 1}\big(R_i(Y_i(t_\ell)-Y_i(t_{\ell'}))^2\big)\wedge r^{-2j} \leq 4\|\Psi_j(\mathcal{S}(t_\ell))-\Psi_j(\mathcal{S}(t_{\ell'}))\|_2^2\,.$$

Let us fix a number p once and for all such that $2^{-p} \leq 1/(32\alpha)$. Then (11.111) implies (rather roughly)

$$\mathsf{P}\Big(\Omega_0 \cap \{\|\Psi_j(\mathcal{S}(t_\ell))-\Psi_j(\mathcal{S}(t_{\ell'}))\|_2 \leq 2^{(n-p)/2}r^{-j}\}\Big) \leq \exp(-2^{n-p+2})\,.$$

Consider the event Ω_1 defined as

$$\Omega_0 \cap \{\exists \ell < \ell' \leq N_{n-p}\,;\, \|\Psi_j(\mathcal{S}(t_\ell))-\Psi_j(\mathcal{S}(t_{\ell'}))\|_2 \leq 2^{(n-p)/2}r^{-j}\}\,.$$

Then

$$\mathsf{P}(\Omega_1) \leq N_{n-p}^2\exp(-2^{n-p+2}) \leq \exp(-2^{n-p+1})\,.$$

In words, even though (a) fails, with overwhelming probability at least the first N_{n-p} of the points u_ℓ are at mutual distances $\geq 2^{(n-p)/2}r^{-j}$. Consider the random subset Z'' defined by $Z'' = T$ if Ω_1 occurs and $Z'' = \emptyset$ otherwise. The purpose of this random set is to exclude the event Ω_1. For any probability μ we have

$$\mathsf{E}\mu(Z'') \leq \mathsf{P}(Z'' \neq \emptyset) = \exp(-2^{n-p+1})\,. \tag{11.112}$$

Next we turn to the realization of (d) above. By definition of B_j, for $t \in H_\ell \subset B_{j+1}(t_\ell, 2^{n+\kappa})$, we have $\varphi_{j+1}(t,t_\ell) \leq 2^{n+\kappa}$. Since $r = 2^{\kappa-2}$, (11.102) implies $\varphi(s,t,r^j) = \varphi_j(t,t_\ell) \leq K2^n r^{-\delta}$. Now we use (11.93) given Ω_0, with $A = K2^n r^\delta$ and since on the event Ω_0 it holds that $\alpha_- \geq 1/\alpha$ we obtain

$$\mathsf{P}\Big(\mathbf{1}_{\Omega_0}\sum_{i\geq 1}\big(R_i(Y_i(t)-Y_i(t_\ell))^2\big)\wedge r^{-2j} \geq K2^n r^{-\delta}r^{-2j}\Big) \leq \exp(-2^n r^{-\delta})\,. \tag{11.113}$$

Using similarly (11.99) with $S = K2^n r^{-\delta}$ and $u = r^j$ we get

$$\mathsf{P}\Big(\mathbf{1}_{\Omega_0}\sum_{i\geq 1}R_i|Y_i(t)-Y_i(t_\ell)|\mathbf{1}_{\{|Y_i(t)-Y_i(t_\ell)|\geq r^{-j}\}} \geq K2^n r^{-\delta}r^{-j}\Big)$$
$$\leq K2^{-n\delta/2}r^{\delta^2/2}\,. \tag{11.114}$$

Consider K_1 large enough that (11.113) and (11.114) hold for $K = K_1$. Let us say that a point $t \in \bigcup_{\ell \le m} H_\ell$ is *regular* if, when ℓ is the index for which $t \in H_\ell$, we have

$$\sum_{i\ge 1}\big(R_i(Y_i(t)-Y_i(t_\ell))^2\big)\wedge r^{-2j} \le K_1 2^n r^{-\delta} r^{-2j} \tag{11.115}$$

and

$$\sum_{i\ge 1} R_i|Y_i(t)-Y_i(t_\ell)|\mathbf{1}_{\{|Y_i(t)-Y_i(t_\ell)|\ge r^{-j}\}} \le K_1 2^n r^{-\delta} r^{-j}\ . \tag{11.116}$$

Thus, (11.113) and (11.114) imply that the probability that Ω_0 occurs and that a given point t is not regular is at most $\exp(-2^n r^{-\delta}) + K2^{-n\delta/2} r^{\delta^2/2}$.

Consider the random set Z''' defined as

$$Z''' = \Big\{t \in \bigcup_{\ell\le m} H_\ell\ ;\ t \text{ is not regular}\Big\}\,,$$

if Ω_0 occurs and $Z''' = \emptyset$ otherwise. Using Fubini's theorem, for each probability measure μ on T we get

$$\mathsf{E}\mu(Z''') = \int \mathsf{P}(\{t\in Z'''\})\mathrm{d}\mu(t) \le \exp(-2^n r^{-\delta}) + K2^{-n\delta/2} r^{\delta^2/2}\ . \tag{11.117}$$

Now we turn to the proof of the growth property itself. Consider $c < \min_{\ell\le m} F_{n+1,j+1}(H_\ell)$. Since $F_{n+1,j+1} \le F_{n+1,j}$, using Lemma 11.6.7 we can find for each $\ell \le m$ a probability measure μ_ℓ with $\mu_\ell(H_\ell) = 1$, a random subset Z_ℓ of H_ℓ with $\mathsf{E}(\mu_\ell(Z_\ell)) \le d(n+1)$ satisfying

$$U \subset H_\ell\backslash Z_\ell\ ,\ \mu_\ell(U) \ge c(n+1) \Rightarrow b_j(\mathcal{S}(U)) \ge c\mathbf{1}_{\Omega_0}\ . \tag{11.118}$$

Assuming $n \ge p$, we define

$$\mu := \frac{1}{N_{n-p}}\sum_{\ell\le N_{n-p}}\mu_\ell\ ;\ Z' := \bigcup_{\ell\le N_{n-p}} Z_\ell\ .$$

Since $\mu_\ell(H_\ell) = 1$ and the sets H_ℓ are disjoint, for any set $Y \subset T$ we have

$$\mu(Y) = \frac{1}{N_{n-p}}\sum_{\ell\le N_{n-p}}\mu_\ell(Y\cap H_\ell)\ . \tag{11.119}$$

In particular since $Z_\ell \subset H_\ell$,

$$\mu(Z') = \frac{1}{N_{n-p}}\sum_{\ell\le N_{n-p}}\mu_\ell(Z_\ell)$$

and thus $\mathsf{E}\mu(Z') \le d(n+1)$. We consider the random set $Z = Z' \cup Z'' \cup Z'''$, so that, combining with (11.112) and (11.117),

$$\mathsf{E}\mu(Z) \le d(n+1)+\exp(-2^{n-p+1})+\exp(-2^n r^{-\delta})+K2^{-n\delta/2}r^{\delta^2/2} \ . \quad (11.120)$$

This finishes the construction of the appropriate random set, and we turn to the question of bounding $F_{n,j}(\bigcup_{\ell \le m} H_\ell)$ from below. Consider a set $U \subset \bigcup_{\ell \le m} H_\ell$, with $U \cap Z = \emptyset$, and assume that

$$\mu(U) \ge c(n+1) + \frac{1}{N_{n-p-1}} \ . \quad (11.121)$$

Our goal is to bound $b_j(\mathcal{S}(U))$ from below. Let us define

$$I = \{\ell \le N_{n-p} \ ; \ \mu_\ell(U) \ge c(n+1)\} \ .$$

Using (11.119) for $Y = U$ implies

$$\mu(U) \le \frac{1}{N_{n-p}} \mathrm{card} I + c(n+1) \ ,$$

so that (11.121) implies $\mathrm{card}\, I \ge N_{n-p}/N_{n-p-1} \ge N_{n-p-1}$. Let $U_\ell = U \cap H_\ell$, so that $\mu_\ell(U) = \mu_\ell(U_\ell)$, and since $U_\ell \cap Z_\ell \subset U \cap Z = \emptyset$, (11.118) yields

$$\forall \ell \in I \ , \ b_j(\mathcal{S}(U_\ell)) \ge c\mathbf{1}_{\Omega_0} \ .$$

We define $H'_\ell = \Psi_j(\mathcal{S}(U_\ell))$, so that $b(H'_\ell) = b_j(\mathcal{S}(U_\ell))$. We define $u_\ell = \Psi_j(\mathcal{S}(t_\ell))$. Since $U \cap Z'' = \emptyset$, U consists only of regular points, so that (11.115), (11.116) and (5.112) imply that if $t \in U_\ell$ then

$$\|\Psi_j(\mathcal{S}(t)) - \Psi_j(\mathcal{S}(t_\ell))\|_2^2 \le K r^{-\delta} 2^n r^{-2j} \ ,$$

so that $\Psi_j(\mathcal{S}(t)) \in B(u_\ell, K r^{-\delta/2} 2^{n/2} r^{-j})$ and consequently

$$H'_\ell \subset B(u_\ell, K r^{-\delta/2} 2^{n/2} r^{-j}) \ . \quad (11.122)$$

Denoting by K_0 the constant of (11.122) and L_0 the constant of (5.76), we now choose once and for all $r = 2^{\kappa-2}$ where κ is the smallest integer such that $K_0 r^{-\delta/2} \le 2^{-p/2}/L_0$, so that (11.122) implies

$$H'_\ell \subset B(u_\ell, 2^{(n-p)/2} r^{-j}/L_0) \ . \quad (11.123)$$

When Ω_1 occurs we have $Z'' = T$. Since $U \ne \emptyset$ and $U \cap Z'' = \emptyset$, the event Ω_1 does not occur, i.e.

$$\forall \ell < \ell' \le N_{n-p} \ , \ \|u_\ell - u_{\ell'}\|_2 \ge 2^{(n-p)/2} r^{-j} \ .$$

We appeal to (5.76) with $m = N_{n-p-1}$ and $a = r^{-j} 2^{(n-p)/2}$ to obtain

$$b\Big(\bigcup_{\ell \in I} H'_\ell\Big) \ge \Big(c + \frac{1}{K_1} 2^n r^{-j}\Big) \mathbf{1}_{\Omega_0} \ .$$

Since

$$b\Big(\bigcup_{\ell\in I} H'_\ell\Big) = b\Big(\Psi_j\big(\mathcal{S}(\bigcup_{\ell\in I} U_\ell)\big)\Big) \le b\big(\Psi_j(\mathcal{S}(U)\big) = b_j(\mathcal{S}(U))\,,$$

this implies

$$b_j(\mathcal{S}(U)) \ge b\Big(\bigcup_{\ell\in I} H'_\ell\Big) \ge \Big(c + \frac{1}{K_1}2^n r^{-j}\Big)\mathbf{1}_{\Omega_0}\,. \tag{11.124}$$

Motivated by (11.121), we now define

$$c(n) = \sum_{q\ge n} \frac{1}{N_{q-p-1}}$$

and, motivated by (11.120),

$$d(n) = \sum_{q\ge n}\Big(\exp(-2^{q-p+1}) + \exp(-2^q r^{-\delta}) + K2^{-q\delta/2} r^{\delta^2/2}\Big)\,.$$

With these choices (11.120) implies that $\mathsf{E}\mu(Z) \le d(n)$, and (11.121) means that $\mu(U) \ge c(n)$. We have proved that this latter conditions, together with $U \subset \bigcup_{\ell\le m} H_\ell$ and $Z \cap U = \emptyset$ implies (11.124). The definition of the functionals $F_{n,j}$ then yields

$$F_{n,j}(\mu) \ge c + \frac{1}{K_1}2^n r^{-j}\,,$$

and therefore

$$F_{n,j}\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge c + \frac{1}{K_1}2^n r^{-j}\,.$$

Thus the growth condition (10.4) holds for the functionals $K_1 F_{n,j}$, and this completes the proof of Proposition 11.6.10 and Theorem 11.2.6. ☐

References

1. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
2. Marcus, M.B., Rosinski, J.: Sufficient conditions for boundedness of moving averages. In: Stochastic Inequalities and Applications. Progr. Probab., vol. 56, pp. 113–128. Birkhäuser, Basel (2003)
3. Marcus, M.B., Rosinski, J.: Continuity and boundedness of infinitely divisible processes: a Poisson point process approach. J. Theor. Probab. **18**, 109–160 (2005)
4. Rosinski, J.: On series representations of infinitely divisible random vectors. Ann. Probab. **18**, 405–430 (1990)

5. Rosinski, J.: Series representation of Lévy processes from the perspective of point processes. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I. (eds.) Lévy Processes, Theory and Applications, pp. 401–415. Birkhäuser, Basel (2001)
6. Rosinski, J.: Representations of Poissonian processes and applications. Manuscript in preparation (2012)
7. Talagrand, M.: Characterization of almost surely continuous 1-stable random Fourier series and strongly stationary processes. Ann. Probab. **18**, 85–91 (1990)
8. Talagrand, M.: Regularity of infinitely divisible processes. Ann. Probab. **21**, 362–432 (1993)

12. The Fundamental Conjectures

12.1 Introduction

Proving the conjectures presented in the present chapter would revolutionize our understanding of stochastic processes in the abstract setting (as developed in this work). The author believes that these conjectures are an excellent long term research project. Some have been formulated very recently, so it is of course a bit dangerous to include them here (even though the author has thinly disguised the most outrageous of these conjectures as research problems). They are however so beautiful that this must be done. Just in case one of these conjectures has been clumsily (and wrongly) formulated let us explain the basic idea (which is likely to be more robust than any precise formulation). It is simply that to bound a process of the type of those considered below, there seems to be only three ways to proceed: using the union bound, taking a convex hull, using positivity, and, of course the combinations of these. Maybe of course this is wishful thinking, and (as e.g. in the case of the example (8.53) about Gaussian chaos) there are fundamentally different methods to bound these processes, but they remain to be imagined.

12.2 Selector Processes

Given a number $0 < \delta < 1$, we consider i.i.d. r.v.s $(\delta_i)_{i \le M}$ with

$$\mathsf{P}(\delta_i = 1) = \delta \; ; \; \mathsf{P}(\delta_i = 0) = 1 - \delta \,. \tag{12.1}$$

The r.v.s δ_i are often called selectors, because they allow to select a random subset J of $\{1, \ldots, M\}$ of cardinal about δM, namely the set $\{i \le M; \delta_i = 1\}$. The most interesting case is $\delta \le 1/2$. These variables will be used in a fundamental way in Sections 16.5 and 16.6. The importance of selector processes goes however well beyond these specific uses. This is because in some sense the preceding method approximates the construction of N independent random points $(X_j)_{j \le N}$ in a probability space (Ω, μ). To explain this, let us assume for clarity that μ has no atoms. Let us then divide Ω into M small pieces Ω_i of equal measure, where M is much larger than N. Consider then selectors $(\delta_i)_{i \le M}$ where δ in (12.1) is given by $\delta = N/M$. When $\delta_i = 1$ let us choose

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*,
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of
Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_12,

a point Y_i in Ω_i. Since Ω_i is small, how we do this is not very important, but let us be perfectionist and choose Y_i according to the conditional probability that $Y_i \in \Omega_i$. Then the collection of points $\{Y_i; \delta_i = 1\}$ resembles a collection $(X_j)_{j\le N'}$ where the points (X_j) are independent uniform in Ω and where $N' = \sum_{i\le M} \delta_i$. For M large, N' is nearly a Poisson r.v. of expectation $\delta M = N$. So a more precise statement is that selector processes approximate the operation of choosing N' independent random points in a probability space, where N' is a Poisson r.v. Many problems are "equivalent", where one considers this random number of points $(X_j)_{j\le N'}$, or a fixed number of points $(X_j)_{j\le N}$ (the so-called Poissonization procedure).

In particular, it is essentially the same problem to relate (at a *given value of* N) the geometry of a class $\mathcal{F}$ of functions on a probability space (Ω, μ) with the quantity

$$\mathsf{E} \sup_{f\in\mathcal{F}} \Big| \sum_{i\le N} (f(X_i) - \mathsf{E} f) \Big| , \tag{12.2}$$

and to try to relate the geometry of a set T of sequences $t = (t_i)_{i\le M}$ with the quantity

$$\mathsf{E} \sup_{t\in T} \Big| \sum_{i\le M} t_i(\delta_i - \delta) \Big| . \tag{12.3}$$

It is really a matter of taste whether to use the "continuous setting" of (12.2) or the "discrete setting" of (12.3). We have chosen mostly to use the discrete setting in view of the combinatorial formulations of Section 12.5. However, since Theorem 11.2.6 is to date the main support for the conjectures of the next section, we have given at the end of this section a similar conjecture in the "continuous" setting that would be a direct generalization of this result.

We will call a family of r.v.s of the type $\sum_{i\le M} t_i(\delta_i - \delta)$ where t varies over a set of sequences a "selector process".

12.3 The Generalized Bernoulli Conjecture

How should one bound a quantity of the type (12.3), and, in particular, how should one bound the tails of a r.v. of the type $\sum_{i\le M} t_i(\delta_i - \delta)$? A first thought is to use Bernstein's inequality (4.59), which in the present case yields in particular

$$\mathsf{P}\bigg(\Big| \sum_{i\ge 1} t_i(\delta_i - \delta) \Big| \ge v \bigg) \le 2\exp\bigg(-\frac{1}{4} \min\bigg(\frac{v^2}{\delta \sum_{i\le M} t_i^2}, \frac{v}{\max_{i\le M} |t_i|} \bigg) \bigg) . \tag{12.4}$$

Combining with Theorem 2.2.23, (12.4) implies a first bound on selector processes. If T is a set of sequences, then

$$\mathsf{E} \sup_{t\in T} \Big| \sum_{i\le M} t_i(\delta_i - \delta) \Big| \le L(\sqrt{\delta}\gamma_2(T, d_2) + \gamma_1(T, d_\infty)) . \tag{12.5}$$

The following simple lemma is very useful when dealing with selector processes.

Lemma 12.3.1. *Consider a fixed set I. If $u \geq 6\delta \operatorname{card} I$ we have*

$$P\Big(\sum_{i\in I}\delta_i \geq u\Big) \leq \exp\Big(-\frac{u}{2}\log\frac{u}{2\delta \operatorname{card} I}\Big) . \tag{12.6}$$

Proof. We are dealing here with the tails of the binomial law and (12.6) follows from the Chernov bounds. For a direct proof, considering $\lambda > 0$ we write

$$\mathsf{E}\exp\lambda\delta_i \leq 1+\delta e^{\lambda} \leq \exp(\delta e^{\lambda})$$

so that we have

$$\mathsf{E}\exp\lambda\sum_{i\in I}\delta_i \leq \exp(\delta e^{\lambda}\operatorname{card} I)$$

and

$$\mathsf{P}\Big(\sum_{i\in I}\delta_i \geq u\Big) \leq \exp(\delta e^{\lambda}\operatorname{card} I - \lambda u) .$$

We then take $\lambda = \log(u/(2\delta\operatorname{card} I))$, so that $\lambda \geq 1$ since $u \geq 6\delta \operatorname{card} I$ and $\delta e^{\lambda}\operatorname{card} I = u/2 \leq \lambda u/2$. □

Exercise 12.3.2. In the setting of Lemma 12.3.1 above, prove that Bernstein's inequality is sharp (in the sense that the logarithm of the bound it provides is of the correct order) only for u of order $\delta \operatorname{card} I$.

After noticing in this exercise that Bernstein's inequality is often not sharp, one is lead to think that the bound (12.5) will not be very useful for studying selector processes. This is wrong, but this misconception delayed the formulation of the following conjectures for many years.

Conjecture 12.3.3. Consider a set T of sequences, and

$$S := \mathsf{E}\sup_{t\in T}\Big|\sum_{i\leq M} t_i(\delta_i-\delta)\Big| . \tag{12.7}$$

Then we can write $T \subset T_1 + T_2$ where

$$\gamma_2(T_1, d_2) \leq \frac{LS}{\sqrt{\delta}} \ ; \ \gamma_1(T_1, d_\infty) \leq LS , \tag{12.8}$$

and

$$\mathsf{E}\sup_{t\in T_2}\sum_{i\leq M}|t_i|\delta_i \leq LS . \tag{12.9}$$

Combining with (12.5), (12.8) implies

$$\mathsf{E}\sup_{t\in T_1}\Big|\sum_{i\leq M} t_i(\delta_i-\delta)\Big| \leq LS . \tag{12.10}$$

In words, chaining as in (12.5) explains the boundedness of the process over the set T_1, while (as follows from (12.9)) the boundedness of the process over T_2 owes nothing to cancellation. The subtle point is that this does *not* say that bounds such as those of Lemma 12.3.1 are not useful in the study of selector processes. These bounds are useful to control quantities such as the left-hand side of (12.9) which can be though of as the "difficult part" of the process, and which we shall try to study later.

Conjecture 12.3.3 implies that every selector process is a mixture of these two types of processes, and is a version of our UNPROVEN BELIEF that for selector processes

chaining explains all the part of
boundedness which is due to cancellation.

This formulation in words does not capture the full strength of Conjecture 12.3.3, since it is a special type of chaining that is used in (12.5).

Exercise 12.3.4. Prove that a positive answer to Conjecture 12.3.3 implies a positive answer to Problem 9.1.3, at least when $\mathcal{F}$ is a finite class of functions and the underlying measure space has no atoms.

As we shall show, Conjecture 12.3.3 is equivalent to the following, which is less intuitive but maybe more amenable to attack.

Conjecture 12.3.5. Consider a number $\delta \le 1/2$ and a set T of sequences $t = (t_i)_{i \le M}$ with $0 \in T$. Define S as in (12.7). Then we can find an admissible sequence $(\mathcal{A}_n)$ of partitions of T and for $A \in \mathcal{A}_n$ a number $j_n(A)$ such that

$$A \in \mathcal{A}_n \ , \ B \in \mathcal{A}_{n-1} \ , \ A \subset B \Rightarrow j_n(A) \ge j_{n-1}(B) \ , \tag{12.11}$$

$$\forall\, x, y \in A \in \mathcal{A}_n \ , \ \sum_{i \le M} |2^{j_n(A)}(x_i - y_i)|^2 \wedge 1 \le \frac{1}{\delta} 2^n \tag{12.12}$$

and

$$\forall\, t \in T \ , \ \sum_{n \ge 0} 2^{n - j_n(A_n(t))} \le LS \ . \tag{12.13}$$

There is a rather subtle point in the formulation of Conjecture 12.3.5. Conditions (12.11) and (12.12) together with

$$\forall\, t \in T \ , \ \sum_{n \ge 0} 2^{n - j_n(A_n(t))} \le C \ . \tag{12.14}$$

do not imply that $S = \mathsf{E} \sup_{t \in T} |\sum_{i \le M} t_i(\delta_i - \delta)| \le LC$. An example is obtained with $T = \{(t_i)_{i \le M}; \sum_{i \le M} |t_i| \le 1\}$. Then Theorem 16.3.1 used with τ such that $2^{-\tau} \simeq \delta$ proves that these conditions hold for $C = L\delta$, while (when $M\delta \ge 1$), S is about 1 here. However knowing that (12.13) holds *for the value* S of (12.7) suffices to prove that T admits the striking decomposition of Conjecture 12.3.3.

Now we explain why when $\delta = 1/2$, Conjecture 12.3.5 is equivalent to the validity of Theorem 5.1.5 and hence holds true. (The name of generalized Bernoulli conjecture arises from the fact that the statement of Theorem 5.1.5 was known as the Bernoulli conjecture before it was proved.) When $\delta = 1/2$ the r.v.s $\varepsilon_i = 2(\delta_i - \delta)$ are independent Bernoulli r.v.s, so

$$S = \frac{1}{2}\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le M} t_i\varepsilon_i\Big| . \tag{12.15}$$

Moreover, taking the expectation inside the supremum, and since $\mathsf{E}\delta_i = 1/2$, (12.9) proves that $\sup_{t\in T}\sum_{i\le M}|t_i| \le LS$, and thus when $\delta = 1/2$, Conjecture 12.3.3 implies Theorem 5.1.5. Conversely, Proposition 5.2.5 shows that when $\delta = 1/2$, Theorem 5.1.5 implies Conjecture 12.3.5 (which as we later show is equivalent to Conjecture 12.3.3).

We now investigate "the stability of Conjecture 12.3.3 with respect to convex hulls". Since

$$\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le M} t_i(\delta_i-\delta)\Big| = \mathsf{E}\sup_{t\in \operatorname{conv} T}\Big|\sum_{i\le M} t_i(\delta_i-\delta)\Big| ,$$

if Conjecture 12.3.3 is true, then the decomposition $T \subset T_1 + T_2$ it provides must also hold for $\operatorname{conv} T$. Combining with (12.5) we see that Conjecture 12.3.3 can be true only if the following problem has a positive answer.

Research problem 12.3.6. Consider a set T of sequences and assume that for a certain number $S > 0$,

$$\gamma_2(T, d_2) \le \frac{S}{\sqrt{\delta}} \ ; \ \gamma_1(T, d_\infty) \le S .$$

Is it true that one can write $\operatorname{conv} T \subset T_1 + T_2$ where T_1 and T_2 satisfy (12.8) and (12.9)?

Corollary 10.2.9 shows that the answer is positive in the case $\delta = 1/2$, even by taking $T_2 = \{0\}$. The following exercise shows however that for a positive solution to Problem 12.3.6, one cannot always choose T_2 reduced to one point.

Exercise 12.3.7. Prove that it is *not* true that if T is a set of sequences for which

$$\gamma_2(T, d_2) \le \frac{A}{\sqrt{\delta}} \ ; \ \gamma_1(T, d_\infty) \le A ,$$

then

$$\gamma_2(\operatorname{conv} T, d_2) \le \frac{LA}{\sqrt{\delta}} \ ; \ \gamma_1(\operatorname{conv} T, d_\infty) \le LA .$$

(Hint: use Exercise 10.2.11 and choose δ appropriately small.)

A negative answer to Problem 12.3.6 would disprove Conjecture 12.3.3, but not the following weaker version.

Conjecture 12.3.8. Consider a set T of sequences, and S as in (12.7). Then it is possible to write $T \subset \mathrm{conv}(T_1 \cup T_2)$ where T_1 and T_2 satisfy (12.8) and (12.9).

Just as Conjecture 12.3.3, Conjecture 12.3.8 is a weaker but still valid formulation of the unproven principle that "chaining explains the part of the process which is due to cancellation".

Proof that Conjecture 12.3.5 implies Conjecture 12.3.3. Assume that (12.11) to (12.13) hold true. Then Theorem 5.2.7 (used for $r = 2$ and $u = 1/\delta$) implies that we can find a decomposition $T \subset T_1 + T_2' + T_3$ such that (12.8) holds while

$$\forall\, t \in T_2'\,,\ \|t\|_1 \le \frac{LS}{\delta}\,, \tag{12.16}$$

and (5.29) holds, i.e.

$$\forall t \in T_3\,, \exists s \in T\,, |t| \le 5|s|\mathbf{1}_{\{2|s| \ge 2^{-j_0(T)}\}}\,. \tag{12.17}$$

Consider $s \in T$ and set $J = \{i \le M; |s_i| \ge 2^{-j_0(T)-1}\}$. Our next goal is to prove that

$$\sum_{i \in J} |s_i| \le \frac{LS}{\delta}\,, \tag{12.18}$$

so that combining with (12.17) we get

$$t \in T_3 \Rightarrow \|t\|_1 \le \frac{LS}{\delta}\,. \tag{12.19}$$

To prove (12.18), we first observe that using (12.12) with $n = 0$, $x = s$ and $y = 0$ yields $\mathrm{card}\, J \le 4/\delta$. Considering a subset J' of J, the definition of S shows that $\mathsf{E}|\sum_{i \le M} s_i(\delta_i - \delta)| \le S$, and Jensen's inequality implies

$$\mathsf{E}\Big|\sum_{i \in J'} s_i(\delta_i - \delta)\Big| \le S\,. \tag{12.20}$$

The event $\Omega = \{\forall j \in J'; \delta_j = 0\}$ satisfies

$$\mathsf{P}(\Omega) = 1 - (1-\delta)^{\mathrm{card}\, J'} \ge 1 - \exp(-\delta\, \mathrm{card}\, J) \ge 1 - \exp(-4)$$

and since

$$\mathsf{E}\Big|\sum_{i \in J'} s_i(\delta_i - \delta)\Big| \ge \mathsf{E}\mathbf{1}_\Omega\Big|\sum_{i \in J'} s_i(\delta_i - \delta)\Big| = \delta \mathsf{P}(\Omega)\Big|\sum_{j \in J'} s_j\Big|\,,$$

(12.20) proves that $|\sum_{j \in J'} s_j| \le LS/\delta$ and since J' is an arbitrary subset of J this implies (12.18) and hence (12.19). Consequently setting $T_2'' = T_2' + T_3$

we then obtain $T \subset T_1 + T_2''$ while (using also (12.16)) $\|t\|_1 \le LS/\delta$ for $t \in T_2''$. Combining (12.10) and the definition of S yields

$$\mathsf{E} \sup_{t \in T - T_1} \Big| \sum_{i \in J} t_i(\delta_i - \delta) \Big| \le LS \ .$$

Setting $T_2 = (T - T_1) \cap T_2''$, we then further have

$$\mathsf{E} \sup_{t \in T_2} \Big| \sum_{i \in J} t_i(\delta_i - \delta) \Big| \le LS \ ,$$

while still $T \subset T_1 + T_2$. Then since $\|t\|_1 \le LS/\delta$ for $t \in T_2$, using (a simple modification of) the Giné-Zinn Theorem 9.1.10 we then obtain (12.9). □

Sketch of proof that Conjecture 12.3.3 implies Conjecture 12.3.5. Consider the sets T_1 and T_2 as in Conjecture 12.3.3. The only property we shall use for those sets is that $\gamma_2(T_1, d_2) \le LS/\sqrt{\delta}$ and $\sup_{t \in T_2} \sum_{i \le N} |t_i| \le LS/\delta$. This second relation follows from (12.9), by moving the expectation inside the supremum. Then if either $T = T_1$ or $T = T_2$ one may find an admissible sequence of partitions as in Conjecture 12.3.5. For T_1 this follows from Proposition 5.2.4, and for T_2 this follows from Theorem 16.3.1 with $\delta \simeq 2^{-\tau}$. To prove that one can find an admissible sequence of partitions which works for $T_1 + T_2$ one reproduces the arguments of Proposition 5.2.5. □

Sets which satisfy (12.9) will be discussed in the next section. Before we do this, we discuss the case where T consists of one single sequence. Of course, such a case is an exercise, but a non trivial one. We find it convenient to move for this to a "symmetrized" version of the process, i.e. to r.v.s of the type $\sum_{i \le M} t_i \varepsilon_i \delta_i$ where (ε_i) are independent Bernoulli r.v.s.

Proposition 12.3.9. *There exists a constant L with the following property. Assume that for a number S we have*

$$\mathsf{P}\Big(\sum_{i \le M} t_i \varepsilon_i \delta_i \ge S \Big) \le \exp(-B) \ , \tag{12.21}$$

for some $B \ge L$. Then we can decompose $t = (t_i)_{i \le M}$ as $t = t^1 + t^2 + t^3$ where

$$\operatorname{card}\{i \le M \ ; \ t_i^3 \ne 0\} \le \frac{L}{\delta} \exp(-B) \ , \tag{12.22}$$

$$\|t^1\|_2 \le \frac{LS}{\sqrt{\delta B}} \ , \ \|t^1\|_\infty \le \frac{LS}{B} \ , \ \|t^2\|_1 \le \frac{LS}{\delta} \ . \tag{12.23}$$

The following proposition explains why this lends credibility to conjecture 12.3.5. (One must be very weary however that, as Theorem 8.2.2 shows, the global behavior of a family of r.v.s may not reflect the tail behavior of the individual r.v.s.)

Proposition 12.3.10. *There exists* $\epsilon_0 > 0$ *such that if for some* $S > 0$ *a sequence* $(t^k) = (t_i^k)_{i \le M}$ *satisfies*

$$\sum_{k \ge 1} \mathsf{P}\Big(\sum_{i \le M} t_i^k \varepsilon_i \delta_i \ge S\Big) \le \epsilon_0 \ , \tag{12.24}$$

then we can find a subset J *of* $\{1, \ldots, M\}$ *with* $\operatorname{card} J \le L/\delta$ *such that* $T = \{t^k \mathbf{1}_{\{i \notin J\}}; k \ge 1\}$ *satisfies the conditions of Conjecture 12.3.3.*

The reason for the set J in the above is that (12.24) implies

$$\mathsf{P}\Big(\sup_{k \ge 1} \sum_{i \le M} t_i^k \varepsilon_i \delta_i \ge S\Big) \le \epsilon_0 \ , \tag{12.25}$$

but provides no control of the quantity $\mathsf{E} \sup_{k \ge 1} \sum_{i \le M} t_i^k \varepsilon_i \delta_i$. This can be seen e.g. in the case where $t_i^k = 0$ for $i \ge 2$ and all k. On the other hand, the contribution of such "small" sets of coordinates as J is essentially trivial, as the following exercise shows.

Exercise 12.3.11. Consider a set T of sequences $t = (t_i)_{i \le N}$ and define $t_i^* = \sup_{t \in T} |t_i|$. If $\operatorname{card} J \le 1/\delta$, prove that

$$\frac{\delta}{L} \sum_{i \in J} t_i^* \le \mathsf{E} \sup_{t \in T} \Big| \sum_{i \in J} t_i \varepsilon_i \delta_i \Big| \le \delta \sum_{i \in J} t_i^* \ . \tag{12.26}$$

(Hint: The right-hand side inequality is obvious. To prove the left-hand side, for $i \in J$ consider the event Ω_i given by $\delta_i = 1$ and $\delta_j = 0$ for $j \ne i$. Observe that these events are disjoint and that $\mathsf{P}(\Omega_i) \ge \delta/L$.)

Proof of Proposition 12.3.10. We choose $\epsilon_0 = \exp(-L)$ where L is the constant of Proposition 12.3.9. For $k \ge 1$ we define B_k by

$$\mathsf{P}\Big(\sum_{i \le M} t_i^k \varepsilon_i \delta_i \ge S\Big) = \exp(-B_k) \ ,$$

so that $\sum_{k \ge 1} \exp(-B_k) \le \epsilon_0$ and in particular B_k is large enough that we can use Proposition 12.3.9. Therefore we can write $t^k = t^{k,1} + t^{k,2} + t^{k,3}$ where

$$\operatorname{card}\{i \le M \ ; \ t_i^{k,3} \ne 0\} \le \frac{L}{\delta} \exp(-B_k) \ , \tag{12.27}$$

$$\|t^{k,1}\|_2 \le \frac{LS}{\sqrt{\delta B_k}} \ , \ \|t^{k,1}\|_\infty \le \frac{LS}{B_k} \ , \ \|t^{k,2}\|_1 \le \frac{LS}{\delta} \ . \tag{12.28}$$

We define

$$J = \{i \le M \ ; \ \exists k \ge 1 \ , \ t_i^{k,3} \ne 0\} \ ,$$

so that (12.25) and (12.27) imply that $\operatorname{card} J \le L/\delta$. Thus it suffices to show that

$$T := \{t^{k,1} + t^{k,2} \ ; \ k \geq 1\}$$

satisfies the conditions of Conjecture 12.3.3. Let $T_1 = \{t^{k,1}; k \geq 1\}$ and $T_2 = \{t^{k,2}; k \geq 1\}$. It is then obvious from (12.28) that $\|t\|_1 \leq LS/\delta$ for $t \in T_2$, so that it suffices to prove (12.8). Assuming without loss of generality that the sequence (B_k) increases, we have $k \exp(-B_k) \leq \sum_{p \geq 1} \exp(-B_p) \leq \varepsilon_0 \leq 1/2$ and thus $B_k \geq \log(2k) \geq \log(k+1)$. Then (12.28) implies

$$\|t^{k,1}\|_2 \leq \frac{LS}{\sqrt{\delta \log(k+1)}} \ ; \ \|t^{k,1}\|_\infty \leq \frac{LS}{\log(k+1)} \ .$$

We are essentially in the situation of Exercise 2.2.16, so it should then be obvious that $\gamma_2(T_1, d_2) \leq LS/\sqrt{\delta}$ and $\gamma_1(T_1, d_\infty) \leq LS$. □

Proof of Proposition 12.3.9. We may assume without loss of generality that the sequence $(|t_i|)_{i \leq M}$ decreases. The simple idea is that for certain integers i_0 and i_1 we have

$$t_i^3 = t_i \text{ for } i \leq i_0 \ ; \ t_i^3 = 0 \text{ for } i > i_0$$

$$t_i^2 = t_i \text{ for } i_0 < i \leq i_1 \ ; \ t_i^2 = 0 \text{ otherwise}$$

$$t_i^1 = t_i \text{ for } i > i_1 \ ; \ t_i^1 = 0 \text{ for } i \leq i_1 \ .$$

We denote by A a parameter that will be chosen later to be a large enough constant. We define i_0 as the largest integer $i_0 \geq 0$ such that

$$i_0 \leq \frac{A}{\delta} \exp(-B) \ . \tag{12.29}$$

We define i_1 as the largest integer for which

$$\sum_{i_0 < i \leq i_1} |t_i| \leq \frac{AS}{\delta} \ . \tag{12.30}$$

The proof is complete if $i_1 = M$, so we assume from now on that this is not the case. Then

$$\sum_{i_0 < i \leq i_1 + 1} |t_i| \geq \frac{AS}{\delta} \ . \tag{12.31}$$

We shall prove that if A is a large enough constant,

$$|t_{i_1+1}| \leq \frac{LS}{B} \tag{12.32}$$

and

$$\sum_{i > i_1} t_i^2 \leq \frac{LS^2}{\delta B} \ , \tag{12.33}$$

and these complete the proof. We shall use many times the following observation. If X and Y are independent symmetric r.v.s then

$$\mathsf{P}(X+Y\geq S)\geq \frac{1}{2}\mathsf{P}(X\geq S)\ . \tag{12.34}$$

The first step of the argument is to prove that

$$|t_{i_0}|\leq S\ . \tag{12.35}$$

Arguing by contradiction, let us assume that $|t_i|>S$ for $i\leq i_0$, and let us consider the event Ω defined by

$$\exists i\leq i_0\ ;\ \delta_i=1\ .$$

Denoting by P_ε the probability given the choice of the r.v.s δ_i, it should be clear by using (12.34) that when Ω occurs one has

$$\mathsf{P}_\varepsilon\Big(\sum_i t_i\varepsilon_i\delta_i\geq S\Big)\geq\frac{1}{4}\ .$$

Consequently,

$$\frac{1}{4}\mathsf{P}(\Omega)=\frac{1}{4}(1-(1-\delta)^{i_0})\leq\mathsf{P}\Big(\sum_i t_i\varepsilon_i\delta_i\geq S\Big)\leq\exp(-B)\ .$$

Since we certainly may assume that $B\geq 2$, this implies $1-(1-\delta)^{i_0}\geq\delta i_0/L$ and thus

$$i_0\delta\leq L\exp(-B)\ ,$$

and this contradicts the definition of i_0 provided that the parameter A has been chosen a large enough constant. Thus (12.35) is proved. We now fix A in this manner, and we assume, as we may, that $A\geq 2$.

Since we assume that the sequence $(|t_i|)$ decreases, we deduce from (12.30) that

$$|t_{i_1}|\leq\frac{AS}{\delta(i_1-i_0)}\ . \tag{12.36}$$

To prove (12.32) (and even the better inequality $|t_{i_1}|\leq LS/B$) we prove that

$$\delta(i_1-i_0)\geq\frac{B}{L}\ . \tag{12.37}$$

Let us lighten notation by setting

$$J=\{i_0+1,\ldots,i_1+1\}\ , \tag{12.38}$$

so that $\operatorname{card} J=i_1-i_0+1$, and (12.31) means

$$\sum_{i\in J}|t_i|\geq\frac{AS}{\delta}\ . \tag{12.39}$$

Consider the r.v.

$$X = \sum_{i \in J} |t_i| \delta_i \ ,$$

so that

$$\mathsf{E}X = \delta \sum_{i \in J} |t_i| \geq AS \ .$$

Now

$$\mathsf{E}X^2 = (\mathsf{E}X)^2 + (\delta - \delta^2) \sum_{i \in J} t_i^2 \leq (\mathsf{E}X)^2 + \delta \sum_{i \in J} t_i^2 \ .$$

Since $|t_i| \leq S$ for $i \in J$, assuming $A \geq 1$, we have

$$\delta \sum_{i \in J} t_i^2 \leq S\delta \sum_{i \in J} |t_i| = S\mathsf{E}X \leq (\mathsf{E}X)^2 \ ,$$

and $\mathsf{E}X^2 \leq 2(\mathsf{E}X)^2$. The Paley-Zygmund inequality then implies

$$\mathsf{P}\Big(X \geq \frac{\mathsf{E}X}{2}\Big) \geq \frac{(\mathsf{E}X)^2}{4\mathsf{E}X^2} \geq \frac{1}{8} \ ,$$

and in particular $\mathsf{P}(X \geq AS/2) \geq 1/8$. Meanwhile, since $\mathsf{E}(\sum_{i \in J} \delta_i) = \delta \operatorname{card} J$ we have

$$\mathsf{P}\Big(\sum_{i \in J} \delta_i \geq 16\delta \operatorname{card} J\Big) \leq \frac{1}{16}$$

by Markov's inequality. Hence the event Ω given by

$$X \geq \frac{AS}{2} \ ; \ \sum_{i \in J} \delta_i \leq 16\delta \operatorname{card} J$$

satisfies $\mathsf{P}(\Omega) \geq 1/16$. Now, given the r.v.s δ_i,

$$\mathsf{P}_\varepsilon\Big(\sum_{i \in J} t_i \varepsilon_i \delta_i \geq X\Big) \geq 2^{-\sum_{i \in J} \delta_i} \ ,$$

and assuming $A \geq 2$, so that $X \geq S$ on Ω,

$$\frac{1}{16} 2^{-16\delta \operatorname{card} J} \leq \mathsf{P}\Big(\sum_{i \in J} t_i \varepsilon_i \delta_i \geq S\Big) \ .$$

Using (12.34) again,

$$\frac{1}{32} 2^{-16\delta \operatorname{card} J} \leq \mathsf{P}\Big(\sum_{i \leq M} t_i \varepsilon_i \delta_i \geq S\Big) \leq \exp(-B) \ .$$

Since $\operatorname{card} J = i_1 - i_0 + 1$ and since $i_1 > i_0$ by (12.35) this proves (12.37) and we turn to the proof of (12.33). Denoting by L_0 the constant of (12.32) we may assume that

$$C := \sum_{i>i_1} t_i^2 \geq \frac{2L_0^2 S^2}{\delta B} , \tag{12.40}$$

for there is nothing to prove otherwise. Consider the smallest integer i_2 such that

$$\sum_{i_1<i<i_2} t_i^2 \geq \frac{C}{2B} .$$

Then, since by (12.32) $|t_i| \leq L_0 S/B$ for $i > i_1$,

$$\sum_{i_1<i<i_2} t_i^2 \leq \sum_{i_1<i<i_2-1} t_i^2 + t_{i_2-1}^2 \leq \frac{C}{2B} + \frac{L_0^2 S^2}{B^2} \leq \frac{C}{B} .$$

Continuing in this manner we construct as long as possible disjoint sets J_ℓ, $\ell \leq m$ such that

$$\frac{C}{2B} \leq \sum_{i\in J_\ell} t_i^2 \leq \frac{C}{B} . \tag{12.41}$$

Then $Cm/2B \leq C$ i.e. $m \leq 2B$. Also, since we construct as many of these sets as possible, there are not enough t_i's left to construct one more set, so that we have

$$C - m\frac{C}{B} \leq \frac{C}{2B} ,$$

and in particular, assuming as we may $B \geq 21$

$$\frac{B}{2} \leq m \leq 2B . \tag{12.42}$$

Consider now the r.v.s $Y_\ell = \sum_{i\in J_\ell} \varepsilon_i \delta_i t_i$, so that

$$\mathsf{E} Y_\ell^2 = \delta \sum_{i\in J_\ell} t_i^2 \geq \frac{\delta C}{2B} .$$

Next we try to bound $\mathsf{E} Y_\ell^4$. First, using the subgaussian inequality, $\mathsf{E}_\varepsilon Y_\ell^4 \leq L(\mathsf{E}_\varepsilon Y_\ell^2)^2 = L(\sum_{i\in J_\ell} \delta_i^2 t_i^2)^2$. Expending the square and taking expectation yields, since $\mathsf{E}\delta_i^2 = \mathsf{E}\delta_i^4 = \delta$,

$$\mathsf{E} Y_\ell^4 \leq L \mathsf{E} Y_\ell^2 + \delta \sum_{i\in J_\ell} t_i^4 .$$

Using (12.32) in the first inequality and (12.40) in the second inequality, for $i > i_1$,

$$t_i^2 \leq \frac{L_0^2 S^2}{B^2} \leq \frac{\delta C}{2B} \leq \mathsf{E} Y_\ell^2 ,$$

so that

$$\delta \sum_{i\in J_\ell} t_i^4 \leq \mathsf{E} Y_\ell^2 \delta \sum_{i\in J_\ell} t_i^2 = (\mathsf{E} Y_\ell^2)^2 ,$$

and therefore $\mathsf{E}Y_\ell^4 \le L^*(\mathsf{E}Y_\ell^2)^2$. The Paley-Zygmund inequality then implies

$$\mathsf{P}\Big(Y_\ell^2 \ge \frac{\delta C}{2L^* B}\Big) \ge \frac{1}{8}\,, \tag{12.43}$$

and moreover the r.v.s (Y_ℓ) are independent. The independent events $\Omega_\ell = \{Y_\ell^2 \ge \delta C/2L^* B\}$ for $\ell \le m$ all have probability $\ge 1/8$. Consider an integer $n \le m/16$. Then (assuming B, and hence m large enough), with probability at least 1/2 at least n of the m events $(\Omega_\ell)_{\ell \le m}$ will occur. That is, the event Ω defined by

$$\exists I \subset \{1, \ldots, m\}\ ;\ \operatorname{card} I = n\ ,\ \ell \in I \Rightarrow |Y_\ell| \ge D := \sqrt{\frac{\delta C}{2L^* B}}$$

satisfies $\mathsf{P}(\Omega) \ge 1/2$.

Let us now consider Bernoulli r.v.s $(\eta_\ell)_{\ell \le m}$ that are independent of all the r.v.s considered so far, so that, with obvious notation, when Ω occurs,

$$\mathsf{P}_\eta\Big(\sum_{\ell \in I} \eta_\ell |Y_\ell| \ge nD\Big) \ge 2^{-\operatorname{card} I} = 2^{-n}\ .$$

Using (12.34) we obtain

$$\mathsf{P}_\eta\Big(\sum_{\ell \le m} \eta_\ell |Y_\ell| \ge nD\Big) \ge 2^{-n-1}\ ,$$

and since $\mathsf{P}(\Omega) \ge 1/2$,

$$\mathsf{P}\Big(\sum_{\ell \le m} \eta_\ell |Y_\ell| \ge nD\Big) \ge 2^{-n-2}\ . \tag{12.44}$$

We observe that in distribution, if $J = \bigcup_{\ell \le m} J_\ell$, we have

$$\sum_{i \in J} t_i \varepsilon_i \delta_i = \sum_{\ell \le m} \eta_\ell |Y_\ell|\ .$$

Let us assume if possible that $nD \ge S$. Then, combining (12.44) with (12.34) we obtain

$$\frac{1}{2} 2^{-n-2} \le \frac{1}{2}\mathsf{P}\Big(\sum_{i \in J} t_i \varepsilon_i \delta_i \ge S\Big) \le \mathsf{P}\Big(\sum_{i \le M} t_i \varepsilon_i \delta_i \ge S\Big) \le \exp(-B)\ .$$

Since $n \le m/16 \le B/16$ this is impossible. Therefore $nD \le S$, which implies (12.33) and completes the proof. □

One cannot argue that Proposition 12.3.9 provides much support for Conjecture 12.3.3. In our view, greater support is provided by Theorem 11.2.6, and we formulate now a conjecture closely related to Conjecture 12.3.3 which is motivated by this theorem. Consider a σ-finite measure space (Ω, ν) and an enumeration (Z_i) of a Poisson point process of intensity measure ν. Consider i.i.d. Bernoulli r.v.s ε_i. For a function t in $L^2(\nu)$ the series $X_t = \sum_i \varepsilon_i t(Z_i)$ converges a.s. We denote by d_2 the distance induced by the norm of $L^2(\nu)$ and by d_∞ the distance induced by the supremum norm.

Conjecture 12.3.12. Given a set T of functions in $L^2(\nu)$, let $S = \mathsf{E} \sup_{t\in T} |X_t|$. Then we can write $T \subset T_1 + T_2$ where

$$\gamma_2(T_1, d_2) \le LS \; ; \; \gamma_1(T_1, d_\infty) \le LS$$

and

$$\mathsf{E} \sup_{t\in T_2} \sum_{i\ge 1} |t(Z_i)| \le LS \; .$$

The condition that $t \in L^2(\nu)$ is inessential, and the r.v. X_t is well defined as soon as $|t| \wedge 1 \in L^2(\nu)$. This condition $t \in L^2(\nu)$ is used in the previous statement only to ensure integrability of X_t and to get a simple formulation.

12.4 Positive Selector Processes

If Conjecture 12.3.3 is true, it reduces the study of selector processes to the study of positive selector processes, as follows from (12.9). That is, we have to understand the quantity

$$\mathcal{E}_\delta(T) := \mathsf{E} \sup_{t\in T} \sum_{i\le M} t_i \delta_i \tag{12.45}$$

where T is a set of sequences $t = (t_i)_{i\le M}$ with $t_i \ge 0$. Whether or not Conjecture 12.3.3 is true, the study of positive selector processes is in any case fundamental, since, following the same steps as in the previous section, it is essentially the same problem as understanding the quantity

$$\mathsf{E} \sup_{f\in\mathcal{F}} \sum_{i\le N} f(X_i)$$

when $\mathcal{F}$ is a class of non-negative functions. The complicated nature of the tails of the r.v.s $\sum_{i\le M} t_i(\delta_i - \delta)$ (as exemplified e.g. in Lemma 12.3.1) makes it difficult to imagine in function of which geometrical characteristics of T one should evaluate the left-hand side of (12.45). We shall discuss in Section 12.6 the case where T consists of indicator of sets, and in particular we shall give concrete examples which illustrate this point.

An important feature of positive selector processes is that we can use *positivity* to construct new processes from processes we already know how to bound. Given a set T we denote by solid T its "solid convex hull", i.e. the set of sequences $(s_i)_{i\le M}$ for which there exists $t \in \operatorname{conv} T$ such that $s_i \le t_i$ for each $i \le M$. It should be obvious that

$$\sup_{t\in \operatorname{solid} T} \sum_{i\le M} t_i\delta_i = \sup_{t\in T} \sum_{i\le M} t_i\delta_i \ , \tag{12.46}$$

so that in particular

$$\mathcal{E}_\delta(\operatorname{solid} T) = \mathsf{E} \sup_{t\in \operatorname{solid} T} \sum_{i\le M} t_i\delta_i = \mathsf{E} \sup_{t\in T} \sum_{i\le M} t_i\delta_i = \mathcal{E}_\delta(T) \ .$$

Thus, to bound $\mathcal{E}_\delta(T)$ is suffices to find a set T' for which $T \subset \operatorname{solid} T'$ and such that we control $\mathcal{E}_\delta(T')$. We know one sure way to bound $\mathcal{E}_\delta(T)$, namely the straightforward use to the union bound, an idea which we formalize now.

Definition 12.4.1. *Given a number $S \ge 0$ we define the class $\mathcal{T}(S)$ of sets T by the following property*

$$\int_S^\infty \sum_{t\in T} \mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge u\Big) \mathrm{d}u \le S \ . \tag{12.47}$$

An important example of set in $\mathcal{T}(S)$ is

$$\Big\{t \ ; \ \sum_{i\le M} t_i \le S\Big\}$$

because then the sum in (12.47) is zero for each $u > S$.

Proposition 12.4.2. *If $T \in \mathcal{T}(S)$ then $\mathcal{E}_\delta(T) \le 2S$.*

Proof.

$$\mathcal{E}_\delta(T) = \int_0^\infty \mathsf{P}\Big(\sup_{t\in T} X_t \ge u\Big)\mathrm{d}u \le S + \int_S^\infty \sum_{t\in T} \mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge u\Big)\mathrm{d}u \le 2S \ . \ \square$$

Consequently, we obtain a method to bound $\mathcal{E}_\delta(T)$ for any set T of sequences. Simply, if $T \subset \operatorname{solid} T'$ for a certain $T' \in \mathcal{T}(S)$, then $\mathcal{E}_\delta(T) \le 2S$. We can believe, optimistically, that this method is optimal, even though at the present time this is little more than wishful thinking, supplemented by a lack of imagination to invent methods to bound positive selector processes.

Research problem 12.4.3. Does there exist a universal constant L such that for any set T of sequences one can find a set $T' \in \mathcal{T}(S)$ for $S = L\mathcal{E}_\delta(T)$ with $T \subset \operatorname{solid} T'$?

12.5 Explicitly Small Events

We would like to explain a new direction of investigation that became apparent during the writing of [3]. Ultimately, we would like to prove theorems of the type of Theorem 2.4.1, which gives a complete description of the quantity $\mathsf{E}\sup_{t\in T} X_t$ as a function of the geometry of the metric space (T,d) for many processes, and this is what motivates the conjectures of the previous sections. But in the mean time is there a way to gather some understanding even if we do not yet have the hope to fully understand the situation? As explained after Theorem 2.4.18, a consequence of this result is that for any Gaussian process we can find a jointly Gaussian sequence (u_k) such that

$$\Big\{\sup_{t\in T}|X_t| \ge L\mathsf{E}\sup_{t\in T}|X_t|\Big\} \subset \bigcup_{k\ge 1}\{u_k \ge 1\} \tag{12.48}$$

and moreover

$$\sum_{k\ge 1}\mathsf{P}(u_k \ge 1) \le \frac{1}{2}\,.$$

The sets $\{u_k \ge 1\}$ are *simple concrete witnesses* that the event on the left-hand side of (12.48) has a probability at most $1/2$. The existence of these witnesses is a non-trivial information, even though it is not as good as the information provided by Theorem 2.4.1. (Let us observe in particular that this information is rather easy to deduce from Theorem 2.4.1, but that it does not seem easy to go the other way around.)

In the setting of positive selector processes, the same idea would require that for some universal constant L and each set T of sequences $t=(t_i)_{i\ge 1}$, $t_i \ge 0$, there exist simple witnesses that the event

$$\sup_{t\in T}\sum_{i\le M}\delta_i t_i \ge L\mathsf{E}\sup_{t\in T}\sum_{i\le M}\delta_i t_i \tag{12.49}$$

has a probability at most $1/2$.

There is a simple and natural choice for these witnesses. For a finite subset I of $\{1,\ldots,M\}$, let us consider the event H_I defined by

$$H_I = \{\forall i\in I, \delta_i = 1\}\,,$$

so that $\mathsf{P}(H_I) = \delta^{\operatorname{card} I}$. The events H_I play the role that the half-spaces play for Gaussian processes in (12.48).

Definition 12.5.1. *An event Ω is δ-small if we can find a family $\mathcal{G}$ of subsets I of $\{1,\ldots,M\}$ with*

$$\sum_{I\in\mathcal{G}}\delta^{\operatorname{card} I} \le 1/2 \tag{12.50}$$

and

$$\Omega \subset \bigcup_{I\in\mathcal{G}} H_I\,. \tag{12.51}$$

Of course, the choice of the constant 1/2 in (12.50) is rather arbitrary. The relation (12.51) makes "explicit" the fact that Ω is small (hence the title of the section). The first thing to understand is that there exists sets of small probability which do not look at all like δ-small sets. A typical example is as follows. Let us consider two integers k, r, and r disjoint subsets $I_1, \ldots, I_r$ of $1, \ldots, M$, each of cardinality k. Let us consider the set

$$A = \{(\delta_i)_{i \le M} \; ; \; \forall \ell \le r \; , \; \exists i \in I_\ell \; , \; \delta_i = 1\} \; . \tag{12.52}$$

It is straightforward to see that $\mathsf{P}(A) = (1-(1-\delta)^k)^r$. In particular, given k, one can chose r large so that $\mathsf{P}(A)$ is small. We leave as teaser to the reader to prove that the set A is not $1/k$ small. (Hint: A carries a probability measure ν such that $\nu(H_I) \le k^{-\operatorname{card} I}$ for each I. A complete proof can be found in [4].)

Research problem 12.5.2. Is it true that we can find a universal constant L such that for any class of sequences T as in (12.49), the event

$$\Big\{\sup_{t \in T} \sum_{i \le M} \delta_i t_i \ge L \mathsf{E} \sup_{t \in T} \sum_{i \le M} \delta_i t_i \Big\} \tag{12.53}$$

is δ-small?

Even proving that the set (12.53) is $\alpha\delta$-small, where α is some universal constant would be of interest. The main result of Section 12.6 is a positive answer to this problem when T consists of indicators of sets.

Proposition 12.5.3. *If Problem 12.4.3 has a positive answer, then so does Problem 12.5.2.*

In view of (12.46) this proposition is an immediate consequence of the following.

Proposition 12.5.4. *For $T \in \mathcal{T}(S)$ the event*

$$\Big\{\sup_{t \in T} \sum_{i \le M} \delta_i t_i \ge LS\Big\}$$

is δ-small.

Lemma 12.5.5. *Consider $t = (t_i)_{i \le M}$ and $v > 0$. Assume that*

$$\mathsf{P}\Big(\sum_{i \le M} t_i \delta_i \ge v\Big) \le \frac{1}{16} \; . \tag{12.54}$$

Then we can find a family $\mathcal{G}$ of subsets of $\{1, \ldots, M\}$ for which

$$\Big\{\sum_{i \le M} t_i \delta_i \ge Lv\Big\} \subset \bigcup_{I \in \mathcal{G}} H_I$$

and

$$\sum_{I\in\mathcal{G}} \delta^{\operatorname{card} I} \le 10\mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge v\Big) .$$

Proof of Proposition 12.5.4. By definition of the statement $T \in \mathcal{T}(S)$, (12.47) holds, so that in particular

$$\sum_{t\in T} \mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge 21S\Big) \le \frac{1}{20} . \tag{12.55}$$

For each $t \in T$ we find a family $\mathcal{G}_t$ as in Lemma 12.5.5, used for $v = 21S$. The union $\mathcal{G}$ of these families as t varies over T satisfies the requirements of Proposition 12.5.4. □

Proof of Lemma 12.5.5. We may and do assume that the sequence (t_i) is non-increasing. Consider the largest integer i_0 such that

$$1-(1-\delta)^{i_0} = \mathsf{P}(\exists i \le i_0, \delta_i = 1) \le \mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge v\Big) ,$$

so that

$$\begin{aligned}\mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge t_{i_0+1}\Big) &\ge \mathsf{P}(\exists i \le i_0+1, \delta_i = 1)\\ &= 1-(1-\delta)^{i_0+1} > \mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge v\Big)\end{aligned}$$

and in particular $t_{i_0+1} < v$ and so $t_i < v$ for $i > i_0$. Also, since $1-(1-\delta)^{i_0} \le \mathsf{P}(\sum_{i\le M} t_i\delta_i \ge v) \le 1/16$, we have $15/16 \le (1-\delta)^{i_0} \le \exp(-\delta i_0)$ and $\delta i_0 \le 1/10$ and thus $1-(1-\delta)^{i_0} \ge \delta i_0/2$ and finally $\delta i_0 \le 2\mathsf{P}(\sum_{i\le M} t_i\delta_i \ge v)$.

Moreover

$$\Big\{\sum_{i\le M} t_i\delta_i \ge v\Big\} \subset \bigcup_{i\le i_0} H_{\{i\}} \cup \Big\{\sum_{i_0<i\le M} t_i\delta_i \ge v\Big\} . \tag{12.56}$$

Since $t_i \le v$ for $i > i_0$, then, as we have used many times, the r.v. $X = \sum_{i_0<i\le M} t_i\delta_i$ satisfies $\mathsf{E}X^2 \le (\mathsf{E}X)^2 + v\mathsf{E}X$. The Paley-Zygmund inequality then shows that we must have $\mathsf{E}X \le 2v$, for otherwise we would have $\mathsf{P}(X \ge v) \ge 1/8$, while in fact we have $\mathsf{P}(X \ge v) \le \mathsf{P}(\sum_{i\le M} t_i\delta_i \ge v) \le 1/16$. It is then proved in [4], Theorem 11.1 that we can find a family $\mathcal{G}$ of subsets of $\{1,\ldots,M\}$ for which $\{X \ge Lv\} \subset \bigcup_{I\in\mathcal{G}} H_I$ and $\sum_{I\in\mathcal{G}} \delta^{\operatorname{card} I} \le 8\mathsf{P}(X \ge v)$. Consider the family $\mathcal{G}' = \mathcal{G} \cup \{\{i\}; i \le i_0\}$. Then (12.56) shows that

$$\Big\{\sum_{i\le M} t_i\delta_i \ge v\Big\} \subset \bigcup_{I\in\mathcal{G}'} H_I ,$$

while

$$\sum_{I\in\mathcal{G}'} \delta^{\operatorname{card} I} \le i_0\delta + \sum_{I\in\mathcal{G}} \delta^{\operatorname{card} I} \le 10\mathsf{P}\Big(\sum_{i\le M} t_i\delta_i \ge v\Big) .$$

□

The proof of Theorem 11.1 of [4] is complicated, so we do not reproduce it here, but rather we hope that a creative reader will invent a clean proof of Lemma 12.5.5.

Part of the beauty of Problem 12.5.2 is that possibly the best way to approach it is through a natural question of a more general nature. To formulate this more general question, we need to consider the law P of the sequence $(\delta_i)_{i\geq 0}$ in $\{0,1\}^M$. With some abuse of notation, we will denote by $(\delta_i)_{i\leq M}$ the generic point of $\{0,1\}^M$. We define an abstract operation as follows. Given a set $D\subset\{0,1\}^M$ and an integer q, let us define the set $D^{(q)}$ as the subset of $\{0,1\}^M$ consisting of the sequences $(\delta_i)_{i\leq M}$ such that the set $\{i\leq M;\delta_i=1\}$ cannot be covered by q sets of the type $\{i\leq M;\delta_i=1\}$ for $\delta\in D$, or more formally,

$$\forall(\delta_i^1)_{i\leq M},\ldots,(\delta_i^q)_{i\leq M}\in D\ ,\ \exists i\leq M\ ,\ \delta_i=1\ ,\ \forall\ell\leq q\ ,\ \delta_i^\ell=0\,.$$

The link with Problem 12.5.2 is that if D is the set consisting of the sequences $(\delta_i)_{i\leq M}$ for which $\sup_{t\in T}\sum_{i\leq M}\delta_i t_i\leq S$, where S is a median of the left-hand side, then $\mathsf{P}(D)\geq 1/2$, while, due to positivity, if $(\delta_i^1)_{i\leq M},(\delta_i^2)_{i\leq M},\ldots,(\delta_i^q)_{i\leq M}\in D$, then for $t\in T$,

$$\sum_{i\leq M}t_i\max_{\ell\leq q}\delta_i^\ell\leq\sum_{\ell\leq q}\sum_{i\leq M}\delta_i^\ell t_i\leq qS\ .$$

Now, if $(\delta_i)\notin D^{(q)}$ we can find $(\delta_i^1)_{i\leq M},(\delta_i^2)_{i\leq M},\ldots,(\delta_i^q)_{i\leq M}\in D$ for which $\delta_i\leq\max_{\ell\leq q}\delta_i^{(\ell)}$, so that $\sum_{i\leq M}\delta_i t_i\leq qS$. Consequently,

$$\Big\{\sup_{t\in T}\sum_{i\leq M}\delta_i t_i>qS\Big\}\subset D^{(q)}\ . \tag{12.57}$$

Research problem 12.5.6. Prove (or disprove) that there exists an integer q with the following property. Consider any value of δ, any value of M and any subset D of $\{0,1\}^M$ with $\mathsf{P}(D)\geq 1-1/q$. Then the set $D^{(q)}$ is δ-small.

If the occurrence of the condition $\mathsf{P}(D)\geq 1-1/q$ puzzles the reader, she should realize that we simply look for $\epsilon>0$ small and q large such that $D^{(q)}$ is δ-small whenever $\mathsf{P}(D)\geq 1-\epsilon$.

To understand this problem, it help to analyze a simple example. Consider the case where, for some integer k, we have $D=\{(\delta_i)_{i\leq M};\sum_{i\leq M}\delta_i=k\}$. Then

$$D^{(q)}=\Big\{(\delta_i)_{i\leq M};\ ;\ \sum_{i\leq M}\delta_i\geq kq+1\Big\}\subset\bigcup_{I\in\mathcal{G}}H_I\ ,$$

where $\mathcal{G}=\{I\subset\{1,\ldots,M\};\operatorname{card}I=kq+1\}$. Thus, using the elementary inequality

$$\binom{n}{k} \le \left(\frac{en}{k}\right)^k \tag{12.58}$$

we obtain

$$\sum_{I \in \mathcal{G}} \delta^{\operatorname{card} I} = \binom{M}{kq+1} \delta^{kq+1} \le \left(\frac{eM\delta}{kq+1}\right)^{kq+1} . \tag{12.59}$$

It is elementary to show that when $\mathsf{P}(D) \ge 1/2$ one has $k \ge \delta M/L$. It then follows that if q is a large enough universal constant, the right-hand side of (12.59) is $\le 1/2$. Of course one has to believe that this simple case is "extremal" i.e. "the worst possible". This is part of the beauty of Problem 12.5.6: it "suffices" to invent a new type of set D to solve it negatively!

A positive solution to Problem 12.5.6 will be rewarded by a \$ 1000 prize, even if it applies only to sufficiently small values of δ. It seems probable that progress on this problem requires methods unrelated to those of this book. A simple positive result in the right direction is provided in the next section.

The author has spent considerable energy on Problem 12.5.6. A sequence of conjectures of increasing strength, of which a positive answer to (a weak version of) Problem 12.5.6 is the weakest, can be found in [4]. Here we shall only state one of them, in an attempt to convey the beauty of these questions. Until the end of the present section, the arguments will be a bit sketchy, since all the details are provided in [4], and our only goal is to advertise for this paper. The reader who finds the material too demanding should simply move to the next section. Let us say that a subset D of $\{0,1\}^M$ is a *downset* if $y = (y_i)_{i \le m} \in D \Rightarrow x = (x_i)_{i \le m} \in D$ whenever $x_i \le y_i$ for all $i \le M$. Consider any set D and the smallest downset D' which contains D. Then it is obvious from the definition that $D'^{(q)} = D^{(q)}$. This implies that it suffices to solve Problem 12.5.6 when D is a downset, and explains why in the next problem we consider only downsets.

In words, one may reformulate Problem 12.5.6 by saying that "the set of points far from D is δ-small", where points are far from D if they belong to $D^{(q)}$ for large q. The purpose of the next construction is introduce a related idea which dispenses with the consideration of the strange sets $D^{(q)}$. Consider a number $0 < \alpha \le 1$, which one should think of as a universal constant. For $x = (x_i)_{i \le M} \in \{0,1\}^M$ let $\theta_{x,\alpha}$ be the product probability on $\{0,1\}^M$ such that on the factor of rank i, the weight of 1 is $x_i \alpha \in \{0, \alpha\}$. Instead of thinking that "the points far away from D are those in $D^{(q)}$" we instead think "that the points x far away from D are those for which $\theta_{x,\alpha}(D)$ is small". The relationship between these two ideas is as follows.

Lemma 12.5.7. *Given $0 < \alpha < 1$ there exists an integer q with the property that for each downset D and any x,*

$$\theta_{x,\alpha}(D) \ge 1 - 1/q \Rightarrow x \notin D^{(q)} . \tag{12.60}$$

Proof. For a finite set I, denote by $\lambda_{I,\alpha}$ the product probability on $\{0,1\}^I$ which gives weight α to 1 on each factor. The key to the proof is the following fact. Given $0 < \alpha < 1$ one can find an integer q with the following property. Given any subset B of $\{0,1\}^I$ one has $\lambda_{I,\alpha}(B) \ge 1 - 1/q \Rightarrow B^{(q)} = \emptyset$. This is not very difficult, and we will leave this as a teaser for the energetic reader. The proof can be found in [4]. As a hint, let us just say that when $\alpha = 1/2$, one may take $q = 2$.

Fixing x, let $I = \{i \le M; x_i = 1\}$, and let D_I be the projection of D onto $\{0,1\}^I$. For $y \in \{0,1\}^I$, let us define $y^* \in \{0,1\}^M$ by $y^*_i = y_i$ if $i \in I$ and $y_i = 0$ otherwise. Let $D^* = \{y \in \{0,1\}^I; y^* \in D\}$. Then the definition of $\theta_{x,\alpha}$ shows that $\theta_{x,\alpha}(D) = \lambda_{I,\alpha}(D^*)$. Also, since D is a downset, we have $D^* = D_I$, and thus $\theta_{x,\alpha}(D) = \lambda_{I,\alpha}(D_I)$.

Using the observation at the beginning of the proof for $B = D_I$, this shows that if $\theta_{x,\alpha}(D) = \lambda_{I,\alpha}(D_I) \ge 1 - 1/q$, then $D_I^{(q)} = \emptyset$. In particular the element y of $\{0,1\}^I$ with all components equal to 1 is not in $D_I^{(q)}$. This means that we can find $y^1, \dots, y^q \in D_I$ such that for all $i \in I$, one of the components y^ℓ_i is 1. Going back to the definition of $D^{(q)}$, this shows in turn that $x \notin D^{(q)}$. □

Research problem 12.5.8. Does there exist a universal constant $\alpha > 0$ with the following property. Consider any values of δ and M, any downset D and any probability measure ν on $\{0,1\}^M$. Assume that

$$\forall\, I \subset \{1,\dots,M\}\ ,\ \nu(H_I) \le (\alpha\delta)^{\operatorname{card} I}\ . \tag{12.61}$$

Then

$$\mathsf{P}(D) \le \prod_{x \in \{0,1\}^M} \theta_{x,\alpha}(D)^{\nu(\{x\})}\ . \tag{12.62}$$

First we explain the nature of this problem. Recalling that $\theta_{x,\alpha}(D) = \lambda_{J,\alpha}(D_J)$ where $J = \{i \le M; x_i = 1\}$ and D_J is the projection of D onto $\{0,1\}^J$, the right-hand side of (12.62) is a product of powers of "measures of projections of D." Of course, the smaller the power attached to a given projection, the closer the corresponding term is to 1 and the less this term matters. The condition (12.61) ensures that the powers attached to these projections "are scattered across very different directions", in the sense that the sum of the powers attached to the projections on $\{0,1\}^J$ where J is a supset of any given set I is $\sum_{x \in H_I} \nu(\{x\}) = \nu(H_I)$ which does not exceed $(\alpha\delta)^{\operatorname{card} I}$ by (12.61). The main difficulty of the problem is to gain an understanding of the probability measures ν which satisfy (12.61). The purpose of this condition is going to be explained soon in more detail.

In some sense Problem 12.5.8 is more appealing than Problem 12.5.6. Not only does it dispense with the strange set $D^{(q)}$ but it requests a single inequality. In the next page or so we investigate the relationship between both problems with the goal of proving that a positive solution to Problem 12.5.8 implies a positive solution of a weaker form of Problem 12.5.6. This weaker form involves a weakening of the notion of δ-small set, and we explain this first. The condition (12.51), i.e. $\Omega \subset \bigcup_{I \in \mathcal{G}} H_I$ implies

$$\mathbf{1}_\Omega \le \sum_{I \in \mathcal{G}} \beta_I((\delta)^{-\operatorname{card} I}\mathbf{1}_{H_I}) , \tag{12.63}$$

where $\beta_I := \delta^{\operatorname{card} I}$ satisfies $\sum_{I \in \mathcal{G}} \beta_I \le 1/2$. This information is not as good as (12.51), but nonetheless it is of the same nature, a concrete witness of the smallness of Ω. (One may observe that the right-hand side has an integral $\le 1/2$ for P.)

Let go back to the problem of comparing Problem 12.5.6 with Problem 12.5.8, and let us assume that the latter has a positive solution. Let $0 < \alpha < 1$ be the number provided by this solution, and let q be then provided by Lemma 12.5.7. Let us consider a down set D with $\mathsf{P}(D) \ge 1 - 1/q$ and a probability measure ν which satisfy (12.61). Since $\theta_{x,\alpha}^{\nu(\{x\})} = 1$ when $\nu(\{x\}) = 0$, and since ν is a probability, (12.62) implies that we can find x with $\nu(\{x\}) > 0$ and $\mathsf{P}(D) \le \theta_{x,\alpha}(D)$. Lemma 12.5.7 shows that then $x \notin D^{(q)}$, and hence that ν is not carried by $D^{(q)}$.

In summary, if Problem 12.5.8 has a positive solution, we find an integer q such that, given a downset with $\mathsf{P}(D) \ge 1 - 1/q$, a probability measure ν supported by $D^{(q)}$ must fail (12.61). We have already observed that if $D^{(q)}$ carries a probability measure which satisfies (12.61) it cannot be δ-small. It turns out that when $D^{(q)}$ carries no such measure, one can find coefficients $\alpha_I \ge 0$ with $\sum_{I \subset \{1,\dots,M\}} \alpha_I \le 1$ and

$$\mathbf{1}_{D^{(q)}} \le \sum_{I \ne \emptyset} \alpha_I(\alpha\delta)^{-\operatorname{card} I}\mathbf{1}_{H_I} , \tag{12.64}$$

and, in particular, setting $\beta_I = 2^{-\operatorname{card} I}\alpha_I$, we have $\sum_{I \ne \emptyset} \beta_I \le 1/2$, and (12.64) implies

$$\mathbf{1}_{D^{(q)}} \le \sum_{I \ne \emptyset} \beta_I(\alpha\delta/2)^{-\operatorname{card} I}\mathbf{1}_{H_I} \tag{12.65}$$

which is exactly the concrete smallness condition (12.63), at the expense of replacing δ by $\alpha\delta/2$ (which arguably is not a dramatic weakening).

The proof of (12.64) relies on the Hahn-Banach theorem. Consider the class $\mathcal{C}$ of functions f on $\{0,1\}^M$ for which there exists coefficients α_I with $\sum_{I \subset \{1,\dots,M\}} \alpha_I \le 1$ and $f \le \sum_I \alpha_I(\alpha\delta)^{-\operatorname{card} I}\mathbf{1}_{H_I}$. We prove by contradiction that we must have $\mathbf{1}_{D^{(q)}} \in \mathcal{C}$. Otherwise we could use the Hahn-Banach theorem to separate $\mathcal{C}$ and $\mathbf{1}_{D^{(q)}}$, that is to find a linear functional φ on the space of all functions on $\{0,1\}^M$ such that, for each $f \in \mathcal{C}$, we have $\varphi(f) < \varphi(\mathbf{1}_{D^{(q)}})$. Since $0 \in \mathcal{C}$ we may as well assume that $\varphi(\mathbf{1}_{D^{(q)}}) = 1$. Since $f \in \mathcal{C}$ whenever $f \le 0$, we see then that φ is given by a positive measure on $D^{(q)}$. The fact that $\varphi((\alpha\delta)^{-\operatorname{card} I}\mathbf{1}_{H_I}) < 1$ for each I shows that $\varphi(\mathbf{1}_{H_I}) < (\alpha\delta)^{\operatorname{card} I}$, so that the restriction ν of φ to $D^{(q)}$ is a probability ν which satisfies (12.61). This contradiction proves that $\mathbf{1}_{D^{(q)}} \in \mathcal{C}$. That is, one can find coefficients α_I as in (12.64).

The main difficulty in Problem 12.5.8 is that it is unclear how to take advantage of condition (12.61). An obvious example of probability measure which satisfies this condition is the product probability when 1 is given weight $\alpha\delta$ on each factor. A less obvious example is given by the following.

Exercise 12.5.9. Prove (12.61) when ν is uniform on the set

$$\Big\{(\delta_i)_{i\le M};\sum_{i\le M}\delta_i=k\Big\}\,,$$

where $k\le\alpha\delta M$. (Hint: try first the case $k=1$, or see [4].)

It is proved in [4] that for such a probability measure ν, (12.62) holds for each downset D.

Exercise 12.5.10. Consider the set $D=\{(\delta_i)_{i\le M};\sum_{i\le M}\delta_i\le\delta M\}$, and assume that $\delta M\in\mathbb{N}$. Consider ν as in Exercise 12.5.9, where $k\le\delta M$.
(a) Prove that $\theta_{x,\alpha}(D)=1$ if and only if $\sum_{i\le M}x_i\le\delta M$.
(b) Observe that $\nu(\{x\})>0$ if and only if $\sum_{i\le M}x_i=k$.
(c) Prove that the right-hand side of (12.62) is equal to 1.

12.6 Classes of Sets

In this section we consider positive selector processes in the simpler case where T consists of indicators of sets. That is, we consider a class $\mathcal{J}$ of subsets of $\{1,\dots,M\}$ and we try to bound the quantity

$$\mathcal{E}_\delta(\mathcal{J}):=\mathsf{E}\sup_{J\in\mathcal{J}}\sum_{i\in J}\delta_i\,.$$

We first give a few simple facts before proving the main result of this section, Theorem 12.6.4.

Proposition 12.6.1. *Assume that for some number S we have*

$$\sum_{J\in\mathcal{J}}\Big(\frac{\delta\operatorname{card}J}{S}\Big)^S\le 1/2\,. \tag{12.66}$$

Then

$$\mathcal{E}_\delta(\mathcal{J})=\mathsf{E}\sup_{J\in\mathcal{J}}\sum_{i\in J}\delta_i\le LS\,. \tag{12.67}$$

Proof. We first observe that by (12.66) we have $\delta\operatorname{card}J\le S$ whenever $J\in\mathcal{J}$, so that $u\ge 6\delta\operatorname{card}J$ whenever $u\ge 6S$. We then simply use Lemma 12.3.1 to obtain that for $u\ge 6S$ we have

$$\mathsf{P}\Big(\sup_{J\in\mathcal{J}}\sum_{i\in J}\delta_i\ge u\Big)\le\sum_{J\in\mathcal{J}}\mathsf{P}\Big(\sum_{i\in J}\delta_i\ge u\Big)\le\sum_{J\in\mathcal{J}}\Big(\frac{2\delta\operatorname{card}J}{u}\Big)^{u/2}\,,$$

and we integrate using (12.66) and simple estimates. □

For a class of sets $\mathcal{J}$ of sets, let us define $\mathcal{S}_\delta(\mathcal{J})$ as the infimum of the numbers S for which (12.66) holds. Thus (12.67) simply means

$$\mathcal{E}_\delta(\mathcal{J}) \le L\mathcal{S}_\delta(\mathcal{J}) \ . \tag{12.68}$$

Exercise 12.6.2. Prove that inequality (12.68) cannot be reversed. That is, given $A > 0$ construct a class $\mathcal{J}$ of sets for which $A\mathcal{E}_d(\mathcal{J}) \le \mathcal{S}_\delta(\mathcal{J})$. (Hint: consider the class $\mathcal{J}$ of subsets of a given set J.)

Given a class $\mathcal{J}$ of sets and two integers n and m let us define the class $\mathcal{J}(n,m)$ as follows:

$$\forall\, J \in \mathcal{J}(n,m)\ ,\ \exists J_1,\dots,J_n \in \mathcal{J}\ ;\ \operatorname{card}\Big(J \setminus \bigcup_{\ell\le n} J_\ell\Big) \le m\ . \tag{12.69}$$

Then for each realization of the r.v.s (δ_i) one has

$$\sum_{i\in J}\delta_i \le m + \sum_{\ell\le n}\sum_{i\in J_\ell}\delta_i$$

and consequently

$$\mathcal{E}_\delta(\mathcal{J}(n,m)) \le n\mathcal{E}_\delta(\mathcal{J}) + m\ . \tag{12.70}$$

Combining (12.70) and (12.68) one obtains

$$\mathcal{E}_\delta(\mathcal{J}(n,m)) \le Ln\mathcal{S}_\delta(\mathcal{J}) + m\ . \tag{12.71}$$

In particular, taking $n = 1$, for two classes $\mathcal{I}$ and $\mathcal{J}$ of sets one has

$$\mathcal{I} \subset \mathcal{J}(1,m) \Rightarrow \mathcal{E}_\delta(\mathcal{I}) \le LS_\delta(\mathcal{J}) + m\ ,$$

and thus

$$\mathcal{E}_\delta(\mathcal{I}) \le L\inf\{\mathcal{S}_\delta(\mathcal{J}) + m\ ;\ \mathcal{I} \subset \mathcal{J}(1,m)\}\ , \tag{12.72}$$

where the infimum is over all classes of sets $\mathcal{J}$ for which $\mathcal{I} \subset \mathcal{J}(1,m)$. The following (very) challenging exercise disproves a most unfortunate conjecture stated in [2] and [4], which overlooked the possibility of taking $n \ge 2$ in (12.71).

Exercise 12.6.3. Using the case $n = 2$, $m = 0$ of (12.71), prove that the inequality (12.72) cannot be reversed. That is, given $A > 0$, construct a class of sets $\mathcal{I}$ such that $A\mathcal{E}_\delta(\mathcal{I}) \le S_\delta(\mathcal{J}) + m$ for each class of sets $\mathcal{J}$ and each m for which $\mathcal{I} \subset \mathcal{J}(1,m)$.

In words, we can prove that (12.72) cannot be reversed because we have found a genuinely different way to bound $\mathcal{E}_\delta(\mathcal{I})$, namely (12.71) for $n = 2$.

In the same line as Exercise 12.6.3 it seems worth investigating whether given a number A we can construct a class of sets $\mathcal{I}$ such that $A\mathcal{E}_\delta(\mathcal{I}) \le n\mathcal{S}(\mathcal{J}) + m$ whenever $\mathcal{I} \subset \mathcal{J}(n,m)$. This seems plausible, because we have a (seemingly) more general way to bound $\mathcal{E}(\mathcal{I})$ than (12.70), namely the "solid convex hull" method of Section 12.4.

In the remainder of this section we prove the following.

Theorem 12.6.4 ([2]). *When T is a class of indicators of sets, Problem 12.5.2 has a positive solution.*

To prove this result, we consider a class $\mathcal{J}$ of subsets of $\{1, \ldots, M\}$ and an integer n. We assume that the event

$$\Big\{\sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i \geq n\Big\} \tag{12.73}$$

is not δ-small. The goal is to prove that

$$\mathsf{E} \sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i \geq n/L_0 \ . \tag{12.74}$$

Before this, we deduce Theorem 12.6.4 from (12.74). We prove that the event

$$\Big\{\sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i \geq 2L_0 \mathsf{E} \sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i\Big\} \tag{12.75}$$

is δ-small. Consider the smallest integer $n \geq 2L_0 \mathsf{E}\sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i > 0$. Then (12.74) fails and the event (12.75) coincides with the event $\{\sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i \geq n\}$. So it must be δ-small since (12.74) fails, proving Theorem 12.6.4.

We turn to the proof of (12.74). We fix n once and for all, and we define

$$\mathcal{J}' = \{J' \subset \{1, \ldots, M\} \ ; \ \operatorname{card} J' = n \ , \ \exists J \in \mathcal{J} \ , \ J' \subset J\} \ . \tag{12.76}$$

We observe that

$$\Big\{\sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i \geq n\Big\} = \Big\{\sup_{J\in\mathcal{J}'} \sum_{i\in J} \delta_i \geq n\Big\} \ . \tag{12.77}$$

For an integer $1 \leq k \leq n$ we set

$$d(k) = 2\Big(\frac{4en\delta}{k}\Big)^k \ . \tag{12.78}$$

Lemma 12.6.5. *Assume that the event (12.77) is not δ-small. Then there exists a probability measure ν on $\mathcal{J}'$ with the following property. For each set A with $1 \leq \operatorname{card} A \leq n$ we have*

$$\nu(\{J \in \mathcal{J}' \ ; \ A \subset J\}) \leq d(\operatorname{card} A) \ . \tag{12.79}$$

Proof. This will follow from the Hahn-Banach theorem. For such a set A consider the function f_A on $\mathcal{J}'$ given by

$$f_A(J) = \frac{1}{d(\operatorname{card} A)} \mathbf{1}_{\{A\subset J\}} \ .$$

To prove the existence of ν it suffices to prove that any convex combination of functions of the type f_A takes at least one value < 1, since then by the

Hahn-Banach theorem there exists a probability measure ν on $\mathcal{J}'$ for which $\int f_A \mathrm{d}\nu \le 1$ for each A, and therefore ν satisfies (12.79). Suppose, for contradiction, that this is not the case, so that there exist coefficients $\alpha_A \ge 0$ of sum 1 for which

$$\forall J \in \mathcal{J}' \, , \; \sum_A \alpha_A f_A(J) = \sum_{A \subset J} \frac{\alpha_A}{d(\operatorname{card} A)} \ge 1 \, . \tag{12.80}$$

For $1 \le k \le n$ let $\mathcal{G}_k$ be the collection of all the sets A for which $\operatorname{card} A = k$ and $\alpha_A \ge 2^{k+1}\delta^k$. Since $\sum_A \alpha_A = 1$ we observe that $\operatorname{card} G_k \le \delta^{-k}2^{-k-1}$, and thus

$$\sum_{k \ge 1} \delta^k \operatorname{card} G_k \le \frac{1}{2} \, . \tag{12.81}$$

We claim that

$$\forall J \in \mathcal{J}' \; ; \; \exists k \le n \, , \; \exists A \in \mathcal{G}_k \; ; \; A \subset J \, . \tag{12.82}$$

Indeed, otherwise we can find $J \in \mathcal{J}'$ for which

$$A \subset J \, , \; \operatorname{card} A = k \, , \; k \le n \Rightarrow \alpha_A < 2^{k+1}\delta^k$$

and thus, using the definition of $d(k)$ and (12.58),

$$\sum_{A \subset J} \frac{\alpha_A}{d(\operatorname{card} A)} < \sum_{1 \le k \le n} \binom{n}{k} \frac{2^{k+1}\delta^k}{d(k)} \le 1 \, .$$

This contradicts (12.80) and proves (12.82). Consider $\mathcal{G} = \bigcup_{1 \le k \le n} \mathcal{G}_k$, which satisfies (12.50) from (12.81). Consider $(\delta_i)_{i \le M}$ such that $\sum_{i \in J} \delta_i \ge n$ for some $J \in \mathcal{J}'$. Then (12.82) proves that J contains a set $A \in \mathcal{G}$, so that $(\delta_i)_{i \le M} \in H_A$, and we have shown that the event (12.77) is contained in $\cup_{I \in \mathcal{G}} H_I$. Thus the event (12.77) is δ-small, a contradiction which finishes the proof. □

Lemma 12.6.6. *Assume that the event (12.77) is not δ-small. Then this event has a probability $\ge \exp(-Ln)$.*

Proof. Consider the probability ν on the set $\mathcal{J}'$ of (12.76) as in (12.79) and the r.v. (depending on the random input $(\delta_i)_{i \le M}$)

$$Y = \nu(\{J \in \mathcal{J}' \; ; \; \forall i \in J \, , \; \delta_i = 1\}) = \nu(\{J \; ; \; (\delta_i) \in H_J\}) = \int \mathbf{1}_{\{(\delta_i) \in H_J\}} \mathrm{d}\nu(J) \, .$$

Obviously the event (12.77) contains the event $Y > 0$. The plan is to use the Paley-Zygmund inequality in the weak form

$$\mathsf{P}(Y > 0) \ge \frac{(\mathsf{E}Y)^2}{\mathsf{E}Y^2} \, , \tag{12.83}$$

which is a simple consequence of the Cauchy-Schwarz inequality. First,

$$\mathsf{E}Y = \mathsf{E}\int \mathbf{1}_{\{(\delta_i)\in H_J\}} \mathrm{d}\nu(J) = \int \mathsf{P}(H_J)\mathrm{d}\nu(J) = \delta^n \ , \tag{12.84}$$

since ν is supported by $\mathcal{J}'$ andcard $J = n$ for $J \in \mathcal{J}'$. Next,

$$\begin{aligned} Y^2 &= \nu^{\otimes 2}(\{(J,J') \ ; \ (\delta_i) \in H_J, (\delta_i) \in H_{J'}\}) \\ &= \nu^{\otimes 2}(\{(J,J') \ ; \ (\delta_i) \in H_J \cap H_{J'}\}) \ , \end{aligned}$$

so that, proceeding as in (12.84), and since $\mathsf{P}((\delta_i) \in H_J \cap H_{J'}) = \delta^{\mathrm{card}(J\cup J')}$,

$$\mathsf{E}Y^2 = \int \delta^{\mathrm{card}(J\cup J')} \mathrm{d}\nu(J)\mathrm{d}\nu(J') \ . \tag{12.85}$$

Now, the choice $A = J \cap J'$ shows that

$$\delta^{\mathrm{card}(J\cup J')} \le \sum_{A\subset J} \delta^{2n-\mathrm{card}\,A} \mathbf{1}_{\{A\subset J'\}}$$

and therefore, using (12.84) and (12.83), and again (12.58),

$$\begin{aligned} \int \delta^{\mathrm{card}(J\cup J')} \mathrm{d}\nu(J') &\le \sum_{A\subset J} \delta^{2n-\mathrm{card}\,A} \nu(\{J' \ ; \ A \subset J'\}) \\ &\le \sum_{0\le k\le n} \binom{n}{k} \delta^{2n-k} d(k) \\ &\le 2\delta^{2n} \sum_{0\le k\le n} \Big(\frac{2en}{k}\Big)^{2k} \ . \end{aligned} \tag{12.86}$$

An elementary computation shows that the last term dominates in the sum, so that the right-hand side of (12.86) is less than $\le \delta^{2n} \exp Ln$, and recalling (12.85) this proves that $\mathsf{E}Y^2 \le \exp(Ln)(\mathsf{E}Y)^2$ and completes the proof using (12.83). □

Proof of (12.74). Consider the r.v. $X = \sup_{J\in\mathcal{J}} \sum_{i\in J} \delta_i$. We assume that the event $\{X \ge n\}$ is not δ-small. Combining Lemmas 12.6.5 and 12.6.6 we have proved that

$$\mathsf{P}(X \ge n) \ge \exp(-L_1 n) \ . \tag{12.87}$$

From this fact alone we shall bound from below $S := \mathsf{E}X$. Using Markov's inequality, we know that $\mathsf{P}(D) \ge 1/2$, where $D = \{X \le 2S\}$. Recalling the set $D^{(q)}$ defined on page 389, given two integers q and $k \ge 0$ we define similarly $D^{(q,k)}$ as the set of sequences $(\delta_i)_{i\le M}$ for such that given any sequences $(\delta_i^\ell)_{i\le M} \in D$ for $\ell \le q$, then

$$\mathrm{card}\{i \le M \ ; \ \delta_i = 1, \forall\, \ell \le q \ , \ \delta_i^\ell = 0\} \ge k + 1 \ .$$

Thus $D^{(q,0)} = D^q$, and as in (12.57) one proves that

$$\{X \geq 2qS + k\} \subset D^{(q,k)} \,. \tag{12.88}$$

The heart of the matter is Theorem 3.3.1 of [1] which asserts that since $\mathsf{P}(D) \geq 1/2$ we have

$$\mathsf{P}(D^{(q,k)}) \leq \frac{2^q}{q^k} \,.$$

Comparing with (12.87) and (12.88) then yields

$$2qS + k \geq n \Rightarrow \exp(-L_1 n) \leq \frac{2^q}{q^k} \,.$$

Let us fix q with $q \geq \exp(2L_1)$, so that q is now an universal constant. Let us assume that $2qS + k \geq n$. Then $\exp(-L_1 n) \leq 2^q/q^k$ so that $\exp(-L_1 n) \leq 2^q \exp(-2L_1 k)$ and thus $2k - n \leq L$, so that $k \leq n/2 + L$. Then $n \leq 2qS + k \leq 2qS + n/2 + L_2$ and thus $S \geq (n - L)/L$. We have proved that $S \geq n/L$ when $n \geq n_0$, where n_0 is a universal constant. This finishes the proof in that case. We finish now the proof when $n \leq n_0$. Since our assumption is that the event $X \geq n$ is not δ-small, the larger event $X \geq 1$ is not δ-small. This implies obviously that the set $I = \bigcup_{J \in \mathcal{J}} J$ satisfies $\delta \operatorname{card} I \geq 1/2$, so that $\mathsf{P}(\Omega) \geq 1/L$, where $\Omega = \{\exists i \in I, \delta_i = 1\}$. Now $X \geq \mathbf{1}_\Omega$ so that taking expectation we get $S \geq \mathsf{P}(\Omega) \geq 1/L$, which finishes the proof. □

References

1. Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. IHÉS **81**, 73–205 (1995)
2. Talagrand, M.: Selector processes on classes of sets. Probab. Theory Relat. Fields **135**(4), 471–486 (2006)
3. Talagrand, M.: The Generic Chaining. Springer Monographs in Mathematics. Springer, Berlin (2005). viii+222 pp. ISBN: 3-540-24518-9
4. Talagrand, M.: Are many small sets explicitly small? In: STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, pp. 13–35. ACM, New York (2010). Available at http://research.microsoft.com/apps/video/dl.aspx?id=137091

13. Convergence of Orthogonal Series; Majorizing Measures

13.1 Introduction

An orthonormal sequence $(\varphi_m)_{m\geq 1}$ on a probability space (Ω, P) is a sequence such that $\mathsf{E}\varphi_m^2 = 1$ for each n and $\mathsf{E}\varphi_m\varphi_n = 0$ for $m \neq n$. A classical question asks which are the sequences (a_m) for which the series

$$\sum_m a_m\varphi_m \tag{13.1}$$

converges a.s. whatever the choice of the orthonormal sequence (φ_m) and of the probability space. (See Section 13.7 for comments on this question.) Since the series $\sum_{m\geq 1} a_m\varepsilon_m$ must converge, where ε_m are independent Bernoulli r.v.s, we have $\sum_{m\geq 1} a_m^2 < \infty$. As we shall see, the condition $\sum_{m\geq 1} a_m^2 < \infty$ is however far from sufficient: there exists an orthonormal sequence (φ_m) and coefficients a_m such that $\sum_{m\geq 1} a_m^2 < \infty$ and the series $\sum_{m\geq 1} a_m\varphi_m$ diverges everywhere.

Let us consider the set

$$T = \Big\{\sum_{m\leq n} a_m^2\ ;\ n \geq 1\Big\}\,. \tag{13.2}$$

Since $\sum_{m\geq 1} a_m^2 < \infty$ we may assume without loss of generality that $T \subset]0,1]$. We may also assume that $a_m \neq 0$ for each m. Let us denote by $\mathcal{I}_n$ the family of the 2^n dyadic intervals $](i-1)2^{-n}, i2^{-n}]$ for $1 \leq i \leq 2^n$. For a point $t \in]0,1]$, we denote by $I_n(t)$ the unique interval of $\mathcal{I}_n$ that contains t.

Theorem 13.1.1 (A. Paszkiewicz [7]). *Given the sequence (a_m), and hence the set T, the following are equivalent.*
(a) The series (13.1) converges a.e. for every choice of the orthonormal sequence (φ_n).
(b) There exists a probability measure μ on T such that

$$\sup_{t\in T}\sum_{n\geq 0}\frac{1}{\sqrt{2^n\mu(I_n(t))}} < \infty\,. \tag{13.3}$$

(c) There exists a number B such that for every probability measure μ on T one has

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_13,

$$\sum_{n\geq 0}\sum_{I\in\mathcal{I}_n}\sqrt{2^{-n}\mu(I)}\leq B\ . \tag{13.4}$$

(d) There exists a number B' such that for each process $(X_t)_{t\in T}$ which satisfies

$$\forall\, s,t\in T\ ,\ \mathsf{E}(X_s-X_t)^2\leq |s-t|\ , \tag{13.5}$$

we have

$$\mathsf{E}\sup_{s,t\in T}|X_s-X_t|\leq B'\ . \tag{13.6}$$

(e) For each process $(X_t)_{t\in T}$ which satisfies (13.5), $\lim_{k\to\infty}X_{t_k}$ exists a.s. where $t_k=\sum_{m\leq k}a_m^2$.

At this stage, this theorem should look completely mysterious, and it is the purpose of this chapter to clarify the underlying issues. Also, it is not immediately obvious how to recover the following classical results on the present formulation. Doing this right away will help us get a feeling for the conditions of Theorem 13.1.1. On a less positive note, it will also illustrate that working with these conditions is not as easy as what one would like it to be.

Corollary 13.1.2 (Radmacher-Menchov [6], [8]). *If*

$$\sum_{m\geq 1}a_m^2(\log m)^2<\infty\ , \tag{13.7}$$

then for each choice of the orthonormal sequence (φ_m) the series $\sum_m a_m\varphi_m$ converges a.s.

Proof. We shall prove that (c) is satisfied. We consider a probability measure μ on T, and we aim to bound the left-hand side of (13.4). The plan is for each n to split the sum $\sum_{I\in\mathcal{I}_n}\sqrt{2^{-n}\mu(I)}$ in several suitable pieces, and to bound each of them using the Cauchy-Schwarz inequality (each time using that for a disjoint family $\mathcal{I}$ of intervals, $\sum_{i\in\mathcal{I}}\mu(I)=\mu(\cup_{I\in\mathcal{I}}I)$). But, first, we must reformulate (13.7). For $n\geq 1$ let $t_n=\sum_{1\leq m\leq n}a_m^2$. For $k\geq 0$ let $u_k=t_{2^{2^k}}$, so that

$$\sum_{k\geq 0}2^{2k}(u_{k+1}-u_k)=\sum_{k\geq 0}\sum_{2^{2^k}<m\leq 2^{2^{k+1}}}2^{2k}a_m^2\leq L\sum_{m\geq 2}a_m^2(\log m)^2<\infty\ . \tag{13.8}$$

In particular, we have $u_{k+1}-u_k\leq C2^{-2k}$ so that if $t^*=\sum_{m\geq 1}a_m^2$, then $t^*-u_k=\sum_{r\geq k}(u_{r+1}-u_r)\leq C2^{-2k}$.

We now fix k and consider $2^k\leq n<2^{k+1}$ and turn to the task of splitting the sum $\sum_{I\in\mathcal{I}_n}\sqrt{2^{-n}\mu(I)}$ in suitable pieces. Consider $I\in\mathcal{I}_n$ with $\mu(I)>0$. We claim that at least one of the following four cases must occur: either I contains a point u_p for $k-1\leq p\leq 2k$ or contains t^*; or else $I\subset]0,u_{k-1}]$; or else $I\subset]u_\ell,u_{\ell+1}]$ for some $k-1\leq\ell\leq 2k$; or finally $I\subset]u_{2k},t^*]$. To see this we simply observe that if the interval I does not contain either the point t^* or one of the points u_p for $k-1 \le p \le 2k$, then it must be contained in one the intervals created when removing these points from $[0,1]$, but, since $\mu(I) > 0$, it cannot be contained in the interval $]t^*, 1]$ because $\mu(]t^*, 1]) = 0$, so that it is contained in one of the other intervals left.

Consequently for $2^k \le n < 2^{k+1}$ we may write

$$\sum_{I \in \mathcal{I}_n} \sqrt{2^{-n}\mu(I)} = \mathrm{I} + \mathrm{II} + \mathrm{III} + \sum_{k-1 \le \ell \le 2k} V(\ell) \ , \tag{13.9}$$

where

- I is the sum over $I \subset]0, u_{k-1}]$. This sum has at most $2^{2^{k-1}}$ non-zero terms, because when $\mu(I) > 0$, I must contain a point t_m with $m \le 2^{2^{k-1}}$. The Cauchy-Schwarz inequality implies then that the sum is $\le 2^{-n/2} 2^{2^{k-2}} \le 2^{-2^{k-2}}$.
- II is the sum over the intervals I that contain a point u_p for $p \le 2k$ or that contain the point t^*. This sum has at most $2k+1$ terms and is bounded as above.
- III is the sum over the intervals contained in $]u_{2k}, t^*]$. Here we use that, if $u \le v$,

$$\sum_{I \in \mathcal{I}_n, I \subset]u,v]} \sqrt{2^{-n}\mu(I)} \le \sqrt{v-u}\sqrt{\mu(]u,v])} \ . \tag{13.10}$$

  This is simply the Cauchy-Schwarz inequality, since the sum has at most $2^n(v-u)$ terms, and since $\sum_{I \subset]u,v]} \mu(I) \le \mu(]u,v])$. Thus III $\le \sqrt{t^* - u_{2k}} \le C 2^{-k}$.
- $V(\ell)$ is the sum over the intervals $I \subset]u_\ell, u_{\ell+1}]$, which, as witnessed by (13.10), is bounded by $\sqrt{u_{\ell+1} - u_\ell}\sqrt{\mu(]u_{\ell+1} - u_\ell])}$.

Summation of the inequalities (13.9) over n with $2^k \le n < 2^{k+1}$ and then over k yields that for a certain number C'

$$\sum_{n \ge 0} \sum_{I \in \mathcal{I}_n} \sqrt{2^{-n}\mu(I)} \le C' + \sum_{k \ge 1} 2^k \sum_{k-1 \le \ell \le 2k} V(\ell) \ .$$

Now,

$$\sum_{k \ge 1} 2^k \sum_{k-1 \le \ell \le 2k} V(\ell) \le \sum_{\ell \ge 0} V(\ell) \sum_{k-1 \le \ell} 2^k \le 4 \sum_{\ell \ge 0} 2^\ell V(\ell) \ ,$$

and

$$\sum_{\ell \ge 0} 2^\ell V(\ell) \le \sum_{\ell \ge 0} 2^\ell \sqrt{u_{\ell+1} - u_\ell}\sqrt{\mu(]u_{\ell+1} - u_\ell])} < \infty$$

using (13.8) and the Cauchy-Schwarz inequality. □

Corollary 13.1.3 (Tandori [10]). *If for each choice of the orthonormal sequence* (φ_m) *the series* $\sum_m a_m \varphi_m$ *converges a.s. then*

$$\sum_{m \geq 1} a_m^2 \log(2/a_m)^2 < \infty \,. \tag{13.11}$$

Before we start the proof, it is useful to spell out an elementary principle that we will use several times during the course of the present chapter. We denote by λ Lebesgue's measure on $[0,1]$.

Lemma 13.1.4. *Consider a collection* $\mathcal{W}$ *of disjoint sub-intervals of* $[0,1]$ *of the type* $]u,v]$*, and assume that for a number* $d > 0$ *we have*

$$\forall\, W \in \mathcal{W} \,,\ \lambda(W) \leq d \,. \tag{13.12}$$

For $W \in \mathcal{W}$*, denote by* λ_W *the restriction of Lebesgue's measure to* W *and consider a measure* θ_W *concentrated on* W*. Assume moreover that for a number* δ *we have*

$$\forall\, W \in \mathcal{W} \,,\ |\theta_W(W) - \lambda(W)| \leq \delta \,. \tag{13.13}$$

Let $\theta_{\mathcal{W}} = \sum_{W \in \mathcal{W}} \theta_W$ *and* $\lambda_{\mathcal{W}} = \sum_{W \in \mathcal{W}} \lambda_W$*. Then for each interval* $I \subset [0,1]$ *we have*

$$|\theta_{\mathcal{W}}(I) - \lambda_{\mathcal{W}}(I)| \leq 2d + \delta \operatorname{card} \mathcal{W} \,. \tag{13.14}$$

Proof. We observe that

$$|\theta_{\mathcal{W}}(I) - \lambda_{\mathcal{W}}(I)| \leq \sum_{W \in \mathcal{W}} |\theta_W(I) - \lambda_W(I)| \,.$$

For each $W \in \mathcal{W}$ we have $|\theta_W(I) - \lambda_W(I)| \leq d + \delta$ since $0 \leq \lambda_W(I) = \lambda(I \cap W) \leq \lambda(W) \leq d$ and $0 \leq \theta_W(I) \leq \theta_W(W) \leq d + \delta$ by (13.13). There are at most two intervals $W \in \mathcal{W}$ which contain an endpoint of I. For the others, we have $\theta_W(I) = \lambda_W(I) = 0$ if $I \cap W = \emptyset$ and

$$|\theta_W(I) - \lambda_W(I)| = |\theta_W(W) - \lambda_W(W)| \leq \delta$$

if W is entirely contained in I. □

Proof of Corollary 13.1.3. The first task is to reformulate (13.11) in a suitable way. Let $t_0 = 0$ and for $n \geq 1$ let $t_n = \sum_{m \leq n} a_m^2$. For $k \geq 0$ let

$$U_k = \{m \,;\ 2^{-2^{k+1}} < a_m^2 \leq 2^{-2^k}\} \,.$$

The point of this definition is of course that $\log(2/a_m)$ is about 2^k for $m \in U_k$. Let

$$b_k := \sum \{a_m^2 \,;\ m \in U_k\} = \sum \{a_m^2 \,;\ 2^{-2^{k+1}} < a_m^2 \leq 2^{-2^k}\} \,.$$

Since $\log(2/a_m) \leq L2^k$ for $m \in U_k$, it suffices to prove that $\sum_{k \geq 0} 2^{2k} b_k < \infty$. Defining $J = \{k \geq 2; b_k \geq 2^{-4k}\}$, it even suffices to prove that

$$\sum_{k\in J} 2^{2k} b_k < \infty \,. \tag{13.15}$$

The principle of the proof is to apply (13.4) to a suitable probability measure, which we construct now. For $k \in J$ consider the probability measure μ_k on T given by $\mu_k(\{t_n\}) = a_n^2/b_k$ if $n \in U_k$ and $\mu_k(\{t_n\}) = 0$ if $n \notin U_k$. Consider numbers $(\alpha_k)_{k\in J}$ with $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$, such that all but finitely many of these numbers are not 0. Consider the probability measure $\mu = \sum_k \alpha_k \mu_k$. Consider n with $2^{k-1} \leq n < 2^k$. We write

$$\sum_{I\in\mathcal{I}_n} \sqrt{2^{-n}\mu(I)} \geq \sqrt{\alpha_k} \sum_{I\in\mathcal{I}_n} \sqrt{2^{-n}\mu_k(I)} \,. \tag{13.16}$$

The key fact is the inequality

$$\sum_{I\in\mathcal{I}_n} \sqrt{2^{-n}\mu_k(I)} \geq \frac{1}{2}\sqrt{b_k} \,. \tag{13.17}$$

Once this is proved, we sum (13.16) over $2^{k-1} \leq n < 2^k$, and then over k to obtain, using also (13.4) in the first inequality:

$$B \geq \sum_{n\geq 0} \sum_{I\in\mathcal{I}_n} \sqrt{2^{-n}\mu(I)} \geq \frac{1}{4} \sum_{k\in J} \sqrt{\alpha_k} 2^k \sqrt{b_k} \,,$$

and since the sequence α_k is arbitrary with $\sum \alpha_k = 1$ then $\sum_{k\in J} 2^{2k} b_k \leq 16B^2$ and (13.15) follows.

To prove (13.17) it suffices to prove that if $2^{k-1} \leq n < 2^k$ then for each $I \in \mathcal{I}_n$

$$\mu_k(I) \leq \frac{2^{-n+1}}{b_k} \,, \tag{13.18}$$

because then $\sqrt{2^{-n}\mu_k(I)} \geq \mu_k(I)\sqrt{b_k}/2$, from which (13.17) follows by summation over $I \in \mathcal{I}_n$. Consider the family $\mathcal{W}$ consisting of the intervals $W =]t_{m-1}, t_m]$ for $m \in U_k$, and let θ_W consist of the mass a_m^2 at t_m. Since $\lambda(]t_{m-1}, t_m]) = a_m^2$, (13.12) holds with $d = 2^{-2^k}$ and (13.13) holds for $\delta = 0$. With the notation of Lemma 13.1.4, we have $b_k\mu_k = \theta_{\mathcal{W}}$. Thus (13.14) implies

$$b_k\mu_k(I) \leq \lambda_{\mathcal{W}}(I) + 2d \leq \lambda(I) + 2d \leq 2^{-n+1}$$

because $\lambda(I) = 2^{-n}$ and $d = 2^{-2^k} \leq 2^{-n-1}$ for $n < 2^k$. This proves (13.18) and hence (13.17) and completes the proof. □

The necessary condition (13.11) is by no means sufficient for the convergence of each series $\sum_{m\geq 1} a_m \varphi_m$. This is obvious from Theorem 13.4.1 below.

Exercise 13.1.5. Prove that the conditions (13.11) and (13.7) are equivalent when the sequence (a_m) is non-increasing. Consequently, (13.11) is a necessary and sufficient condition so that one can find a permutation π such that the series $\sum_m a_{\pi(m)}\varphi(m)$ converges a.s. for each orthonormal sequence (φ_m).

Exercise 13.1.6. For a finite subset T of $]0,1]$, consider the following quantity $M(T)$. If $\operatorname{card} T = 1$, we set $M(T) = 0$. Otherwise, let $n(T)$ be the largest integer such that there exists $I \in \mathcal{I}_{n(T)}$ for which $T \subset I$. Call I_T this interval. Define then

$$M(T) = \inf \sup_{t \in T} \sum_{n \geq n(T)} \frac{1}{\sqrt{2^n \mu(I_n(t))}} \ ,$$

where the infimum is computed over all choices of probability measures on T. Now, I_T is the union of two intervals I_1 and I_2 of $\mathcal{I}_{n(T)+1}$. Explain how to compute $M(T)$ when you know $M(T \cap I_j)$ for $j = 1, 2$. In this manner the quantity $M(T)$ can be "computed recursively".

We now describe our approach to Theorem 13.1.1. The following is an obvious consequence of orthonormality.

Lemma 13.1.7. *For $t = \sum_{m \leq n} a_m^2 \in T$, let us define*

$$X_t = \sum_{m \leq n} a_m \varphi_m \ . \tag{13.19}$$

Then

$$\forall\, s, t \in T \ , \ \mathsf{E}(X_s - X_t)^2 = |s - t| \ . \tag{13.20}$$

This makes it obvious that (e) implies (a). It also motivates the following.

Definition 13.1.8. *If T is a subset of $[0,1]$ we say that the process $(X_t)_{t \in T}$ is* orthonormal *if it satisfies (13.20) and if moreover $\mathsf{E}X_t = 0$ for each t.*

The main ingredient in the proof of Theorem 13.1.1 is the following result.

Theorem 13.1.9 (W. Bednorz [4]). *Consider a finite subset T of $[0,1]$ and define*

$$F^*(T) = \sup \mathsf{E} \sup_{t \in T} X_t \ , \tag{13.21}$$

where the supremum is taken over all orthonormal processes indexed by T. Then for each probability measure μ on T we have

$$\sum_{n \geq 0} \sum_{I \in \mathcal{I}_n} \sqrt{2^{-n} \mu(I)} < L(1 + F^*(T)) \ . \tag{13.22}$$

Our first task it to make the link between Theorem 13.1.1 and Theorem 13.1.9.

Lemma 13.1.10. *If the process $(X_t)_{t\in T}$ is orthonormal, then*

$$t_1 \le t_2 \le t_3 \le t_4 \in T \Rightarrow \mathsf{E}(X_{t_4} - X_{t_3})(X_{t_2} - X_{t_1}) = 0 \,. \tag{13.23}$$

Proof. Consider $t_1 \le t_2 \le t_3 \in T$. Then

$$\begin{aligned} t_3 - t_1 &= \mathsf{E}(X_{t_3} - X_{t_1})^2 \\ &= \mathsf{E}(X_{t_3} - X_{t_2})^2 + \mathsf{E}(X_{t_2} - X_{t_1})^2 + 2\mathsf{E}(X_{t_3} - X_{t_2})(X_{t_2} - X_{t_1}) \\ &= t_3 - t_2 + t_2 - t_1 + 2\mathsf{E}(X_{t_3} - X_{t_2})(X_{t_2} - X_{t_1}) \,, \end{aligned}$$

so that we have proved

$$t_1 \le t_2 \le t_3 \in T \Rightarrow \mathsf{E}(X_{t_3} - X_{t_2})(X_{t_2} - X_{t_1}) = 0 \,. \tag{13.24}$$

We use (13.24) to write

$$\begin{aligned} 0 &= \mathsf{E}(X_{t_4} - X_{t_3})(X_{t_3} - X_{t_1}) \\ &= \mathsf{E}(X_{t_4} - X_{t_3})(X_{t_2} - X_{t_1}) + \mathsf{E}(X_{t_4} - X_{t_3})(X_{t_3} - X_{t_2}) \\ &= \mathsf{E}(X_{t_4} - X_{t_3})(X_{t_2} - X_{t_1}) \,, \end{aligned}$$

using again (13.24) in the third inequality. □

We will also need a classical result of Tandori [11]. This lemma really brings out the strength of the statement "for every orthonormal sequence..."

Lemma 13.1.11. *Assume that the sequence (a_n) has the property that for every orthonormal sequence (φ_n), the series $\sum_{m\ge1} a_m\varphi_m$ converges a.e. Then there exists a number A such that for each orthonormal sequence (φ_n) we have*

$$\mathsf{E}\sup_{n\ge1}\Big(\sum_{1\le m\le n} a_m\varphi_m\Big)^2 \le A \,. \tag{13.25}$$

Proof. For $1 \le p \le q$ let us define

$$V(p,q) = \mathsf{E}\sup_{p\le n\le q,\varphi}\Big(\sum_{p\le m\le n} a_m\varphi_m\Big)^2 \,, \tag{13.26}$$

where the supremum is also over all orthonormal sequences. Let us assume for contradiction that (13.25) fails i.e. that $\lim_{q\to\infty} V(1,q) = \infty$. Then for each p we have $\lim_{q\to\infty} V(p,q)^2 = \infty$ and therefore we can find an increasing sequence $(p(k))$ such that $V(p(k), p(k+1)) \ge 2$ for each k. By definition of $V(p,q)$ we can then find an orthonormal sequence $(\varphi_{m,k})_{m\ge1}$ for which

$$W(k) := \max_{p(k)\le n\le p(k+1)}\Big|\sum_{p(k)\le m\le n} a_m\varphi_{m,k}\Big|$$

satisfies $\mathsf{E}W(k)^2 \ge 1$. Let us define the function

$$\theta_k = \frac{W(k)^2}{\mathsf{E}W(k)^2} ,$$

so that $\mathsf{E}\theta_k = 1$ and θ_k is a density. We observe that for $p(k) < m \le p(k+1)$ we have $|a_m \varphi_{m,k}| \le 2W(k)$, since $a_m \varphi_{m,k} = \sum_{p(k)\le s\le m} a_s \varphi_{s,k} - \sum_{p(k)\le s\le m-1} a_s \varphi_{s,k}$. Consequently $\varphi_{m,k} = 0$ when $\theta_k = 0$. For $p(k) < m \le p(k+1)$ we may then define $\varphi'_m = \varphi_{m,k}/\sqrt{\theta_k}$. Since

$$\int \varphi'_m \varphi'_{m'} \theta_k \mathrm{dP} = \int \varphi_m \varphi_{m'} \mathrm{dP} ,$$

the functions $(\varphi'_n)_{p(k)\le n<p(k+1)}$ still form an orthonormal sequence after a suitable change of density, and satisfy

$$\max_{p(k)\le n\le p(k+1)} \Big| \sum_{p(k)\le m\le n} a_m \varphi'_m \Big| = \frac{W(k)}{\sqrt{\theta_k}} = (\mathsf{E}W(k)^2)^{1/2} \ge 1 . \tag{13.27}$$

We can moreover assume that $\mathsf{E}\varphi'_m = 0$. To see this we replace the sequence $(\varphi'_m)_{p(k)\le m<p(k+1)}$ by the sequence $(\varepsilon\varphi'_m)_{p(k)\le m<p(k+1)}$ where ε is a Bernoulli r.v. independent of all the r.v.s φ'_m.

We consider now a sequence $(\psi_m)_{m\ge 1}$ with the following properties. For each $k \ge 1$, the sequence $(\psi_m)_{p(k)\le m<p(k+1)}$ is a copy of the sequence $(\varphi'_m)_{p(k)\le m<p(k+1)}$. Moreover these sequences are globally independent as k varies. Since $\mathsf{E}\varphi'_m = 0$ for each m, the sequence $(\psi_m)_{m\ge k(1)}$ is orthonormal. We complete in any way we like in an orthonormal sequence $(\psi_m)_{m\ge 1}$. It follows from (13.27) that the series $\sum_{m\ge 1} a_m \psi_m$ diverges a.s. This contradiction shows that (13.25) holds and concludes the proof. □

Corollary 13.1.12. *Under the hypothesis of Lemma 13.1.11 for each orthonormal process $(X_t)_{t\in T}$ on has*

$$\mathsf{E} \sup_{s,t\in T_k} |X_s - X_t| \le 2\sqrt{A} . \tag{13.28}$$

Proof. We set $t_p = \sum_{1\le m\le p} a_m^2$, so that $T_k = \{t_1, t_2, \ldots, t_k\}$. It follows from Lemma 13.1.10 that the sequence $(\varphi_m)_{2\le m\le k}$ given by

$$\varphi_m = a_m^{-1}(X_{t_m} - X_{t_{m-1}})$$

is an orthonormal sequence. It can be extended in an orthonormal sequence $(\varphi_m)_{m\ge 1}$. If $p \le q$ then $X_{t_q} - X_{t_p} = \sum_{p<m\le q} a_m \varphi_m$, so that

$$\sup_{s,t\in T_k} |X_s - X_t| \le \sup_{p,q} \Big| \sum_{p<m\le q} a_m \varphi_m \Big| \le 2 \sup_n \Big| \sum_{1\le m\le n} a_m \varphi_m \Big|$$

and (13.25) implies (13.28). □

Assuming Theorem 13.1.9 we are now ready to prove the "hard part" of Theorem 13.1.1.

Theorem 13.1.13. *If the series (13.1) converges a.e. for each choice of the orthonormal sequence* (φ_n) *or if (d) holds, then there exists a number* B *such that for each probability measure* μ *on* T *one has*

$$\sum_{n\geq 0}\sum_{I\in\mathcal{I}_n}\sqrt{2^{-n}\mu(I)}\leq B\,. \tag{13.29}$$

Proof. Assuming first that the series (13.1) converges a.e. for each choice of the orthonormal sequence (φ_n), it follows from (13.28) that the quantity $F^*(T)$ of (13.21) satisfies $F^*(T)\leq 2\sqrt{A}$.

Consequently Theorem 13.1.9 implies that for each probability measure μ on T_k one has

$$\sum_{n\geq 0}\sum_{I\in\mathcal{I}_n}\sqrt{2^{-n}\mu(I)}\leq B:=L(1+\sqrt{A})\,.$$

It then should be obvious that this implies the same inequality for each probability measure μ on T.

Assuming now that (d) holds, one has $F^*(T)\leq B'$ and the proof is the same. □

We shall also use the following, which is a special case of an important result of Fernique, of which we will prove the general form in Lemma 13.5.10 below.

Lemma 13.1.14. *Consider a finite set* $T\subset[0,1]$, *and assume that for each probability measure* μ *on* T *one has*

$$\sum_{n\geq 0}\sum_{I\in\mathcal{I}_n}\sqrt{2^{-n}\mu(I)}\leq B\,. \tag{13.30}$$

Then there is a probability measure μ *on* T *for which*

$$\sup_{t\in T}\sum_{n\geq 0}\frac{1}{\sqrt{2^n\mu(I_n(t))}}\leq B\,. \tag{13.31}$$

Proof. Let us denote by $\mathcal{M}(T)$ the set of probability measures on T, and for $\mu\in\mathcal{M}(T)$ let us consider the function

$$f_\mu(t):=\sum_{n\geq 0}\frac{1}{\sqrt{2^n\mu(I_n(t))}}\,.$$

Since the function $x\mapsto 1/\sqrt{x}$ is convex, the map $\mu\mapsto f_\mu$ is convex. Consequently the class $\mathcal{C}$ of functions f on T that satisfy

$$\exists \mu \in \mathcal{M}(T) \; ; \; \forall t \in T \; , \; f_\mu(t) \le f(t)$$

is convex. For each probability measure ν on T, there exists f in $\mathcal{C}$ with $\int f \mathrm{d}\nu \le B$. This is because this is true for $f = f_\nu$ by (13.30). Consequently by the Hahn-Banach theorem, there exists $f \in \mathcal{C}$ such that $f \le B$, which is the content of the lemma. □

Proposition 13.1.15. *If condition (c) of Theorem 13.1.1 holds, so does condition (b).*

Proof. Consider the subset T_k of T that consists of the k smallest points of T. Combining (c) with Lemma 13.1.14 we obtain a probability measure μ_k on T_k for which

$$\sup_{t \in T_k} \sum_{n \ge 0} \frac{1}{\sqrt{2^n \mu_k(I_n(t))}} \le LB \; . \tag{13.32}$$

From here the proof is basically a compactness argument. Let us consider $t^* = \sum_{m \ge 1} a_m^2$, so that $T^* = T \cup \{t^*\}$ is compact. Taking a subsequence if necessary, we may assume that the sequence (μ_k) converges weakly as $k \to \infty$ to a probability measure μ' on T^*. Since the sets $I \cap T^*$ for $I \in \mathcal{I}_n$ are compact, it follows from (13.32) that

$$\sup_{t \in T} \sum_{n \ge 0} \frac{1}{\sqrt{2^n \mu'(I_n(t))}} \le LB \; . \tag{13.33}$$

It might happen that $\mu'(\{t^*\}) > 0$, and then μ' is not supported by T. We modify μ' to take care of this problem. Consider a sequence $t^p \in T$ with $t^p \in I_p$, where I_p is the unique interval of $\mathcal{I}_p$ that contains t^* (here we use the fact that the interval of $\mathcal{I}_p$ are of the type $]u, v]$, so that an interval of this type that contains t^* must meet T). Consider a probability measure μ on T such that for each $C \subset T$ we have

$$\mu(C) \ge \frac{1}{2}\mu'(C \cap T) + \sum \{2^{-p-1} \; ; \; p \ge 1 \; , \; t^p \in C\} \; .$$

Then, for $I \in \mathcal{I}_n$ we have $\mu(I) \ge \mu'(I)/2$ if $t^* \notin I$, while if $t^* \in I$ then $\mu(I) \ge 2^{-n-1}$. It is then immediate to check that μ satisfies (13.3). □

13.2 Chaining, I

To complete the proof of Theorem 13.1.1 (still assuming Theorem 13.1.9) we would like to control the supremum of a stochastic process under condition (13.5). Rather than controlling the increments of the process "exponentially well" as in the case e.g. of Gaussian processes, we only have a "second moment control". As we shall see this creates significant differences. In this section we develop a chaining scheme adapted to this case, in a considerably more general setting than that required for the purpose of Theorem 13.1.1. We consider a finite metric space (T,d) and we try to bound a process $(X_t)_{t\in T}$ which satisfies

$$\forall\, s,t \in T\ ,\ \mathsf{E}(X_s - X_t)^2 \le d(s,t)^2\ . \tag{13.34}$$

Consider a sequence $(T_n)_{n\ge 0}$ of subsets of T. We assume that $\operatorname{card} T_0 = 1$, and we denote by t_0 the unique element of T_0. We assume that for each $n \ge 1$ we are given a map $\theta_n : T_n \to T_{n-1}$. Since we assume that T is finite, it is no much of a restriction to assume that $T_m = T$ for a certain (large) integer m. We define $\pi_m(t) = t$ for each t and recursively $\pi_{n-1}(t) = \theta_n(\pi_n(t))$. First, as usual, we write

$$|X_t - X_{t_0}| \le \sum_{1\le n\le m} |X_{\pi_n(t)} - X_{\pi_{n-1}(t)}|\ . \tag{13.35}$$

Using the inequality $xy \le x^2+y^2$ it is rather natural to write that, for $s \in T_n$, and introducing a parameter $c_n(s)$

$$|X_s - X_{\theta_n(s)}| \le \frac{d(s,\theta_n(s))}{c_n(s)} + d(s,\theta_n(s))c_n(s)\Big(\frac{X_s - X_{\theta_n(s)}}{d(s,\theta_n(s))}\Big)^2\ .$$

Let us assume for simplicity that for numbers $\epsilon_n > 0$ we have

$$\forall\, s \in T_n\ ,\ d(s,\theta_n(s)) \le \epsilon_n\ . \tag{13.36}$$

Then

$$|X_s - X_{\theta_n(s)}| \le \frac{\epsilon_n}{c_n(s)} + \epsilon_n c_n(s)\Big(\frac{X_s - X_{\theta_n(s)}}{d(s,\theta_n(s))}\Big)^2\ .$$

Using this for $s = \pi_n(t)$, and recalling that $\pi_{n-1}(t) = \theta_n(\pi_n(t))$, we obtain

$$|X_{\pi_{n-1}(t)} - X_{\pi_n(t)}| \le \frac{\varepsilon_n}{c_n(\pi_n(t))} + \sum_{s\in T_n} \epsilon_n c_n(s)\Big(\frac{X_s - X_{\theta_n(s)}}{d(s,\theta_n(s))}\Big)^2\ .$$

We then deduce from (13.35)

$$|X_t - X_{t_0}| \le \sum_{1\le n\le m} \frac{\epsilon_n}{c_n(\pi_n(t))} + \sum_{1\le n\le m}\sum_{s\in T_n} \epsilon_n c_n(s)\Big(\frac{X_s - X_{\theta_n(s)}}{d(s,\theta_n(s))}\Big)^2 . \tag{13.37}$$

Let us now set

$$S = \sup_{t\in T} \sum_{1\le n\le m} \frac{\epsilon_n}{c_n(\pi_n(t))}\ , \tag{13.38}$$

$$S^* = \sum_{1\le n\le m}\sum_{s\in T_n} \epsilon_n c_n(s)\ . \tag{13.39}$$

Then (13.37) yields

$$\sup_{t\in T}|X_t - X_{t_0}| \le S + \sum_{1\le n\le m}\sum_{s\in T_n}\epsilon_n c_n(s)\left(\frac{X_s - X_{\theta_n(s)}}{d(s,\theta_n(s))}\right)^2 . \tag{13.40}$$

Taking expectation and using (13.34) we obtain the following important relation.

Lemma 13.2.1. *Recalling (13.38) and (13.39) we have*

$$\mathsf{E}\sup_{t\in T}|X_t - X_{t_0}| \le S + S^* . \tag{13.41}$$

Corollary 13.2.2. *We have*

$$\mathsf{E}\sup_{t\in T}|X_t - X_{t_0}| \le L\sum_{n\ge 1}\epsilon_n\sqrt{\operatorname{card} T_n} . \tag{13.42}$$

Proof. Choose $c_n(t) = 1/\sqrt{\operatorname{card} T_n}$ for $t \in T_n$. □

Exercise 13.2.3. We recall that $N(T,d,\epsilon)$ denotes the smallest number of balls of radius $\le \epsilon$ needed to cover T. Deduce from (13.42) that if the process $(X_t)_{t\in T}$ satisfies (13.34) then

$$\mathsf{E}\sup_{s,t\in T}|X_s - X_t| \le L\int_0^{\Delta(T,d)}\sqrt{\operatorname{card} N(T,d,\epsilon)}\mathrm{d}\epsilon .$$

We recall the notation $I_n(t)$ of Theorem 13.1.1.

Corollary 13.2.4. *Consider a countable subset T of $[0,1]$. Assume that for a certain integer $n_0 \ge 0$ and a certain $I_0 \in \mathcal{I}_{n_0}$ we have $T \subset I_0$. Consider a probability measure μ on T such that*

$$A := \sup_{t\in T}\sum_{n\ge n_0}\frac{1}{\sqrt{2^n\mu(I_n(t))}} < \infty .$$

Then for each process $(X_t)_{t\in T}$ that satisfies (13.5) we have

$$\mathsf{E}\sup_{s,t\in T}|X_s - X_t| \le LA. \tag{13.43}$$

Proof. Since the process satisfies (13.5) it satisfies (13.34) for $d(s,t) = \sqrt{|s-t|}$. The plan is to use (13.41) and we construct the relevant chaining. We construct inductively for $n \ge n_0$ a set $T_n \subset T$ such that $\operatorname{card} T_n \cap I = 1$ whenever $I \in \mathcal{I}_n$ and $I \cap T \ne \emptyset$, and such that moreover $T_{n-1} \subset T_n$. When s is the unique point of $T_n \cap I$, let us then set

$$c_n(s) = \sqrt{\mu(I)} .$$

Let us moreover define the map $\theta_n : T_n \to T_{n-1}$ in the canonical manner. That is, if s is the unique point of $T_n \cap I$ where $I \in \mathcal{I}_n$, there is a unique

$I' \in \mathcal{I}_{n-1}$ with $I \subset I'$, and a unique point s' in $T_{n-1} \cap I'$. We then set $\theta_n(s) = s'$. We have $|s - \theta_n(s)| = |s - s'| \le 2^{-(n-1)}$, so that $d(s, \theta_n(s)) = \sqrt{|s, \theta_n(s)|} \le \epsilon_n := 2^{-(n-1)/2}$, i.e. (13.36) holds for this value of ϵ_n.

Considering an arbitrary integer m, we now use the bound (13.40) for T_m rather than T. Then, for $t \in T_m$, we have

$$\sum_{1 \le n \le m} \frac{\epsilon_n}{c_n(\pi_n(t))} = \sum_{1 \le n \le m} \frac{2^{-(n-1)/2}}{\sqrt{\mu(I_n(t))}} \le 2A ,$$

so that $S \le 2A$. Also, integrating the inequality

$$\forall t \in T , \sum_{n \ge n_0} \frac{1}{\sqrt{2^n \mu(I_n(t))}} \le A$$

with respect to μ we obtain

$$\sum_{n \ge n_0} \sum_{I \in \mathcal{I}_n} 2^{-n/2} \sqrt{\mu(I)} \le A .$$

This means that $S^* \le LA$. Consequently the bound (13.41) implies

$$\mathsf{E} \sup_{s,t \in T_m} |X_s - X_t| \le LA ,$$

and since m is arbitrary, this proves (13.43). □

Proof that (b) implies (e). Let us consider the point $t^* = \sum_{m \ge 1} a_m^2 = \lim_{k \to \infty} t_k$, the supremum of T. Let us consider an integer n_0 and the unique $I_0 \in \mathcal{I}_{n_0}$ with $t^* \in I_0$. Consider the set $T' = T \cap I_0$, so that $t_k \in T'$ for k large enough. Then

$$\sup_{t \in T'} \sum_{n \ge n_0} \frac{1}{\sqrt{\mu(I_n(t))}} \le A^* := \sup_{t \in T} \sum_{n \ge 0} \frac{1}{\sqrt{\mu(I_n(t))}} .$$

Consequently, the probability measure μ' on T' given for $B \subset T'$ by $\mu'(B) = \mu(B \cap T')/\mu(T') = \mu(B \cap I_0)/\mu(I_0)$ satisfies

$$\sup_{t \in T'} \sum_{n \ge n_0} \frac{1}{\sqrt{\mu'(I_n(t))}} \le A^* \sqrt{\mu(I_0)} .$$

The bound (13.43) used for T' and μ' then implies that for each process $(X_t)_{t \in T}$ which satisfies (13.5) satisfies

$$\mathsf{E} \sup_{s,t \in T'} |X_s - X_t| \le LA^* \sqrt{\mu(I_0)} .$$

Now for n_0 large enough $\mu(I_0)$ is arbitrarily small since $\cap_{\epsilon > 0}(T \cap]t^* - \epsilon, t^*]) = \emptyset$. Consequently,

$$\lim_{n \to \infty} \mathsf{E} \sup_{k, \ell \ge n} |X_{t_k} - X_{t_\ell}| = 0 .$$

This concludes the proof. □

13.3 Proof of Bednorz's Theorem

The next paragraph attempts to take some mystery out of the following proofs. In case the reader finds it confusing rather than helpful, she should just try to read the proofs themselves.

One might think that it is very difficult to construct processes that satisfy as precise a constraint as being orthonormal, but this is not quite true. Consider the probability space $[0,1]$ provided with Lebesgue's measure. A canonical orthonormal process is given by $B_t = \mathbf{1}_{[0,t]} \in L^2([0,1])$ (forgetting for the moment the requirement that $\mathsf{E}B_t = 0$). Consider an orthonormal basis $(\psi_n)_{n\geq 1}$ (the Haar basis is particularly appropriate), and a subset S of $\mathbb{N}$. Let us denote by $\langle\cdot,\cdot\rangle$ the canonical duality of L^2 with itself. A process which satisfies

$$\forall\, s,t \in [0,1]\ ,\ \mathsf{E}(X_s - X_t)^2 = \sum_{n\in S} \langle \psi_n, B_t - B_s\rangle^2\ , \tag{13.44}$$

is just as good for our purposes as on orthonormal process, because one can very easily find a process (Y_t) for which

$$\forall\, s,t \in [0,1]\ ,\ \mathsf{E}(Y_s - Y_t)^2 = \sum_{n\notin S} \langle \psi_n, B_t - B_s\rangle^2\ .$$

Assuming without loss of generality that the process (Y_t) is independent of the process X_t, the process $(X_t + Y_t)_{t\in[0,1]}$ is then orthonormal. Moreover the operation of adding (Y_t) behaves well with respect to supremum. In words, one "completes" a process as in (13.44) to make it orthonormal. A special case of this procedure occurs in Lemma 13.3.2 below. It is possible to formulate the constructions below as the constructions of processes satisfying a condition of the type (13.44), and to "complete them into orthonormal processes" at the end of the construction. We have chosen another route, which is to "complete these processes along the way". This is at first a bit more mysterious, but allows for simpler notation.

The main step in the proof of Bednorz's theorem is, given a finite subset T of $]0,1]$, to relate the "size" of T with the size of the four sets $T_j = T \cap I_j$ where for $1 \leq j \leq 4$, I_j is the interval $](j-1)/4, j/4]$. The reason why we use 4-adic partitions is that we are certain that "T_1 is far apart from T_3"(etc.). On the other hand we cannot say the same about, say, T_1 and T_2 since T_1 might be located to the very right of I_1 and T_2 might be located to the very left of I_2. (This is why dyadic partitions would not work.)

Definition 13.3.1. *Consider an interval* $J =]c,d] \subset [0,1]$*, and* $\overline{J} = [c,d]$*. We say that the process* $(X_t)_{t\in\overline{J}}$ *is* normalized *if* $\mathsf{E}X_t = 0$*,* $X_c = X_d = 0$ *and*

$$\forall\, s,t \in \overline{J}\ ,\ s < t\ ,\ \mathsf{E}(X_s - X_t)^2 = t - s - (d-c)^{-1}(t-s)^2\ . \tag{13.45}$$

The reason behind the formula in the right-hand side of (13.45) will be explained soon. We fix the finite set $T \subset]0,1]$ once and for all. For an interval $J =]c,d] \subset [0,1]$ we consider the quantity

$$F(J) = \sup \mathsf{E} \sup_{t \in T \cap J} X_t \ , \tag{13.46}$$

where the first supremum is taken over all normalized processes indexed by $\overline{J} = [c,d]$. Although $X_c = 0$ is defined, in (13.46) the supremum is only over $T \cap J$, not over $T \cap \overline{J}$. We define $F(J) = 0$ when $T \cap J = \emptyset$. We recall the quantity $F^*(T)$ of (13.21).

Lemma 13.3.2. *We have* $F(]0,1]) \le F^*(T)$.

Proof. Consider a normalized process $(X_t)_{t \in [0,1]}$. Consider a centered r.v. Z, independent of this process, and such that $\mathsf{E}Z^2 = 1$. Then the process $Y_t = X_t + tZ$ is orthonormal. Using the definition of F^* in the first inequality and using Jensen's inequality (taking the expectation in Z inside the supremum) in the second inequality yields

$$F^*(T) \ge \mathsf{E} \sup_{t \in T} Y_t \ge \mathsf{E} \sup_{t \in T} X_t \ ,$$

and since the normalized process (X_t) is arbitrary this proves that $F(]0,1]) \le F^*(T)$. $\square$

We state now the main step in the proof of Bednorz's theorem.

Proposition 13.3.3. *Consider* $I =]0,1]$ *and for* $j = 1,2,3,4$ *consider* $I_j =](j-1)/4, j/4]$ *and numbers* $\alpha_j \ge 0$ *such that* $\sum_{j \le 4} \alpha_j = 1$. *Then*

$$F(I) \ge \sum_{1 \le j \le 4} \sqrt{\alpha_j} F(I_j) \ . \tag{13.47}$$

Moreover if for each $1 \le j \le 4$ *we have* $\alpha_j \ge 1/400$ *and* $T \cap I_j \ne \emptyset$ *then*

$$F(I) \ge \sum_{1 \le j \le 4} \sqrt{\alpha_j} F(I_j) + \frac{1}{80} \ . \tag{13.48}$$

The first task is to understand how we relate normalized processes on the intervals I_j with a normalized process on I. Consider the probability space $[0,1]$ provided with Lebesgue's measure. The archetypical example of normalized process is given by the formula

$$W_t = \mathbf{1}_{[0,t]} - t \ . \tag{13.49}$$

Consider now the algebra $\mathcal{S}$ of subsets of $[0,1]$ generated by the intervals I_j for $1 \le j \le 4$, and denote by $\mathsf{E}_{\mathcal{S}}$ conditional expectation with respect to this algebra. We define

$$V'_t = W_t - \mathsf{E}_{\mathcal{S}} W_t \ ; \ V_t = \mathsf{E}_{\mathcal{S}} W_t \ , \tag{13.50}$$

so that $W_t = V'_t + V_t$ and V_t is $\mathcal{S}$-measurable, while $\mathsf{E}_{\mathcal{S}}(V'_t) = 0$. We observe that if $\mathsf{E}_{\mathcal{S}} f = 0$ and f' is $\mathcal{S}$-measurable, then $\mathsf{E}(f+f')^2 = \mathsf{E}f^2 + \mathsf{E}f'^2$. Thus, if $s \le t$,

$$t - s - (t-s)^2 = \mathsf{E}(W_s - W_t)^2 = \mathsf{E}(V'_s - V'_t)^2 + \mathsf{E}(V_s - V_t)^2 \ . \tag{13.51}$$

Let us consider the function

$$\varphi(x) = x - 4x^2 \ , \tag{13.52}$$

and for $t \in I$ and $1 \le j \le 4$ let us define

$$t^j = \max(\min(t, j/4), (j-1)/4) \in \overline{I}_j = [(j-1)/4, j/4] \ .$$

We observe that for $0 \le s \le t \le 1$ the interval $]s,t]$ is the disjoint union of the intervals $]s^j, t^j]$ for $1 \le j \le 4$. In particular $t - s = \sum_{j \le 4} t^j - s^j$.

We claim that

$$\mathsf{E}(V'_s - V'_t)^2 = \sum_{1 \le j \le 4} \varphi(t^j - s^j) \ . \tag{13.53}$$

This is because, assuming without loss of generality that $s \le t$,

$$\mathsf{E}(V'_s - V'_t)^2 = \mathsf{E}(\mathbf{1}_{[s,t]} - \mathsf{E}_{\mathcal{S}}\mathbf{1}_{[s,t]})^2 = \mathsf{E}\mathbf{1}_{[s,t]} - \mathsf{E}(\mathsf{E}_{\mathcal{S}}\mathbf{1}_{[s,t]})^2 \ ,$$

while $t - s = \sum_{j \le 4}(t^j - s^j)$ and $\mathsf{E}_{\mathcal{S}}\mathbf{1}_{[s,t]} = 4\sum_{j \le 4}(t^j - s^j)\mathbf{1}_{I_j}$, so that $\mathsf{E}(\mathsf{E}_{\mathcal{S}}\mathbf{1}_{[s,t]})^2 = 4\sum_{j \le 4}(t^j - s^j)^2$. This proves the claim. We then conclude from (13.51) that

$$t - s - (t-s)^2 = \sum_{1 \le j \le 4} \varphi(s^j - t^j) + \mathsf{E}(V_s - V_t)^2 \ . \tag{13.54}$$

In due time we shall use the following elementary property of the process $(V_t)_{t \in I}$, which is a consequence of the definition $V_t = \mathsf{E}_{\mathcal{S}}(\mathbf{1}_{[0,t]} - t)$.

Lemma 13.3.4. *We have*

$$\inf_{t \in I_1} \mathsf{E}(V_t \mathbf{1}_{I_4}) \ge -\frac{1}{16} \ ; \ \inf_{t \in I_2} \mathsf{E}(V_t \mathbf{1}_{I_1}) \ge \frac{1}{8}$$
$$\inf_{t \in I_3} \mathsf{E}(V_t \mathbf{1}_{I_2}) \ge \frac{1}{16} \ ; \ \inf_{t \in I_4} \mathsf{E}(V_t \mathbf{1}_{I_3}) \ge 0 \ . \tag{13.55}$$

Proof. We simply observe that for $t \in I_1$ and $x \in I_4$ we have $V_t(x) = -t \ge -1/4$, for $x \in I_1$ and $t \in I_2$ we have $V_t(x) = 1 - t \ge 1/2$ etc. □

To prove Proposition 13.3.3 we consider normalized processes $(Y_t^j)_{t\in \overline{I}_j}$ for $1 \le j \le 4$. (We remind the reader that in particular $\mathsf{E}Y_t^j = 0$.) From these we shall construct a suitable normalized process on I. The construction will involve an auxiliary process $(Z_t)_{t\in T}$ and a r.v $\tau \in \{1,2,3,4\}$. Throughout the proof we assume the following:

$$\text{The processes } Y_t^j \text{ are independent of each other and of the r.v.s } Z_t \text{ and } \tau \;, \tag{13.56}$$

$$\forall\, j \le 4\;;\; \mathsf{P}(\tau = j) = \alpha_j \;, \tag{13.57}$$

and

$$\mathsf{E}Z_t = 0\;;\; \mathsf{E}(Z_s - Z_t)^2 = \mathsf{E}(V_s - V_t)^2 \;. \tag{13.58}$$

We do *not* assume that Z_t and τ are independent. When $\alpha_j = 0$, for $t \in I$ let us define $U_t^j = Y_{t^j}^j$. Otherwise we define

$$U_t^j = \frac{1}{\sqrt{\alpha_j}}\mathbf{1}_{\{\tau=j\}} Y_{t^j}^j \;, \tag{13.59}$$

and we observe that, using the independence of τ and Y in the second equality, and that the process Y^j is normalized in the third one,

$$\mathsf{E}(U_s^j - U_t^j)^2 = \frac{1}{\alpha_j}\mathsf{E}\mathbf{1}_{\{\tau=j\}}(Y_{t^j}^j - Y_{s^j}^j)^2 = \mathsf{E}(Y_{t^j}^j - Y_{s^j}^j)^2 = \varphi(t^j - s^j) \;. \tag{13.60}$$

This formula remains true even when $\alpha_j = 0$, since then $U_t^j = Y_{t^j}^j$. We define

$$S_t = \sum_{1\le j\le 4} U_t^j \;. \tag{13.61}$$

It follows from (13.56) and the fact that Y^j and $Y^{j'}$ are independent for $j \ne j'$ that then $\mathsf{E}U_s^j U_t^{j'} = 0$, so that

$$\mathsf{E}(S_s - S_t)^2 = \sum_{1\le j\le 4}\mathsf{E}(U_s^j - U_t^j)^2 = \sum_{1\le j\le 4}\varphi(t^j - s^j) \;. \tag{13.62}$$

The process (S_t) is the important part of this construction. We then transform this process into a normalized process by adding (Z_t).

Lemma 13.3.5. *The process*

$$X_t = S_t + Z_t \tag{13.63}$$

is normalized.

Proof. It follows from (13.56) that $\mathsf{E}S_s Z_t = 0$, so that

$$\mathsf{E}(X_s - X_t)^2 = \mathsf{E}(S_s - S_t)^2 + \mathsf{E}(Z_s - Z_t)^2 \;,$$

and the result follows from (13.54), (13.58) and (13.62). □

Lemma 13.3.6. *Assume that $T \cap I_j \neq \emptyset$ for each $j \leq 4$. Then*

$$\mathsf{E}\sup_{t\in T} X_t \geq \sum_{1\leq j\leq 4}\Big(\sqrt{\alpha_j}\,\mathsf{E}\sup_{t\in T\cap I_j} Y_t^j + \inf_{t\in T\cap I_j} \mathsf{E}\mathbf{1}_{\{\tau=j\}} Z_t\Big)\,. \tag{13.64}$$

Proof. First, we observe that, using that $T \cap I_j \neq \emptyset$ in the last inequality,

$$\begin{aligned}\mathsf{E}\sup_{t\in T} X_t &= \sum_{1\leq j\leq 4} \mathsf{E}\mathbf{1}_{\{\tau=j\}}\sup_{t\in T} X_t \\ &= \sum_{1\leq j\leq 4} \mathsf{E}\sup_{t\in T}\mathbf{1}_{\{\tau=j\}} X_t \\ &\geq \sum_{1\leq j\leq 4} \mathsf{E}\sup_{t\in T\cap I_j}\mathbf{1}_{\{\tau=j\}} X_t\,. \end{aligned} \tag{13.65}$$

Let us fix $j \leq 4$ with $\alpha_j \neq 0$ and denote by E^j conditional expectation given the r.v.s Y_t^j. Then Jensen's inequality implies

$$\mathsf{E}\sup_{t\in T\cap I_j}\mathbf{1}_{\{\tau=j\}} X_t \geq \mathsf{E}\sup_{t\in T\cap I_j}\mathsf{E}^j\mathbf{1}_{\{\tau=j\}} X_t\,. \tag{13.66}$$

Let us fix $t \in I_j$, so that then $t^j = t$. Since $\mathbf{1}_{\{\tau=j\}}\mathbf{1}_{\{\tau=j'\}} = 0$ for $j' \neq j$ we have by definition of X_t

$$\mathbf{1}_{\{\tau=j\}} X_t = \frac{1}{\sqrt{\alpha_j}}\mathbf{1}_{\{\tau=j\}} Y_t^j + \mathbf{1}_{\{\tau=j\}} Z_t\,.$$

Using the independence of τ and Y_j in the first equality, and the independence of Y_t^j from τ and Z in the second equality, we get

$$\mathsf{E}^j\mathbf{1}_{\{\tau=j\}} X_t = \sqrt{\alpha_j} Y_t^j + \mathsf{E}^j\mathbf{1}_{\{\tau=j\}} Z_t = \sqrt{\alpha_j} Y_t^j + \mathsf{E}\mathbf{1}_{\{\tau=j\}} Z_t\,. \tag{13.67}$$

To conclude we simply use that $\sup_t(y_t + z_t) \geq \sup_t y_t + \inf_t z_t$, and thus

$$\mathsf{E}\sup_{t\in T\cap I_j}\mathsf{E}^j\mathbf{1}_{\{\tau=j\}} X_t \geq \sqrt{\alpha_j}\,\mathsf{E}\sup_{t\in T\cap I_j} Y_t^j + \inf_{t\in T\cap I_j}\mathsf{E}\mathbf{1}_{\{\tau=j\}} Z_t\,. \qquad \square$$

Lemma 13.3.7. *Even when $T \cap I_j = \emptyset$ for some $j \leq 4$, if the process (Z_t) is independent of τ we have*

$$\mathsf{E}\sup_{t\in T} X_t \geq \sum_{j\in J}\sqrt{\alpha_j}\,\mathsf{E}\sup_{t\in T\cap I_j} Y_t^j\,, \tag{13.68}$$

where $J = \{j \leq 4; I_j \neq \emptyset\}$.

Proof. Since $\mathsf{E}\sup_{t\in T}\mathbf{1}_{\{\tau=j\}} X_t \geq 0$ because $\mathsf{E}\mathbf{1}_{\{\tau=j\}} X_t = 0$ for each t, as in (13.65) we obtain

$$\mathsf{E}\sup_{t\in T} X_t \geq \sum_{j\in J}\mathsf{E}\sup_{t\in T\cap I_j}\mathbf{1}_{\{\tau=j\}} X_t\,.$$

Since τ and Z_t are independent, and since $\mathsf{E}Z_t = 0$, then (13.67) implies $\mathsf{E}^j\mathbf{1}_{\{\tau=j\}} X_t = \sqrt{\alpha_j} Y_t^j$, and the conclusion from (13.66). $\square$

Proof of Proposition 13.3.3. To prove (13.47) we simply choose Z_t independent of τ and we use (13.68) since by definition $F(I) = 0$ when $I \cap T = \emptyset$. It remains only to prove (13.48). We shall use (13.64) with an appropriate choice of the process (Z_t). To simplify the notation we assume without loss of generality that the underlying probability space is $[0,1]$ provided with Lebesgue's measure, and that for $j \le 4$,

$$]0,1/100] \cap \{\tau = j\} =](n(j)-1)/400, n(j)/400] \;, \tag{13.69}$$

where $n(1) = 4, n(2) = 1, n(3) = 2, n(4) = 3$. This is possible since we assume $P(\tau = j) = \alpha_j \ge 1/400$ for each j. Let us then define $Z_t(x) \equiv 0$ for $x > 1/100$ and for $x \le 1/100$ let us define $Z_t(x) = 10 V_t(100x)$, where V_t is defined in (13.50).

The fundamental relation is, recalling that $I_j =](j-1)/4, j/4]$,

$$\mathsf{E}\mathbf{1}_{\{\tau=j\}} Z_t = \frac{1}{10} \mathsf{E}\mathbf{1}_{I_{n(j)}} V_t \;.$$

The proof is straight forward by change of variable:

$$\begin{aligned} \mathsf{E}\mathbf{1}_{\{\tau=j\}} Z_t &= 10 \int_{\{t=j\}\cap]0,1/100]} V_t(100x)\mathrm{d}x = 10 \int_{I_{n(j)}/100} V_t(100x)\mathrm{d}x \\ &= \frac{1}{10} \int_{I_{n(j)}} V_t(x)\mathrm{d}x = \frac{1}{10} \mathsf{E}\mathbf{1}_{I_{n(j)}} V_t \;. \end{aligned} \tag{13.70}$$

It then follows from Lemma 13.3.4 that

$$\sum_{1\le j\le 4} \inf_{t\in I_j} \mathsf{E}\mathbf{1}_{\{\tau=j\}} Z_t = \frac{1}{10} \sum_{1\le j\le 4} \inf_{t\in I_j} \mathsf{E}\mathbf{1}_{I_{n(j)}} V_t \ge \frac{1}{10}\Big(-\frac{1}{16} + \frac{1}{8} + \frac{1}{16}\Big) = \frac{1}{80} \;,$$

and combining with (13.64) this completes the proof, since it is also obvious (using again change of variable) that (13.58) holds. □

Corollary 13.3.8. *Consider $I \in \mathcal{I}_m$ and the four intervals I_j of $\mathcal{I}_{m+2}$ for $j = 1,2,3,4$ which it contains. Consider numbers $\alpha_j \ge 0$ such that $\sum_{j\le 4} \alpha_j = 1$. Then*

$$F(I) \ge \sum_{1\le j\le 4} \sqrt{\alpha_j} F(I_j) \;. \tag{13.71}$$

Moreover if for each $1 \le j \le 4$ we have $\alpha_j \ge 1/400$ and $T \cap I_j \ne \emptyset$ then

$$F(I) \ge \sum_{1\le j\le 4} \sqrt{\alpha_j} F(I_j) + 2^{-m/2} \frac{1}{80} \;. \tag{13.72}$$

Proof. A first method is to repeat the proof of Proposition 13.3.3. There is really nothing to change, and we chose to prove Proposition 13.3.3 simply because the notation is simpler. A second method is to deduce Corollary 13.3.8 from Proposition 13.3.3 using a scaling argument. Namely, if we denote by $F_T(J)$ the quantity (13.46) to indicate the dependence in T, it suffices to prove that for $a > 0$ and $b \in \mathbb{R}$ we have, with obvious notation $F_{aT+b}(aJ+b) = \sqrt{a}F_T(J)$. This follows from the fact that if the process $(X_t)_{t \in aJ+b}$ is normalized on the interval $aJ+b$ then the process $Y_t = a^{1/2}X_{(t-b)/a}$ is normalized on the interval J. □

We now turn to the task of combining the estimates (13.48). Our next goal is as follows. We recall that $\mathcal{I}_n$ denotes the family of 2^n dyadic intervals of length 2^{-n}. Theorem 13.1.9 is a consequence of Lemma 13.3.2 and the following.

Proposition 13.3.9. *Consider a finite set $T \subset [0,1]$. Then, given a probability measure μ on T we have*

$$\frac{1}{80}\sum_{n\geq 0}\sum_{I\in\mathcal{I}_n} 2^{-n/2}\sqrt{\mu(I)} \leq F(]0,1]) \,. \tag{13.73}$$

As a preparation for the proof, we fix a probability measure μ on T and for $n \geq 0$ we define $\mathcal{I}^*_{2n}$ as the collection of intervals $I \in \mathcal{I}_{2n}$ that have the following property:

$$I' \in \mathcal{I}_{2n+2}\ ,\ I' \subset I \Rightarrow \mu(I') \geq \mu(I)/400 \,. \tag{13.74}$$

We then define

$$M_n = 2^{-n}\sum\left\{\sqrt{\mu(I)}\ ;\ I \in \mathcal{I}^*_{2n}\right\} .$$

Lemma 13.3.10. *We have*

$$\sum_{n\geq 0} M_n \leq 80F(]0,1]) \,. \tag{13.75}$$

Proof. We define $F(J) = 0$ when $J \cap T = \emptyset$. We prove that for each $n \geq 0$ we have

$$\sum_{I\in\mathcal{I}_{2n}} \sqrt{\mu(I)}F(I) \geq \frac{1}{80}M_n + \sum_{I\in\mathcal{I}_{2n+2}} \sqrt{\mu(I)}F(I) \,. \tag{13.76}$$

Summation of these inequalities over $n \geq 0$ then yields (13.75). To prove (13.76) we observe first that if $I \in \mathcal{I}_{2n}$ then

$$\sqrt{\mu(I)}F(I) \geq \sum_{I'\subset I, I'\in\mathcal{I}_{2n+2}} \sqrt{\mu(I')}F(I') \,. \tag{13.77}$$

This simply follows from (13.71) with $\alpha_j = \mu(I_j)/\mu(I)$, where the elements of $\mathcal{I}_{2n+2}$ contained in I are denoted by I_1, I_2, I_3, I_4. On the other hand, when $I \in \mathcal{I}^*_{2n}$ we can now use (13.72) with $m = 2n$ to obtain, since $2^{-n} = 2^{-m/2}$,

$$\sqrt{\mu(I)}F(I) \geq \sum_{I'\subset I, I'\in\mathcal{I}_{2n+2}} \sqrt{\mu(I')}F(I') + \frac{1}{80}2^{-n}\sqrt{\mu(I)} \,. \tag{13.78}$$

Summation of the inequalities (13.77) and (13.78) over $I \in \mathcal{I}_{2n}$ completes the proof of (13.76). □

The following will complete the proof of Proposition 13.3.9.

Lemma 13.3.11. *We have*

$$\sum_{n\geq 0}\sum_{I\in\mathcal{I}_n} 2^{-n/2}\sqrt{\mu(I)} \leq 2S(\mu) := 2\sum_{n\geq 0}\sum_{I\in\mathcal{I}_{2n}} 2^{-n}\sqrt{\mu(I)}\ , \tag{13.79}$$

and

$$S(\mu) \leq 10 + 10\sum_{n\geq 0} M_n\ . \tag{13.80}$$

Lemma 13.3.12. *Consider numbers $\alpha_j \geq 0$ for $j = 1, 2, 3, 4$ such that $\sum_{1\leq j\leq 4}\alpha_j = 1$. Then*

$$\min_{1\leq j\leq 4}\alpha_j \leq \frac{1}{400} \Rightarrow \frac{1}{2}\sum_{1\leq j\leq 4}\sqrt{\alpha_j} \leq \frac{9}{10}\ . \tag{13.81}$$

Proof. Assume for example that $\alpha_1 \leq 1/400$. Then since $\sqrt{\alpha_2}+\sqrt{\alpha_3}+\sqrt{\alpha_4} \leq \sqrt{3(\alpha_2+\alpha_3+\alpha_4)} \leq \sqrt{3}$,

$$\frac{1}{2}\sum_{1\leq j\leq 4}\sqrt{\alpha_j} \leq \frac{1}{2}\Big(\frac{1}{20}+\sqrt{3}\Big) \leq \frac{9}{10}\ .$$

□

Proof of Lemma 13.3.11. The proof of (13.79) is immediate because the quantity $\sum_{I\in\mathcal{I}_n} 2^{-n/2}\sqrt{\mu(I)}$ decreases with n (by the inequality $\sqrt{a}+\sqrt{b} \leq \sqrt{2}\sqrt{a+b}$). For $I \in \mathcal{I}_{2n}$ let

$$w(I) = 2^{-n-1}\sum\Big\{\sqrt{\mu(J)}\ ;\ J\subset I\ ,\ J\in\mathcal{I}_{2n+2}\Big\}\ .$$

First, we observe that

$$S(\mu) = 1 + \sum_{n\geq 0}\sum_{I\in\mathcal{I}_{2n}} w(I)\ . \tag{13.82}$$

This is simply because on the right all the terms in the summation that defines $S(\mu)$ occur in one of the terms $w(I)$, except the term for $n = 0$, which is equal to 1 since μ is a probability. Let us now observe that, given $n \geq 0$,

$$\sum_{I\in\mathcal{I}_{2n}} w(I) = \sum_{I\in\mathcal{I}^*_{2n}} w(I) + \sum_{I\notin\mathcal{I}^*_{2n}} w(I)\ . \tag{13.83}$$

It follows from Lemma 13.3.11 that for $I \notin \mathcal{I}^*_{2n}$ we have

$$w(I) < \frac{9}{10}2^{-n}\sqrt{\mu(I)}\ . \tag{13.84}$$

Consequently, by summation of the relations (13.84),

$$\sum_{n\geq 0}\sum_{I\notin \mathcal{I}^*_{2n}} w(I) \leq \frac{9}{10}S(\mu)\ .$$

Combining with (13.82) and (13.83) we thus obtain

$$S(\mu) \leq 1+\frac{9}{10}S(\mu)+\sum_{n\geq 0}\sum_{I\in \mathcal{I}^*_{2n}} 2^{-n}\sqrt{\mu(I)} \leq 1+\frac{9}{10}S(\mu)+\sum_{n\geq 0} M_n\ .$$

This completes the proof. □

13.4 Permutations

One may also ask the following question. What are the sequences (a_m) such that for any permutation π and any orthonormal sequence (φ_m) the series $\sum_m a_{\pi(m)}\varphi_m$ converges a.e.? The answer to this question was also discovered by A. Paszkiewicz, and is announced in [7]. Given the sequence (a_m) and the permutation π of $\mathbb{N}$ we define the set

$$T_\pi = \Big\{ \sum_{1\leq m\leq n} a^2_{\pi(m)}\ ;\ n\geq 1\Big\}\ . \tag{13.85}$$

We also consider the numbers

$$b_k := \sum\{a^2_m\ ;\ 2^{-2^{k+1}} < a^2_m \leq 2^{-2^k}\}\ . \tag{13.86}$$

Without loss of generality we assume that $\sum_m a^2_m \leq 1$.

Theorem 13.4.1. *For a sequence (a_m) the following are equivalent.*
(f) For every permutation π and every orthonormal sequence (φ_m) the series $\sum_m a_{\pi(m)}\varphi_m$ converges a.e.
(g) We have

$$\sum_{k\geq 1} 2^k\sqrt{b_k} < \infty\ . \tag{13.87}$$

(h) There exists π such that

$$\sum_{n\geq 0}\sqrt{2^{-n}\,\mathrm{card}\{I\in \mathcal{I}_n\ ;\ I\cap T_\pi \neq \emptyset\}} < \infty\ . \tag{13.88}$$

(i) Condition (13.88) holds for each π.

This should be compared with Corollary 13.1.3. When the series $\sum_m a_m\varphi_m$ converges a.s. whatever the choice of the orthonormal sequence φ_m, Corollary 13.1.3 asserts that $\sum_{k\geq 1} 2^{2k} b_k < \infty$. The stronger hypothesis of Theorem 13.4.1 implies the stronger conclusion (13.87). One should also note that (13.88) is basically a condition on the covering numbers of T_π, where the distance $d(s,t) = \sqrt{|s-t|}$ has been replaced by a slightly larger distance associated to the sequence of partitions $\mathcal{I}_n$.

Theorem 13.4.1 will turn out to be a corollary of Theorem 13.1.1. Moreover, as we explain now, the idea behind it is very clear, which greatly facilitates the discovery of its proof. It is not very difficult to prove the equivalence of (g) to (i), and this is what we will do first. Condition (13.88) is the natural "covering number condition". The basic idea is that the more sophisticated conditions of Theorem 13.1.1 are equivalent to this condition "when the set T is homogeneous". Therefore to prove that (f) implies (h) the idea will be to construct π so that T_π is "as homogeneous as possible".

Let us prove "the easy part" of Theorem 13.4.1.

Proposition 13.4.2. *Conditions (g) to (i) of Theorem 13.4.1 are equivalent.*

Proof. It suffices to prove that conditions (g) and (h) are equivalent when π is the identity, because the value of b_k does not depend on which order we consider the elements a_m.

We first prove that (g) implies (h). For $n \geq 1$ define $\mathcal{J}_n$ as the set of dyadic intervals $I \in \mathcal{I}_n$ for which $T \cap I \neq \emptyset$. We shall prove that

$$\sum_{n \geq 0} \sqrt{2^{-n} \operatorname{card} \mathcal{J}_n} < \infty \,. \tag{13.89}$$

Let us as usual enumerate T as a sequence $t_n = \sum_{1 \leq m \leq n} a_m^2$, and let $t^* = \sum_{m \geq 1} a_m^2$. Define

$$W_k = \{t_n \,;\, \max(a_n^2, a_{n+1}^2) > 2^{-2^k}\} \cup \{t_1, t^*\} \,,$$

$$V_k = \bigcup \{[t_n, t_{n+1}] \,;\, a_{n+1}^2 = t_{n+1} - t_n \leq 2^{-2^k}\} \subset [0,1] \,.$$

Denoting Lebesgue's measure by λ, we deduce from (13.86) that

$$\lambda(V_k) \leq \sum_{r \geq k} b_r \,, \tag{13.90}$$

while, since $\sum_{m \geq 1} a_m^2 \leq 1$,

$$\operatorname{card} W_k \leq 2 + 2 \cdot 2^{2^k} \,. \tag{13.91}$$

Consider $2^{k+1} \leq \ell < 2^{k+2}$ and an interval $I \in \mathcal{J}_\ell$, so that $I \cap T \neq \emptyset$. The basic observation is that one of the following occurs: either

$$I \cap W_k \neq \emptyset \tag{13.92}$$

or else

$$I \subset V_k \,. \tag{13.93}$$

To prove this, we assume $I \cap W_k = \emptyset$ and we prove (13.93). We observe first that if $I \in \mathcal{I}_\ell$ then I meets T by definition, so that either $t_1 \in I$, or $t^* \in I$, or else

$$I \subset \bigcup_{n \geq 1} [t_n, t_{n+1}] \,. \tag{13.94}$$

Since we assume that $I \cap W_k = \emptyset$, we have in particular that $t_1, t^* \notin I$ and thus (13.94) holds. Consider an interval $[t_n, t_{n+1}]$ which meets I. Then either t_n or t_{n+1} belongs to I, for otherwise $I \subset]t_n, t_{n+1}[$, contradicting the fact that $I \cap T \neq \emptyset$. Since $I \cap W_k = \emptyset$ it cannot happen that both t_n and t_{n+1} belong to W_k. Thus $a_{n+1}^2 \leq 2^{-2^k}$, so that $[t_n, t_{n+1}] \subset V_k$ by definition of V_k. Thus every interval $[t_n, t_{n+1}]$ which meets I is a subset of V_k. Then (13.94) proves (13.93).

There are at most $\operatorname{card} W_k$ intervals I which satisfy (13.93). Consequently, and since $\lambda(I) = 2^{-\ell}$ for $I \in \mathcal{I}_\ell$,

$$\operatorname{card} \mathcal{J}_\ell \leq \operatorname{card} W_k + 2^\ell \lambda(V_k) \leq 2 + 2 \cdot 2^{2^k} + 2^\ell \sum_{r \geq k} b_r \,,$$

so that

$$\sqrt{\operatorname{card} \mathcal{J}_\ell} \leq 2 + 2 \cdot 2^{2^{k-1}} + 2^{\ell/2} \sqrt{\sum_{r \geq k} b_r} \,.$$

This holds whenever $2^{k+1} \leq \ell < 2^{k+2}$ and therefore

$$\sum_{2^{k+1} \leq \ell < 2^{k+2}} 2^{-\ell/2} \sqrt{\operatorname{card} \mathcal{J}_\ell} \leq L 2^{-2^k} 2^{2^{k-1}} + L 2^k \sqrt{\sum_{r \geq k} b_r} \,,$$

from which (13.89) follows by summation over $k \geq 1$. This concludes the proof that (g) implies (h).

Now we prove that (h) implies (g). For this we recall (13.17) and the measure μ_k there, and using the Cauchy-Schwarz inequality we obtain that for $2^{k-1} \leq n < 2^k$ we have

$$\frac{1}{2}\sqrt{b_k} \leq \sum_{I \in \mathcal{I}_n} \sqrt{2^{-n} \mu_k(I)} = \sum_{I \in \mathcal{J}_n} \sqrt{2^{-n} \mu_k(I)} \leq \sqrt{2^{-n} \operatorname{card} \mathcal{J}_n} \,,$$

and we sum over n and then over k. □

The next goal is to complete the proof of Theorem 13.4.1. First we prove that (i) implies (f). This is because each set T_π satisfies condition (c) of Theorem 13.1.1, as follows from the Cauchy-Schwarz inequality:

$$\sum_{I \in \mathcal{I}_n} \sqrt{2^{-n} \mu(I)} \leq \sqrt{2^{-n} \operatorname{card}\{I \in \mathcal{I}_n \,;\, I \cap T_\pi \neq \emptyset\}} \,.$$

Now we come to the main argument, the proof that (f) implies (g). The proof is based on a recursive construction. This construction is performed by induction and requires the formulation of an adequate induction hypothesis, to which we turn now.

Given a finite set U let us denote by $\mathcal{T}(U)$ the collection of sets of the type

$$T = \Big\{ \sum_{1 \le m \le n} \rho(m) \; ; \; 1 \le n \le q \Big\}$$

where ρ is a one-to map from $\{1, \dots, q\}$ into U. In other words, to construct $T \in \mathcal{T}(U)$ we pick distinct elements $u_1, u_2, \dots, u_q$ of U and we take $T = \{u_1, u_1 + u_2, \dots, u_1 + u_2 + \cdots + u_q\}$. For such a set T we denote by $t^*(T)$ its largest element. We prove the following.

Proposition 13.4.3. *There exists a universal constant L^*, an integer n_0 and a sequence $(\epsilon_k)_{k \ge n_0}$ of positive numbers such that $\sum_{k \ge n_0} \epsilon_k < \infty$ with the following property. Consider a finite set $J \subset \{n_0, n_0 + 1, \dots, \}$ with the property that*

$$k < k' \in J \Rightarrow k' - k \ge 4 \, . \tag{13.95}$$

Consider a finite set $U \subset \mathbb{R}^+$, and assume that

$$\sum \{u \in U\} \le 1/2 \, . \tag{13.96}$$

Consider a partition of U in sets $(U_k)_{k \in J}$. Assume that

$$u \in U_k \Rightarrow 2^{-2^{k+1}} \le u \le 2^{-2^k} \, , \tag{13.97}$$

and set

$$b_k = \sum \{u \; ; \; u \in U_k\} \, . \tag{13.98}$$

Then there exists a set $T \in \mathcal{T}(U)$ and a probability measure μ on T with the following property. Consider $x \in [0,1]$ with $x + t^*(T) \le 1$. Then

$$\sum_{n \ge 0} \sum_{I \in \mathcal{I}_n, I - x \subset [0, t^*(T)]} \sqrt{2^{-n} \mu(I - x)} \ge \frac{1}{L^*} \sum_{k \in J} (2^k \sqrt{b_k} - \epsilon_k) \, . \tag{13.99}$$

The number 4 in (13.95) is simply a convenient choice whose relevance will became apparent at the very end of the present section.

Let us first complete the proof of Theorem 13.4.1 i.e. prove that (f) implies (g). We argue by contradiction, and we assume that (g) fails, i.e. $\sum_{k \ge 1} 2^k \sqrt{b_k} = \infty$. Consider the set $V = \{a_m^2 ; m \ge 1\}$.

Next we construct by induction finite sets $V_s \subset V$ with $\max V_{s+1} < \min V_s$, sets $T_s \in \mathcal{T}(V_s)$ and probability measures μ_s on T_s with the following property: Consider $x \in [0,1]$ with $x + t^*(T_s) \le 1$. Then

$$\sum_{n \ge 0} \sum_{I \in \mathcal{I}_n, I - x \subset [0, t^*(T)]} \sqrt{2^{-n} \mu_s(I - x)} \ge 2^s \, . \tag{13.100}$$

To perform the construction, having constructed V_s, consider k_s such that $2^{-2^{k_s}} < u$ for $u \in V_p$. Then find a subset J of $\{k_s, k_{s+1}, \ldots\}$ which satisfies (13.95) and $\sum_{k \in J}(2^k\sqrt{b_k} - \varepsilon_k) \geq 2^{s+1}L^*$. To find V_{s+1} we then use Proposition 13.4.3 with $U_k = \emptyset$ for $k < k_s$ and $U_k = \{a_m^2 ; 2^{-2^{k+1}} < a_m^2 \leq 2^{-2^k}\}$ for $k \geq k_s$.

By construction for $s \geq 1$ we have $T_s = \{\sum_{1 \leq m \leq n} a^2_{r_{m,s}} ; n \leq q_s\}$ where the integers $r_{m,s}$ for $s \geq 1$ and $n \leq q_s$ are all distinct. Consider then a permutation π with the property that for each s the integers $r_{m,s}, 1 \leq m \leq q_s$ occur as $\pi(j_s + 1), \ldots, \pi(j_s + q_s)$ for consecutive integers $j_s + 1, \ldots, j_s + q_s$. Then for each s we have $x_s + T_s \subset T_\pi$ for a certain $x_s \in [0, 1]$. In particular $x_s + t^*(T_s) \leq 1$. Given s, (13.100) proves that the probability measure ν on T_π given by $\nu(C) = \mu_s(C - x_s)$ satisfies

$$\sum_{n \geq 0} \sum_{I \in \mathcal{I}_n} \sqrt{2^{-n}\nu(I)} \geq 2^s .$$

By Theorem 13.1.1, there exists an orthonormal sequence (φ_m) such that the series $\sum_m a_{\pi(m)}\varphi_m$ does not converge a.s. so that (f) fails and the proof of Theorem 13.4.1 is complete.

Proof of Proposition 13.4.3. The proof is by induction over card J. We choose

$$\epsilon_k = 2^{k+1} \cdot 2^{-2^{k-3}} . \tag{13.101}$$

The reason for this choice will of course become apparent in due time, but let us mention that there is plenty of room in the construction.

When card $J = 1$, i.e. $J = \{k\}$, all the elements u of U satisfy $2^{-2^{k+1}} \leq u \leq 2^{-2^k}$. We built T by using the elements of U in any order we wish. So for convenience let us assume that $U = \{u_1, \ldots, u_q\}$, and let $t_n = \sum_{m \leq n} u_m$. Consider the probability measure μ on T such that $\mu(\{t_n\}) = u_n/b$, where b is the sum of the elements of U (so that $b = t^*(T)$). The idea is that for $n \leq 2^k - 2$, at the scale 2^{-n} the probability μ looks uniform on the interval $[0, b]$ because the distance between two consecutive elements of T is smaller than 2^{-n-2}. In particular for $I \in \mathcal{I}_n$ and any x we should have

$$b\mu(I - x) \simeq \lambda((I - x) \cap [0, b]) .$$

To implement the idea we appeal to Lemma 13.1.4, for the family $\mathcal{W}$ of intervals of the type $W =]t_{n-1}, t_n]$ for $1 \leq n \leq q$ (and $t_0 = 0$). We define θ_W as consisting of the mass $u_n = t_n - t_{n-1}$ at the point t_n. Then (13.12) holds for $d = 2^{-2^k}$ and (13.13) holds for $\delta = 0$. Since $\theta_{\mathcal{W}} = b\mu$ and $\lambda_{\mathcal{W}}$ is the restriction of λ to $[0, b]$, (13.14) implies that for any interval J

$$|b\mu(J) - \lambda(J \cap [0, 1])| \leq 2d . \tag{13.102}$$

Moreover for $n \leq 2^k - 2$ we have $d = 2^{-2^k} \leq 2^{-n-2} = \lambda(I - x)/4$ and (13.102) used for $J = I - x$ implies that if $I - x \subset [0, b]$

$$b\mu(I-x) \geq \lambda(I-x) - 2d \geq 2^{-n-1}\,,$$

so that

$$\sqrt{2^{-n}\mu(I-x)} \geq \frac{2^{-n-1}}{\sqrt{b}}\,. \tag{13.103}$$

Assume now that

$$b \geq 4 \cdot 2^{-2^{k-1}} \tag{13.104}$$

and that $n \geq 2^{k-1}$. Thus $b \geq 4 \cdot 2^{-n}$. We claim that when $x + b \leq 1$ then

$$\operatorname{card}\{I \in \mathcal{I}_n\,;\, I - x \subset [0,b]\} = \operatorname{card}\{I \in \mathcal{I}_n\,;\, I \subset [x, x+b]\} \geq 2^{n-1}b\,. \tag{13.105}$$

To see this we first observe that $[x, x+b] \subset [0,1]$ since $x + b \leq 1$. Next, consider the smallest integer i_1 with $x \leq i_1 2^{-n}$ and the largest integer i_2 with $i_2 2^{-n} \leq x + b$. Then $i_1 2^{-n} - x \leq 2^{-n}$ and $x + b - i_2 2^{-n} \leq 2^{-n}$ so that

$$b - (i_2 - i_1)2^{-n} \leq 2 \cdot 2^{-n} \leq \frac{b}{2}\,,$$

and thus $i_2 - i_1 \geq 2^{n-1}b$, which implies (13.105). Consequently, using (13.103) yields

$$\sum_{I \in \mathcal{I}_n, I - x \subset [0,b]} \sqrt{2^{-n}\mu(I-x)} \geq \frac{1}{4}\sqrt{b}\,. \tag{13.106}$$

Summation over $2^{k-1} \leq n \leq 2^k - 2$ then yields

$$\sum_{n \geq 0} \sum_{I \in \mathcal{I}_n, I - x \subset [0,b]} \sqrt{2^{-n}\mu(I-x)} \geq 2^{k-5}\sqrt{b} \geq 2^{-5}(2^k\sqrt{b} - \epsilon_k)\,. \tag{13.107}$$

Now, it follows from (13.101) that even when (13.104) fails, (13.107) is still satisfied since then the right-hand side is ≤ 0. This finishes the proof in the case where $\operatorname{card} J = 1$.

Let us now turn to the induction step. To bring out the idea, we first give the proof without checking all the details, to which we will come back later. We denote by k_0 the smallest element of J and $J^* = J \setminus \{k_0\}$, the set to which we will apply the induction hypothesis. We note in particular that $k_0 \geq n_0$, and the purpose of this parameter n_0 is simply to ensure that k_0 is large enough. For $k \in J$ let b_k be the sum of the elements of U_k. If

$$b_{k_0} < 4 \cdot 2^{-2^{k_0 - 2}}\,, \tag{13.108}$$

we simply apply the induction hypothesis to J^*, and this completes the induction in this case because then $2^{k_0}\sqrt{b_{k_0}} - \epsilon_{k_0} \leq 0$ from (13.101). So we assume that (13.108) fails, i.e.

$$b_{k_0} \geq 4 \cdot 2^{-2^{k_0 - 2}}\,. \tag{13.109}$$

(Here we should point out that there is huge room in the choice of the value $4\cdot 2^{-2^{k_0-2}}$.) Let us enumerate $U_{k_0} = \{u_1, \dots, u_q\}$. To each $m \le q$ we associate the weight

$$\alpha_m = u_m / b_{k_0} \; . \tag{13.110}$$

We (shall prove that we can) partition each of the sets U_k, $k \in J^*$ into sets $U_{k,m}$ for $m \le q$ in such a way that the sums $\sum\{v; v \in U_{k,m}\}$ are essentially proportional to α_m, i.e.

$$\alpha_m b_k - 2^{-2^k} \le b_{k,m} := \sum\{v \; ; \; v \in U_{k,m}\} \le \alpha_m b_k + 2^{-2^k} \; . \tag{13.111}$$

We then apply the induction hypothesis to each set $V_m = \bigcup_{k \in J^*} U_{k,m}$ to obtain a probability measure μ_m on a certain set $T_m \in \mathcal{T}(V_m)$ with the property that for $t_m^* = t^*(T_m)$ and any x with $x + t_m^* \le 1$, then

$$\sum_{n \ge 0} \sum_{I \in \mathcal{I}_n, I - x \subset [0, t_m^*]} \sqrt{2^{-n} \mu_m(I - x)} \ge \frac{1}{L^*} \sum_{k \in J^*} (2^k \sqrt{b_{k,m}} - \epsilon_k) \; . \tag{13.112}$$

Let us next describe in words the construction of T. We denote by $t_m^* = t^*(T_m)$ the largest element of T_m. We construct the elements of T in turn, each time adding a new element of U to the previously constructed element of T. We start with u_1, the first element of U_1. We then number the elements of V_1 as $v_1, v_2, \dots$ in the order in which we use them to construct T_1, and the first elements of T are $u_1, u_1 + v_1, u_1 + v_1 + v_2, \dots$, until we use up all the elements of V_1 and obtain the element $y_1 = u_1 + t_1^*$. We next add up u_2 to obtain $y_1 + u_2$. We then add the elements of V_2, in the order that they are used to construct T_2, until we reach the point $y_2 := y_1 + u_2 + t_2^*$. We then add u_3 and then start adding the elements of V_3, etc.

Formally, we construct points y_m and x_m as follows. First, we set $y_0 = 0$, and $x_1 = u_1$. Assuming that y_{m-1} has been constructed and $m \le q$, we set $x_m = y_{m-1} + u_m$ and $y_m = x_m + t_m^* = y_{m-1} + u_m + t_m^*$. We consider then the set T that consists of all the points x_m for $1 \le m \le q$ and $x_m + w_m$ where $1 \le m \le q$ and $w_m \in T_m$. We note that $y_q = t^*(T)$ is the largest element of T. It should be obvious by construction that $T \in \mathcal{T}(U)$. We then consider the probability measure $\mu = \sum_{1 \le m \le q} \alpha_m \nu_m$, where ν_m is the image of μ_m under the map $y \mapsto x_m + y$. Thus μ_m is supported by the interval $]x_m, y_m] \subset]y_{m-1}, y_m]$.

We now claim that, given x with $x + t^*(T) \le 1$,

$$\sum_{n \ge 0} \sum_{I \in \mathcal{I}_n, I - x \subset [0, t^*(T)]} \sqrt{2^{-n} \mu(I - x)} \ge \mathrm{I} + \sum_{1 \le m \le q} \mathrm{II}(m) \; , \tag{13.113}$$

where

$$\mathrm{I} = \sum_{2^{k_0 - 2} \le n < 2^{k_0 - 1}} \sum_{I \in \mathcal{I}_n, I - x \subset [0, t^*(T)]} \sqrt{2^{-n} \mu(I - x)} \tag{13.114}$$

and

$$\mathrm{II}(m) = \sum_{n>0} \sum_{I \in \mathcal{I}_n, I - x \subset [x_m, x_m + t_m^*]} \sqrt{2^{-n}\mu(I - x)} \,. \tag{13.115}$$

As we shall see,

$$t_m^* \le 2^{-2^{k_0 - 1}} \,, \tag{13.116}$$

so that none of the intervals $x + [x_m, x_m + t_m^*]$ can contain an interval $I \in \mathcal{I}_n$ for $n < 2^{k_0 - 1}$. As the intervals $x + [x_m, x_m + t_m^*]$ are disjoint as m varies, the terms in the summations (13.114) and (13.115) are all different, and each occurs in the left-hand side of (13.113). This proves (13.113).

Now we use (13.112) to obtain that, since $\mu \ge \alpha_m \nu_m$,

$$\begin{aligned}
\mathrm{II}(m) &\ge \sqrt{\alpha_m} \sum_{n>0} \sum_{I \in \mathcal{I}_n, I - x \subset [x_m, x_m + t_m^*]} \sqrt{2^{-n}\nu_m(I - x)} \\
&= \sqrt{\alpha_m} \sum_{n>0} \sum_{I \in \mathcal{I}_n, I - x - x_m \subset [0, t_m^*]} \sqrt{2^{-n}\mu_m(I - x - x_m)} \\
&\ge \frac{\sqrt{\alpha_m}}{L^*} \sum_{k \in J^*} (2^k \sqrt{b_{k,m}} - \epsilon_k) \,.
\end{aligned} \tag{13.117}$$

Using (13.111) and the inequality $\sqrt{x - y} \ge \sqrt{x} - \sqrt{y}$ yields

$$\mathrm{II}(m) \ge \frac{\sqrt{\alpha_m}}{L^*} \sum_{k \in J^*} (2^k \sqrt{\alpha_m b_k} - 2^k \cdot 2^{-2^{k-1}} - \epsilon_k) \,,$$

so that (since $\sum_{m \le q} \sqrt{\alpha_m} \le \sqrt{q}$),

$$\sum_{m \le q} \mathrm{II}(m) \ge \frac{1}{L^*} \sum_{k \in J^*} 2^k \sqrt{b_k} - \frac{\sqrt{q}}{L^*} \sum_{k \in J^*} (\epsilon_k + 2^k 2^{-2^{k-1}}) \,. \tag{13.118}$$

To conclude the proof we shall show that at scale 2^{-n} where $n < 2^{k_0 - 1}$ then μ looks uniform on the interval $[0, t^*]$, so that the same argument as in the case card $J = 1$ proves that $\mathrm{I} \ge 2^{k_0}\sqrt{b_k}/L$, and combining with (13.118) finishes the argument.

Let us now complete the details. First we fix k and we construct the sets $U_{k,m}$ in (13.111). We recall that $b_k = \sum\{v ; v \in U_k\}$. We proceed recursively. We consider a $U'_{k,1}$ a subset of U_k as large a possible with respect to the fact that $\sum\{v ; v \in U'_{k,1}\} \le \alpha_1 b_k$. Since each element of U_k is $\le 2^{-2^k}$ this implies that

$$\alpha_1 b_k - 2^{-2^k} \le \sum\{v \,;\, v \in U'_{k,1}\} \le \alpha_1 b_k$$

We then repeat this procedure on $U_k \setminus U'_{k,1}$ to construct $U'_{k,2}$, etc. In this manner we may not use all the elements of U_k. We label the remaining elements in any order, and we add as many as we can to $U'_{k,1}$ to get a set $U_{k,1}$ which still satisfies (13.111). We then add as many possible elements to $U'_{k,2}$, etc, and it should be clear that there is enough space so that we can find a proper place for each element of U_k.

Next, since

$$t_m^* = \sum\{v\ ;\ v \in U_m\} = \sum_{k\in J^*}\sum\{v\ ;\ U_{k,m}\} = \sum_{k\in J^*} b_{k,m}\ ,$$

it follows from (13.111) that

$$\Big|t_m^* - \alpha_m \sum_{k\in J^*} b_k\Big| \le \delta := \sum_{k\in J^*} 2^{-2^k}\ . \tag{13.119}$$

Since $\sum_k b_k \le 1/2$ by (13.96), this implies (for n_0 and hence k_0 large enough)

$$t_m^* + u_m \le u_m + \frac{\alpha_m}{2} + \delta \le \frac{1}{4}2^{-2^{k_0-1}}\ , \tag{13.120}$$

because

$$\alpha_m = \frac{u_m}{b_{k_0}} \le \frac{1}{4}2^{-2^{k_0-1}}\ , \tag{13.121}$$

since $u_m \le 2^{-2^{k_0}}$ while b_{k_0} satisfies (13.109). In particular we have proved (13.116).

Next, we set $b = \sum_{k\in J} b_k = b_{k_0} + \sum_{k\in J^*} b_k$, and our goal is to prove that

$$n < 2^{k_0-1}\ ;\ I \in \mathcal{I}_n\ ,\ I - x \subset [0,b] \Rightarrow b\mu(I-x) \ge 2^{-n-1}\ . \tag{13.122}$$

The proof relies on Lemma 13.1.4. Since $u_m = \alpha_m b_{k_0}$, (13.119) implies

$$|t_m^* + u_m - \alpha_m b| \le \delta\ . \tag{13.123}$$

Consider the family $\mathcal{W}$ of the intervals of the type $W =]y_{m-1}, y_m]$ for $1 \le m \le q$, and define $\theta_W = b\nu_m$, which is supported by $]x_m, y_m]$ and hence by W. Since $\lambda(W) = y_m - y_{m-1} = u_m + t_m^*$, (13.123) implies (13.13). Moreover, from (13.120), (13.12) holds for $d := 2^{-2^{k_0-1}-2}$. Since $\lambda_{\mathcal{W}}$ is the restriction of Lebesgue's measure to $[0,b]$, it follows from Lemma 13.1.4 that

$$|b\mu(I-x) - \lambda((I-x)\cap[0,b])| \le 2d + q\delta\ .$$

To prove (13.122) it suffices to prove that when n_0 is large enough the right-hand side $2d + q\delta$ is $\le 2^{-n-1}$. First since $n < 2^{k_0-1}$ we have $2d \le 2^{-n-2}$. Since we assume that $u \ge 2^{-2^{k+1}}$ for $u \in U_k$, we have

$$q = \operatorname{card} U_0 \le 2^{2^{k_0+1}} b_{k_0} \le 2^{2^{k_0+1}}\ , \tag{13.124}$$

and (13.95) implies that (when n_0 large enough),

$$q\delta \le 2^{2^{k_0+1}} \sum_{k\ge k_0+4} 2^{-2^k} \le 2\cdot 2^{2^{k_0+1}-2^{k_0+4}} \le 2\cdot 2^{-2^{k_0+1}} \le 2^{-n-2}\ ,$$

and this completes the proof of (13.122).

We can now finish the proof of Proposition 13.4.3. To bound from below the term I of (13.114), we proceed as in the case $k = 1$. Condition (13.109) ensures that $b \geq b_{k_0} \geq 4 \cdot 2^{-2^{k_0-2}}$, and for $n \geq 2^{k_0-1}$, (13.105) implies that there exist at least $b2^{n-1}$ many intervals $I \in \mathcal{I}_n$ with $I - x \subset [0, b]$. If now $n < 2^{k_0-1}$, (13.122) proves that for $I - x \subset [0, b]$ we have $\sqrt{2^{-n}\mu(I-x)} \geq 2^{-n-1}/\sqrt{b}$. Consequently, for $2^{k_0-2} \leq n < 2^{k_0-1}$ we have

$$\sum_{I \in \mathcal{I}_n, I-x \subset [0, t^*(T)]} \sqrt{2^{-n}\mu(I-x)} \geq \frac{1}{L}\sqrt{b}$$

and (13.114) yields

$$\mathrm{I} \geq \frac{1}{L_0} 2^{k_0}\sqrt{b} \geq \frac{1}{L_0} 2^{k_0}\sqrt{b_{k_0}}\ .$$

Combining with (13.118) and choosing $L^* = L_0$ we obtain

$$\mathrm{I} + \sum_{m \leq q} \mathrm{II}(m) \geq \frac{1}{L^*}\sum_{k \in J} 2^k \sqrt{b_k} - \frac{1}{L^*}\sqrt{q}\sum_{k \in J^*}(\epsilon_k + 2^k 2^{-2^k})\ .$$

This completes the induction provided

$$\epsilon_{k_0} \geq \sqrt{q}\sum_{k \in J^*}(\epsilon_k + 2^k 2^{-2^k})\ ,$$

for which it suffices, recalling (13.124) and (13.95) that

$$\epsilon_{k_0} \geq 2^{2^{k_0}} \sum_{k \geq k_0+4}(\epsilon_k + 2^k 2^{-2^k})\ ,$$

a relation which is satisfied for n_0 large enough by the choice $\epsilon_k = 2^{k+1} \cdot 2^{-2^{k-3}}$ of (13.101). □

13.5 Chaining, II

We have seen the relevance of the processes (X_t) indexed by a subset of $[0, 1]$ and for which $\mathsf{E}(X_s - X_t)^2 \leq |s - t|$. Other conditions than control of the second moment are also natural.

Definition 13.5.1. *We say that a function $\varphi : \mathbb{R} \to \mathbb{R}$ is a Young function if $\varphi(0) = 0$, $\varphi(-x) = \varphi(x)$, φ is convex and $\varphi \not\equiv 0$.*

On a metric space (T, d) one may then consider processes $(X_t)_{t \in T}$ that satisfy the condition

$$\forall\, s, t \in T\ ,\ \mathsf{E}\varphi\Big(\frac{X_s - X_t}{d(s,t)}\Big) \leq 1\ . \tag{13.125}$$

This condition is quite natural, because given the process $(X_t)_{t\in T}$, it is simple to show that the quantity

$$d(s,t) = \inf\Big\{u > 0\,;\ \mathsf{E}\varphi\Big(\frac{X_s - X_t}{u}\Big) \le 1\Big\} \tag{13.126}$$

is a quasi-distance on T, for which (13.125) is satisfied. On the other hand, it would also be natural to consider processes where the size of the "increments" $X_s - X_t$ is controlled by a distance d in a different manner, e.g. for all $u > 0$

$$\mathsf{P}(|X_s - X_t| \ge ud(s,t)) \le \psi(u)\ ,$$

for a given function ψ, see [12]. This question has received considerably less attention than the condition (13.125), but we leave its investigation for another day.

What are natural conditions that will ensure that we control the size of the process $(X_t)_{t\in T}$ under (13.125)? In the remainder of this chapter we shall briefly consider this question. The material of this section is self contained, but the reader might do well to master first the simpler ideas of Section B.2 to provide perspective. For simplicity we consider only the case where T is finite.

We say that a sequence $\mathcal{T} = (T_n)_{n\ge 0}$ of subsets of T is *admissible* if it satisfies

$$\operatorname{card} T_0 = 1 \tag{13.127}$$

and

$$\operatorname{card} T_n \le \varphi(4^n)\ . \tag{13.128}$$

Let us consider the following quantities

$$S(\mathcal{T}) = \sup_{t\in T} \sum_{n\ge 0} 4^n d(t, T_n)\ , \tag{13.129}$$

and

$$S^*(\mathcal{T}) = \sum_{n\ge 1} \sum_{s\in T_n} \frac{4^n d(s, T_{n-1})}{\varphi(4^n)}\ . \tag{13.130}$$

In the case where $\varphi(x) = \exp(x^2) - 1$, which corresponds to Gaussian processes, we have $\operatorname{card} T_n \le \exp(4^{2n})$, and the quantity (13.129) is then basically the right-hand side of (2.32) (the difference is that we change n into $4n$). The new feature here is the quantity $S^*(\mathcal{T})$, which was not needed in the Gaussian case, or more generally in the case where one has "exponential tails". The formulation of the following theorem is due again to W. Bednorz, although statements of this type have a long history.

Theorem 13.5.2. *Consider a process that satisfies (13.125). Then, for each sequence $\mathcal{T}$ of admissible sets we have*

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \le L(S(\mathcal{T}) + S^*(\mathcal{T}))\ . \tag{13.131}$$

It is not required here that $\mathsf{E}X_t = 0$.

Proof. For $n \geq 1$ let us define a map $\theta_n : T_n \to T_{n-1}$ such that for $s \in T_n$ one has

$$d(s, \theta_n(s)) = d(s, T_{n-1}) \ . \tag{13.132}$$

We may assume that $S(\mathcal{T}) < \infty$ for otherwise there is nothing to prove. This implies that for large m, T_m is a good approximation of T and in particular since T is finite, there exists m with $T = T_m$. Let us consider such a value of m. For $t \in T$ we define $\pi_m(t) = t$, and we define recursively $\pi_{n-1}(t) = \theta_n(\pi_n(t))$, so that (13.132) implies

$$d(\pi_n(t), \pi_{n-1}(t)) = d(\pi_n(t), T_{n-1}) \ . \tag{13.133}$$

We observe the following inequality: for $x, y > 0$,

$$\frac{y}{x} \leq 1 + \frac{\varphi(y)}{\varphi(x)} \ . \tag{13.134}$$

This is obvious if $y \leq x$, and if $x \leq y$ this follows from the fact that $\varphi(x) \leq x\varphi(y)/y$ by convexity of φ. We use (13.134) with $y = |X_s - X_{\theta_n(s)}|/d(s, \theta_n(s))$ and $x = 4^n$ to obtain (since $\varphi(y) = \varphi(|y|)$),

$$|X_s - X_{\theta_n(s)}| \leq 4^n d(s, \theta_n(s)) + \frac{4^n d(s, \theta_n(s))}{\varphi(4^n)} \varphi\Big(\frac{X_s - X_{\theta_n(s)}}{d(s, \theta_n(s))}\Big) \ .$$

Using this for $s = \pi_n(t)$ and (13.35) yields, if $T_0 = \{t_0\}$,

$$\begin{aligned} |X_t - X_{t_0}| \leq & \sum_{n \geq 1} 4^n d(\pi_{n-1}(t), \pi_n(t)) \\ & + \sum_{n \geq 1} \sum_{s \in T_n} \frac{4^n d(s, \theta_n(s))}{\varphi(4^n)} \varphi\Big(\frac{X_s - X_{\theta_n(s)}}{d(s, \theta_n(s))}\Big) \ , \end{aligned} \tag{13.135}$$

and consequently,

$$\begin{aligned} \sup_{t \in T} |X_t - X_{t_0}| \leq & \sup_{t \in T} \sum_{n \geq 1} 4^n d(\pi_{n-1}(t), \pi_n(t)) \\ & + \sum_{n \geq 1} \sum_{s \in T_n} \frac{4^n d(s, \theta_n(s))}{\varphi(4^n)} \varphi\Big(\frac{X_s - X_{\theta_n(s)}}{d(s, \theta_n(s))}\Big) \ . \end{aligned} \tag{13.136}$$

Taking expectation and using (13.125) yields

$$\mathsf{E} \sup_{t \in T} |X_t - X_{t_0}| \leq \sup_{t \in T} \sum_{n \geq 1} 4^n d(\pi_{n-1}(t), \pi_n(t)) + S^*(\mathcal{T}) \ . \tag{13.137}$$

Now, recalling (13.133),

$$\begin{aligned} d(\pi_{n-1}(t),\pi_n(t)) &= d(\pi_n(t),T_{n-1}) \\ &\le d(t,T_{n-1}) + d(t,\pi_n(t)) \\ &\le d(t,T_{n-1}) + \sum_{k\ge n} d(\pi_k(t),\pi_{k+1}(t)) \ . \end{aligned}$$

Thus, using that $\sum_{n\le k} 4^n \le 4^{k+1}/2$ we get

$$\begin{aligned} \sum_{n\ge1} 4^n d(\pi_{n-1}(t),\pi_n(t)) &\le \sum_{n\ge1} 4^n d(t,T_{n-1}) + \sum_{n\ge1} 4^n \sum_{k\ge n} d(\pi_k(t),\pi_{k+1}(t)) \\ &= \sum_{n\ge1} 4^n d(t,T_{n-1}) + \sum_{k\ge1}\Big(\sum_{n\le k} 4^n\Big) d(\pi_k(t),\pi_{k+1}(t)) \\ &\le \sum_{n\ge1} 4^n d(t,T_{n-1}) + \frac{1}{2}\sum_{k\ge1} 4^{k+1} d(\pi_k(t),\pi_{k+1}(t)) \\ &\le \sum_{n\ge1} 4^n d(t,T_{n-1}) + \frac{1}{2}\sum_{n\ge1} 4^n d(\pi_{n-1}(t),\pi_n(t)) \ , \end{aligned}$$

so that recalling (13.129) we get

$$\sum_{n\ge1} 4^n d(\pi_{n-1}(t),\pi_n(t)) \le 2\sum_{n\ge1} 4^n d(t,T_{n-1}) = 8\sum_{n\ge0} 4^n d(t,T_n) \le 8S(\mathcal{T}) \ . \tag{13.138}$$

Combining with (13.137) this finishes the proof. □

Interestingly, the previous proof does not use (13.128)!

Corollary 13.5.3. *Define* $e_0^* = \Delta(T,d)$ *and for* $n \ge 1$ *define*

$$e_n^* = \inf\{\epsilon > 0 \ ; \ \exists U \subset T \ , \ \operatorname{card} U \le \varphi(4^n) \ , \ \forall\, t \in T \ , \ d(t,U) \le \epsilon\} \ . \tag{13.139}$$

Then

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \le L \sum_{n\ge0} 4^n e_n^* \ . \tag{13.140}$$

Proof. Consider an arbitrary point t_0 of T, and for $n \ge 1$ consider a subset T_n of T with $\operatorname{card} T_n \le \varphi(4^n)$ and $d(t,T_n) \le 2e_n^*$. It is then obvious that the quantities $S(\mathcal{T})$ and $S^*(\mathcal{T})$ of (13.129) and (13.130) satisfy

$$S(\mathcal{T}) \le L\sum_{n\ge0} 4^n e_n^* \ ; \ S^*(\mathcal{T}) \le L\sum_{n\ge0} 4^n e_n^* \ .$$ □

Exercise 13.5.4. We recall that the covering number $N(T,d,\epsilon)$ is the smallest number of balls of radius ϵ that covers T. Deduce from Corollary 13.5.3 that

$$\mathsf{E} \sup_{s,t\in T} |X_s - X_t| \le L \int_0^{\Delta(T,d)} \varphi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon \ . \tag{13.141}$$

The bound (13.141) is called the metric entropy bound. It is of fundamental importance. A more direct proof of this inequality is given in Section B.2. Numerous applications can be found e.g. in the recent book of Michel Weber [13].

The bound of Theorem 13.5.2 raises two questions: How to construct admissible sequences? How sharp is this result?

Definition 13.5.5. *For a metric space* (T,d) *let*

$$S(T,d,\varphi) = \sup\Big\{\mathsf{E}\sup_{s,t\in T}|X_s - X_t|\Big\}\ , \tag{13.142}$$

where the supremum is taken over all the processes which satisfy (13.125).

The reader will need this definition throughout the rest of this chapter. We reformulate (13.131) as

$$S(T,d,\varphi) \le L(S(\mathcal{T}) + S^*(\mathcal{T}))\ , \tag{13.143}$$

and the question arises to which extent this inequality is sharp for the best possible choice of $\mathcal{T}$. W. Bednorz has recently discovered a rather general setting where this is the case.

Definition 13.5.6. *Consider* $p > 1$*. A distance* d *on a metric space is called* p*-concave if* d^p *is still a distance, i.e.*

$$d(s,t)^p \le d(s,v)^p + d(v,t)^p\ . \tag{13.144}$$

A p-concave distance satisfies an improved version of the triangle inequality.

Lemma 13.5.7. *If the distance* d *is* p*-concave, then for* $s,t,v\in T$ *we have*

$$d(s,v) - d(t,v) \le d(s,t)\Big(\frac{d(s,t)}{d(t,v)}\Big)^{p-1}\ . \tag{13.145}$$

Proof. We have

$$d(s,v)^p \le d(t,v)^p + d(s,t)^p\ ,$$

so that since (crudely) $(1+x)^{1/p} \le 1+x$ for $x\ge 0$,

$$d(s,v) \le d(t,v)\Big(1+\frac{d(s,t)^p}{d(t,v)^p}\Big)^{1/p} \le d(t,v) + d(s,t)\Big(\frac{d(s,t)}{d(t,v)}\Big)^{p-1}\ . \qquad \square$$

Theorem 13.5.8 (W. Bednorz [2]). *If the distance* d *is* p*-concave, then for each probability measure* μ *on* T *one has*

$$\int_T \mathrm{d}\mu(t)\int_0^{\Delta(T,d)} \varphi^{-1}\Big(\frac{1}{\mu(B(t,\varepsilon))}\Big)\mathrm{d}\epsilon \le K(p)S(T,d,\epsilon)\ , \tag{13.146}$$

where $S(T,d,\epsilon)$ *is defined in (13.142).*

The relevance of this result to the question of whether (13.143) can be reversed will only become apparent later. We prepare the proof with the following.

Lemma 13.5.9. *Consider $s, t \in T$. Then for each probability measure μ on T one has*

$$\int_T \mathrm{d}\mu(\omega) \int_{\min(d(s,\omega),d(t,\omega))}^{\max(d(s,\omega),d(t,\omega))} \frac{1}{\mu(B(\omega,3\epsilon))} \mathrm{d}\epsilon \le K(p) d(s,t) \ . \tag{13.147}$$

Proof. Let $A = \{\omega \in T \ ; \ d(t,\omega) \le d(s,\omega)\}$. Since $B(\omega, 3\epsilon) \supset B(t, 2\epsilon)$ for $\epsilon \ge d(t,\omega)$, it suffices to prove that

$$\int_A \mathrm{d}\mu(\omega) \int_{d(t,\omega)}^{d(s,\omega)} \frac{1}{\mu(B(t,2\epsilon))} \mathrm{d}\epsilon \le K(p) d(s,t) \ . \tag{13.148}$$

Let

$$A_0 = \{\omega \in A \ ; \ d(t,\omega) \le 2d(s,t)\} \ .$$

Then, since $d(s,\omega) \le d(s,t) + d(t,\omega)$,

$$\begin{aligned} &\int_{A_0} \mathrm{d}\mu(\omega) \int_{d(t,\omega)}^{d(s,\omega)} \frac{1}{\mu(B(t,2\epsilon))} \mathrm{d}\epsilon \\ &= \int \mathbf{1}_{\{d(t,\omega)\le\epsilon\le d(s,\omega)\}} \mathbf{1}_{\{d(t,\omega)\le 2d(s,t)\}} \frac{1}{\mu(B(t,2\epsilon))} \mathrm{d}\epsilon \mathrm{d}\mu(\omega) \\ &\le \int \mathbf{1}_{\{d(t,\omega)\le\epsilon\}} \mathbf{1}_{\{\epsilon\le 3d(s,t)\}} \frac{1}{\mu(B(t,2\epsilon))} \mathrm{d}\epsilon \mathrm{d}\mu(\omega) \\ &\le \int \mathbf{1}_{\{\epsilon\le 3d(s,t)\}} \frac{\mu(B(t,\epsilon))}{\mu(B(t,2\epsilon))} \mathrm{d}\epsilon \le 3d(s,t) \ . \end{aligned} \tag{13.149}$$

Next, for $n \ge 1$ let

$$A_n = \{\omega \in A \ ; \ 2^n d(s,t) \le d(t,\omega) \le 2^{n+1} d(s,t)\} \ .$$

It follows from (13.145) that for $\omega \in A_n$,

$$d(s,\omega) \le d(t,\omega) + 2^{-n(p-1)} d(s,t) \ ,$$

so that

$$\int_{d(t,\omega)}^{d(s,\omega)} \frac{1}{\mu(B(t,2\epsilon))} \mathrm{d}\epsilon \le 2^{-n(p-1)} d(s,t) \frac{1}{\mu(B(t, 2^{n+1} d(s,t)))} \ ,$$

and consequently

$$\int_{A_n} \mathrm{d}\mu(\omega) \int_{d(t,\omega)}^{d(s,\omega)} \frac{1}{\mu(B(t,2\epsilon))} \mathrm{d}\epsilon \le 2^{-n(p-1)} d(s,t) \ .$$

The result follows by summation over $n \ge 1$ and combining with (13.149).

□

Proof of Theorem 13.5.8. On the probability space (T,μ) consider the process $(X_t)_{t\in T}$ given by

$$X_t(\omega) = c\int_{d(t,\omega)}^{\Delta(T,d)/2} \varphi^{-1}\Big(\frac{1}{\mu(B(\omega,3\epsilon))}\Big)\mathrm{d}\epsilon\ , \tag{13.150}$$

where the constant $c \le 1$ will be determined later. Choosing $s=\omega$ and t with $d(s,t)\ge \Delta(T,d)/2$ we first observe that

$$\sup_{s,t\in T}|X_s(\omega)-X_t(\omega)| \ge c\int_0^{\Delta(T,d)/2}\varphi^{-1}\Big(\frac{1}{\mu(B(\omega,3\epsilon))}\Big)\mathrm{d}\epsilon\ ,$$

and thus

$$\mathsf{E}\sup_{s,t\in T}|X_s-X_t| \ge c\int_T \mathrm{d}\mu(\omega)\int_0^{\Delta(T,d)/2}\varphi^{-1}\Big(\frac{1}{\mu(B(\omega,3\epsilon))}\Big)\mathrm{d}\epsilon\ . \tag{13.151}$$

Next, letting $a(\omega)=\min(d(s,\omega),d(t,\omega))$ and $b(\omega)=\max(d(s,\omega),d(t,\omega))$, we have

$$|X_s(\omega)-X_t(\omega)| = c\int_{a(\omega)}^{b(\omega)}\varphi^{-1}\Big(\frac{1}{\mu(B(\omega,3\epsilon))}\Big)\mathrm{d}\epsilon\ .$$

Since $b(\omega)-a(\omega)\le d(s,t)$, we have $c(b(\omega)-a(\omega))/d(s,t)\le 1$. Using the convexity of φ in the second inequality, and Jensen's inequality in the third inequality,

$$\begin{aligned}
&\varphi\Big(\frac{X_s(\omega)-X_t(\omega)}{d(s,t)}\Big)\\
&=\varphi\bigg(\frac{c(b(\omega)-a(\omega))}{d(s,t)}\frac{1}{b(\omega)-a(\omega)}\int_{a(\omega)}^{b(\omega)}\varphi^{-1}\Big(\frac{1}{\mu(B(\omega,3\epsilon))}\Big)\mathrm{d}\epsilon\bigg)\\
&\le\frac{c(b(\omega)-a(\omega))}{d(s,t)}\varphi\bigg(\frac{1}{b(\omega)-a(\omega)}\int_{a(\omega)}^{b(\omega)}\varphi^{-1}\Big(\frac{1}{\mu(B(\omega,3\epsilon))}\Big)\mathrm{d}\epsilon\bigg)\\
&\le\frac{c}{d(s,t)}\int_{a(\omega)}^{b(\omega)}\frac{1}{\mu(B(\omega,3\epsilon))}\mathrm{d}\epsilon\ .
\end{aligned} \tag{13.152}$$

Lemma 13.5.9 implies that we may choose $c=c(p)$ depending on p only such (13.125) holds. Combining (13.151) with the definition of $S(T,d,\varphi)$ we then obtain

$$\sup_{\omega\in T}\int_0^{\Delta(T,d)/2}\varphi^{-1}\Big(\frac{1}{\mu(B(\omega,3\epsilon))}\Big)\mathrm{d}\epsilon \le K(p)\mathcal{S}(T,d,\varphi)\ ,$$

and change of variable then completes the proof. □

Our next result is a more general form of Lemma 13.1.14.

Lemma 13.5.10. *Assume that*

$$\text{the function } x \mapsto \varphi^{-1}(1/x) \text{ is convex} \ . \tag{13.153}$$

Assume that for each probability measure μ on T one has

$$\int_T \mathrm{d}\mu(t) \int_0^{\Delta(T,d)} \varphi^{-1}\Big(\frac{1}{\mu(B(t,\varepsilon))}\Big) \mathrm{d}\epsilon \le B \ . \tag{13.154}$$

Then there exists a probability measure μ on T for which

$$\sup_{t \in T} \int_0^{\Delta(T,d)} \varphi^{-1}\Big(\frac{1}{\mu(B(t,\varepsilon))}\Big) \mathrm{d}\epsilon \le B \ . \tag{13.155}$$

Proof. We copy the proof of Lemma 13.1.14. Let us denote by $\mathcal{M}(T)$ the set of probability measures on T. The class $\mathcal{C}$ of functions on T that satisfy

$$\exists \mu \in \mathcal{M}(T) \ ; \ \forall t \in T \ , \ f_\mu(t) := \int_0^{\Delta(T,d)} \varphi^{-1}\Big(\frac{1}{\mu(B(t,\epsilon))}\Big) \mathrm{d}\epsilon \le f(t)$$

is convex. For each probability measure ν on T, there exists f in $\mathcal{C}$ with $\int f \mathrm{d}\nu \le B$. This is because this is true for $f = f_\nu$ by (13.154). Consequently by the Hahn-Banach theorem, there exists $f \in \mathcal{C}$ such that $f \le B$, which is the content of the lemma. □

Condition (13.153) is inessential and is imposed only for simplicity. It is the behavior of φ^{-1} at zero that matters.

Our next result uses a probability measure as in (13.135) to construct a suitable admissible net. There is a genuine difficulty in this construction, namely that the measure of the balls $B(t, \epsilon)$ can greatly vary for a small variation of t. This difficulty has been bypassed in full generality by an argument of W. Bednorz, which we present now. This argument is so effective that the difficulty might no longer be noticed. Without loss of generality, we assume

$$\varphi(1) = 1 \ , \tag{13.156}$$

but (13.153) is not required.

Theorem 13.5.11 (W. Bednorz, [1]). *Consider a probability measure μ on T, and let*

$$B = \sup_{t \in T} \int_0^{\Delta(T,d)} \varphi^{-1}\Big(\frac{1}{\mu(B(t,\varepsilon))}\Big) \mathrm{d}\epsilon \ . \tag{13.157}$$

Then there is an admissible sequence $\mathcal{T}$ of subsets of T for which

$$S(\mathcal{T}) \le LB \ ; \ S^*(\mathcal{T}) \le LB \ . \tag{13.158}$$

A probability μ on (T,d) as in (13.157) is traditionally called a *majorizing measure*. Theorems 13.5.2 and 13.5.11 below show that indeed such a measure can be used to "majorize" processes on T. Technically *any* probability measure on T is a majorizing measure, although when using this name one unusually implicitly assume that the resulting bound on the process is of interest. The importance of majorzing measures has greatly decreased since the invention of the generic chaining, since they seem no longer of any use in the context of Gaussian processes, see Section 6.2.

The proof of Theorem 13.5.11 is based on the functions $\epsilon_n(t)$ defined for $n \geq 0$ as

$$\epsilon_n(t) = \inf\Big\{\epsilon > 0 \; ; \; \mu(B(t,\epsilon)) \geq \frac{1}{\varphi(4^n)}\Big\} \; . \tag{13.159}$$

This quantity is well defined since $\varphi(4^n) \geq 1$ for $n \geq 0$.

Lemma 13.5.12. *We have*

$$\mu(B(t,\epsilon_n(t))) \geq \frac{1}{\varphi(4^n)} \; ; \tag{13.160}$$

$$|\epsilon_n(s) - \epsilon_n(t)| \leq d(s,t) \; , \tag{13.161}$$

$$\forall\, t \in T \; , \; \sum_{n\geq 0} 4^n \epsilon_n(t) \leq 2B \; . \tag{13.162}$$

Proof. First, (13.160) is obvious, and since $B(t,\epsilon) \subset B(s,\epsilon + d(s,t))$, $\epsilon_n(s) \leq \epsilon_n(t) + d(s,t)$ and (13.161) follows. Next, since

$$\epsilon < \epsilon_n(t) \Rightarrow \varphi^{-1}\Big(\frac{1}{\mu(B(t,\epsilon))}\Big) > 4^n \; ,$$

we have

$$B \geq \sum_{n\geq 0} \int_{\epsilon_{n+1}(t)}^{\epsilon_n(t)} \varphi^{-1}\Big(\frac{1}{\mu(B(t,\epsilon))}\Big) \mathrm{d}\epsilon \geq \sum_{n\geq 1} 4^n(\epsilon_n(t) - \epsilon_{n+1}(t)) \; .$$

Now,

$$\sum_{n\geq 0} 4^n(\epsilon_n(t) - \epsilon_{n+1}(t)) = \sum_{n\geq 0} 4^n \epsilon_n(t) - \sum_{n\geq 0} 4^{n-1}\epsilon_n(t) \geq \frac{1}{2}\sum_{n\geq 0} 4^n \epsilon_n(t) \; . \quad \square$$

Lemma 13.5.13. *For each $n \geq 0$ there exists a subset T_n of T that satisfies the following conditions:*

$$\operatorname{card} T_n \leq \varphi(4^n) \; . \tag{13.163}$$

$$\text{The balls } B(t,\epsilon_n(t)) \text{ for } t \in T_n \text{ are disjoint} \; . \tag{13.164}$$

$$\forall\, t \in T \; , \; d(t,T_n) \leq 4\epsilon_n(t) \; . \tag{13.165}$$

$$\forall\, t \in T_n \; , \; \forall\, s \in B(t,\epsilon_n(t)) \; , \; \epsilon_n(s) \geq \frac{1}{2}\epsilon_n(t) \; . \tag{13.166}$$

Proof. We define $D_0 = T$ and we choose $t_1 \in D_0$ such that $\epsilon_n(t_1)$ is as small as possible. Assuming that we have constructed $D_{k-1} \neq \emptyset$, we choose $t_k \in D_{k-1}$ such that $\epsilon_n(t_k)$ is as small as possible and we define

$$D_k = \left\{ t \in D_{k-1} \; ; \; d(t, t_k) \geq 2(\epsilon_n(t) + \epsilon_n(t_k)) \right\} .$$

The construction continues as long as possible. It stops at the first integer p for which $D_p = \emptyset$. We define $T_n = \{t_1, t_2 \dots, t_p\}$. Consider $t_k, t_{k'} \in T_n$ with $k < k'$. Then by construction, and since the sequence (D_k) decreases, $t_{k'} \in D_k$, so that

$$d(t_{k'}, t_k) \geq 2(\epsilon_n(t_{k'}) + \epsilon_n(t_k)) ,$$

and therefore the balls $B(t_k, \epsilon_n(t_k))$ and $B(t_{k'}, \epsilon_n(t_{k'}))$ are disjoint. This proves (13.164) and (13.160) implies (13.163). To prove (13.165) consider $t \in T$ and the largest $k \geq 1$ such that $t \in D_{k-1}$. Then by the choice of t_k we have $\epsilon_n(t) \geq \epsilon_n(t_k)$. Since by definition of k we have $t \notin D_k$ the definition of D_k shows that

$$d(t, t_k) \leq 2(\epsilon_n(t) + \epsilon_n(t_k)) \leq 4\epsilon_n(t) ,$$

and since $t_k \in T_n$ this proves (13.165).

Finally, consider t_k and $s \in B(t_k, \epsilon_n(t_k))$. If $s \in D_{k-1}$ then $\epsilon_n(s) \geq \epsilon_n(t_k)$ and (13.166) is proved. Otherwise, the unique k' such that $s \in D_{k'-1}$ and $s \notin D_{k'}$ satisfies $k' < k$. Since $s \in D_{k'-1}$ but $s \notin D_{k'}$, the definition of this set shows that

$$d(s, t_{k'}) \leq 2(\epsilon_n(s) + \epsilon_n(t_{k'})) ,$$

and since $d(s, t_k) \leq \epsilon_n(t_k)$ we get

$$d(t_k, t_{k'}) \leq d(s, t_k) + d(s, t_{k'}) \leq \epsilon_n(t_k) + 2(\epsilon_n(s) + \epsilon_n(t_{k'})) . \tag{13.167}$$

On the other hand, since $k' < k$ then $t_k \in D_{k-1} \subset D_{k'}$ so the definition of this set implies

$$d(t_k, t_{k'}) \geq 2(\epsilon_n(t_k) + \epsilon_n(t_{k'})) ,$$

and comparing with (13.167) completes the proof of (13.166). □

Proof of Theorem 13.5.11. For $n \geq 0$ we consider the set T_n provided by Lemma 13.5.13, so $\operatorname{card} T_0 = 1$. Combining (13.162) and (13.165) we obtain

$$\sum_{n \geq 0} 4^n d(t, T_n) \leq 8B ,$$

and this proves that $S(\mathcal{T}) \leq 8B$.

Next, since $\mu(B(s, \epsilon_n(s)) \geq 1/\varphi(4^n)$ by (13.160) and since $d(s, T_{n-1}) \leq 4\epsilon_{n-1}(s)$ by (13.165), for $n \geq 1$ we have

$$\sum_{s \in T_n} \frac{d(s, T_{n-1})}{\varphi(4^n)} \leq 4 \sum_{s \in T_n} \int_{B(s, \epsilon_n(s))} \epsilon_{n-1}(s) \mathrm{d}\mu(t) .$$

Now, combining (13.161) and (13.166) for $t \in B(s, \epsilon_n(s))$ implies

$$\epsilon_{n-1}(s) \le \epsilon_{n-1}(t) + \epsilon_n(s) \le \epsilon_{n-1}(t) + 2\epsilon_n(t) ,$$

and since the balls $B(s, \epsilon_n(s))$ are disjoint for $s \in T_n$ this yields

$$\sum_{s \in T_n} \frac{4^n d(s, T_{n-1})}{\varphi(4^n)} \le 4^{n+1} \int_T (\epsilon_{n-1}(t) + 2\epsilon_n(t)) \mathrm{d}\mu(t) .$$

Summation over $n \ge 1$ and use of (13.162) conclude the proof. □

For a metric space (T, d) we define $\mathcal{M}(T, d, \varphi)$ as the infimum over all probability measures μ on T of the quantity

$$\sup_{t \in T} \int_0^{\Delta(T,d)} \varphi^{-1}\Big(\frac{1}{\mu(B(t, \varepsilon))}\Big) \mathrm{d}\epsilon . \tag{13.168}$$

Combining Theorems 13.5.2, 13.5.8 and 13.5.11, we have proved the following.

Theorem 13.5.14. *If the distance d is p-concave, then*

$$S(T, d, \varphi) \le L\mathcal{M}(T, d, \varphi) \le K(p) S(T, d, \varphi) . \tag{13.169}$$

This allows in principle to compute $S(T, d, \varphi)$ although the determination of the quantity $\mathcal{M}(T, d, \varphi)$ is by no means easy.

What happens when we do not assume that the distance is p-concave? This situation will be briefly discussed in the next section, and we end up the present section by discussing two more specialized questions. A striking feature of Theorem 13.1.1 is that even though we studied processes that satisfied $\mathsf{E}(X_s - X_t)^2 = d(s, t)$ where d is the usual distance on the unit interval, we ended up considering the sequence $\mathcal{I}_n$ of partitions of this unit interval, and, implicitly, the distance δ given by $\delta(s, t) = 2^{-n}$ where n is the largest integer for which s, t belong to the same element of $\mathcal{I}_n$. This distance is *ultrametric*, i.e. it satisfies

$$\forall\, s, t, v \in T \ , \ \delta(s, t) \le \max(\delta(s, v), \delta(t, v)) . \tag{13.170}$$

Note that in particular a distance is ultrametric if and only if it is p-concave for all p. Ultrametric distances are intimately connected to increasing sequences of partitions, because the balls of a given radius form a partition in a ultrametric space. As the following shows, the occurrence of an ultrametric structure is very frequent.

Theorem 13.5.15 (W. Bednorz [3]). *Let us assume that the Young function φ satisfies*

$$\forall\, k \ge 1 \ , \ \sum_{n > k} \frac{4^n}{\varphi(4^n)} \le C \frac{4^k}{\varphi(4^k)} . \tag{13.171}$$

Consider an admissible sequence $\mathcal{T}$ of subsets of (T,d). Then there exists a ultrametric distance $\delta \geq d$ and an admissible sequence $\mathcal{T}^*$ of subsets of (T,δ) such that

$$S(\mathcal{T}^*) + S^*(\mathcal{T}^*) \leq K(C)(S(\mathcal{T}) + S^*(\mathcal{T})) , \tag{13.172}$$

where $K(C)$ depends on C only.

In words, this means that if the existence of an admissible net provides a bound for processes that satisfy the increment condition (13.125), then there is an ultrametric distance δ greater than d such that the processes satisfying the increment condition (13.125) for this greater distance basically still satisfy the same bound.

Proof. Let $\mathcal{T} = (T_n)_{n\geq 0}$. In a first step we prove that we may assume that the sequence (T_n) increases. Define $T'_0 = T_0$ and for $n \geq 1$ define $T'_n = \cup_{k<n} T_k$. Thus

$$\operatorname{card} T'_n \leq \sum_{k<n} \varphi(4^k) \leq \sum_{k<n} 4^{k-n}\varphi(4^n) \leq \varphi(4^n) ,$$

so that the sequence $\mathcal{T}' = (T'_n)_{n\geq 1}$ is admissible. Since $d(t,T'_n) \leq d(t,T_{n-1})$ for $n \geq 1$, it follows that $S(\mathcal{T}') \leq 4S(\mathcal{T})$. Next, we observe that for $n \geq 2$, and since $T'_n \subset \cup_{k<n} T_k$,

$$\sum_{s\in T'_n} d(s,T'_{n-1}) \leq \sum_{k<n}\sum_{s\in T_k} d(s,T'_{n-1}) \leq \sum_{k<n}\sum_{s\in T_k} d(s,T_{k-1}) ,$$

because $T_{k-1} \subset T'_{n-1}$ for $k < n$. For $n = 1$, $d(s,T'_{n-1}) = 0$ for $s \in T'_1 = T_0$. Thus, using (13.171) in the last line,

$$\begin{aligned} S^*(\mathcal{T}') &= \sum_{n\geq 1}\sum_{s\in T'_{n-1}} \frac{4^n d(s,T'_{n-1})}{\varphi(4^n)} \\ &\leq \sum_{n\geq 2}\sum_{k<n}\sum_{s\in T_k} \frac{4^n d(s,T_{k-1})}{\varphi(4^n)} \\ &= \sum_{k\geq 1}\sum_{s\in T_k} d(s,T_{k-1}) \sum_{n>k} \frac{4^n}{\varphi(4^n)} \\ &\leq CS(\mathcal{T}) . \end{aligned}$$

In summary, the sequence $\mathcal{T}'$ is admissible and increasing, and satisfies $\mathcal{S}(\mathcal{T}') \leq 4\mathcal{S}(\mathcal{T})$ and $\mathcal{S}^*(\mathcal{T}') \leq C\mathcal{S}^*(\mathcal{T})$. Therefore replacing $\mathcal{T}$ by $\mathcal{T}'$ we now assume that the sequence (T_n) increases. Let us consider the points $\pi_n(t)$ as in the proof of Theorem 13.5.2. Since the sequence (T_n) increases, we have $\pi_k(t) = t$ for $t \in T_n$ and $k \geq n$. Given $s,t \in T$, let us consider the largest integer m for which $\pi_m(s) = \pi_m(t)$ and define

$$\delta(s,t) = 2\max\Big(\sum_{k\geq m} d(\pi_k(t),\pi_{k+1}(t)), \sum_{k\geq m} d(\pi_k(s),\pi_{k+1}(s))\Big) . \tag{13.173}$$

It is straightforward to check that this defines a ultrametric distance. Moreover, since $t = \pi_n(t)$ for n large enough, the triangle inequality implies

$$d(t, \pi_m(t)) \le \sum_{k \ge m} d(\pi_k(t), \pi_{k+1}(t)) ,$$

so that

$$\delta(s,t) \ge 2 \max\bigl(d(t, \pi_m(t)), d(s, \pi_m(s))\bigr) \ge d(s,t) ,$$

because $d(s,t) \le d(s, \pi_m(s)) + d(t, \pi_m(s)) = d(s, \pi_m(s)) + d(t, \pi_m(t))$.

Consider now $t \in T$ and $s = \pi_n(t) \in T_n$. Then $\pi_k(t) = \pi_k(s)$ for $k \le n$, and $\pi_k(s) = s$ for $k \ge n$. Consequently the definition of δ shows that

$$\delta(t, T_n) \le 2 \sum_{k \ge n} d(\pi_k(t), \pi_{k+1}(t)) . \tag{13.174}$$

Interchanging as usual the sums over k and n

$$\sum_{n \ge 0} 4^n \delta(t, T_n) \le \sum_{k \ge 0} d(\pi_k(t), \pi_{k+1}(t)) \sum_{n \le k} 4^n \le \sum_{k \ge 0} 4^k d(\pi_k(t), \pi_{k+1}(t)) ,$$

and (13.138) proves that if we denote by $\mathcal{T}^*$ the sequence (T_n) seen as an admissible sequence in the metric space (T, δ), then $S(\mathcal{T}^*) \le LS(\mathcal{T})$.

Now, if $t \in T_{n+1}$ we have $\pi_k(t) = t$ for $k \ge n+1$ and thus (13.174) and (13.133) yields $\delta(t, T_n) \le 2d(\pi_{n+1}(t), \pi_n(t)) = 2d(\pi_{n+1}(t), T_n) = 2d(t, T_n)$. This implies that $S^*(\mathcal{T}^*) \le 2S^*(\mathcal{T})$. □

The conclusion of Theorem 13.5.15 is not true without some kind of condition on φ such as (13.171). A counter example is provided in [9] in the case $\varphi(x) = x$.

Another topic that we would like to briefly investigate is to which extend we can improve (13.131) by requiring a stronger integrability condition on $\sup_{s,t} |X_s - X_t|$. For a Young function φ, and a r.v. X let us define

$$\|X\|_\varphi = \inf\{u > 0 \; ; \; \mathsf{E}\varphi(X/u) \le 1\} , \tag{13.175}$$

so that the distance of (13.126) is simply $\|X_s - X_t\|_\varphi$. It would be nice if we could replace the left-hand side of (13.131) by

$$\Bigl\| \sup_{s,t \in T} |X_s - X_t| \Bigr\|_\varphi ,$$

but unfortunately this is not true. However we have the following (which is a special case of a general principle, see [9]).

Proposition 13.5.16. *Assume that for a Young function ψ we have*

$$x \ge \varphi^{-1}(1) = 1 \; , \; y \ge 1 \Rightarrow \varphi(xy) \ge \varphi(x)\psi(y) . \tag{13.176}$$

Then we may replace (13.131) by

$$\Bigl\| \sup_{s,t \in T} |X_s - X_t| \Bigr\|_\psi \le L(S(\mathcal{T}) + S^*(\mathcal{T})) . \tag{13.177}$$

In particular we may improve the metric entropy bound (13.141) into

$$\Big\| \sup_{s,t\in T} |X_s - X_t| \Big\|_\psi \le L \int_0^{\Delta(T,d)} \varphi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon \,. \tag{13.178}$$

Proof. Proceeding as in (13.135) we observe that for each number $a > 0$ we have

$$\begin{aligned} \sup_{t\in T} |X_t - X_{t_0}| \le a \sup_{t\in T} \sum_{n\ge 1} 4^n d(\pi_{n-1}(t), \pi_n(t)) \\ + \sum_{n\ge 1} \sum_{s\in T_n} b(n,s)\varphi\Big(\frac{Y_{n,s}}{a}\Big) \,, \end{aligned} \tag{13.179}$$

where we lighten notation by writing

$$b(n,s) = \frac{4^n d(s,\theta_n(s))}{\varphi(4^n)} \; ; \; Y_{n,s} = \frac{|X_s - X_{\theta_n(s)}|}{d(s,\theta_n(s))} \,.$$

Let us define

$$h(\omega) = \inf\Big\{ a > 0 \; ; \; \sum_{n\ge 1} \sum_{s\in T_n} b(n,s)\varphi\Big(\frac{Y_{n,s}(\omega)}{a}\Big) \le 2S^*(\mathcal{T}) \Big\} \,, \tag{13.180}$$

so that (13.179) implies

$$\sup_{t\in T} |X_t(\omega) - X_{t_0}(\omega)| \le h(\omega) \sup_{t\in T} \sum_{n\ge 1} 4^n d(\pi_{n-1}(t), \pi_n(t)) + 2S^*(\mathcal{T}) \,,$$

and recalling (13.138) it suffices to prove that $\|h\|_\psi \le 2$. We deduce from (13.180)

$$\sum_{n\ge 1} \sum_{s\in T_n} b(n,s)\varphi\Big(\frac{Y_{n,s}(\omega)}{h(\omega)}\Big) = 2S^*(\mathcal{T}) \,. \tag{13.181}$$

Recalling that $\varphi(1) = 1$ (so that $\varphi(x) \le 1$ for $|x| \le 1$) and that the sum of the coefficients $b(n,s)$ for $s \in T_n$ and $n \ge 1$ is $S^*(\mathcal{T})$,

$$\sum_{n\ge 1} \sum_{s\in T_n} b(n,s)\varphi\Big(\frac{Y_{n,s}(\omega)}{h(\omega)}\Big) \mathbf{1}_{\{Y_{n,s} \le h(\omega)\}} \le S^*(\mathcal{T}) \,,$$

and comparing with (13.181),

$$\sum_{n\ge 1} \sum_{s\in T_n} b(n,s)\varphi\Big(\frac{Y_{n,s}(\omega)}{h(\omega)}\Big) \mathbf{1}_{\{Y_{n,s}(\omega)| > h(\omega)\}} \ge S^*(\mathcal{T}) \,. \tag{13.182}$$

Now, (13.176) implies that $\varphi(y/h(\omega)) \le \varphi(y)/\psi(h(\omega))$ for $y \ge 1$ and $h(\omega) \ge 1$. Using this for $y = Y_{s,n}(\omega)$, (13.182) implies

$$\mathbf{1}_{\{h(\omega)\geq 1\}}\psi(h(\omega))S^*(\mathcal{T}) \leq \sum_{n\geq 1}\sum_{s\in T_n} b(n,s)\varphi(Y_{s,n}(\omega)) \; .$$

Since the expected value of the right-hand side is $\leq S^*(\mathcal{T})$, taking expectation implies $\mathsf{E}\mathbf{1}_{\{h(\omega)\geq 1\}}\psi(h(\omega)) \leq 1$. Since taking $x = y = 1$ in (13.176) proves that $\psi(1) \leq 1$, it follows that $\mathsf{E}\psi(h) \leq 2$, so $\mathsf{E}\psi(h/2) \leq 1$ and $\|h\|_\psi \leq 2$. □

Condition (13.176) is essentially optimal, as the following challenging exercise shows.

Exercise 13.5.17. Investigate the necessary conditions on the function ψ so that for any metric space and any process $(X_t)_{t\in T}$ that satisfies (13.125) one has

$$\Big\|\sup_{s,t\in T}|X_s - X_t|\Big\|_\psi \leq L\int_0^{\Delta(T,d)} \varphi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon \; . \tag{13.183}$$

(Hint. Consider N and the space T of cardinality N where any two distinct points are at distance 1. Consider $\epsilon < 1$ and consider disjoint events $(\Omega_t)_{t\in T}$ with $\mathsf{P}(\Omega_t) = \epsilon/N$. Apply (13.183) to the process $(X_t)_{t\in T}$ given by $X_t = \varphi^{-1}(N/\epsilon)\mathbf{1}_{\Omega_t}$.)

13.6 Chaining, III

We now briefly discuss the problem of the boundedness of processes that satisfy (13.125) in a general metric space, when the distance is not assumed to be p-concave.

In all the examples of chaining we have used, the interpolation points $\pi_n(t)$ converge geometrically towards t. This feature is not always optimal. To understand this, consider a toy example, the unit interval with the usual distance.

Proposition 13.6.1. *Consider a process $(X_t)_{t\in[0,1]}$ that satisfies*

$$\forall\, s,t \in [0,1]\;,\; \mathsf{E}|X_s - X_t| \leq |s-t| \; . \tag{13.184}$$

Then

$$\mathsf{E}\sup_{0\leq s,t\leq 1}|X_s - X_t| \leq 1 \; . \tag{13.185}$$

Proof. If $0 \leq t_1 < \ldots < t_n \leq 1$, then

$$\mathsf{E}\sup_{\ell<\ell'}|X_{t_\ell} - X_{t'_\ell}| \leq \mathsf{E}\sum_{1\leq \ell<n}|X_{t_{\ell+1}} - X_{t_\ell}| \leq \sum_{1\leq\ell\leq n} t_{\ell+1} - t_\ell \leq 1 \; . \qquad \square$$

Exercise 13.6.2. Prove that if μ is a probability measure on $[0,1]$ then

$$\int_0^1 \mathrm{d}t \int_0^1 \frac{1}{\mu(B(t,\epsilon))}\mathrm{d}\epsilon = \infty \; .$$

Exercise 13.6.3. Review the proof of Theorem 13.5.2 to show that when $\varphi(x) = |x|$, then one can improve (13.131) into

$$\mathsf{E} \sup_{s,t \in T} |X_s - X_t| \le LS^*(\mathcal{T}) .$$

Exercise 13.6.4. On $[0,1]$ consider the distance δ given by $d(s,t) = 2^{-n}$, where n is the largest such that s, t belong to the same dyadic interval of length 2^{-n}. Construct an unbounded process (X_t) on $([0,1], \delta)$ that satisfies (13.125). Compare the covering numbers of (T, δ) and (T, d) where $T = [0,1]$ and d is the usual distance.

What happens in Proposition 13.6.1 is that one can joint two elements of the space by a long chain of small steps (and this is not the case in the setting of Exercise 13.6.4). A somewhat similar phenomenon occurs when $T = [0,1]^p$ is provided with the usual distance d. Which are the functions φ such that the processes satisfying (13.125) are bounded? In that situation the covering numbers $N(T, d, \epsilon)$ behave like ϵ^{-p}, so Corollary 13.5.3 implies that it suffices that

$$\sum_n 2^{-n} \varphi^{-1}(2^{np}) < \infty .$$

This, however, is not sharp. It is proved in [9] that the necessary and sufficient condition is $\sum_n 2^{-n} \varphi'(2^{n(1-p)}) < \infty$ (and that this condition is weaker than the previous one, which is not obvious). We do not reproduce these results, which unfortunately are still waiting to receive their first application.

There is some kind of "connectivity" in the structure of $[0,1]^p$ that explains the previous results. This phenomenon does not exist when the distance is ultrametric, as Exercise 13.6.4 shows. There are of course situations where both aspects are present, e.g. if one takes a product of $[0,1]^p$ with a ultrametric space. Not surprisingly, these situations are "intermediate" between the "connected case" and the ultrametric case. Complete computations are performed in [9] in such a genuinely non-trivial instance. The complicated necessary and sufficient conditions found in this case probably indicate that no simple complete description of the metric spaces for which condition (13.125) implies boundedness can be found, even in the "homogeneous situation" where covering numbers suffice.

13.7 Notes and Comments

Il y a les questions qui se posent, et les questions que l'on se pose.
Henri Poincaré

This is a fundamental thought. I feel that unfortunately many of the problems considered in the chapter belong to the second category rather than the first. What is the point of determining the sequences (a_m) as in Theorem 13.2?

Of course one might say (and the author agrees) that it is worth the effort to determine the best possible condition on these sequences (a_m) just to make sure that we understand what is going on. As for solving any practical problem, that is, determining whether an actual orthonormal series converges, it is very unlikely that this series has been chosen by a clever adversary making matters as bad as possible. Matters should typically be much better than this “worst case” study.

Similarly, how important is the question of studying boundedness of processes under the increment conditions (13.125)? The author undertook the systematic study [9] because he wanted to work on the problem of boundedness of Gaussian processes, and felt that he had no chance unless he understood really well majorizing measures. The strategy worked, but the author does not feel that these are important questions, and covered only the part of the theory which is made irresistibly attractive by the recent work of W. Bednorz, who found these very clean and seemingly final arguments. Section 13.5 follow very closely those of his corresponding papers.

In his formidable paper [7], A. Paszkiewicz writes: “ Only classical methods will be used. A reader who knows only the classical proofs of Rademacher-Menshov and Tandori theorems ([5] Chap. 8) is as much prepared to study this paper as a reader who knows contemporary theory, e.g. the generic chaining.” It is tempting to paraphrase him, and to say that the reader who know only contemporary theory is as much prepared to study Bednorz's theorem than a reader who also knows classical theory. The reader will of course decide by herself which approach she prefers.

Let us stress that a particularly important contribution of W. Bednorz is to have brought to light the technical importance of (13.4), after which everything becomes much easier.

Let us finally point out that when the Young function φ as in Section 13.5 has “polynomial growth” rather than “exponential growth”, it does not seem possible to characterize the size of T according to majorizing measures in terms of the size of the trees it contains, as we did in Section 6.2.

References

1. Bednorz, W.: On Talagrand's admissible net approach to majorizing measures and boundedness of stochastic processes. Bull. Pol. Acad. Sci., Math. **56**(1), 83–91 (2008)
2. Bednorz, W.: Majorizing measures on metric spaces. C. R. Math. Acad. Sci. Paris **348**(1–2), 75–78 (2006)
3. Bednorz, W.: Majorizing measures and ultrametric spaces. Bull. Pol. Acad. Sci., Math. **60**(1), 91–100 (2012)
4. Bednorz, W.: The complete characterization of a.s. convergence of orthogonal series. Ann. Probab. **41**(2), 1055–1071 (2013)
5. Kashin, B.S., Saakyan, A.A.: Orthogonal Series. Translations of Mathematical Monographs, vol. 75. Am. Math. Soc., Providence (1989)

6. Menchov, D.E.: Sur les séries de fonctions orthogonales. Fundam. Math. **4**, 82–105 (1923)
7. Paszkiewicz, A.: A complete characterization of coefficients of a.e. convergent orthogonal series and majorizing measures. Invent. Math. **180**(1), 50–110 (2010)
8. Radmacher, H.: Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. Math. Ann. **87**(1–2), 112–138 (1922)
9. Talagrand, M.: Sample boundedness of stochastic processes under increment conditions. Ann. Probab. **18**, 1–49 (1990)
10. Tandori, K.: Über die divergenz der Orthogonalreihen. Publ. Math. Drecen **8**, 291–307 (1961)
11. Tandori, K.: Über die Konvergenz der Orthogonalreihen. Acta Sci. Math. (Szeged) **24**, 131–151 (1963)
12. Weber, M.: Analyse infiniésimale de fonctions aléatoires. In: Eleventh Saint Flour Probability Summer School, Saint Flour, 1981. Lecture Notes in Math., vol. 976, pp. 383–465. Springer, Berlin (1981)
13. Weber, M.: Dynamical Systems and Processes. IRMA Lectures in Mathematics and Theoretical Physics, vol. 14. European Mathematical Society (EMS), Zürich (2009). xii+761 pp. ISBN: 978-3-03719-046-3

14. Matching Theorems, II: Shor's Matching Theorem

14.1 Introduction

This chapter continue Chapter 4, which should be fresh in the reader's mind before attempting to penetrate the more difficult material presented here. In particular the notion of "evenly spread" points is explained on page 101. The main result is as follows.

Theorem 14.1.1 (P. Shor). *Consider evenly spread points* $(Y_i)_{i\le N}$ *of* $[0,1]^2$*. Set* $Y_i = (Y_i^1, Y_i^2)$*. Consider i.i.d. points* $(X_i)_{i\le N}$ *uniform over* $[0,1]^2$ *and set* $X_i = (X_i^1, X_i^2)$*. Then with probability* $\ge 1 - LN^{-10}$ *there exists a matching* π *such that*

$$\sum_{i\le N} |X_i^1 - Y_{\pi(i)}^1| \le L\sqrt{N\log N} \tag{14.1}$$

$$\sup_{i\le N} |X_i^2 - Y_{\pi(i)}^2| \le L\sqrt{\frac{\log N}{N}}\ . \tag{14.2}$$

The power N^{10} plays no special role, and (14.1) and (14.2) show that Theorem 14.1.1 improves upon Theorem 4.3.1. The difference of course is that in Theorem 4.3.1 we know only that $|X_i^2 - Y_{\pi(i)}^2| \le L\sqrt{\log N/N}$ in average over i while now we know this for each $i \le N$.

A remarkable feature of Theorem 14.1.1 is that both coordinates do not play the same role. Following this idea, one may ask the following.

Research problem 14.1.2 (The ultimate matching conjecture). Prove or disprove the following. Consider $\alpha_1, \alpha_2 > 0$ with $1/\alpha_1 + 1/\alpha_2 = 1/2$. Then with high probability we can find a matching π such that, for $j = 1, 2$, we have

$$\sum_{i\le N} \exp\left(\sqrt{\frac{N}{\log N}}\frac{|X_i^j - Y_{\pi(i)}^j|}{L}\right)^{\alpha_j} \le 2N\ .$$

Noting that

$$\sum_{i\le N} \exp a_i^4 \le 2N \Rightarrow \max_{i\le N} |a_i| \le L(\log N)^{1/4}\,,$$

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_14,

shows that the case $\alpha_1 = \alpha_2 = 4$ would provide a very neat common generalization of Theorems 4.3.1 and 4.4.1, while the case "$\alpha_1 = 2, \alpha_2 = \infty$" (when suitably formulated) would be much stronger than Theorem 14.1.1. In Chapter 15 we shall prove a suitable version of this conjecture in dimension $d \geq 3$. In dimension 2, a partial result in the direction of Problem 14.1.2 is as follows.

Theorem 14.1.3. *Consider a number* $0 < \alpha < 1/2$*, an integer* $N \geq 2$*, and evenly spread points* $(Y_i)_{i \leq N}$ *of* $[0,1]^2$*. Set* $Y_i = (Y_i^1, Y_i^2)$*. Consider i.i.d..d points* $(X_i)_{i \leq N}$ *uniform over* $[0,1]^2$ *and set* $X_i = (X_i^1, X_i^2)$*. Then with probability* $\geq 1 - LN^{-10}$ *there exists a matching* π *such that*

$$\sum_{i \leq N} \exp\left(\sqrt{\frac{N}{\log N}} \frac{|X_i^1 - Y_{\pi(i)}^1|}{K(\alpha)}\right)^\alpha \leq 2N \tag{14.3}$$

$$\sup_{i \leq N} |X_i^2 - Y_{\pi(i)}^2| \leq K(\alpha) \sqrt{\frac{\log N}{N}} \,. \tag{14.4}$$

Since $\exp |x|^\alpha \geq |x|/K(\alpha)$, it follows from (14.3) that

$$\sum_{i \leq N} |X_i^1 - Y_{\pi(i)}^1| \leq L\sqrt{N \log N} \,. \tag{14.5}$$

Of course (14.4) and (14.5) show that Theorem 14.1.3 improves upon Theorem 14.1.1. We leave it to the reader to show that when α increases, the conclusion of Theorem 14.1.3 becomes stronger. A special case of the ultimate matching conjecture is to prove Theorem 14.1.3 for $\alpha = 2$ (which is a nice research problem by itself). We shall not prove Theorem 14.1.3 here. The proof is based on the same ideas as the proof of Theorem 14.1.1 but is more technical. It can be found in [2]. It is quite simpler than the original proof of this result [1] (which did not reach as good a value of α). We decided this time to present the simpler case of Theorem 14.1.1, since the reader really desperate to see the proof of Theorem 14.1.3 can probably find it easily.

14.2 The Discrepancy Theorem

The proof of Theorem 14.1.1 relies again on Proposition 4.2.1 and a "discrepancy theorem" of the same nature as (4.55), but for a more complicated class of functions. This is Theorem 14.2.1 below. It requires some preparations to state this discrepancy theorem.

To prove Theorem 14.1.1 we do not care about what happens at a scale less than $\sqrt{\log N}/\sqrt{N}$. We shall later choose an integer p with 2^{-p} about $\sqrt{\log N}/\sqrt{N}$, with the idea of dividing the unit square into 2^{2p} equal little squares, and we think of each of these little squares a single point. This motivates to introduce the set $G = \{1, \ldots, 2^p\}^2$, each point of G corresponding to a little square. So G is now our model for the unit square $[0,1]^2$. (The reader observes that the notation G does not have the same meaning as in Chapter 4.) We forget about the unit square until the near end of the present section, on page 470, when we will come back to it and complete the proof of Theorem 14.1.1 by easy arguments. We turn our attention to proving matching theorems in G.

We want to match random points $(U_i)_{i \le N}$ of G to "evenly spread" points $(Z_i)_{i \le N}$ of G. Since 2^{2p} may not divide N, we may not be able to put the same number of points Z_i at each point of G. For $(k, \ell) \in G$, let us introduce

$$n(k,\ell) = \mathrm{card}\{i \le N \; ; \; Z_i = (k,\ell)\} \; , \tag{14.6}$$

so that

$$\sum_{(k,\ell) \in G} n(k,\ell) = N \; . \tag{14.7}$$

Since 2^{-p} is about $\sqrt{\log N}/\sqrt{N}$, $N/2^{2p}$ is about $\log N$, and hence large, and since we try to make the points (Z_i) evenly spread, we can certainly arrange that for a certain integer m_0 we have

$$m_0 \le n(k,\ell) \le 2m_0 \; . \tag{14.8}$$

We note that (14.7) implies

$$N2^{-2p-1} \le m_0 \le N2^{-2p} \; . \tag{14.9}$$

We shall always assume that

$$m_0 \ge p \; . \tag{14.10}$$

It is natural to consider on G the probability measure μ given by

$$\mu(\{(k,\ell)\}) = \frac{n(k,\ell)}{N} \; . \tag{14.11}$$

Thus for a function h on G, we have

$$\int h \mathrm{d}\mu = \frac{1}{N} \sum_{(k,\ell) \in G} n(k,\ell) h(k,\ell) \; . \tag{14.12}$$

For all practical purposes, one may think to μ as the uniform measure on G. The reader should not be disturbed that we shall consider i.i.d. r.v.s (U_i) of law μ rather than uniform. This simply corresponds to the fact that each point of G will not represent exactly a little square of side 2^{-p} but rather a slightly different domain, as will become clear later.

We consider the class $\mathcal{H}$ of functions $h : G \to \mathbb{R}$ such that

$$\sum |h(k, \ell + 1) - h(k, \ell)| \le 2^{2p} \tag{14.13}$$

$$\forall\, k, \ell\,,\ |h(k+1, \ell) - h(k, \ell)| \le 1\,. \tag{14.14}$$

The summation in (14.13) is over $1 \le k \le 2^p$, $1 \le \ell \le 2^p - 1$. To lighten notation we will not mention any more that it is always understood that when a quantity such as $h(k, \ell+1) - h(k, \ell)$ occurs in a summation, we consider only the values of ℓ with $\ell + 1 \le 2^p$. In a similar manner, when the quantity $|h(k+1, \ell) - h(k, \ell)|$ occurs in a condition, it is alway understood that we consider only the values of k for which $k + 1 \le 2^p$.

It is not obvious yet that the class $\mathcal{H}$ of functions is related to a matching problem, although of course the discussion at the end of Section 4.2 makes it less surprising. Let us also keep in mind that we are dealing with a difficult problem, so patience is required from the reader in this section, and things can become clear only gradually.

The central ingredient to our approach is the following.

Theorem 14.2.1. *Consider independent r.v.s U_i valued in G, of law μ. Then, with probability $\ge 1 - \exp(-46p)$, we have*

$$\forall h \in \mathcal{H}\,,\ \Big|\sum_{i \le N} (h(U_i) - \int h \mathrm{d}\mu)\Big| \le L\sqrt{p m_0}\, 2^{2p}\,. \tag{14.15}$$

We shall explain soon how to turn this result in a matching theorem. The larger the class $\mathcal{H}$ in (14.15), the better the matching theorem one gets. It is therefore a natural question to wonder for which classes of functions a result such as Theorem 14.2.1 might be true.

Research problem 14.2.2. Consider two functions $\theta_1(x) \ge x, \theta_2(x) \ge x$. Consider the class $\mathcal{H}$ of functions $h : G \to \mathbb{R}$ such that

$$\sum \theta_1(|h(k+1, \ell) - h(k, \ell)|) + \sum \theta_2(|h(k, \ell+1) - h(k, \ell)|) \le 2^{2p}\,. \tag{14.16}$$

What are the conditions on θ_1 and θ_2 so that

$$\mathsf{E} \sup_{h \in \mathcal{H}} \Big|\sum_{i \le N} (h(U_i) - \int h \mathrm{d}\mu)\Big| \le K\sqrt{p m_0}\, 2^{2p} \tag{14.17}$$

for a constant K independent of p?

Of particular interest is the case $\theta_1(x) = x(\log(3 + x))^{1/2}$ and $\theta_2(x) = x$. A positive answer (and significant extra work) would allow to prove Theorem 14.1.3 for $\alpha = 2$.

We shall outline the proof of Theorem 14.2.1 on page 453, when we start its proof, but we first prove that it implies a matching theorem (from which Theorem 14.1.1 will easily follow). In the following statement we denote by U_i^1 and U_i^2 the components of U_i, and similarly for Z_i.

Theorem 14.2.3. *There exists a number* L_0 *with the following property. Assume that*

$$p \le \frac{m_0}{L_0} \, . \tag{14.18}$$

Consider points $(Z_i)_{i \le N}$ *in* G *and assume that for each* $(k, \ell) \in G$ *we have* $\operatorname{card}\{i \le N; Z_i = (k, \ell)\} = n(k, \ell) = N\mu(\{(k, \ell)\})$. *Consider points* $(U_i)_{i \le N}$ *in* G *and assume that (14.15) holds. Then we can find a permutation* π *of* $\{1, \ldots, N\}$ *for which*

$$\sum_{i \le N} |U_i^1 - Z_{\pi(i)}^1| \le N \, , \tag{14.19}$$

$$\forall i \le N \, , \, |U_i^2 - Z_{\pi(i)}^2| \le 1 \, . \tag{14.20}$$

It is unimportant to have N rather than LN in (14.19).

We now follow the steps which we outlined at the end of Section 4.2. In the specific situation here it cleans up matters to take advantage of the fact that a number of the points Z_i are located at a given point of G (which is what we do in (14.24) below), so we repeat the general argument from the beginning. Let us then start the proof of Theorem 14.2.3 until we run into the main difficulty.

Beginning of the proof of Theorem 14.2.3. The first steps of the proof are somehow canonical. The author cannot imagine how one could possible proceed otherwise. Proposition 4.2.1 (used with $c_{ij} = |U_i^1 - Z_j^1|$ if $|U_i^2 - Z_j^2| \le 1$ and c_{ij} very large otherwise) implies that the smallest value of the left-hand side of (14.19) among all permutations that satisfy (14.20) is given by

$$M_1 = \sup \sum_{i \le N} (w_i + w'_i) \, , \tag{14.21}$$

where the supremum is taken over all families $(w_i), (w'_i)$ for which

$$\forall i, j \le N \, , \, |U_i^2 - Z_j^2| \le 1 \Rightarrow w_i + w'_j \le |U_i^1 - Z_j^1| \, . \tag{14.22}$$

We fix families $(w_i), (w'_i)$ satisfying (14.22), and such that the supremum is attained in (14.21). We consider the function h' on G given by

$$h'(k, \ell) = \min_j \bigl\{ |k - Z_j^1| - w'_j \; ; \; |\ell - Z_j^2| \le 1 \bigr\} \, .$$

When $\tau = (k, \ell) \in G$ we define $h'(\tau) = h'(k, \ell)$. By (14.22) we have $h'(U_i) \ge w_i$ and thus (14.21) implies

$$M_1 \le \sum_{i \le N} (h'(U_i) + w'_i) \, . \tag{14.23}$$

For $(k, \ell) \in G$, we define

$$u(k, \ell) = -\frac{1}{n(k, \ell)} \sum \{ w'_i \; ; \; Z_i = (k, \ell) \} \, , \tag{14.24}$$

so that

$$\sum_{i\le N} w'_i = -\sum_G n(k,\ell)u(k,\ell)\ . \tag{14.25}$$

Consider the function h on G given by

$$h(k,\ell) = \inf\bigl\{|k-r| + u(r,s)\ ;\ |\ell - s| \le 1\bigr\}\ . \tag{14.26}$$

Given (r,s) and j with $Z_j = (r,s)$ we can find j' with $Z_{j'} = Z_j = (r,s)$ and $-w'_{j'} \le u(r,s)$. Comparing the definitions of h and h' proves that $h' \le h$. Consequently (14.23) and (14.25) imply

$$M_1 \le \sum_{i\le N} h(U_i) - \sum_{k,\ell} n(k,\ell)u(k,\ell)\ . \tag{14.27}$$

Using (14.12) and (14.27) we get

$$M_1 \le \Bigl|\sum_{i\le N}\Bigl(h(U_i) - \int h\mathrm{d}\mu\Bigr)\Bigr| - \sum_{k,\ell} n(k,\ell)(u(k,\ell) - h(k,\ell))\ . \tag{14.28}$$

To pursue now it must occur that

$$\sum_{k,\ell} n(k,\ell)(u(k,\ell)-h(k,\ell)) \text{ small } \Rightarrow \text{ the function } h \text{ behaves well,} \tag{14.29}$$

so that we may have a chance that with high probability, for all possible functions h arising in this manner, the quantity $|\sum_{i\le N}(h(U_i) - \int h\mathrm{d}\mu)|$ is small, and consequently the right-hand side of (14.28) remains bounded. The difficulty (which is generic when deducing matching theorems from Proposition 4.2.1) is to find a usable way to express that "h behaves well". In the present case, this difficulty is solved by the following result.

Proposition 14.2.4. *Consider numbers $u(k,\ell)$ for $(k,\ell) \in G = \{1,\ldots,2^p\}^2$, and consider the function h of (14.26), i.e.*

$$h(k,\ell) = \inf\bigl\{u(r,s) + |k-r|\ ;\ (r,s)\in G\,,\ |\ell - s| \le 1\bigr\}\ . \tag{14.30}$$

Then

$$\forall\, k,\ell\ ,\quad |h(k+1,\ell) - h(k,\ell)| \le 1 \tag{14.31}$$

and

$$m_0 \sum_{k,\ell} |h(k,\ell+1) - h(k,\ell)| \le L \sum_{k,\ell} n(k,\ell)(u(k,\ell) - h(k,\ell))\ . \tag{14.32}$$

So, when the left-hand side of (14.29) is small, h behaves well in the sense that $m_0\sum_{k,\ell}|h(k,\ell+1) - h(k,\ell)|$ is also small. This is of course what motivated the introduction of the class $\mathcal{H}$ and of (14.13). The proof of Proposition 14.2.4 is elementary, and is rather unrelated with the main ideas of this work. It is given at the very end of the present chapter, in Section 14.6.

End of the proof of Theorem 14.2.3. We continue from (14.28). Define

$$B = 2^{-2p} \sum |h(k, \ell+1) - h(k, \ell)| \,, \tag{14.33}$$

and $B' = B + 1$ so that $B' \geq 1$ and $h/B' \in \mathcal{H}$. Then (14.15) implies

$$\Big| \sum_{i \leq N} (h(U_i) - \int h \mathrm{d}\mu) \Big| \leq L\sqrt{pm_0}\, 2^{2p} B' \,, \tag{14.34}$$

whereas (14.32) and (14.33) imply

$$\begin{aligned} \sum n(k,\ell)(u(k,\ell) - h(k,\ell)) &\geq \frac{m_0}{L} \sum |h(k,\ell+1) - h(k,\ell)| \\ &= \frac{m_0 2^{2p}}{L} B \,. \end{aligned}$$

Combining with (14.28) and (14.34) we get, since $B' = B + 1$,

$$\begin{aligned} M_1 &\leq L\sqrt{pm_0}\, 2^{2p} B' - \frac{m_0}{L} 2^{2p} B \\ &\leq B 2^{2p} \Big(L\sqrt{pm_0} - \frac{m_0}{L} \Big) + L\sqrt{pm_0} 2^{2p} \,. \end{aligned} \tag{14.35}$$

Consequently if the constant L_0 in (14.18) is large enough, the first term is negative, so that (14.35) implies as desired that $M_1 \leq L\sqrt{pm_0} 2^{2p} \leq m_0 2^{2p} \leq N$ using (14.18) and (14.9). □

14.3 Decomposition of Functions of $\mathcal{H}$

We now outline the proof of Theorem 14.2.1. The main difficulty is to control $\gamma_2(\mathcal{H}, d_2)$, where d_2 is the distance in $\ell^2(G)$, or, equivalently the Euclidean distance on $\mathbb{R}^G$. Ideally this should be done by using a suitable sequence of functionals and Theorem 2.3.16. However, to find these functionals one needs to understand the underlying geometry, which unfortunately the author does not. The difficulty is to figure out how to use condition (14.13). (This is a good place to repeat that as of today, the Ellipsoid theorem is only instance of a clear geometric picture in this circle of ideas.) Being unable to produce the "correct" argument, we must then resort to "ad hoc" arguments. These are sufficient to prove Theorem 14.2.1, but they entail a small loss of information, which prevents using these arguments to provide the solution to Problem 14.2.2. To keep things simple we of course keep our outline of proof a bit imprecise. The basic idea is to decompose a function h of $\mathcal{H}$ as a sum $\sum_{j \geq 0} h_j$ where the functions h_j satisfy (14.14), and rather than (14.13), the condition

$$|h(k, \ell+1) - h(k, \ell)| \leq 2^j \,. \tag{14.36}$$

This is condition becomes less stringent as j grows, but we also manage to achieve that the support of h_j is small, with a cardinal not larger than $2^{-2j}\,\mathrm{card}\,G$. Proving this decomposition requires only a discrete version of the classical Vitali's covering lemma.

We have different coefficients in (14.14) and (14.36), but this is only a matter of scaling, and it is appropriate to think of a function satisfying these two conditions as a kind of Lipschitz function. How do we use that the support of h_j is small? We all know that a subset of small measure of the unit square is contained in a disjoint union of rectangles of small measure, and it is not difficult to prove a discrete version of this principle (which is given on page 459). In this manner we can decompose each function h_j as a sum of functions on small sub-rectangles of G, each of them being a kind a Lipschitz function, with the essential information that the sum of the cardinalities of these sub-rectangles is at most about $2^{-2j}\,\mathrm{card}\,G$. For each of these pieces we are basically in the situation of Theorem 4.3.2 (for a smaller value of N). We cannot use that theorem directly, but we essentially copy its proof. This is done in the central Proposition 14.3.3. The Fourier transform is then replaced by the discrete Fourier transform. Unfortunately, as in the case of Theorem 4.3.2, we also have to spend a significant part of the proof controlling lower order terms. The proofs contain a number of small technical ideas, the mastery of which the reader should eventually find most profitable. One of these ideas in particular (symmetrization) is not small at all but fundamental, and the author feels that the two different proofs he presents of Proposition 14.5.1 below are quite instructive as to the efficiency of this device.

We start now the detailed proofs. We recall the constant L_0 of (14.18). We define p as the largest integer for which

$$p2^{2p} \le \frac{N}{2L_0}\ , \tag{14.37}$$

so that $p2^{2p} \ge N/L$. We note that (14.18) is a consequence of (14.9). Moreover, taking logarithm yields $p \le L\log N$ and therefore since $2^{-p} \le L\sqrt{p}/\sqrt{N}$,

$$2^{-p} \le L\frac{\sqrt{\log N}}{\sqrt{N}}\ , \tag{14.38}$$

and also since $\log 2 < 1$ and $\exp(-2p) = (2^{-2p})^{1/\log 2}$,

$$\exp(-2p) \le \frac{L}{N}\ . \tag{14.39}$$

We consider the class $\mathcal{H}_1$ consisting of the functions $h : G \to \mathbb{R}$ such that

$$\forall k,\ell\,,\ |h(k+1,\ell) - h(k,\ell)| \le 1\ ;\ |h(k,\ell+1) - h(k,\ell)| \le 1\ . \tag{14.40}$$

Given an integer $j \ge 2$, for a number $V > 0$ we consider the class $\mathcal{H}_j(V)$ of functions $h : G \to \mathbb{R}$ such that

$$\forall k, \ell\,,\ |h(k+1,\ell)-h(k,\ell)| \le 1\,,\ |h(k,\ell+1)-h(k,\ell)| \le 2^j \tag{14.41}$$

$$\operatorname{card}\{(k,\ell)\in G\ ;\ h(k,\ell)\neq 0\} \le V\ . \tag{14.42}$$

Proposition 14.3.1. *If $h \in \mathcal{H}$ we can decompose*

$$h = \sum_{j\ge 1} h_j \text{ where } h_1 \in L\mathcal{H}_1 \text{ and } h_j \in L\mathcal{H}_j(2^{2p-j}) \text{ for } j \ge 2\ . \tag{14.43}$$

Thus, we can decompose h as a sum of terms that satisfy simple conditions, and that will be studied separately.

We will denote by I an interval of $\{1,\ldots,2^p\}$, that is a set of the type

$$I = \{k\ ;\ k_1 \le k \le k_2\}\ .$$

Lemma 14.3.2. *Consider a map $f : \{1,\ldots,2^p\} \to \mathbb{R}^+$, a number $a > 0$ and*

$$A = \Big\{k\ ;\ \exists I\,,\ k\in I\,,\ \sum_{k'\in I} f(k') \ge a\operatorname{card} I\Big\}\ .$$

Then

$$\operatorname{card} A \le \frac{L}{a}\sum_{k\in A} f(k)\ .$$

Proof. This uses a discrete version of the classical Vitali covering theorem (with the same proof). Namely, a family $\mathcal{I}$ of intervals contains a disjoint family $\mathcal{I}'$ such that

$$\operatorname*{card}_{I\in\mathcal{I}} \bigcup I \le L \operatorname*{card}_{I\in\mathcal{I}'}\bigcup I = L\sum_{I\in\mathcal{I}'} \operatorname{card} I\ .$$

We use this for $\mathcal{I} = \{I; \sum_{k'\in I} f(k') \ge a\operatorname{card} I\}$, so that $A = \bigcup_{I\in\mathcal{I}} I$ and $\operatorname{card} A \le L\sum_{I\in\mathcal{I}'}\operatorname{card} I$. Since $\sum_{k'\in I} f(k') \ge a\operatorname{card} I$ for $I\in\mathcal{I}'$, and since the intervals of $\mathcal{I}'$ are disjoint and contained in A, we have $a\sum_{I\in\mathcal{I}'}\operatorname{card} I \le \sum_{k'\in A} f(k')$. ☐

Proof of Proposition 14.3.1. We consider $h\in\mathcal{H}$, and for $j\ge 2$ we define

$$B(j) = \Big\{(k,\ell)\ ;\ \exists I\,,\ \ell\in I\,,\ \sum_{\ell'\in I} |h(k,\ell'+1)-h(k,\ell')| \ge 2^j\operatorname{card} I\Big\}\ .$$

We claim that when $r, s, \ell \le 2^p$, then

$$(r,s)\notin B(j) \Rightarrow |h(r,\ell)-h(r,s)| \le 2^j|\ell - s|\ . \tag{14.44}$$

To see this, assuming for specificity that $s < \ell$, we note that

$$|h(r,\ell)-h(r,s)| \le \sum_{\ell'\in I} |h(r,\ell'+1)-h(r,\ell')| < 2^j\operatorname{card} I$$

where $I = \{s, s+1, \ldots, \ell-1\}$, and where the last inequality follows from the fact that $s \in I$ and $(r,s) \notin B(j)$.

Using Lemma 14.3.2 for each ℓ and summing over ℓ we obtain

$$\operatorname{card} B(j) \le \frac{L}{2^j} \sum_{(k,\ell)\in B(j)} |h(k,\ell+1) - h(k,\ell)| \;.$$

Now (14.13) implies $\sum_{k,\ell} |h(k,\ell+1) - h(k,\ell)| \le 2^{2p}$, and therefore we get

$$\operatorname{card} B(j) \le L_1 2^{2p-j} \;. \tag{14.45}$$

We consider the smallest integer j_0 such that $L_1 2^{-j_0} < 1/4$, so that $L_1 \le 2^{j_0-2}$ and hence for $j \ge j_0$ we have

$$\operatorname{card} B(j) \le 2^{2p-j+j_0-2} \;, \tag{14.46}$$

and in particular $B(j) \ne G$. For $j \ge j_0$ we define

$$g_j(k,\ell) = \min\bigl\{h(r,s) + |k-r| + 2^j|\ell - s| \;;\; (r,s) \notin B(j)\bigr\} \;.$$

The idea here is that g_j is a regularization of h. The larger j, the better g_j approximates h, but this comes at the price that the larger j, the less regular g_j is. We will simply use these approximations to write

$$h = (h - g_{j_0}) + (g_{j_0} - g_{j_0+1}) + \cdots$$

to obtain the desired decomposition (14.43).

It is obvious that for $(k,\ell) \notin B(j)$ we have $g_j(k,\ell) \le h(k,\ell)$, and that

$$|g_j(k+1,\ell) - g_j(k,\ell)| \le 1 \tag{14.47}$$

$$|g_j(k,\ell+1) - g_j(k,\ell)| \le 2^j \;, \tag{14.48}$$

since g_j is the minimum of functions that satisfy the same properties. Consider $(r,s) \notin B(j)$. Then (14.44) yields

$$|h(r,\ell) - h(r,s)| \le 2^j|\ell - s| \;,$$

while the first part of (14.41) yields

$$|h(r,\ell) - h(k,\ell)| \le |k - r| \;,$$

and thus we have proved that

$$(r,s) \notin B(j) \Rightarrow |h(k,\ell) - h(r,s)| \le |k-r| + 2^j|\ell - s| \;. \tag{14.49}$$

This implies that $g_j(k,\ell) \ge h(k,\ell)$. Consequently, since we already observed that $g_j(k,\ell) \le h(k,\ell)$ for $(k,\ell) \notin B(j)$, we have proved that

$$(k,\ell) \notin B(j) \Rightarrow g_j(k,\ell) = h(k,\ell) \; . \tag{14.50}$$

We define $h_1 = h - g_{j_0}$, and we prove first that $h_1 \in L\mathcal{H}_1$. We want to bound $|h(k,\ell) - h(r,s)|$. From (14.50) this quantity is 0 if both (k,ℓ) and (r,s) belong to $B(1)$, so we may assume that say $(r,s) \notin B(1)$. Combining (14.47) and (14.48) as in the proof of (14.49) we obtain that $|g_1(k,\ell) - g_1(r,s)| \le |k-r| + 2|\ell - s|$, and combining with (14.49) for $j = 1$ we obtain that $|h_1(k,\ell) - h_1(r,s)| \le 2|k-r| + 4|\ell - s|$, and we have proved that $h_1 \in 4\mathcal{H}_1$. For $j > 1$ we define $h_j = g_{j+j_0-2} - g_{j+j_0-1}$ so that, by (14.50), and since $B(j+j_0-1) \subset B(j+j_0-2)$,

$$h_j(k,\ell) \neq 0 \Rightarrow (k,\ell) \in B(j+j_0-2) \; ,$$

and thus from (14.46) that

$$\operatorname{card}\{(k,\ell) \; ; \; h_j(k,\ell) \neq 0\} \le 2^{2p-j} \; .$$

Combining with (14.47) and (14.48) we obtain $h_j \in L\mathcal{H}_j(2^{2p-j})$.

Now for $j > 2p$ we have $B(j) = \emptyset$ (since for each k and ℓ we have $|h(k,\ell+1) - h(k,\ell)| \le 2^{2p}$ by (14.13)), so that then $g_j = h$ from (14.50). Consequently $h_j = 0$ for large j and thus $h = \sum_{j\ge1} h_j$. The proof is complete. □

The central step in the proof of Theorem 14.2.1 is as follows.

Proposition 14.3.3. *Consider* $1 \le k_1 \le k_2 \le 2^p$, $1 \le \ell_1 \le \ell_2 \le 2^p$ *and* $R = \{k_1, \ldots, k_2\} \times \{\ell_1, \ldots, \ell_2\}$. *Assume that*

$$\ell_2 - \ell_1 + 1 = 2^{-j}(k_2 - k_1 + 1) \; . \tag{14.51}$$

Consider independent r.v.s U_i *valued in* G, *of law* μ. *Then, with probability at least* $1 - L\exp(-50p)$, *the following occurs. Consider any function* $h : G \to \mathbb{R}$, *and assume that*

$$h(k,\ell) = 0 \quad \text{unless} \quad (k,\ell) \in R \; . \tag{14.52}$$

$$(k,\ell)\, , \, (k+1,\ell) \in R \Rightarrow |h(k+1,\ell) - h(k,\ell)| \le 1 \tag{14.53}$$

$$(k,\ell)\, , \, (k,\ell+1) \in R \Rightarrow |h(k,\ell+1) - h(k,\ell)| \le 2^j \tag{14.54}$$

$$\forall (k,\ell) \in R \; , \; |h(k,\ell)| \le 2(k_2 - k_1) \; . \tag{14.55}$$

Then

$$\Big|\sum_{i\le N} \Big(h(U_i) - \int h \mathrm{d}\mu\Big)\Big| \le L 2^{j/2} \sqrt{p m_0} \operatorname{card} R \; , \tag{14.56}$$

where m_0 *is as in (14.8) and (14.9).*

In order to understand the absolute necessity of condition (14.55) in this theorem, the reader should consider the case where $j = 0$, $k_1 = k_2$ and $\ell_1 = \ell_2$. In that case h satisfies (14.52) to (14.54) if (and only if) $h(k,\ell) = 0$ unless $(k,\ell) = (k_1,\ell_1) = (k_2,\ell_2)$. The function h is then determined by the value a it takes at the point (k_1,ℓ_1) and the left-hand side of (14.56) is $|a||\operatorname{card}\{i \le N; U_i = (k_0,\ell_0)\} - N\mu(\{(k,\ell)\})|$, so that (14.56) holds for every value of a only when $N\mu(\{(k,\ell)\}) = \operatorname{card}\{i \le N, U_i = (k_0,\ell_0)\}$, and this is not true in general.

We will also use the following, in the same spirit as Proposition 14.3.3, but very much easier.

Proposition 14.3.4. *Consider* $1 \le k_1 \le k_2 \le 2^p$, $1 \le \ell_0 \le 2^p$ *and* $R = \{k_1,\dots,k_2\} \times \{\ell_0\}$. *Assume that*

$$k_2 - k_1 + 1 \le 2^j \ . \tag{14.57}$$

Consider independent r.v.s U_i *valued in* G*, of law* μ*. Then, with probability at least* $1 - L\exp(-50p)$*, the following occurs. Consider any function* $h : G \to \mathbb{R}$*, and assume that*

$$h(k,\ell) = 0 \quad \text{unless} \quad (k,\ell) \in R \ . \tag{14.58}$$

$$(k,\ell)\,,\ (k+1,\ell) \in R \Rightarrow |h(k+1,\ell) - h(k,\ell)| \le 1 \tag{14.59}$$

$$\forall (k,\ell) \in R\ ,\ |h(k,\ell)| \le 2(k_2 - k_1)\ . \tag{14.60}$$

Then

$$\Big|\sum_{i\le N}(h(U_i) - \int h \mathrm{d}\mu)\Big| \le L\sqrt{m_0}(k_2 - k_1 + 1)^{3/2} \le L2^{j/2}\sqrt{m_0}\operatorname{card} R\ . \tag{14.61}$$

Proof of Theorem 14.2.1. In Proposition 14.3.3 there are (crudely) at most 2^{4p} choices for the quadruplet (k_1,k_2,ℓ_1,ℓ_2). Thus with probability at least $1 - L\exp(-46p)$, the conclusions of Proposition 14.3.3 are true for all values of k_1, k_2, ℓ_1 and ℓ_2, and the conclusions of Proposition 14.3.4 hold for all values of k_1, k_2 and ℓ_0. We assume that this is the case in the remainder of the proof.

First we show that if $h \in \mathcal{H}_1$ then

$$\Big|\sum_{i\le N}(h(U_i) - \int h \mathrm{d}\mu)\Big| \le L\sqrt{pm_0}\,2^{2p}\ . \tag{14.62}$$

The proof relies on the case $k_1 = \ell_1 = 1$ and $k_2 = \ell_2 = 2^p$ of Proposition 14.3.4. The function h satisfies (14.53) and (14.54) and hence $|h(k,\ell) - h(1,1)| \le 2^{p+1}$ for each $(k,\ell) \in G$. Consequently the function $h^*(k,\ell) = h(k,\ell) - h(1,1)$ satisfies (14.55), (14.53) and (14.54). Therefore h^* satisfies (14.62) and consequently this is also the case for h.

Using now the decomposition (14.43) of a function in $\mathcal{H}$ provided by Proposition 14.3.1, it suffices to show that if $h \in \mathcal{H}_j(2^{2p-j})$ then

$$\Big|\sum_{i\le N}(h(U_i)-\int h\mathrm{d}\mu)\Big|\le L\sqrt{pm_0}\,2^{2p-j/2}\,. \tag{14.63}$$

We think of j as being fixed once and for all. The idea is to use (14.56) for the functions $h\mathbf{1}_R$ where R is a suitable rectangle, and to recover (14.63) by summation of the corresponding inequalities over a suitable disjoint family of rectangles. The most difficult point is to ensure that the functions $h\mathbf{1}_R$ satisfy (14.55). In fact, rather than (14.55) we shall prove that

$$\forall(k,\ell)\in R\,,\ |h(k,\ell)|\le L(k_2-k_1)\,, \tag{14.64}$$

which suffices by homogeneity.

For $j\le q\le p$ we consider the partition $\mathcal{D}(q)$ of G consisting of the sets of the type

$$\{a2^q+1,\dots,(a+1)2^q\}\times\{b2^{q-j}+1,\dots,(b+1)2^{q-j}\}\,, \tag{14.65}$$

where a and b are integers with $0\le a<2^{p-q}$ and $0\le b<2^{p-q+j}$. For $3\le q\le j$, we define $\mathcal{D}(q)$ as the partition consisting of the sets of the type

$$\{a2^q+1,\dots,(a+1)2^q\}\times\{b\} \tag{14.66}$$

where $0\le a<2^{p-q}$ and $1\le b\le 2^p$.

We observe that if $q'>q$, $R'\in\mathcal{D}(q')$ and $R\in\mathcal{D}(q)$, then either $R\subset R'$ or $R\cap R'=\emptyset$.

Fixing a function $h\in\mathcal{H}_j(2^{2p-j})$, we consider the set $C=\{(k,\ell);h(k,\ell)\ne 0\}$ so $\operatorname{card}C\le 2^{2p-j}$. We proceed to the following construction. Keeping in mind that the sequence $(\mathcal{D}(q))$ of partitions increases, so that $\mathcal{D}(p)$ consists of the largest rectangles, we first consider the set $U(p)$ that is the union of all rectangles $R\in\mathcal{D}(p)$ such that

$$\operatorname{card}(R\cap C)\ge\frac{1}{8}\operatorname{card}R\,. \tag{14.67}$$

Then we consider the union $U(p-1)$ of all the rectangles $R\in\mathcal{D}(p-1)$ that are not contained in $U(p)$ and that satisfy (14.67), and we continue in this manner until we construct $U(3)$. Since the sets $U(p)\,,\,\dots\,,\,U(3)$ are disjoint and each is a union of disjoint sets satisfying (14.67), we get

$$\sum_{3\le q\le p}\operatorname{card}U(q)\le 8\operatorname{card}C\le 2^{2p-j+3}\,. \tag{14.68}$$

Moreover

$$C\subset\sum_{1\le q\le p}U(q)\,. \tag{14.69}$$

This is simply because if $(k,\ell)\in C$ and $(k,\ell)\in R\in\mathcal{D}(3)$ then if $(k,\ell)\notin\bigcup_{q\ge 4}U(q)$ we have $R\subset U(3)$ since (14.67) holds because $\operatorname{card}R=8$. We also note that

$$R \in \mathcal{D}(q)\,,\; q \le p-1\,,\; R \subset U(q) \Rightarrow \operatorname{card}(R \cap C) \le \frac{1}{2}\operatorname{card} R\,. \tag{14.70}$$

Indeed if $R' \supset R$ and $R' \in \mathcal{D}(q+1)$, then $\operatorname{card} R' \le 4\operatorname{card} R$. Since $R \subset U(q)$ we have $R' \not\subset U(q+1)$, so that

$$\operatorname{card}(R \cap C) \le \operatorname{card}(R' \cap C) \le \frac{1}{8}\operatorname{card} R' \le \frac{1}{2}\operatorname{card} R\,.$$

Now (14.69) implies

$$h = \sum h\mathbf{1}_R\,, \tag{14.71}$$

where the summation is over $3 \le q \le p$, $R \in \mathcal{D}(q)$ and $R \subset U(q)$. Writing $R = \{k_1, \ldots, k_2\} \times \{\ell_1, \ldots, \ell_2\}$ as in Proposition 14.3.3, we observe that by construction (14.51) holds for $q \ge j$, that (14.57) holds for $3 \le q \le j$, and that the function $h\mathbf{1}_R$ satisfies (14.52) to (14.54).

We turn to the proof of (14.64). We start by the typical case, $R \in \mathcal{D}(q)\,,\; 3 \le q < p$. Then (14.70) implies that there exists $(k_0, \ell_0) \in R$ with $h(k_0, \ell_0) = 0$, and (14.41) implies, using also (14.51),

$$\begin{aligned}|h(k,\ell)| = |h(k,\ell) - h(k_0,\ell_0)| &\le |h(k,\ell) - h(k,\ell_0)| + |h(k,\ell_0) - h(k_0,\ell_0)| \\ &\le 2^j|\ell - \ell_0| + |k - k_0| \le 2(k_2 - k_1)\,,\end{aligned}$$

and this proves (14.64).

Next we consider the case $q = p$ so that $R \in \mathcal{D}(p)$ and $R = \{1, \ldots, 2^p\} \times \{b2^{p-j}+1, \ldots, (b+1)2^{p-j}+1\}$. Given an integer r, define

$$R' = G \cap (\{1, \ldots, 2^p\} \times \{b2^{p-j}+1-r, \ldots, (b+1)2^{p-j}+1+r\})\,.$$

Then, for $r \le 2^p$, we have $\operatorname{card} R' \ge 2^p r/L$, so that if $2^p r/L > 2^{2p-j}$, R' contains a point (k, ℓ') with $h(k, \ell') = 0$. Then R contains a point (k, ℓ) with $|\ell - \ell'| \le r$, so that the second part of (14.41) implies

$$|h(k,\ell)| \le r2^j\,.$$

Assuming that we choose r as small as possible with $2^p r/L > 2^{2p-j}$, we then have

$$|h(k,\ell)| \le L2^{2p-j}2^{-p}2^j \le L2^p\,,$$

and (14.41) shows that this remains true for each point (k, ℓ) of R, completing the proof of (14.64).

Consequently for $R \in \mathcal{D}(q)\,,\; R \subset U(q)$ and $j \le q \le p$ we can use (14.56), which implies

$$\Big|\sum_{i \le N} (h\mathbf{1}_R(U_i) - \int h\mathbf{1}_R \mathrm{d}\mu)\Big| \le L\sqrt{pm_0}\,2^{j/2}\operatorname{card} R\,. \tag{14.72}$$

Moreover for $3 \le q \le j$ this inequality remains true from (14.61). Recalling (14.71), summation of these inequalities over $R \in \mathcal{D}(q)\,,\; R \subset U(q)$ yields (14.63) and completes the proof. □

14.4 Discrete Fourier Transform

We turn to the proof of Proposition 14.3.3. This proposition is a close cousin of Theorem 4.3.2. There is a "main contribution" that comes from the functions h for which

$$\forall k, \ell\,,\ h(k_1, \ell) = h(k_2, \ell)\ ;\ h(k, \ell_1) = h(k, \ell_2) \tag{14.73}$$

and there are "second-order contributions". To bring this out, one uses a discrete version of the decomposition (4.77). Since this is really tedious, we shall not give all the details, which present no difficulty.

The following is closely related to Proposition 4.3.8 and its relevance to Theorem 14.2.1 should be obvious.

Proposition 14.4.1. *Consider integers $q_1, q_2 \le 2^p$, and the class $\mathcal{G}$ of functions $h : G' = \{0, \ldots, q_1\} \times \{0, \ldots, q_2\} \to \mathbb{R}$ that satisfy*

$$\forall \ell \le q_2\,,\ h(0, \ell) = h(q_1, \ell) \tag{14.74}$$

$$\forall k \le q_1\,,\ h(k, 0) = h(k, q_2)\,, \tag{14.75}$$

as well as

$$\forall k, \ell\,,\ |h(k+1, \ell) - h(k, \ell)| \le 1\ ;\ |h(k, \ell+1) - h(k, \ell)| \le 2^j \tag{14.76}$$

$$\forall k, \ell\,, |h(k, \ell)| \le 2q_1\ . \tag{14.77}$$

Assume that

$$2^{-j-1} q_1 \le q_2 \le 2^{-j+1} q_1\ . \tag{14.78}$$

Then

$$\gamma_2(\mathcal{G}, d) \le L\sqrt{p}\, 2^{j/2} q_1 q_2 \tag{14.79}$$

and

$$e_n(\mathcal{G}, d) \le 2^{-n/2} 2^{j/2} q_1 q_2\ , \tag{14.80}$$

where d is the Euclidean distance on $\mathbb{R}^{G'}$.

Proof. The idea is (again) to use the Fourier transform to reduce to the study of certain ellipsoids. Consider the groups $H_1 = \mathbb{Z}/q_1\mathbb{Z}$, $H_2 = \mathbb{Z}/q_2\mathbb{Z}$, and their product $H := H_1 \otimes H_2$. Consider the canonical map ψ from G' to H. Given a function h on H the function $h \circ \psi$ on G' satisfies (14.74) and (14.75).

Consider the class $\mathcal{G}^*$ of functions h from $H = H_1 \otimes H_2$ to $\mathbb{R}$ that satisfy (14.76) and (14.77), where now $k \in H_1, \ell \in H_2$ and where 1 denotes the image of $1 \in \mathbb{Z}$ in either H_1 or H_2. Then $\mathcal{G}$ is exactly the class of functions $h \circ \psi$ for $h \in \mathcal{G}^*$. For a function h on H let us denote by $\|h\|_*$ its norm in $L^2(\mu)$, where μ is the uniform probability measure of H, and by d_* the corresponding distance. It is straightforward that for two functions h_1 and h_2 on H one has

$$\|h_1 \circ \psi - h_2 \circ \psi\|_2^2 \le 2q_1q_2\|h_1 - h_2\|_*^2 \,,$$

where on the left-hand side the norm is for the Euclidean distance. Therefore, it suffices to prove the estimates

$$\gamma_2(\mathcal{G}^*, d_*) \le L2^{j/2}\sqrt{pq_1q_2} \tag{14.81}$$

and

$$e_n(\mathcal{G}^*, d_*) \le 2^{-n/2}2^{j/2}\sqrt{q_1q_2} \,. \tag{14.82}$$

For this we use the Fourier transform in the group $H = H_1 \otimes H_2$. For integers r_1 and r_2, we define

$$c_{r_1r_2}(h) = \frac{1}{q_1q_2}\sum_{(k,\ell)\in H_1\otimes H_2} \exp\Big(2i\pi\big(\frac{r_1}{q_1}k + \frac{r_2}{q_2}\ell\big)\Big)h(k,\ell)\,, \tag{14.83}$$

and we have the Plancherel formula

$$\|h\|_*^2 = \sum_{0\le r_1<q_1, 0\le r_2<q_2} |c_{r_1r_2}(h)|^2 \,. \tag{14.84}$$

Changing k into $k+1$ in (14.83) we get

$$c_{r_1r_2}(h) = \exp\Big(2i\pi\frac{r_1}{q_1}\Big)\frac{1}{q_1q_2}\sum_{H_1\otimes H_2} \exp\Big(2i\pi\big(\frac{r_1}{q_1}k + \frac{r_2}{q_2}\ell\big)\Big)h(k+1,\ell)$$

and thus

$$\begin{aligned}&\Big(\exp\big(-2i\pi\frac{r_1}{q_1}\big)-1\Big)c_{r_1r_2}(h)\\ &= \frac{1}{q_1q_2}\sum_{(k,\ell)\in H_1\otimes H_2} \exp\Big(2i\pi\big(\frac{r_1}{q_1}k + \frac{r_2}{q_2}\ell\big)\Big)(h(k+1,\ell) - h(k,\ell))\,.\end{aligned}$$

Using (14.84) for the function $h'(k,\ell) = h(k+1,\ell) - h(k,\ell)$ and the first part of (14.76) we get

$$\sum_{0\le r_1<q_1, 0\le r_2<q_2} \Big|1 - \exp\Big(-2i\pi\frac{r_1}{q_1}\Big)\Big|^2 |c_{r_1r_2}(h)|^2 = \|h'\|_*^2 \le 1\,.$$

We now use that for $0 \le r_1 < q_1$ we have

$$\Big|1 - \exp\Big(-2i\pi\frac{r_1}{q_1}\Big)\Big| \ge \frac{1}{Lq_1}\min(r_1, q_1 - r_1)$$

to get

$$\sum_{0\le r_1<q_1, 0\le r_2<q_2} \frac{1}{Lq_1^2}\min(r_1, q_1-r_1)^2\big|c_{r_1r_2}(h)\big|^2 \le 1\,.$$

Proceeding in the same manner with the second variable, and combining with the previous inequality we get

$$\sum_{0\le r_1<q_1, 0\le r_2<q_2} d_{r_1,r_2}|c_{r_1r_2}(h)|^2 \le 1 \; , \tag{14.85}$$

where

$$d_{r_1,r_2} = \frac{1}{Lq_1^2}\min(r_1, q_1 - r_1)^2 + \frac{1}{L2^{2j}q_2^2}\min(r_2, q_2 - r_2)^2 \; .$$

Since $d_{0,0} = 0$ condition (14.85) does not control $c_{0,0}$, but (14.77) shows that $|c_{0,0}| \le 2q_1$ and thus $c_{0,0}^2/8b_{0,0}^2 \le 1/2$. Therefore, since $2^{2j}q_2^2 \le Lq_1^2$ by (14.78), we have

$$\sum_{0\le r_1<q_1, 0\le r_2<q_2} \frac{|c_{r_1r_2}(h)|^2}{b_{r_1,r_2}^2} \le 1 \; , \tag{14.86}$$

were $b_{0,0}^2 = 8q_1^2$ while for $(r_1, r_2) \ne (0, 0)$,

$$\frac{1}{b_{r_1,r_2}^2} = \frac{1}{Lq_1^2}\min(r_1, q_1 - r_1)^2 + \frac{1}{Lq_1^2}\min(r_2, q_2 - r_2)^2 \; .$$

The Plancherel formula (14.84) describes $\mathcal{G}^*$ as isometric to a subset of the ellipsoid $\mathcal{E}$ in $\mathbb{C}^{q_1q_2}$ defined by (14.86) so that it suffices to prove the estimates (14.81) and (14.82) for $\mathcal{E}$ rather than $\mathcal{G}^*$. The key point is the following: for each $t > 0$,

$$\mathrm{card}\{(r_1, r_2)\; ;\; b_{r_1,r_2}^2 \ge t\} \le \frac{Lq_1^2}{t} \; . \tag{14.87}$$

To prove this we observe first that $b_{r_1,r_2}^2 \le L_0q_1^2$ for all values of r_1 and r_2, so that it suffices to prove (14.87) when $t < L_0q_1^2$, since the left-hand side is 0 otherwise. When $b_{r_1,r_2}^2 \ge t$ then

$$\min(r_1, q_1 - r_1) \le \frac{Lq_1}{\sqrt{t}} \text{ and } \min(r_2, q_1 - r_2) \le \frac{Lq_1}{\sqrt{t}} \; .$$

There are at most $L(1+q_1/\sqrt{t})$ choices for both r_1 and r_2. But since $t < L_0q_1^2$ and hence $1 \le L_0^{1/2}q_1/\sqrt{t}$, we have

$$1 + \frac{q_1}{\sqrt{t}} \le (1 + L_0^{1/2})\frac{q_1}{\sqrt{t}} \; ,$$

and thus there are at most $L'q_1/\sqrt{t}$ choices for each of r_1 and r_2, and this proves (14.87).

It follows from (14.87) that if we reorder the numbers b_{r_1,r_2}^2 as a non-increasing sequence $(a_i^2)_{1\le i\le q_1q_2}$ then for each $i \ge 1$ (using (14.78) in the second inequality) we have

$$a_i^2 \le \frac{Lq_1^2}{i} \le \frac{L2^j q_1 q_2}{i} \;, \tag{14.88}$$

and moreover the ellipsoid $\mathcal{E}$ defined by (14.86) is isometric to the ellipsoid of $\mathbb{C}^{q_1 q_2}$ given by

$$\sum_{1 \le i \le q_1 q_2} \frac{|x_i|^2}{a_i^2} \le 1 \;.$$

We leave it to the reader to check that in the case of *complex* ellipsoids (2.118) yields

$$e_{n+3}(\mathcal{E}) \le 3 \max_{k \le n} a_{2^{k+1}} 2^{k-n} \;.$$

Now (14.88) implies that $a_{2^k} \le L2^{-k/2} 2^{j/2} \sqrt{q_1 q_2}$ and therefore $e_{n+3}(\mathcal{E}) \le L2^{j/2}\sqrt{q_1 q_2} \max_{k \le n} 2^{k/2-n}$ and this proves (14.82). Moreover since $q_1 q_2 \le 2^{2p}$, we have $\sum_{i \le q_1 q_2} 1/i \le Lp$ and (14.88) implies

$$\sum_{1 \le i \le q_1 q_2} a_i^2 \le Lp 2^j q_1 q_2 \;,$$

and (2.115) proves (14.81). □

14.5 Main Estimates

The following formally states the control of the main contribution in Proposition 14.3.3.

Proposition 14.5.1. *Consider* $1 \le k_1 \le k_2 \le 2^p$, $1 \le \ell_1 \le \ell_2 \le 2^p$ *and assume that (14.51) holds. Let* $R = \{k_1, \ldots, k_2\} \times \{\ell_1, \ldots, \ell_2\}$*. Consider independent r.v.* U_i *valued in* G*, with* $\mathsf{P}(U_i = (k,\ell)) = n(k,\ell)/N$*. Then, with probability at least* $1 - L\exp(-50p)$*, the following occurs. Consider any function* $h : G \to \mathbb{R}$ *that satisfies (14.73), i.e.*

$$\forall k, \ell \;,\; h(k_1, \ell) = h(k_2, \ell) \;;\; h(k, \ell_1) = h(k, \ell_2) \;. \tag{14.73}$$

and assume that moreover

$$h(k,\ell) = 0 \quad \text{unless} \quad (k,\ell) \in R \;, \tag{14.89}$$

$$(k,\ell)\,,\,(k+1,\ell) \in R \Rightarrow |h(k+1,\ell) - h(k,\ell)| \le 1 \;, \tag{14.90}$$

$$(k,\ell)\,,\,(k,\ell+1) \in R \Rightarrow |h(k,\ell+1) - h(k,\ell)| \le 2^j \;, \tag{14.91}$$

$$\forall (k,\ell) \in R\,,\, |h(k,\ell)| \le 2(k_2 - k_1 + 1) \;. \tag{14.92}$$

Then

$$\Big|\sum_{i \le N} (h(U_i) - \int h \mathrm{d}\mu)\Big| \le L2^{j/2} \sqrt{pm_0}\, \operatorname{card} R \;. \tag{14.93}$$

We shall present two proofs of this proposition. The first one relies on Bernstein's inequality. Is pretty straightforward, but the control in the supremum norm becomes a bit of a nuisance, and although all the required estimates are essentially trivial, putting all of them together takes a page or so of calculations. After this first proof, we shall present a second proof using a fundamental method called symmetrization. This method replaces (most of) the use of Bernstein's inequality by the use of the subgaussian inequality. The comparison of both proofs is quite instructive.

Lemma 14.5.2. *With probability* $\geq 1 - \exp(-100p)$ *we have*

$$\forall\,(k,\ell) \in G\ ,\ \operatorname{card}\{i \leq N\ ;\ U_i = (k,\ell)\} \leq Lm_0\ . \tag{14.94}$$

Proof. We are here talking about the tails of the Binomial law. The r.v.s $Y_i = \mathbf{1}_{\{U_i=(k,\ell)\}} - \mu(\{(k,\ell)\})$ satisfy $\mathsf{E}Y_i = 0$, $|Y_i| \leq 1$ and $\mathsf{E}Y_i^2 \leq \mu(\{(k,\ell)\})$, where

$$\mu(\{(k,\ell)\}) = \frac{n(k,l)}{N} \leq 2m_0\ . \tag{14.95}$$

Thus Bernstein's inequality (4.59) implies

$$\mathsf{P}\Big(\sum_{i\leq N} Y_i \geq u\Big) \leq \exp\Big(-\frac{1}{L}\min\Big(\frac{u^2}{2m_0}, u\Big)\Big)\ .$$

Since $m_0 \geq p$, the choice $u = Lm_0$ implies that with probability at least $1 - \exp(-102p)$ we have $\sum_{i\leq N} Y_i \leq Lm_0$, and using (14.95) again (14.94) holds for a given value of (k,ℓ), and the result follows. □

First proof of Proposition 14.5.1. We denote by $\mathcal{G}$ the class of functions on G that satisfy conditions (14.89) to (14.92). To lighten notation we write

$$s = k_2 - k_1 + 1\ .$$

Consider the distance δ on R given by

$$\delta((k,\ell),(k',\ell')) = |k - k'| + 2^j|\ell - \ell'|\ .$$

Then the functions of $\mathcal{G}$ are 1-Lipschitz for this distance. We leave it to the reader to check that the space (R,δ) satisfies (4.67) with $\alpha = 2$ and $B = Ls$, so that (4.69) implies $e_n(\mathcal{G}, d_\infty) \leq Ls2^{-n/2}$. Consider then the largest integer m such that

$$2^m \leq pm_0 \operatorname{card} R \tag{14.96}$$

and a subset T of $\mathcal{G}$ with $\operatorname{card} T \leq N_m$ and

$$\forall t \in \mathcal{G}\ ,\ d_\infty(t,T) \leq Ls2^{-m/2}\ . \tag{14.97}$$

The core of the proof is to deduce from Theorem 2.2.28 that with probability $\geq 1 - \exp(-100p)$ we have

$$\sup_{h\in T}\Big|\sum_{i\le N}\Big(h(U_i)-\int h\mathrm{d}\mu\Big)\Big|\le L2^{j/2}\sqrt{pm_0}\ \mathrm{card}\,R\ . \tag{14.98}$$

A first observation is that (14.8) implies

$$\int h^2\mathrm{d}\mu=\sum_{k,\ell}\frac{n(k,\ell)}{N}h^2(k,\ell)\le\frac{2m_0}{N}\sum_{k,\ell}h^2(k,\ell)\ .$$

Consequently, given h and h', the r.v.s $Z_i=(h(U_i)-\int h\mathrm{d}\mu)-(h'(U_i)-\int h'\mathrm{d}\mu)$ satisfy $\mathsf{E}Z_i=0$, $\mathsf{E}Z_i^2\le 2m_0d(h,h')^2/N$ and $|Z_i|\le 2d_\infty(h,h')$. We then deduce from Bernstein's inequality (4.59) that the r.v.s $Y_h=\sum_{i\le N}(h(U_i)-\int h\mathrm{d}\mu)$ satisfy the condition

$$\forall u>0,\mathsf{P}(|Y_h-Y_{h'}|\ge u)\le 2\exp\Big(-\min\Big(\frac{u^2}{Lm_0d(h,h')^2},\frac{u}{Ld_\infty(h,h')}\Big)\Big)\ .$$

In other words, the condition (2.50) of Theorem 2.2.28 is satisfied with $d_1:=Ld_\infty$ and $d_2=L\sqrt{m_0}d$. In order to apply this theorem we estimate the various quantities it involves.

Since $e_n(\mathcal{G},d_\infty)\le Ls2^{-n/2}$, we have $e_n(T,d_\infty)\le Ls2^{-n/2}$ and moreover $e_n(T,d_\infty)=0$ for $n\ge m$, so that Corollary 2.3.2 implies $\gamma_1(T,d_\infty)\le Ls2^{m/2}$. Now, since $d_1=Ld_\infty$, we have $\gamma_2(T,d_1)\le L\gamma_2(T,d_\infty)$ and $e_n(T,d_1)\le Le_n(T,d_\infty)$ so that

$$\gamma_1(T,d_1)\le Ls2^{m/2}\ ;\ D_1=\sum_{n\ge 0}e_n(T,d_1)\le Ls\ . \tag{14.99}$$

Proposition 14.4.1 implies

$$\gamma_2(T,d)\le L2^{j/2}\sqrt{p}\ \mathrm{card}\,R$$
$$\forall n\ge 0\,,\ e_n(T,d)\le L2^{j/2}2^{-n/2}\ \mathrm{card}\,R\,.$$

Thus, and since $d_2=L\sqrt{m_0}d$,

$$\gamma_2(T,d_2)\le L2^{j/2}\sqrt{pm_0}\ \mathrm{card}\,R\ ;\ D_2=\sum_{n\ge 0}e_n(T,d_2)\le L2^{j/2}\sqrt{m_0}\ \mathrm{card}\,R\ . \tag{14.100}$$

Consequently, using (14.99) and (14.100), Theorem 2.2.28 implies

$$\mathsf{P}\Big(\sup_{h,h'\in T}|Y_h-Y_{h'}|\ge LU\Big)\le\exp(-\min(u_2^2,u_1))\ , \tag{14.101}$$

where

$$U=s2^{m/2}+2^{j/2}\sqrt{pm_0}\ \mathrm{card}\,R+u_1s+u_22^{j/2}\sqrt{m_0}\ \mathrm{card}\,R\ .$$

We choose $u_1=100p$ and $u_2=10\sqrt{p}$, so that the right-hand side of (14.101) is $\le\exp(-100p)$. We observe from (14.51) that

$$s = k_2 - k_1 + 1 = 2^{j/2}\sqrt{\operatorname{card} R}\,,$$

and using (14.96) we obtain that $s2^{m/2} \le L2^{j/2}\sqrt{m_0 p}\,\operatorname{card} R$. Also, and since $p \le m_0$, we have $p \le \sqrt{m_0 p}$ and $sp \le L2^{j/2}\sqrt{m_0 p}\sqrt{\operatorname{card} R}$. Consequently $U \le L2^{j/2}\sqrt{m_0 p}\,\operatorname{card} R$. Since $0 \in \mathcal{C}$ we may assume that $0 \in T$, and we have proved that (14.98) holds with the required probability.

In the last step we deduce from (14.98) that we can actually control the supremum over all $h \in \mathcal{G}$. Consider two functions $h, h^* \in \mathcal{G}$. Since h and h^* are 0 outside R, we have

$$\begin{aligned}
\Big|\sum_{i\le N}(h(U_i) - \int h\mathrm{d}\mu) - \sum_{i\le N}(h^*(U_i) - \int h^*\mathrm{d}\mu)\Big| \\
\le \sum_{i\le N}|h(U_i) - h^*(U_i)| + N\int |h-h^*|\mathrm{d}\mu \\
\le \|h-h^*\|_\infty(\operatorname{card}\{i\le N\,;\,U_i\in R\} + NA)\,, \quad (14.102)
\end{aligned}$$

where $A = \mathsf{P}(U_i \in R) = \mu(R)$. Therefore, since for each $h \in \mathcal{G}$ we can find h^* in T with $\|h - h^*\|_\infty \le Ls2^{m/2}$, we obtain

$$\begin{aligned}
\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i) - \int h\mathrm{d}\mu)\Big| \le \sup_{h\in T}\Big|\sum_{i\le N}(h(U_i) - \int h\mathrm{d}\mu)\Big| \quad (14.103)\\
+ Ls2^{-m/2}(\operatorname{card}\{i\le N\,;\,U_i\in R\} + NA)\,.
\end{aligned}$$

Since $\mathsf{P}(U_i = (k,\ell)) = \mu(\{(k,\ell)\}) = n(k,\ell)/N \le 2m_0/N$, we have $NA \le 2m_0\operatorname{card} R$. With probability at least $1-\exp(-100p)$ the event (14.94) occurs and then

$$\operatorname{card}\{i\le N\,;\,U_i\in R\} \le Lm_0\operatorname{card} R.$$

The last term in the right-hand side of (14.103) is then at most

$$Ls2^{-m/2}m_0\operatorname{card} R = L2^{-m/2}2^{j/2}m_0(\operatorname{card} R)^{3/2}\,.$$

Since $L2^{m/2} \ge \sqrt{pm_0\operatorname{card} R}$ by definition of m, the above quantity is $\le L2^{j/2}\sqrt{m_0/p}\operatorname{card} R \le L2^{j/2}\sqrt{pm_0}\,\operatorname{card} R$. This finishes the proof. □

The second proof of Proposition 14.5.1 relies on the following, where ε_i denote independent Bernoulli r.v.s independent of the r.v.s U_i.

Proposition 14.5.3. *Using the notation of Proposition 14.5.1, let us denote by $\mathcal{G}$ the class of functions on G that satisfy conditions (14.89) to (14.92). Then with probability $\ge 1 - L\exp(-100p)$, we have*

$$\sup_{g\in\mathcal{G}}\Big|\sum_{i\le N}\varepsilon_i h(U_i)\Big| \le L2^{j/2}\sqrt{pm_0}\,\operatorname{card} R\,. \quad (14.104)$$

Proof. Proposition 14.4.1 implies

$$\gamma_2(\mathcal{G},d) \le L2^{j/2}\sqrt{p}\,\operatorname{card} R\ ;\ \Delta(\mathcal{G},d) \le L2^{j/2}\operatorname{card} R\ . \tag{14.105}$$

Let us assume that the event of Lemma 14.5.2 occurs. Let us think of the points U_i as being fixed and consider the random distance d_ω on $\mathcal{G}$ given by

$$d_\omega(h,h')^2 = \sum_{i\le N}(h(U_i)-h'(U_i))^2\ , \tag{14.106}$$

so that (14.94) implies $d_\omega(h,h') \le L\sqrt{m_0}d(h,h')$ and we get from (14.105)

$$\gamma_2(\mathcal{G},d_\omega) \le L2^{j/2}\sqrt{pm_0}\,\operatorname{card} R\ ;\ \Delta(\mathcal{G},d_\omega) \le L2^{j/2}\sqrt{m_0}\,\operatorname{card} R\ . \tag{14.107}$$

Denoting by P_ε the conditional probability given the r.v.s U_i, the subgaussian inequality (3.2.2) implies that the process $X_h = \sum_{i\le N}\varepsilon_i h(U_i)$ satisfies the increment condition

$$\mathsf{P}_\varepsilon(|X_h - X_{h'}| \ge u) \le 2\exp\Big(-\frac{u^2}{d_\omega(h,h')^2}\Big)\ .$$

Since $0\in\mathcal{G}$, Theorem 2.2.27 then implies

$$\mathsf{P}_\varepsilon\Big(\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}\varepsilon_i h(U_i)\Big| \ge L(\gamma_2(\mathcal{G},d_\omega)+u\Delta(\mathcal{G},d_\omega))\Big) \le \exp(-u^2)\ .$$

Consequently, taking $u = L\sqrt{p}$, we obtain

$$\mathsf{P}_\varepsilon\Big(\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}\varepsilon_i h(U_i)\Big| \ge L2^{j/2}\sqrt{pm_0}\,\operatorname{card} R\Big) \le \exp(-100p)\ . \tag{14.108}$$

This inequality holds provided the event of Lemma 14.5.2 occurs, which is the case with probability $\ge 1-\exp(-100p)$. This completes the proof. □

This of course was simpler than the first proof of Proposition 14.5.1, but the difficult point, i.e. the control of $\gamma_2(\mathcal{G},d)$ is the same. Moreover, some work remains to be done.

Second proof of Proposition 14.5.1. We consider a sequence (U_i') that is distributed like the sequence (U_i), but is independent of this sequence and of the sequence (ε_i). Then, with probability $\ge 1-\exp(-100p)$ we have both (14.104) and

$$\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}\varepsilon_i h(U_i')\Big| \le L2^{j/2}\sqrt{pm_0}\,\operatorname{card} R\ ,$$

and consequently

$$\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}\varepsilon_i (h(U_i)-h(U_i'))\Big| \le L2^{j/2}\sqrt{pm_0}\,\operatorname{card} R\ .$$

Now comes the beautiful part: the families $(\sum_{i\le N}\varepsilon_i(h(U_i)-h(U_i')))_h$ and $(\sum_{i\le N}(h(U_i)-h(U_i')))_h$ have the same distributions, so that with probability $\ge 1-L\exp(-100p)$ we have

$$\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i)-h(U_i'))\Big|\le L2^{j/2}\sqrt{pm_0}\;\mathrm{card}\,R\;. \tag{14.109}$$

The rest of the proof is basically to integrate in U' and to use Jensen's inequality to move this integral inside the supremum and the absolute value. For the argument it is clearer to assume that the underlying probability space is a product with a generic point (ω,ω') provided with a product probability $\mathsf{P}\otimes\mathsf{P}'$ so that the r.v.s U_i depend only on ω while the r.v.s U_i' depend only on ω'. The exceptional event Ξ where (14.109) fails satisfies $\mathsf{P}\otimes\mathsf{P}'(\Xi)\le L\exp(-100p)$. By Fubini's theorem there exists an event Ω, depending only on ω such that $\mathsf{P}(\Omega)\ge 1-L\exp(-50p)$ and, for each $\omega\in\Omega$ we have

$$\mathsf{P}'(\Xi_\omega)\le L\exp(-50p)\;, \tag{14.110}$$

where

$$\Xi_\omega=\{\omega'\;;\;(\omega,\omega')\in\Xi\}\;.$$

We fix $\omega\in\Omega$ and we denote by E' expectation in ω' only. Thus

$$\mathsf{E}'\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i)-h(U_i'))\Big|\le \mathrm{I}+\mathrm{II}\;,$$

where

$$\mathrm{I}=\mathsf{E}'\mathbf{1}_{\Xi_\omega^c}\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i)-h(U_i'))\Big|\le L2^{j/2}\sqrt{pm_0}\;\mathrm{card}\,R\;,$$

whereas, using that $|h(U_i)|\le L2^p$ by (14.92),

$$\mathrm{II}=\mathsf{E}'\mathbf{1}_{\Xi_\omega}\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i)-h(U_i'))\Big|\le LN2^p\mathsf{P}'(\Xi_\omega)\;.$$

Combining (14.39) with (14.110) we obtain that $\mathrm{II}\le L$ and thus, using Jensen's inequality, we prove that for $\omega\in\Omega$,

$$\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i)-\int h\mathrm{d}\mu)\Big|\le \mathsf{E}'\sup_{h\in\mathcal{G}}\Big|\sum_{i\le N}(h(U_i)-h(U_i'))\Big|\le L2^{j/2}\sqrt{pm_0}\;\mathrm{card}\,R\;.$$

□

The second-order contributions in Proposition 14.3.3 can e.g. be taken care of by combining as previously the use of Bernstein's inequality with the following easy result.

Proposition 14.5.4. *Consider an integer q, and the class $\mathcal{G}$ of functions $h : \{1, \ldots, q\} \to \mathbb{R}$ that satisfy*

$$\forall k \le q-1\,, \quad |h(k+1) - h(k)| \le 1$$
$$\forall k \le q\,, \quad |h(k)| \le 2q\,.$$

Then

$$N(\mathcal{G}, d_\infty, \epsilon) \le \exp\Big(\frac{Lq}{\epsilon}\Big) \tag{14.111}$$

and

$$\gamma_2(\mathcal{G}, d) \le Lq^{3/2}\,, \tag{14.112}$$

where d denotes the Euclidean distance in $\mathbb{R}^q$ and d_∞ the supremum distance.

Proof. First, (4.69) implies (14.111). This implies in turn that $e_n(\mathcal{G}, d_\infty) \le Lq2^{-n}$, and Corollary 2.3.2 yields $\gamma_2(\mathcal{G}, d_\infty) \le Lq$. Since $d \le \sqrt{q} d_\infty$ on $\mathcal{G}$, this implies (14.112). □

Proof of Theorem 14.1.3. In this last part of the proof we go back from our model G of the unit square to the unit square itself. This takes some care but is completely elementary. Consider an integer $N \ge 2$ and an integer p, that will be determined later. For $(k, \ell) \in G$ we consider the point

$$a(k, \ell) = ((2k-1)2^{-p-1}\,, (2\ell-1)2^{-p-1}) \in [0,1]^2\,.$$

These are the centers of 2^{2p} little squares $C(k, \ell)$ of side 2^{-p} that divide $[0,1]^2$. To lighten notation, for $\tau = (k, \ell) \in G$ we write $a(\tau) = a(k, \ell)$ and $C(\tau) = C(k, \ell)$. This notation will be used in particular for $\tau = U_i$ or $\tau = Z_i$.

Consider evenly spread points $(Y_j)_{j \le N}$, a map $\eta : \{1, \ldots, N\} \to G$ such that

$$Y_j \in C(\eta(j))\,, \tag{14.113}$$

so that

$$d(Y_j, a(\eta(j))) \le \sqrt{2} 2^{-p-1} \le 2^{-p} \tag{14.114}$$

and set

$$n(k, \ell) = \operatorname{card}\{j \le N\,;\, \eta(j) = (k, \ell)\}\,, \tag{14.115}$$

the number of points Y_j that belong to $C(k, \ell)$. To avoid trivial complications, we assume that no point Y_j belongs to the boundary of a little square $C(k, \ell)$, so that $\sum n(k, \ell) = N$. The points Y_j are evenly spread, so that these points are centers of non-overlapping rectangles of area $1/N$ and diameter at most $20/\sqrt{N}$. It should be clear that for N and $N2^{-2p}$ large enough, each square $C(k, \ell)$ contains about the same number of points Y_i, so that, for a certain integer m_0, we have

$$m_0 \le n(k, \ell) \le 2m_0\,. \tag{14.116}$$

Since with our choice of p, the quantity $N2^{-2p}$ is about $\log N$, (14.116) holds for N large enough. We consider N points (Z_j) of G such that exactly $n(k,\ell)$ of them are located at the point (k,ℓ). We can assume by (14.114) that these points are labeled in a way that $d(Y_j, a(Z_j)) \le 2^{-p}$.

Consider points X_i independently uniformly distributed over $[0,1]^2$. We claim that we can find independently distributed points U_i of G such that $\mathsf{P}(U_i = (k,\ell)) = n(k,\ell)/N$ and $d(X_i, a(U_i)) \le L2^{-p}$. To see this we recall that by our definition, the fact that the points $(Y_j)_{j\le N}$ are uniformly spread means there exists a partition of $[0,1]^2$ into N rectangles $(R_j)_{j\le N}$ of area $1/N$, each with a width and a height of order $1/\sqrt{N}$, and each containing exactly one point Y_j. For $(k,\ell) \in G$ we define the domain $\mathcal{D}(k,\ell)$ as the union of the sets R_j for which $\eta(j) = (k,\ell)$. It follows from (14.115) that there are $n(k,\ell)$ such rectangles, so that $\mathcal{D}(k,\ell)$ has area $n(k,\ell)/N$. We define $U_i = (k,\ell)$ when $X_i \in \mathcal{D}(k,\ell)$, so that the r.v.s U_i are i.i.d. and $P(U_i = (k,\ell)) = n(k,\ell)/N$. Moreover, when $U_i = (k,\ell)$ then $X_i \in \mathcal{D}(k,\ell)$ so that there exists j with $X_i \in R_j$ and $\eta(j) = (k,\ell)$, and

$$d(X_i, a(U_i)) = d(X_i, a(k,\ell)) \le d(X_i, Y_j) + d(Y_j, a(\eta(j)) \le L2^{-p},$$

using (14.113) and that $d(X_i, Y_j) \le \Delta(R_j) \le L/\sqrt{N}$.

Let us write $X_i = (X_i^1, X_i^2)$, $Y_i = (Y_i^1, Y_i^2)$, and for $\tau \in G$ let us write $a(\tau) = (a(\tau)^1, a(\tau)^2)$. Thus, by definition of $a(\tau)$, if $\tau = (k,\ell)$ we have $a(\tau)^1 = 2^{-p-1}(2k-1)$ and $a(\tau)^2 = 2^{-p-1}(2\ell-1)$. Thus, for $j = 1,2$ we have

$$|a(U_i)^j - a(Z_{i'})^j| = 2^{-p}|U_i^j - Z_{i'}^j| \; . \tag{14.117}$$

For $j = 1,2$ we have

$$\begin{aligned} |X_i^j - Y_{i'}^j| &\le |X_i^j - a(U_i)^j| + |a(U_i)^j - a(Z_{i'})^j| + |a(Z_{i'})^j - Y_{i'}^j| \\ &\le d(X_i, a(U_i)) + |a(U_i)^j - a(Z_{i'})^j| + d(a(Z_{i'}), Y_{i'}) \\ &\le L2^{-p} + 2^{-p}|U_i^j - Z_{i'}^j| \; , \end{aligned} \tag{14.118}$$

using (14.117) in the last line.

Theorems 14.2.1 and 14.2.3 imply that with probability $\ge 1 - LN^{-10}$, there is a permutation π of $\{1, \ldots, N\}$ for which

$$\sum_{i \le N} |U_i^1 - Z_{\pi(i)}^1| \le LN \tag{14.119}$$

$$\forall i \le N \, , \; |U_i^2 - Z_{\pi(i)}^2| \le 1 \; . \tag{14.120}$$

Recalling that $2^{-p} \le L\sqrt{\log N}/\sqrt{N}$ by (14.38) this proves Theorem 14.1.1.

□

14.6 Proof of Proposition 14.2.4

It remains only to prove Proposition 14.2.4, and the next lemmas prepare for this.

Lemma 14.6.1. *Consider numbers $(v_k)_{k\le 2^p}$ and $(v'_k)_{k\le 2^p}$. Consider the numbers*

$$g(k) = \inf\bigl\{v_r + |k-r|\ ;\ 1\le r\le 2^p\bigr\}\ , \tag{14.121}$$

and the numbers $g'(k)$ defined from the sequence (v'_k) the way the numbers $g(k)$ are defined from the sequence (v_k). Then

$$\sum_{k\le 2^p} |g(k)-g'(k)| \le \sum_{k\le 2^p}\bigl(v_k + v'_k - g(k) - g'(k) + |v_k - v'_k|\bigr)\ . \tag{14.122}$$

Proof. Obviously $g(k)\le v_k$ and $g'(k)\le v'_k$. If $g'(k)\ge g(k)$, then

$$\begin{aligned} g'(k)-g(k) \le v'_k - g(k) &= v'_k - v_k + v_k - g(k)\\ &\le |v'_k - v_k| + v_k - g(k) + v'_k - g'(k)\ . \end{aligned}$$

A similar argument when $g(k)\ge g'(k)$ and summation finish the proof. □

We consider numbers $u(k,\ell)$ for $(k,\ell)\in G$, and $h(k,\ell)$ as in (14.30). We set

$$v(k,\ell) = \min\{u(k,s)\ ;\ |\ell - s|\le 1\}\ , \tag{14.123}$$

so that

$$h(k,\ell) = \inf\bigl\{v(r,\ell) + |k-r|\ ;\ 1\le r\le 2^p\bigr\}\ . \tag{14.124}$$

We observe that $v(k,\ell)\le u(k,\ell)$.

Lemma 14.6.2. *We have*

$$m_0 \sum_{k\le 2^p,\ell<2^p} |v(k,\ell+1) - v(k,\ell)| \le 10 \sum_{k,\ell\le 2^p} n(k,\ell)(u(k,\ell) - v(k,\ell))\ . \tag{14.125}$$

Proof. We observe that $|a-b| = a+b-2\min(a,b)$, and that

$$\begin{aligned} v(k,\ell) &\le \min(u(k,\ell+1),u(k,\ell))\\ v(k,\ell+1) &\le \min(u(k,\ell+1),u(k,\ell))\ . \end{aligned}$$

Thus

$$\begin{aligned} |u(k,\ell+1)-u(k,\ell)| &= u(k,\ell)+u(k,\ell+1) - 2\min(u(k,\ell+1)\,,\,u(k,\ell))\\ &\le u(k,\ell) - v(k,\ell) + u(k,\ell+1) - v(k,\ell+1)\ . \end{aligned}$$

By summation we get

$$\sum_{k\le 2^p,\ell<2^p}|u(k,\ell+1)-u(k,\ell)|\le 2\sum_{k,\ell\le 2^p}(u(k,\ell)-v(k,\ell))$$

and since $m_0\le n(k,\ell)$,

$$m_0\sum_{k\le 2^p,\ell<2^p}|u(k,\ell+1)-u(k,\ell)|\le 2\sum_{k,\ell\le 2^p}n(k,\ell)(u(k,\ell)-v(k,\ell))\ . \quad (14.126)$$

Now

$$|v(k,\ell)-u(k,\ell)|\le|u(k,\ell+1)-u(k,\ell)|+|u(k,\ell-1)-u(k,\ell)|$$

so that

$$\begin{aligned}|v(k,\ell+1)-v(k,\ell)|&\le|v(k,\ell+1)-u(k,\ell+1)|+|u(k,\ell+1)-u(k,\ell)|\\&\quad+|u(k,\ell)-v(k,\ell)|\\&\le|u(k,\ell-1)-u(k,\ell)|+3|u(k,\ell+1)-u(k,\ell)|\\&\quad+|u(k,\ell+2)-u(k,\ell+1)|\ .\end{aligned}$$

Then (14.125) follows by summation from (14.126). □

Proof of Proposition 14.2.4. Given $1\le\ell<2^p$, we use Lemma 14.6.1 for $v_k=v(k,\ell)$, and $v'_k=v(k,\ell+1)$, where $v(k,\ell)$ is given by (14.123). Thus $g(k)=h(k,\ell)$ and $g'(k)=h(k,\ell+1)$. Summing the inequalities (14.122) for $1\le k\le 2^p$ we get

$$\begin{aligned}\sum_{k\le 2^p,\ell<2^p}|h(k,\ell+1)-h(k,\ell)|&\le 2\sum_{k,\ell}(v(k,\ell)-h(k,\ell))\\&\quad+\sum_{k,\ell}|v(k,\ell)-v(k,\ell+1)|\ .\end{aligned}$$

Using (14.125), and since $m_0\le n(k,\ell)$ we get

$$\begin{aligned}m_0\sum_{k\le 2^p,\ell<2^p}|h(k,\ell+1)-h(k,\ell)|&\le 2\sum_{k,\ell}n(k,\ell)(v(k,\ell)-h(k,\ell))\\&\quad+10\sum_{k,\ell}n(k,\ell)(u(k,\ell)-v(k,\ell))\\&\le 12\sum_{k,\ell}n(k,\ell)(u(k,\ell)-h(k,\ell))\ ,\end{aligned}$$

using that $h(k,\ell)\le v(k,\ell)\le u(k,\ell)$ in the last line. This proves (14.32), and (14.31) is obvious. □

14.7 Notes and Comments

Shor's original proof of Theorem 14.1.1 establishes more than this theorem, since it produces a matching "on line". The point of presenting our arguments is of course that basically the same approach proves Theorem 14.1.3.

The expert might wonder why in the proof of Proposition 14.5.1, I have first monkeyed around with Bernstein's inequality instead of using symmetrization right away. The power of the symmetrization method has been understood a long time ago through the work of many (and it is used in many places in this book). However this method seemed really like an overkill to prove something like Theorem 4.3.2, and this gave rise to the temptation of proving all the matching theorems without using it. In the end I decided in the case of Proposition 14.5.1 to keep both the approach using Bernstein's inequality and the approach using symmetrization, if only to demonstrate the power of the symmetrization method.

References

1. Talagrand, M.: Matching theorems and discrepancy computations using majorizing measures. J. Am. Math. Soc. **7**, 455–537 (1994)
2. Talagrand, M.: The Generic Chaining. Springer Monographs in Mathematics. Springer, Berlin (2005). viii+222 pp. ISBN: 3-540-24518-9

15. The Ultimate Matching Theorem in Dimension ≥ 3

15.1 Introduction

In this chapter we continue the study of matchings, but in dimension $d \geq 3$ rather than 2. We consider i.i.d. r.v.s $(X_i)_{i \leq N}$ uniformly distributed over the set $[0,1]^d$. We want to match these points to "evenly spread" points $(Y_i)_{i \leq N}$. Here we say that $(Y_i)_{i \leq N}$ are *evenly spread* if one can cover $[0,1]^d$ with N rectangular boxes with disjoint interiors, such that each box R has a d-dimensional volume $1/N$, contains exactly one point Y_i, and is such that $R \subset B(Y_i, 10\sqrt{d}/N^{1/d})$. The cost of a matching will be measured by a rather general function φ. For simplicity of notation we state and prove our results only for $d = 3$. No further ideas are needed to cover the case $d \geq 4$. We denote by $\lambda(A)$ the Lebesgue measure of a subset A of $\mathbb{R}^3$. The entire chapter is devoted to the proof of the following result.

Theorem 15.1.1 ([2]). *Consider a convex function $\varphi \geq 0$ on $\mathbb{R}^3$, with $\varphi(0) = 0$, which is allowed to take infinite values. Assume that it satisfies the following conditions.*

$$\forall\, u \geq 1\ ,\ \lambda(\{\varphi \leq u\}) \geq \log u\ , \tag{15.1}$$

$$\varphi(\pm x_1, \pm x_2, \pm x_3) = \varphi(x_1, x_2, x_3)\ , \tag{15.2}$$

$$\varphi(2^6, 0, 0) \leq 2^{2^{25}}\ ;\ \varphi(0, 2^6, 0) \leq 2^{2^{25}}\ ;\ \varphi(0, 0, 2^6) \leq 2^{2^{25}}\ . \tag{15.3}$$

Then, with probability $\geq 1 - LN^{-10}$ there exists a permutation π of $\{1, \ldots, N\}$ such that

$$\frac{1}{N} \sum_{i \leq N} \varphi\Big(\frac{X_i - Y_{\pi(i)}}{LN^{-1/3}}\Big) \leq 1\ . \tag{15.4}$$

We shall start to discuss how to prove this result in Section 15.2, but in the rest of the present section we discuss the meaning of the theorem itself.

Let us first give an example. Consider $\alpha_1, \alpha_2, \alpha_3 \in]0, \infty]$ with

$$\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \frac{1}{\alpha_3} = 1\ ,$$

and the function

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*,
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_15,

$$\psi(x_1, x_2, x_3) = \exp(|x_1|^{\alpha_1} + |x_2|^{\alpha_2} + |x_3|^{\alpha_3}) - 1 \ .$$

Here, we define $|x|^\infty = 0$ if $|x| < 1$ and $|x|^\infty = \infty$ if $|x| \geq 1$. We have

$$\{\psi \leq u\} \supset \Big\{(x_1, x_2, x_3) \ ; \ \forall j \leq 3 \ , \ |x_j| < \Big(\frac{1}{3}\log(1+u)\Big)^{1/\alpha_j}\Big\} \ ,$$

and consequently

$$\lambda(\{\psi \leq u\}) \geq \frac{1}{3}\log(1+u) \ .$$

The function $\varphi(x) = \psi(x/2^7)$ satisfies $\varphi(2^6, 0, 0) \leq e - 1$ (etc.) and since $\lambda(\{\varphi \leq u\}) = (2^7)^3 \lambda(\{\psi \leq u\})$ it satisfies the conditions of Theorem 15.1.1. The corresponding result proves in this setting the "ultimate matching conjecture" of Problem 14.1.2.

The special case $\alpha_1 = \alpha_2 = \alpha_3 = 3$ is essentially the case where $\varphi(x) = \exp(\|x\|^3) - 1$. It was proved earlier by J. Yukich using the so-called transportation method (unpublished), but the transportation method seems powerless to prove anything close to Theorem 15.1.1. This special case shows that with probability $\geq 1 - N^{-10}$ we can find a matching for which

$$\sum_{i \leq N} \exp(N d(X_i, Y_{\pi(i)})^3 / L) \leq 2N \ ,$$

so that in particular $\sum_{i \leq N} d(X_i, Y_{\pi(i)}) \leq LN^{2/3}$ (since $x \leq \exp x$) and for each i, $\exp(N d(X_i, Y_{\pi(i)})^3/L) \leq 2N$, so that $\max_{i \leq N} d(X_i, Y_{\pi(i)}) \leq LN^{-1/3}(\log N)^{1/3}$ (a result first obtained by J. Yukich and P. Shor in [1]).

Let us try to explain in words the difference between the situation in dimension 3 and in dimension 2. In dimension 2, there are irregularities at all scales in the distribution of a random sample $(X_i)_{i \leq N}$ of $[0,1]^2$, and these irregularities combine to create the mysterious fractional powers of $\log N$. In dimension 3, no such phenomenon occurs, but there are still irregularities at many different scales. Cubes of volume about A/N with a deficit of points X_i exist for A up to about $\log N$. The larger A, the fewer such cubes, in a way which is captured in Theorem 15.1.1 in an essentially optimal manner. The essential feature of dimension ≥ 3 is that, as we will detail below, irregularities at different scales *cannot combine*.

Let us now discuss the conditions of Theorem 15.1.1. Condition (15.2) is satisfied in the most natural examples. We feel that it might well be only a matter of technical work to remove this condition altogether, but since the elementary nature of this work would not be in the line of the main ideas of this book, we have not attempted this. It is good to observe that (15.2) implies

$$\forall j \leq 3 \ , \ |x_j| \leq y_j \Rightarrow \varphi(x) \leq \varphi(y) \tag{15.5}$$

because then x is in the convex hull of the points $(\pm y_1, \pm y_2, \pm y_3)$.

Condition (15.3) is rather mild. It prevents the set $\{\varphi \leq 2^{2^{25}}\}$ to be too thin in one direction. It is in some sense necessary. To prove this we show the following. Given any number A, we can find a number $a > 0$ such that if a function φ satisfies $\varphi(x) = \infty$ for $|x_1| > a$ then (15.4) where $L = A$ fails with probability $\geq 1/2 \geq AN^{-10}$ for some value of N. Indeed, given N and Y, the probability that $\varphi(|X_i - Y|/AN^{-1/3}) < \infty$ is $\lambda([0,1]^3 \cap (Y + AN^{-1/3}\{\varphi < \infty\})$ and this is $\leq 2AaN^{-1/3}$. Consequently if a is small enough, we can choose $N \geq 1$ for which $N2AaN^{-1/3} \leq 1/2$ and $AN^{-10} < 1/2$. Then with probability $\geq 1/2 > AN^{-10}$ there exists no $i \leq N$ for which $\varphi(|X_i - Y_1|/AN^{-1/3}) < \infty$ and then no permutation π for which (15.4) holds.

The crucial condition of Theorem 15.1.1 is (15.1). This condition is also necessary. That is, if (15.4) holds for a constant L_0, then for u large enough we must have

$$\lambda(\{\varphi \leq u\}) \geq \frac{\log u}{L} . \tag{15.6}$$

The idea to prove this is very simple, at least when explained at the most basic level. Let us fix u and set $U = \{\varphi \leq u\}$. If the set U has too small measure, there are so many disjoint sets of the type $x + N^{-1/3}U$ which are entirely contained in $[0,1]^3$, that one of them will contain no point X_i. On the other hand there is a point Y_j inside $x + (1/2)N^{-1/3}U$ and the cost of matching this very point to a point X_i is already sufficient to make (15.4) impossible. Of course there are many details to take care of, and, most importantly, one must get really quantitative. Nonetheless it is perfectly true that (15.6) holds because its failure creates obstacles to the existence of a good matching at a given scale. The situation is then as follows: if there in no obstacle to a good matching at any given scale, then (15.6) holds, and then Theorem 15.1.1 proves that good matchings exist. In word this justifies the second part of the claim (4.29):

obstacles to matchings at different scales do not combine if $d \geq 3$.

This is fundamentally different from what happens in dimension 2.

We now start the rigorous proof that when (15.4) holds for a constant L_0 and each N this implies (15.6). This proof occupies the rest of the present section. Consider N large (to be specified later) and a subset F of $[0,1]^3$, maximal with respect to the property that the sets $x + 2N^{-1/3}L_0U$ for $x \in F$ are disjoint and entirely contained in $[0,1]^3$. Let us assume for the moment that

$$\operatorname{card} F \geq \exp(2(2L_0)^3\lambda(U)) . \tag{15.7}$$

Consider the event Ω_N defined by

$$\exists x \in F \; ; \; \forall i \leq N \; , \; X_i \notin x + 2N^{-1/3}L_0U \; .$$

We are first going to prove that (provided N is large enough) we have

$$\mathsf{P}(\Omega_N) \geq 1/4 \; . \tag{15.8}$$

For this we use the standard device of comparing the set $\{X_1, \ldots, X_N\}$ with a realization of a Poisson point process Π on $[0,1]^3$ of intensity $2N$. For each subset A of $[0,1]^3$ we have $\mathsf{P}(\Pi \cap A = \emptyset) = \exp(-2N\lambda(A))$. Now for $x \in F$ we have $x + 2N^{-1/3}L_0 U \subset [0,1]^3$ and $N\lambda(x + 2N^{-1/3}L_0 U) = (2L_0)^3 \lambda(U)$ and therefore

$$\mathsf{P}((x + 2N^{-1/3}L_0 U) \cap \Pi = \emptyset) = \exp(-2(2L_0)^3\lambda(U)) \geq \frac{1}{\operatorname{card} F} . \tag{15.9}$$

The events (15.9) are independent as x varies over F, because the corresponding sets $x + 2N^{-1/3}L_0 U$ are disjoint. When considering $m (= \operatorname{card} F)$ independent events, each occurring with probability $\geq 1/m$, the probability that at least one of then occurs is $\geq 1 - (1 - 1/m)^m \geq 1 - 1/e \geq 1/2$. Thus $\mathsf{P}(\Omega) \geq 1/2$, where the event Ω is defined by

$$\exists x \in F \; , \; (x + 2N^{-1/3}L_0 U) \cap \Pi = \emptyset \; . \tag{15.10}$$

Consider a sequence $(X_i)_{i \geq 1}$ that is i.i.d. uniformly distributed over $[0,1]^3$. For $M \geq 1$ define

$$p_M := \mathsf{P}(\exists x \in F \; ; \; \forall i \leq M \; , \; X_i \notin x + 2N^{-1/3}L_0 U) \; ,$$

so that $p_N = \mathsf{P}(\Omega_N)$. Now we have

$$\mathsf{P}(\Omega | \operatorname{card} \Pi = M) = p_M \; ,$$

because given that $\operatorname{card} \Pi = M$ the set Π has the same distribution as the set $\{X_1, \ldots, X_M\}$. Thus

$$\mathsf{P}(\Omega \cap \{\operatorname{card} \Pi = M\}) = p_M \mathsf{P}(\operatorname{card} \Pi = M) \; .$$

Obviously the sequence (p_M) decreases. Consequently

$$\begin{aligned} p_N \mathsf{P}(\operatorname{card} \Pi \geq N) &= p_N \sum_{M \geq N} \mathsf{P}(\operatorname{card} \Pi = M) \\ &\geq \sum_{M \geq N} p_M \mathsf{P}(\operatorname{card} \Pi = M) \\ &= \sum_{M \geq N} \mathsf{P}(\Omega \cap \{\operatorname{card} \Pi = M\}) \\ &= \mathsf{P}(\Omega \cap \{\operatorname{card} \Pi \geq N\}) \\ &\geq \mathsf{P}(\Omega) - \mathsf{P}(\operatorname{card} \Pi < N) \; . \end{aligned}$$

Since $\operatorname{card} \Pi$ is a Poisson r.v. of expectation $2N$, for N large enough $\mathsf{P}(\operatorname{card} \Pi < N) \leq 1/4$. Since $\mathsf{P}(\Omega) \geq 1/2$ this implies (15.8).

Let us now assume that Ω_N occurs, and let $x \in F$ such that $x + 2N^{-1/3}L_0 U$ contains no point X_i for $i \leq N$. Using (15.3), we observe that for

$u \geq 2^{2^{25}}$ the set U contains the points $(\pm 2^6, 0, 0), (0, \pm 2^6, 0, 0)$ and $(0, 0 \pm 2^6)$. Since U is convex it contains the set $[-2^4, 2^4]^3$. Consequently (assuming without loss of generality that $L_0 \geq 10$), the set $x + N^{-1/3} L_0 U$ contains the point Y_j which belongs to the small box R that contains x. Consider a permutation π of $\{1, \ldots, N\}$. Then for $i = \pi^{-1}(j)$ we have $X_i \notin x + 2N^{-1/3} L_0 U$ and $Y_{\pi(i)} = Y_j \in x + N^{-1/3} L_0 U$, so that $X_i - Y_{\pi(i)} \notin N^{-1/3} L_0 U$, and this implies

$$\varphi\Big(\frac{X_i - Y_{\pi(i)}}{L_0 N^{-1/3}}\Big) \geq u \;.$$

In particular for each permutation π of $\{1, \ldots, N\}$ we have

$$\frac{1}{N} \sum_{i \leq N} \varphi\Big(\frac{X_i - Y_{\pi(i)}}{L_0 N^{-1/3}}\Big) \geq \frac{u}{N} \;.$$

Thus, since we assume that (15.4) occurs with probability $\geq 1 - LN^{-10}$ we must have $u \leq N$.

In conclusion we have shown that (15.7) implies $u \leq N$, and thus

$$N < u \Rightarrow \operatorname{card} F \leq \exp(2(2L_0)^3 \lambda(U)) \;. \tag{15.11}$$

To use this we fix N as the largest integer for which $N < u$ (so that $N \geq u - 1 \geq u/2$) and we turn towards the study of lower bounds for $\operatorname{card} F$. Since we assume that the cardinality of F is maximal, for every point $y \in [0,1]^3$, either the set $y + 2N^{-1/3} L_0 U$ intersects the boundary of $[0,1]^3$ or it intersects a set $x + 2N^{-1/3} L_0 U$ for some x in F. Consequently if we denote by $D(u)$ the diameter of U, every point y of $[0,1]^3$ is either within distance $2N^{-1/3} L_0 D(u)$ of the boundary of $[0,1]^3$ or belongs to a set $x + 4N^{-1/3} L_0 U$. The volume of the set of points within distance $2N^{-1/3} L_0 D(u)$ of the boundary of $[0,1]^3$ is $\leq L L_0 D(u) N^{-1/3}$; the volume of the union of the sets $x + 4N^{-1/3} L_0 U$ is $\leq (4L_0)^3 N^{-1} \lambda(U) \operatorname{card} F$. Since the union of these two domains cover $[0,1]^3$, the sum of their volumes is ≥ 1, and thus

$$L L_0 N^{-1/3} D(u) + (4L_0)^3 N^{-1} \lambda(U) \operatorname{card} F \geq 1 \;. \tag{15.12}$$

We recall from (15.3) that for $u \geq 2^{2^{25}}$ we have $U \supset V := [-2^4, 2^4]^3$. It is then clear that for any $z \in \mathbb{R}^3$ the convex hull W of z and U satisfies $\lambda(W) \geq 2d(0, z)$. Since U is convex for any $z \in U$ we have $\lambda(U) \geq \lambda(W) \geq 2d(0, z)$ i.e. $U \subset B(0, \lambda(U)/2)$ and hence $D(u) \leq \lambda(U)$. If it happens that $L L_0 N^{-1/3} D(u) \geq 1/2$ then $\lambda(U) \geq D(u) \geq N^{1/3}/L$, i.e. $\lambda(U)$ is of order $u^{1/3}$ and the proof is finished. Otherwise (15.12) implies

$$\operatorname{card} F \geq \frac{N}{2^7 L_0^3 \lambda(U)} \;,$$

and comparing with (15.11) and since $N \geq u/2$ we obtain

$$u \leq L L_0^3 \lambda(U) \exp(2(2L_0)^3 \lambda(U)) \;,$$

and this finishes the proof of (15.6).

15.2 The Crucial Discrepancy Bound

The proof shares some superficial features with the approach of Section 14.1. It is certainly not needed for the reader to have looked at this section before reading the present chapter, but it is essential to review the discussion at the end of Section 4.2, since the difficulty described there is probably the main obstacle we face.

We are not interested in what happens at scales less that $N^{-1/3}$ so that we replace $[0,1]^3$ with the set $G = \{1, \ldots, 2^p\}^3$, and we think of each point of G as representing a small cube of side 2^{-p} (although in the end, as in Section 14.1 it will represent a slightly different region). Here we will have $2^{3p} \simeq N/L$, where L is a large constant that we shall choose later. The generic element of G will be denoted by τ. We will try to match (random) points $(U_i)_{i \le N}$ in G to given "evenly spread" points $(Z_i)_{i \le N}$. Since 2^{3p} may not divide N, we may not put the same number of points Z_i at each point of G, so we denote

$$n(\tau) = \operatorname{card}\{i \le N \; ; \; Z_i = \tau\} \; . \tag{15.13}$$

Since we try to make the points Z_i evenly spread, and since the ratio $N/2^{3p}$ will be a large number, we certainly may assume that there exists a certain integer m_0 for which

$$\forall \tau \in G \; , \; m_0 \le n(\tau) \le 2m_0 \; . \tag{15.14}$$

Let us observe that by summation of the relations (15.14) over $\tau \in G$ we get

$$m_0 2^{3p} \le N \le 2m_0 2^{3p} \; . \tag{15.15}$$

We consider the probability measure μ on G given by

$$\forall \tau \in G \; , \; \mu(\{\tau\}) = \frac{n(\tau)}{N} \; , \tag{15.16}$$

so that

$$\forall \tau \in G \; , \; \frac{m_0}{N} \le \mu(\{\tau\}) \le \frac{2m_0}{N} \; . \tag{15.17}$$

Thus μ is nearly uniform. To each function $w : G \to \mathbb{R}$ we associate the function $h_w : G \to \mathbb{R}$ given by

$$h_w(\tau) = \inf\{w(\tau') + \varphi(\tau - \tau') \; ; \; \tau' \in G\} \; . \tag{15.18}$$

We observe that since $\varphi(0) = 0$ we have

$$h_w \le w \; , \tag{15.19}$$

and we define

$$\Delta(w) = \int (w - h_w) \mathrm{d}\mu \ge 0 \; . \tag{15.20}$$

The crucial ingredient for Theorem 15.1.1 is the following discrepancy bound.

Theorem 15.2.1. *Consider an i.i.d. sequence of r.v.s* $(U_i)_{i\le N}$ *distributed like* μ*. Then with probability* $\ge 1 - L\exp(-100p)$ *the following occurs:*

$$\forall\, w\ ,\ \Big|\sum_{i\le N}\Big(h_w(U_i) - \int h_w \mathrm{d}\mu\Big)\Big| \le L\sqrt{m_0}2^{3p}(\Delta(w)+1)\ . \tag{15.21}$$

As we explained in Section 4.2 the essential difficulty in a statement of this type is to understand which kind of information on the function h_w we may obtain from the fact that $\Delta(w)$ is given. In very general terms there is not choice: we must extract information to the effect that such functions "do not vary widely" so that we may bound the left-hand side of (15.21) with overwhelming probability. It is precisely because the whole procedure goes through several non-trivial stages that it occupies all the forthcoming sections of this chapter. In still rather general terms, we shall prove that control of $\Delta(w)$ implies a kind of local Lipschitz condition on h_w. This is the goal of Section 15.4. This local Lipschitz condition implies in turn a suitable control on the coefficients of a Haar basis expansion of h_w, and this will allow us to conclude. The proof does not explicitly use chaining, although it is in a similar spirit. The formulation in an abstract setting of the principle behind this proof is a possible topic for further research.

In the remainder of this section, we first prove a matching theorem related to the bound (15.21), and we then use Theorem 15.2.1 to complete the proof of Theorem 15.1.1.

Theorem 15.2.2. *There exists a constant* L_1 *such that the following occurs. Assume that*

$$m_0 \ge L_1\ . \tag{15.22}$$

Consider points $(U_i)_{i\le N}$ *as in (15.21). Then there exists a permutation* π *of* $\{1,\dots,N\}$ *for which*

$$\sum_{i\le N}\varphi(U_i - Z_{\pi(i)}) \le N\ . \tag{15.23}$$

For further use, the reader will observe that the following proof does not use that φ is convex, but only that $\varphi(0)=0$ and $\varphi\ge 0$.

Proof. First we deduce from Proposition 4.2.1 that

$$\inf_\pi \sum_{i\le N}\varphi(U_i - Z_\pi(i)) = \sup\sum_{i\le N}(w_i + w_i')\ , \tag{15.24}$$

where the supremum is over all families $(w_i)_{i\le N}$ and $(w_i')_{i\le N}$ for which

$$\forall\, i,j\le N\ ,\ w_i + w_j' \le \varphi(U_i - Z_j)\ . \tag{15.25}$$

Given such families (w_i) and (w_i'), for $\tau\in G$ let us then define

$$h(\tau) = \inf_{j\le N}(-w_j' + \varphi(\tau - Z_j))\ ,$$

so that from (15.25) we obtain $w_i \le h(U_i)$ and thus

$$\sum_{i\le N}(w_i + w'_i) \le \sum_{i\le N}(h(U_i) + w'_i) . \tag{15.26}$$

For $\tau \in G$ we define

$$w(\tau) = \inf\{-w'_j\ ;\ Z_j = \tau\} ,$$

so that

$$h(\tau) = \inf\{w(\tau') + \varphi(\tau - \tau')\ ;\ \tau' \in G\} , \tag{15.27}$$

and consequently, recalling the notation (15.18),

$$h(\tau) = h_w(\tau) .$$

Also,

$$-w(\tau) = \sup\{w'_j\ ;\ Z_j = \tau\} ,$$

so that, using (15.16),

$$-N\mu(\{\tau\})w(\tau) \ge \sum\{w'_j\ ;\ Z_j = \tau\} ,$$

and thus by summation of these inequalities over $\tau \in G$,

$$\sum_{i\le N} w'_i \le -N\int w\mathrm{d}\mu . \tag{15.28}$$

Consequently

$$\begin{aligned}\sum_{i\le N}(h(U_i) + w'_i) &\le \sum_{i\le N} h(U_i) - N\int w\mathrm{d}\mu \\ &\le \sum_{i\le N}\Big(h(U_i) - \int h\mathrm{d}\mu\Big) - N\int (w-h)\mathrm{d}\mu \\ &= \sum_{i\le N}\Big(h(U_i) - \int h\mathrm{d}\mu\Big) - N\Delta(w) . \end{aligned} \tag{15.29}$$

Now (15.21) implies, since $h = h_w$,

$$\sum_{i\le N}\Big(h(U_i) - \int h\mathrm{d}\mu\Big) \le L\sqrt{m_0}2^{3p}(\Delta(w) + 1) , \tag{15.30}$$

and combining with (15.26) and (15.29) we have proved that all families (w_i) and (w'_i) as in (15.25) satisfy

$$\sum_{i\le N}(w_i + w'_i) \le L\sqrt{m_0}2^{3p}(\Delta(w) + 1) - N\Delta(w) .$$

Recalling that $N \ge m_0 2^{3p}$ by (15.15), we obtain that for $m_0 \ge L_1$ the right-hand side is $\le N$. □

We end this section by completing the proof of Theorem 15.1.1. The argument basically reproduces the proof of Theorem 14.1.3 on page 470.

Proof of Theorem 15.1.1. Consider an integer $N \geq 2$ and an integer p, that will be determined later. For $(\tau) = (\tau^j)_{j \leq 3} \in G$ we consider the point

$$a(\tau) = ((2\tau^j - 1)2^{-p-1})_{j \leq 3} \in [0,1]^3 .$$

These points are the centers of 2^{3p} little cubes $C(\tau)$ of side 2^{-p} that divide $[0,1]^3$.

Consider evenly spread points $(Y_i)_{i \leq N}$ and a map $\eta : \{1, \ldots, N\} \to G$ such that

$$Y_i \in C(\eta(i)) . \tag{15.31}$$

Consequently,

$$d(Y_i, a(\eta(i))) \leq \Delta(C(\eta(i))) \leq \sqrt{3} 2^{-p-1} \leq 2^{-p} . \tag{15.32}$$

We denote

$$n(\tau) = \operatorname{card}\{i \leq N \,;\, \eta(i) = \tau\} , \tag{15.33}$$

the number of points Y_i "assigned to τ". (In the generic case where no point Y_i belongs to the boundary of a little cube $C(\tau)$ this is simply the number of points Y_i contained in $C(\tau)$.) Thus $\sum_\tau n(\tau) = N$. The points Y_i are evenly spread, so that by definition these points are centers of non-overlapping rectangular boxes R_i of volume $1/N$ and diameter at most $20/N^{1/3}$. It should be clear that for $N \geq L$ and $N 2^{-3p} \geq L$, each cube $C(\tau)$ contains about the same number of points Y_i. To give a formal proof one simply bounds from below (respectively form above) the number of points Y_i assigned to $C(\tau)$ by the number of boxes R_i which are entirely contained (respectively which intersect) $C(\tau)$, and the difference of these two numbers is not large because the cubes $C(\tau)$ are then much bigger that the boxes R_i. Consequently, for a certain integer m_0, we have

$$m_0 \leq n(\tau) \leq 2m_0 , \tag{15.34}$$

and (15.15) implies $m_0 \geq N 2^{-3p-1}$. Let us then choose p as the largest integer for which $N 2^{-3p-1} \geq L_1$, where L_1 is the constant of (15.22), so that (15.22) holds and

$$2^{3p} \geq \frac{N}{L} . \tag{15.35}$$

This implies in particular that $1 - L\exp(-100p) \geq 1 - LN^{-1/10}$. We consider N points (Z_i) of G such that exactly $n(\tau)$ of them are located at the point τ. We may assume by (15.32) that these points are labeled in such a way that

$$d(Y_i, a(Z_i)) \leq 2^{-p} . \tag{15.36}$$

Consider points X_i independently uniformly distributed over $[0,1]^3$. We claim that we can find independently distributed points U_i of G such that

$\mathsf{P}(U_i = \tau) = \mu(\{\tau\}) = n(\tau)/N$ and $d(X_i, a(U_i)) \le L2^{-p}$. For $\tau \in G$ we define the domain $\mathcal{D}(\tau)$ as the union of the sets R_i (the little box containing Y_i) for which $\eta(i) = \tau$. From (15.33) there are $n(\tau)$ such boxes, so that $\mathcal{D}(\tau)$ has volume $n(\tau)/N$. We define $U_i = \tau$ when $X_i \in \mathcal{D}(\tau)$, so that the r.v.s U_i are i.i.d. and $P(U_i = \tau) = \mu(\{\tau\}) = n(\tau)/N$. Moreover, when $U_i = \tau$ then $X_i \in \mathcal{D}(\tau)$ so that there exists ℓ with $X_i \in R_\ell$ and $\eta(\ell) = \tau$, and

$$d(X_i, a(U_i)) = d(X_i, a(\tau)) \le d(X_i, Y_\ell) + d(Y_\ell, a(\eta(\ell)) \le L2^{-p} \ , \tag{15.37}$$

using (15.32), that $d(X_i, Y_\ell) \le \Delta(R_\ell) \le L/N^{1/3}$ and (15.35).

Let us write $U_i = (U_i^j)_{j\le 3}$, $V_i = (V_i^j)_{j\le 3}$, and for $\tau \in G$ let us write $a(\tau) = (a(\tau)^j)_{j\le 3}$. Thus, by definition of $a(\tau)$, if $\tau = (\tau^j)_{j\le 3}$ we have $a(\tau)^j = 2^{-p-1}(2\tau^j - 1)$. Consequently, for $j \le 3$ we have

$$|a(U_i)^j - a(Z_{i'})^j| = 2^{-p}|U_i^j - Z_{i'}^j| \ . \tag{15.38}$$

Let us write $X_i = (X_i^j)_{j\le 3}$, $Y_i = (Y_i^j)_{j\le 3}$. For $j \le 3$ we have

$$\begin{aligned} |X_i^j - Y_{i'}^j| &\le |X_i^j - a(U_i)^j| + |a(U_i)^j - a(Z_{i'})^j| + |a(Z_{i'})^j - Y_{i'}^j| \\ &\le d(X_i, a(U_i)) + |a(U_i)^j - a(Z_{i'})^j| + d(a(Z_{i'}), Y_{i'}) \\ &\le \frac{L}{N^{1/3}}(1 + |U_i^j - Z_{i'}^j|) \ , \end{aligned} \tag{15.39}$$

using (15.36) and (15.37) in the last inequality, and since $2^{-p} \le LN^{-1/3}$ by (15.35).

Theorems 15.2.1 and 15.2.2 imply that with probability $\ge 1 - LN^{-10}$, there is a permutation π of $\{1, \dots, N\}$ for which

$$\sum_{i\le N} \varphi(U_i - Z_{\pi(i)}) \le N \ . \tag{15.40}$$

Using (15.5) and convexity, (15.39) implies

$$\varphi\Big(\frac{X_i - Y_{\pi(i)}}{2LN^{-1/3}}\Big) \le \frac{1}{2}\varphi((1,1,1)) + \frac{1}{2}\varphi(U_i - Z_{\pi(i)}) \ .$$

We recall that $\varphi((1,1,1)) \le L$ from (15.3). Summation over $i \le N$ and the fact that $\varphi(x/L) \le \varphi(x)/L$ by convexity of φ proves Theorem 15.1.1. □

15.3 Cleaning up φ

It is difficult to work with a function φ as general as that of Theorem 15.1.1. In this section we extract the information we shall use about φ. We prove that we can replace φ by a certain function φ^*, with the properties given in Proposition 15.3.1 below. This will be very convenient but is purely a technical point, and the present section does not contain any essential idea towards the proof of Theorem 15.2.1.

Throughout the rest of this chapter we define

$$M_k = 2^{2^{k+25}} (= N_{k+25}) \ .$$

Since $\log 2 \geq 1/2$ we deduce from (15.1) that

$$\lambda(\{\varphi \leq M_k\}) \geq 2^{k+24} \ . \tag{15.41}$$

For $k \geq 0$ let us define $m_1(k)$ as the largest integer in $\mathbb{Z}$ for which

$$\varphi(2^{m_1(k)+6}, 0, 0) \leq M_k \ ,$$

so that $m_1(k) \geq 0$ by (15.3). Also, the sequence $(m_1(k))$ is obviously non-decreasing. We define $m_2(k)$ and $m_3(k)$ similarly, and by (15.2) and convexity we obtain

$$\varphi(\pm 2^{m_1(k)+4}, \pm 2^{m_2(k)+4}, \pm 2^{m_3(k)+4}) \leq M_k \ ,$$

so that

$$16 \prod_{j \leq 3} [-2^{m_j(k)}, 2^{m_j(k)}] \subset \{\varphi \leq M_k\} \ . \tag{15.42}$$

The definition of $m_1(k)$ implies $\varphi(2^{m_1(k)+7}, 0, 0) \geq M_k$, so that (15.5) imply $\varphi(x) > M_k$ whenever $|x_1| > 2^{m_1(k)+7}$ and consequently

$$\{\varphi \leq M_k\} \subset \prod_{j \leq 3} [-2^{m_j(k)+7}, 2^{m_j(k)+7}] \ .$$

Therefore

$$\lambda(\{\varphi \leq M_k\}) \leq 2^{24 + \sum_{j \leq 3} m_j(k)} \ .$$

Comparing with (15.41) we obtain

$$\sum_{j \leq 3} m_j(k) \geq k \ . \tag{15.43}$$

For each $k \geq 0$ we construct for $j \leq 3$ numbers $n_j(k) \geq 0$ with the following properties: each sequence $(n_j(k))_{k \geq 0}$ is non-decreasing and

$$n_j(k) \leq m_j(k) \ ; \ \sum_{j \leq 3} n_j(k) = k \ . \tag{15.44}$$

The construction is immediate using (15.43). Let us observe that

$$n_j(k) \leq n_j(k+1) \leq n_j(k) + 1 \ . \tag{15.45}$$

We define

$$S_k = \prod_{j\le 3} [-2^{n_j(k)}, 2^{n_j(k)}] \,, \tag{15.46}$$

so that (15.42) implies

$$16S_k \subset \{\varphi \le M_k\} \,. \tag{15.47}$$

We also observe from (15.45) that

$$S_{k+1} \subset 2S_k \,. \tag{15.48}$$

Proposition 15.3.1. *There exists a function φ^* with $\varphi^*(0) = 0$ that satisfies the following properties:*

$$\forall\, x \,,\ \varphi^*(2x) \ge \varphi(x) \,, \tag{15.49}$$

$$\{\varphi^* \le M_k\} \supset 8S_k \,, \tag{15.50}$$

$$\{\varphi^* \le M_k\} \subset 16S_k \,, \tag{15.51}$$

$$\textit{the set } \{\varphi^* \le u\} \textit{ is convex for each } u > 0 \,, \tag{15.52}$$

$$\forall\, x \,,\ \varphi^*(x) = \varphi^*(-x) \,, \tag{15.53}$$

$$u \ge 4M_1 \Rightarrow \frac{3}{4}\{\varphi^* \le u\} \subset \Big\{\varphi^* \le \frac{u}{4}\Big\} \,. \tag{15.54}$$

The function φ^* (that need not be convex) crystallizes the properties of φ we need.

For $j \le 3$ we construct a function θ_j as follows. For $q \ge n_j(1)$ let us denote by $k_j(q)$ the smallest integer k such that $n_j(k) = q$. When $q > \lim_{k\to\infty} n_j(k)$, no such integer exists, and we set $k_j(q) = \infty$. We observe using (15.45) that $k_j(q+1) \ge k_j(q) + 1$, and, from the last part of (15.44), that $k_j(q) \ge q$. We define the function θ_j on the interval $[2^{n_j(1)}, \infty[$ such that $\theta_j(2^q) = \log M_{k_j(q)}$ for $q \ge n_j(1)$ and that θ_j is affine between the values 2^q and 2^{q+1}. When $k_j(q+1) = \infty$ this means that $\theta_j(u) = \infty$ for $u > 2^q$. We claim that

$$\theta_j \text{ has a slope } \ge 2^{24} \,. \tag{15.55}$$

To see this we simply observe that $M_{k_j(q+1)} \ge M_{k_j(q)+1} = M^2_{k_j(q)}$, so that the slope of θ_j on the interval $[2^q, 2^{q+1}]$ is at least

$$2^{-q} \log M_{k_j(q)} \ge 2^{-q} \log M_q \ge 2^{24} \,.$$

Let us then define the function $\psi_j : \mathbb{R} \to \mathbb{R}^+ \cup \{\infty\}$ as follows:

$$\psi_j(x) = |x| 2^{-n_j(1)-3} M_1 \quad \text{if} \quad |x| \le 2^{n_j(1)+3} \,, \tag{15.56}$$

$$\psi_j(x) = \exp \theta_j\Big(\frac{|x|}{8}\Big) \quad \text{if} \quad |x| \ge 2^{n_j(1)+3} \,. \tag{15.57}$$

In particular we observe that

$$|x| = 2^{n_j(1)+3} \Rightarrow \psi_j(x) = M_1 \,. \tag{15.58}$$

Lemma 15.3.2. *We have*

$$\psi_j\Big(\frac{3}{4}x\Big) \le \max\Big(M_1, \frac{1}{4}\psi_j(x)\Big)\ . \tag{15.59}$$

Proof. We may assume that $\psi_j(3x/4) \ge M_1$, for there is nothing to prove otherwise. The definition of ψ_j implies that $y := 3|x|/4$ satisfies $y \ge 2^{n_j(1)+3}$, so that $y/8 \ge 2^{n_j(1)}$. Since we have shown that θ_j has a slope $\ge 2^{24}$ on the interval $[2^{n_j(1)}, \infty[$, we have

$$\theta_j\Big(\frac{4}{3}\frac{y}{8}\Big) \ge 2^{24}\Big(\frac{1}{3}\frac{y}{8}\Big) + \theta_j\Big(\frac{y}{8}\Big) = 2^{19}|x| + \theta_j\Big(\frac{y}{8}\Big)\ ,$$

so that

$$\psi_j(x) = \exp\theta_j\Big(\frac{x}{8}\Big) = \exp\theta_j\Big(\frac{4}{3}\frac{y}{8}\Big) \ge \exp(2^{19}|x|)\psi_j\Big(\frac{3x}{4}\Big)\ .$$

Now,

$$\psi_j\Big(\frac{3}{4}x\Big) > M_1 \Rightarrow \psi_j(x) > M_1 \Rightarrow |x| \ge 2^{n_j(1)+3} \ge 8\ ,$$

so that

$$\psi_j(x) \ge \exp(2^{22})\psi_j\Big(\frac{3}{4}x\Big) \ge 4\psi_j\Big(\frac{3}{4}x\Big)\ . \qquad \square$$

Proof of Proposition 15.3.1. We define

$$\varphi^*(x_1, x_2, x_3) = \max_{j\le 3}\psi_j(x_j)\ , \tag{15.60}$$

so that $\varphi^*(0) = 0$ and (15.52) holds since the sets $\{\psi_j \le u\}$ are intervals. Next, to prove (15.50) it suffices to prove that

$$\{\theta_j \le \log M_k\} \supset [-2^{n_j(k)}, 2^{n_j(k)}]\ . \tag{15.61}$$

This is true because if $q = n_j(k)$, then $k_j(q) \le k$ by definition of $k_j(q)$, and then $\theta_j(2^q) = \log M_{k_j(q)} \le \log M_k$. To prove (15.51) it suffices to prove that

$$\{\theta_j \le \log M_k\} \subset [-2^{n_j(k)+1}, 2^{n_j(k)+1}]\ . \tag{15.62}$$

This holds because if $q = n_j(k) + 1$ then $k_j(q) > k$, and then $\theta_j(2^q) = \log M_{k_j(q)} > M_k$.

To prove (15.54) we observe simply from Lemma 15.3.2 that

$$\varphi^*\Big(\frac{3}{4}x\Big) \le \max\Big(M_1, \frac{1}{4}\varphi^*(x)\Big)\ .$$

We turn to the proof of (15.49). Consider $x \in \mathbb{R}^3$ and let $a = \varphi(x)$. We assume first that $a > M_1$, and let $k \ge 1$ such that $M_k < a \le M_{k+1}$. Then (15.47) yields

$$16S_k \subset \{\varphi \le M_k\} \subset \{\varphi < a\}\ ,$$

so that, using (15.48),

$$8S_{k+1} \subset \{\varphi < a\} ,$$

and thus $x \notin 8S_{k+1}$ and $2x \notin 16S_{k+1}$. Since $\{\varphi^* \le M_{k+1}\} \subset 16S_{k+1}$ we obtain that $\varphi^*(2x) \ge M_{k+1} > \varphi(x)$.

Assume next that $a \le M_1$. We may assume that $x \in 8S_1$ for otherwise $2x \notin 16S_1$ and $\varphi^*(2x) > M_1$ by (15.51). We then write $x = \alpha y$ with $0 \le \alpha \le 1$ and y on the boundary of $8S_1$. That is, $y = (y_1, y_2, y_3)$ where $|y_j| \le 2^{n_j(1)+3}$ and equality for at least one $1 \le j \le 3$. Then $\varphi^*(y) = M_1$ from (15.58) and definition of φ^* and $\varphi(y) \le M_1$ by (15.47), so that $\varphi^*(y) \ge \varphi(x)$. Moreover (15.56) implies $\varphi^*(x) = \alpha\varphi^*(y) \ge \alpha\varphi(y) \ge \varphi(x)$ using convexity of φ and since $\varphi(0) = 0$. Finally it is obvious that $\varphi^*(2x) \ge \varphi^*(x)$ by construction. □

In the remainder of this chapter we prove that Theorem 15.2.1 holds when φ is replaced by φ^*. Since $\varphi^*(x) \ge \varphi(x/2)$, and since as noted the proof of Theorem 15.2.2 does not use the convexity of φ, we may then replace (15.40) by $\sum_{i \le N} \varphi((U_i - Z_{\pi(i)})/2) \le N$, and this suffices instead of (15.40) to prove Theorem 15.1.1.

To lighten notation **we assume from now on that $\varphi = \varphi^*$ satisfies (15.50) to (15.54).**

15.4 Geometry

The real work towards the proof of Theorem 15.2.1 starts here. In this section we carry out the task of finding some "regularity" of the functions h_w (defined in (15.18)) for which $\Delta(w)$ is not too large. In other words, we understand some of the underlying "geometry" of this class of functions.

This regularity is a kind of "local Lipschitz condition", which is expressed in Theorem 15.4.2 below. We will comment on this theorem in more details when we state it.

We define

$$s_j(k) = \min(p, n_j(k)) \; ; \; s(k) = \sum_{j \le 3} s_j(k) \; . \tag{15.63}$$

It follows from (15.44) that $n_j(k) \le k$, so that

$$k \le p \Rightarrow s_j(k) = n_j(k) \tag{15.64}$$

and also using (15.44),

$$k \le p \Rightarrow s(k) = k \; . \tag{15.65}$$

We consider the collection $\mathcal{P}_k$ of subsets of G of the form

$$\prod_{j \le 3} \{1 + b_j 2^{s_j(k)}, \ldots, (b_j + 1) 2^{s_j(k)}\} \tag{15.66}$$

for $b_j \in \mathbb{N}$, $0 \le b_j \le 2^{p-s_j(k)} - 1$. There are $2^{3p-s(k)}$ such sets, which form a partition of G. Each of this sets has cardinality $2^{s(k)}$. Let us say that two such sets are **adjacent** if for all $j \le 3$ the corresponding values of b_j differ by at most 1. Thus a set is adjacent to itself, and to at most 26 other sets. Given an integer q, let us say that two sets of $\mathcal{P}_k$ are **q-adjacent** if for each $j \le 3$ the corresponding values of b_j differ by at most q. Thus that are at most $(2q+1)^3$ sets which are q-adjacent to a given set. The elementary proof of the following is better left to the reader.

Lemma 15.4.1. *(a) If C, C' in $\mathcal{P}_k$ are adjacent, and if $\tau \in C, \tau' \in C'$ then $\tau - \tau' \in 2S_k$.*
(b) If $\tau \in C \in \mathcal{P}_k$, $A \subset \tau + qS_k$, $A \in \mathcal{P}_k$, A and C are q-adjacent.

Let us also observe that

$$C \in \mathcal{P}_k \Rightarrow \operatorname{card} C = 2^{s(k)} . \tag{15.67}$$

Given $\Delta \ge 0$ we define the class $\mathcal{S}(\Delta)$ of functions on G by

$$\mathcal{S}(\Delta) = \left\{ h_w \ ; \ \Delta(w) = \int (w - h_w) \mathrm{d}\mu \le \Delta \right\} . \tag{15.68}$$

We say that a subset A of G is $\mathcal{P}_k$-measurable if it is a union of sets belonging to $\mathcal{P}_k$. The main result of this section is as follows.

Theorem 15.4.2. *Given $\Delta \ge 1$, for every function $h \in \mathcal{S}(\Delta)$ we can find a partition $(B_k)_{k \ge 1}$ of G such that B_k is $\mathcal{P}_k$-measurable, and such that for each $C \in \mathcal{C}_k := \{C \in \mathcal{P}_k ; C \subset B_k\}$, we can find a number $z(C)$ such that the following properties hold:*

$$\sum_{k \ge 1} \sum_{C \in \mathcal{C}_k} 2^{s(k)} z(C) \le L 2^{3p} \Delta , \tag{15.69}$$

$$k \ge 1 \ , \ C \in \mathcal{C}_k \Rightarrow M_k \le z(C) \le M_{k+1} , \tag{15.70}$$

For every $k \ge 1$, if $C \in \mathcal{C}_k$ and if $C' \in \mathcal{P}_k$ is adjacent to C, then

$$\tau \in C \ , \ \tau' \in C' \Rightarrow |h(\tau) - h(\tau')| \le z(C) . \tag{15.71}$$

Let us stress that (15.71) holds in particular for $C' = C$.

The idea of the partition B_k is simply that it locally defines a "scale" at which we study the behavior of h. Any point τ of G belongs to some B_k, and the scale relevant to τ is given by the partition $\mathcal{P}_k$. defines the scale relevant to τ. We should think of condition (15.71) as locally controlling the variations of h, by the quantities $z(C)$. The scale at which this occurs is local. Condition (15.70) controls the size of the numbers $z(C)$, depending on the local scale at which the bound (15.71) holds. It is the restriction $z(C) \le M_{k+1}$ which is essential, the lower bound is purely technical. Finally the global size of the weighted quantities $z(C)$ is controlled by (15.69). Of course each of these precise quantitative controls will become essential in its turn. Probably the reader finds this is terribly complicated, but it seems to be in the very nature of the problem that one must use this idea of varying local scale in some form.

In the remainder of this chapter, we shall use the information provided by Theorem 15.4.2 to prove Theorem 15.2.1.

We now start the proof of Theorem 15.4.2, which will occupy us until the end of the present section. This proof is fortunately not as formidable as the statement of the theorem itself, and again we have tried to write it in great detail as well as we could. By definition of $\mathcal{S}(\Delta)$ we can find a function $w : \mathbb{R} \to G$ such that

$$\forall\, \tau \in G \;,\; h(\tau) = \inf\{w(\tau') + \varphi(\tau - \tau') \;;\; \tau' \in G\} \;, \tag{15.72}$$

while

$$\int (w - h) \mathrm{d}\mu \le \Delta \;. \tag{15.73}$$

For each τ and τ' in G we have

$$h(\tau) \le w(\tau') + \varphi(\tau - \tau') \;,$$

so that

$$w(\tau') \ge h(\tau) - \varphi(\tau - \tau') \;.$$

Let us then define

$$\widehat{h}(\tau') = \sup\{h(\tau) - \varphi(\tau - \tau') \;;\; \tau \in G\} \;,$$

so that

$$h \le \widehat{h} \le w \;. \tag{15.74}$$

Moreover,

$$\int (\widehat{h} - h) \mathrm{d}\mu \le \int (w - h) \mathrm{d}\mu \le \Delta \;. \tag{15.75}$$

For C in $\mathcal{P}_k$ let us define

$$y(C) = \min_C \widehat{h}(\tau) - \max_C h(\tau) \;. \tag{15.76}$$

Thus, for $\tau \in C$ we have $y(C) \le \widehat{h}(\tau) - h(\tau)$ and, using (15.67),

$$2^{s(k)} y(C) \le \sum_{\tau \in C} (\widehat{h}(\tau) - h(\tau)) \;. \tag{15.77}$$

Let $q = 32$. For $C \in \mathcal{P}_k$ we set

$$x(C) = \max\{y(C') \;;\; C' \in \mathcal{P}_k \;,\; C' \text{ is } q\text{-adjacent to } C\} \;. \tag{15.78}$$

For $k \geq 2$ we denote by D_k the union of all the sets in $\mathcal{P}_k$ for which $x(C) \geq M_k/2$. Thus D_k is $\mathcal{P}_k$-measurable and D_ℓ is empty for ℓ large enough. For $k \geq 2$ we define

$$B_k = D_k \setminus \bigcup_{\ell > k} D_\ell \; .$$

We define $B_1 = G \setminus \bigcup_{\ell \geq 2} D_\ell$. Thus B_k is $\mathcal{P}_k$-measurable (B_1 is even $\mathcal{P}_2$-measurable). Let us note from the definition

$$\ell > k \Rightarrow B_k \cap D_\ell = \emptyset. \tag{15.79}$$

The sets B_k form the partition we are looking for. We recall the notation $\mathcal{C}_k = \{C \in \mathcal{P}_k, C \subset B_k\}$. Our next goal is to prove that following, which is the main step towards (15.69).

Lemma 15.4.3. *We have*

$$\sum_{k \geq 2} \sum_{C \in \mathcal{C}_k} 2^{s(k)} x(C) \leq L q^3 2^{3p} \Delta \; . \tag{15.80}$$

Proof. When $C \in \mathcal{C}_k$, we have $x(C) \geq M_k/2$ and by definition of $x(C)$ there exists $\theta(C) \in \mathcal{P}_k$ which is q-adjacent to C and such that $x(C) = y(\theta(C)) \geq M_k/2$. Thus

$$\sum_{k \geq 2} \sum_{C \in \mathcal{C}_k} 2^{s(k)} x(C) = \sum_{k \geq 2} \sum_{C \in \mathcal{C}_k} 2^{s(k)} y(\theta(C)) \; . \tag{15.81}$$

Let us define

$$L_k = \bigcup \{\theta(C) \; ; \; C \in \mathcal{C}_k\} \; ,$$

so that L_k is $\mathcal{P}_k$-measurable. Let us observe an important property: if $C' \in \mathcal{P}_k$ and $C' \subset L_k$, then $y(C') \geq M_k/2$ and C' is q-adjacent to a certain $C \in \mathcal{C}_k$.

Next, consider $C' \in \mathcal{P}_k$, $C' \subset L_k$, and assume that $C' = \theta(C)$ for a certain $C \subset \mathcal{C}_k$. Then C is q-adjacent to $C' = \theta(C)$. Thus there at most $(2q+1)^3$ such sets C. In summary, given $C \in \mathcal{C}_k$, we have $C' = \theta(C) \in \mathcal{P}_k$ and $\mathbb{C}' \subset L_k$, and moreover there are at most $(2q+1)^3$ sets C'' for which $\theta(C'') = C'$. Consequently,

$$\sum_{k \geq 2} \sum_{C \in \mathcal{C}_k} 2^{s(k)} y(\theta(C)) \leq (2q+1)^3 \sum_{k \geq 2} \sum_{C \in \mathcal{P}_k, C \subset L_k} 2^{s(k)} y(C) \; . \tag{15.82}$$

Now, it follows from (15.77) that

$$\sum_{C \in \mathcal{P}_k, C \subset L_k} 2^{s(k)} y(C) \leq \sum_{\tau \in L_k} (\widehat{h}(\tau) - h(\tau)) \; . \tag{15.83}$$

Our next (and crucial) step is to prove that the sets L_k are disjoint. For this, we consider $C \in \mathcal{P}_k$, $C \subset L_k$, and we prove that if $k' < k$ we have $C \cap L_{k'} = \emptyset$.

We recall that $y(C) \geq M_k/2$, and we proceed by contradiction. Assume if possible that we can find $C' \in \mathcal{P}_{k'}$, $C' \subset C$, $C' \subset L_{k'}$. Then by definition, C' is q-adjacent to an element $C^{\sim}$ of $B_{k'}$. Consider the element C^* of $\mathcal{P}_k$ that contains $C^{\sim}$. Then since $k' \leq k$, C^* is q-adjacent to C, because $\mathcal{P}_k$ is a coarser partition of G than $\mathcal{P}_{k'}$. Since $y(C) \geq M_k/2$, this proves by definition of D_k that $C^* \subset D_k$. Then $C^* \cap B_{k'} = \emptyset$ by (15.79). Since $C^{\sim} \subset C^*$ this contradicts the fact that $C^{\sim} \subset B_{k'}$, and completes the proof that $C \cap L_{k'} = \emptyset$, and hence $L_k \cap L_{k'} = \emptyset$.

Since the sets L_k are disjoint,

$$\sum_{k \geq 2} \sum_{C \in \mathcal{P}_k, C \subset L_k} 2^{s(k)} y(C) \leq \sum_{\tau} (\widehat{h}(\tau) - h(\tau)) \leq L 2^{3p} \int (\widehat{h} - h) \mathrm{d}\mu \,. \tag{15.84}$$

Combining with (15.81) and (15.82) proves (15.80). □

For $C \in \mathcal{C}_k$, $k \geq 2$, we set

$$z(C) = \min(2x(C), M_{k+1}) \geq M_k \,.$$

If $C \in \mathcal{C}_1$ we set $z(C) = M_2$. Thus (15.70) holds, and since $2^{s(1)} \operatorname{card} \mathcal{C}_1 \leq 2^{s(1)} \operatorname{card} \mathcal{P}_1 = 3^p$, (15.69) follows from (15.80) since $\Delta \geq 1$.

We turn to the proof of (15.71), the core of Theorem 15.4.2. In that part of the argument, we view G as a subset of $\mathbb{R}^3$. We start with a preliminary result.

Lemma 15.4.4. *If $x \in [1, 2^p]^3$ we can find $A \in \mathcal{P}_k$ such that $A \subset x + S_k$.*

Proof. If $x = (x_j)_{j \leq 3}$, consider for $j \leq 3$ an integer b_j with $0 \leq b_j \leq 2^{p - s_j(k)} - 1$ and

$$1 + b_j 2^{s_j(k)} \leq x_j \leq 1 + (b_j + 1) 2^{s_j(k)} \,.$$

Then

$$\{1 + b_j 2^{s_j(k)}, \ldots, (b_j + 1) 2^{s_j(k)}\} \subset [x_j - 2^{s_j(k)}, x_j + 2^{s_j(k)}] \,,$$

so that the set (15.66) is entirely contained in $x + S_k$. □

We are now ready to prove "half" of (15.71).

Lemma 15.4.5. *If $k \geq 1$, $C \in \mathcal{C}_k$, $C' \in \mathcal{P}_k$ are adjacent and $\tau \in C, \tau' \in C'$, then*

$$h(\tau') \leq h(\tau) + z(C) \,. \tag{15.85}$$

Proof. Since $z(C) \geq M_{\max(k,2)}$ there is nothing to prove unless

$$h(\tau') - h(\tau) \geq M_{\max(k,2)} \,,$$

so we assume that this is the case in the rest of the argument.

It follows from (15.72) that for some $\rho \in G$ we have

$$h(\tau) = w(\rho) + \varphi(\tau - \rho) . \tag{15.86}$$

We define

$$u = \max(h(\tau') - h(\tau), \varphi(\tau - \rho)) ,$$

so that $u \ge M_{\max(k,2)}$. We set

$$U = \{\varphi \le u\} .$$

This is a convex set, and (15.53) implies that $U = -U$. Since $u \ge M_k$, (15.50) implies $8S_k \subset U$. Also, since C and C' are adjacent by Lemma 15.4.1 (a), we have $\tau' - \tau \in 2S_k$, and consequently,

$$\tau' - \tau \in \frac{U}{4} .$$

Since $\varphi(\tau - \rho) \le u$ we have $\tau - \rho \in U$, so that $\tau' - \rho = \tau' - \tau + \tau - \rho \in 5U/4$ and therefore $\rho - \tau' \in 5U/4$ by (15.53). Consequently,

$$\frac{\rho + \tau'}{2} \in \tau' + \frac{5U}{8} \; ; \; \frac{\rho + \tau'}{2} \in \rho + \frac{5U}{8} . \tag{15.87}$$

Here of course $(\rho + \tau')/2$ need not be a point of G. We define

$$V = \frac{\rho + \tau'}{2} + \frac{U}{8} ,$$

so that (15.87) implies

$$V \subset \left(\tau' + \frac{3U}{4}\right) \cap \left(\rho + \frac{3U}{4}\right) . \tag{15.88}$$

Since $u \ge M_2 \ge 4M_1$ it follows from (15.54) that $\varphi(x) \le u/4$ for $x \in 3U/4$. Consequently if $\rho' \in G \cap V$ we have $\varphi(\rho' - \tau') \le u/4$ and $\varphi(\rho' - \rho) \le u/4$. Thus

$$\widehat{h}(\rho') \ge h(\tau') - \varphi(\tau' - \rho') \ge h(\tau') - \frac{u}{4} .$$

Also, by (15.72) we have

$$h(\rho') \le w(\rho) + \varphi(\rho' - \rho) \le w(\rho) + \frac{u}{4} .$$

Thus, using (15.86) in the second line,

$$\begin{aligned} \min_{\rho' \in V \cap G} \widehat{h}(\rho') - \max_{\rho' \in V \cap G} h(\rho') &\ge h(\tau') - w(\rho) - \frac{u}{2} \\ &= h(\tau') - h(\tau) + \varphi(\tau - \rho) - \frac{u}{2} \\ &\ge \max(h(\tau') - h(\tau), \varphi(\tau - \rho)) - \frac{u}{2} \\ &= \frac{u}{2} . \end{aligned} \tag{15.89}$$

Consider the largest integer $\ell \ge 1$ such that $u \ge M_\ell$. Since $u \ge M_{\max(k,2)}$ we have $\ell \ge \max(k,2)$, and by definition of ℓ we have $u < M_{\ell+1}$. We then use (15.50), (15.51) and (15.48) to obtain

$$8S_\ell \subset U \subset 32S_\ell \; . \tag{15.90}$$

Thus

$$V = \frac{\rho+\tau'}{2} + \frac{U}{8} \supset \frac{\rho+\tau'}{2} + S_\ell \; ,$$

and Lemma 15.4.4 shows that there exists $A \in \mathcal{P}_\ell$ with $A \subset V$. It then follows from (15.76) and (15.89) that $y(A) \ge u/2$. Also, since $\tau' - \tau \in U/4$, we have, using (15.88) in the first inclusion,

$$V \subset \tau' + \frac{3U}{4} \subset \tau + U \subset \tau + 32S_\ell \; . \tag{15.91}$$

Since $\ell \ge k$, and since $C \in \mathcal{P}_k$, there is a unique set $D \in \mathcal{P}_\ell$ with $C \subset D$, and since $A \subset V$ and $\tau \in C$, (15.91) and Lemma 15.4.1 imply that D and A are q-adjacent for $q = 32$. Consequently the definition of $x(D)$ shows that $x(D) \ge y(A) \ge u/2 \ge M_\ell/2$, and therefore $D \subset D_\ell$ by definition of D_ℓ. It is impossible that $\ell > k$, because the inclusion $D \subset D_\ell$ contradicts the fact that $C \subset D$ and $C \subset B_k$ (see (15.79)). When $k = 1$, since $\ell \ge \max(k,2) = 2 > k$, we have already obtained a contradiction proving that $u \le M_2$. When $k \ge 2$, since $\ell \ge k$, we then have $\ell = k$. Thus $D = C$ and then $x(C) = x(D) \ge u/2$. Also, $u < M_{\ell+1} = M_{k+1}$, so that $u \le \min(2x(C), M_{k+1}) \le z(C)$, completing the proof of (15.85). □

Finally it remains to prove that, with the notation of (15.71),

$$h(\tau) \le h(\tau') + z(C) \; . \tag{15.92}$$

For this, we repeat the previous argument, exchanging the roles of τ and τ', up to (15.91), which we replace by

$$V \subset \tau + \frac{3U}{4} \subset \tau + U \subset \tau + 32S_\ell \; ,$$

and we finish the proof in exactly the same manner. □

15.5 Probability, I

To prove a discrepancy bound involving the functions of the class $\mathcal{S}(\Delta)$ of Theorem 15.4.2, we must understand "how they oscillate". They are two sources for such oscillations.

- The function h oscillates within each set $C \in \mathcal{C}_k$
- The function h oscillates when it changes set C.

In the present section, we take care of the first type of oscillation. We show that we can reduce the proof of Theorem 15.2.1 to that of Theorem 15.5.7 below, that is to the case where h is constant on each set $C \in \mathcal{C}_k$. This is significantly easier then the proof of Theorem 15.5.7 itself, but already bring to light the use of the various conditions of Theorem 15.4.2. We elaborate more about this fundamental point after the statement of Proposition 15.5.1 below. Throughout the rest of the proof, for $\tau \in G$ we consider the r.v.

$$Y_\tau = \operatorname{card}\{i \le N\ ;\ U_i = \tau\} - N\mu(\{\tau\})\ , \tag{15.93}$$

where U_i are i.i.d. r.v.s on G with $\mathsf{P}(U_i = \tau) = \mu(\tau)$. We recall the number m_0 of (15.34). It is good to note right away that

$$\mathsf{E} Y_\tau^2 \le N\mu(\{\tau\}) \le 2m_0\ ,$$

so that we may think of $|Y_\tau|$ as being typically of size about $\sqrt{m_0}$.

We state the main result of this section.

Proposition 15.5.1. *With probability* $\ge 1 - L\exp(-100p)$ *the following happens. Consider any* $\Delta \ge 1$*, and a partition* $(B_k)_{k\ge 1}$ *of* G *such that* B_k *is* $\mathcal{P}_k$*-measurable, and for each* $C \in \mathcal{C}_k := \{C \in \mathcal{P}_k; C \subset B_k\}$*, consider a number* $z(C)$ *such that the following properties hold:*

$$\sum_{k\ge 1}\sum_{C\in\mathcal{C}_k} 2^{s(k)} z(C) \le L2^{3p}\Delta\ , \tag{15.94}$$

$$k \ge 1\ ,\ C \in \mathcal{C}_k \Rightarrow z(C) \le M_{k+1}\ . \tag{15.95}$$

Then

$$\sum_{k\ge 1}\sum_{C\in\mathcal{C}_k} z(C) \sum_{\tau\in C} |Y_\tau| \le L2^{3p}\sqrt{m_0}\Delta\ . \tag{15.96}$$

This result is closely connected to Theorem 15.4.2. It will allow to reduce the proof of this theorem to the case where h is constant on each $C \in \mathcal{C}_k$. Let us explain in words a very simple, yet central idea about why Proposition 15.5.1 is true. First, since $\operatorname{card} C = 2^{s(k)}$ for $C \in \mathcal{C}_k$, the expectation of the left-hand side of (15.96) is about $\sum_{k\ge1}\sum_{C\in\mathcal{C}_k} 2^{s(k)} z(C)\sqrt{m_0}$, so to obtain (15.96) it is needed to control the weighted averages of the numbers $z(C)$ as in (15.94). This however does not suffice, because for a few values of C the corresponding quantities $\sum_{\tau\in C}|Y_\tau|$ are going to be abnormally large. This however does not influence too much the sum (15.96) because the corresponding coefficient $z(C)$ cannot be too large by (15.95). The proof consists in quantifying this idea, and one may wonder why everything fits so well in the computations.

We need first to understand the properties of this family (Y_τ) of r.v.s and other related entities. This is the motivation behind the following definition.

Definition 15.5.2. *Consider a finite set V. We say that a family $(Y_v)_{v\in V}$ is of type $\mathcal{B}(N)$ if there exists a probability space (Ω,θ), i.i.d. r.v.s $(W_i)_{i\le N}$ valued in Ω, of law θ, and functions ψ_v on Ω, $|\psi_v|\le 1$, with* disjoint support, *such that*

$$\frac{1}{2\operatorname{card} V}\le \theta(\{\psi_v\neq 0\})\le \frac{2}{\operatorname{card} V}\,, \tag{15.97}$$

and for which

$$Y_v=\sum_{i\le N}\Big(\psi_v(W_i)-\int \psi_v \mathrm{d}\theta\Big)\,. \tag{15.98}$$

The family $(Y_\tau)_{\tau\in G}$ is of type $\mathcal{B}(N)$. The following crucial property will simply follow from Bernstein's inequality.

Lemma 15.5.3. *Consider any family $(Y_v)_{v\in V}$ of type $\mathcal{B}(N)$ and numbers $(\eta_v)_{v\in V}$ with $|\eta_v|\le 1$. Then for $u>0$ we have*

$$\mathsf{P}\Big(\Big|\sum_{v\in V}\eta_v Y_v\Big|\ge u\Big)\le 2\exp\Big(-\frac{1}{L}\min\Big(u,\frac{u^2\operatorname{card} V}{N\sum_{v\in V}\eta_v^2}\Big)\Big)\,. \tag{15.99}$$

Proof. Consider the r.v.s W_i and the functions ψ_v as in Definition 15.5.2. We define

$$S_i=\sum_{v\in V}\eta_v\psi_v(W_i)\,.$$

Since the functions ψ_v have disjoint support, we have $|S_i|\le 1$ and also

$$\mathsf{E}(S_i-\mathsf{E}S_i)^2\le \mathsf{E}S_i^2\le \frac{2}{\operatorname{card} V}\sum_{v\in V}\eta_v^2\,.$$

Since $\sum_{v\in V}\eta_v Y_v=\sum_{i\le N}(S_i-\mathsf{E}S_i)$, and since $|S_i|\le 1$, (15.99) follows from Bernstein's inequality (4.59). □

We prepare the proof with two lemmas which limit the possible size of the quantities $\sum_{\tau\in C}|Y_\tau|$ for $C\in\mathcal{C}_k$. It is convenient to distinguish the cases of "small k" and "large k". These are the cases $k\le k_0$ and $k\ge k_0$ for an integer k_0 which we define now. If $2^{3p}<M_2$ we define $k_0=0$. Otherwise we define k_0 as the largest integer with $M_{k_0+2}\le 2^{3p-k_0}$, so that $M_{k_0+3}\ge 2^{3p-k_0-1}$. Therefore $2^{k_0+27}\le 3p-k_0$ and $2^{k_0+28}\ge 3p-k_0-1$, so that

$$\frac{p}{L}\le 2^{k_0}\le Lp\,. \tag{15.100}$$

In particular for N large enough, and hence for p large enough, we have $k_0\le p$. Since $s(k)=k$ for $k\le p$ we then have $s(k)=k$ for $k\le k_0$.

We first take care of the values $k\ge k_0$. This is the easiest case because for these values the individual sums $\sum_{\tau\in C}|Y_\tau|$ for $C\in\mathcal{P}_k$ are not really larger than their expectations, as the following shows.

Lemma 15.5.4. *With probability $\geq 1 - L\exp(-100p)$ the following occurs. Consider $k_0 \leq k \leq p$. Then if $C \in \mathcal{P}_k$ we have*

$$\sum_{\tau \in C} |Y_\tau| \leq L 2^{s(k)} \sqrt{m_0}\,. \tag{15.101}$$

Proof. For τ in C consider $\eta_\tau = \pm 1$, and for $\tau \notin C$ let $\eta_\tau = 0$. Then since $\operatorname{card} C = 2^{s(k)}$, we have $\sum_{\tau \in G} \eta_\tau^2 = 2^{s(k)}$ so that, using (15.99) for $V = G$, and since $\operatorname{card} V = 2^{3p}$ and $N \leq 2m_0 2^{3p}$,

$$\mathsf{P}\Big(\sum_{\tau \in G} \eta_\tau Y_\tau \geq u\Big) \leq 2\exp\Big(-\frac{1}{L}\min\Big(u, \frac{u^2}{m_0 2^{s(k)}}\Big)\Big)\,.$$

Thus, if $A \geq 1$ is a parameter,

$$\mathsf{P}\Big(\sum_{\tau \in G} \eta_\tau Y_\tau \geq A\sqrt{m_0} 2^{s(k)}\Big) \leq 2\exp\Big(-\frac{A 2^{s(k)}}{L}\Big)\,.$$

Crudely, there are most 2^{3p} choices for $C \in \mathcal{P}_k$. Using that $k_0 = s(k_0)$ and (15.100),

$$p \leq L 2^{k_0} = L 2^{s(k_0)} \leq L 2^{s(k)}\,. \tag{15.102}$$

Given $C \in \mathcal{P}_k$, there are $2^{\operatorname{card} C} = 2^{2^{s(k)}}$ choices for the signs $(\eta_\tau)_{\tau \in C}$. Consequently, with probability at least

$$1 - L\exp\Big(\Big(L - \frac{A}{L}\Big) 2^{s(k)}\Big)$$

we have $\sum_{\tau \in C} \eta_\tau Y_\tau \leq A\sqrt{m_0} 2^{s(k)}$ whatever the choice of $C \in \mathcal{P}_k$ and of the signs $(\eta_\tau)_{\tau \in C}$, and thus $\sum_{\tau \in C} |Y_\tau| \leq A\sqrt{m_0} 2^{s(k)}$ for each $C \in \mathcal{P}_k$. Since $2^{s(k)} \geq p/L$ by (15.102) we then obtain that if A is a large enough constant this holds for all $k_0 \leq k \leq p$ with probability $\geq 1 - L\exp(-100p)$. □

We now turn to the case $k \leq k_0$. The situation is more complicated because some of the quantities $\sum_{\tau \in C} |Y_\tau|$ can be quite larger than their expectations, and we have to quantify this. We shall use the following well known elementary fact:

$$\binom{n}{k} \leq \Big(\frac{en}{k}\Big)^k\,. \tag{15.103}$$

Lemma 15.5.5. *With probability $\geq 1 - L\exp(-100p)$ the following occurs. Consider $k \leq k_0$ and let $q_k = \lfloor 2^{3p-k}/M_{k+2} \rfloor$. Then if B is the union of q_k sets in $\mathcal{P}_k$ we have*

$$\sum_{\tau \in B} |Y_\tau| \leq L q_k 2^k \sqrt{m_0}\,. \tag{15.104}$$

We may think of (15.104) as allowing $\sum_{\tau \in C} |Y_\tau|$ for $C \in \mathcal{P}_k$, $C \subset B$ to be (in average over $C \subset B$) about 2^k times larger than what it should be. Fortunately, the larger k, the smaller q_k.

Proof. Since $k \leq k_0$ we have $M_{k+2} \leq 2^{3p-k}$ and thus $q_k \geq 1$. Since $\lfloor x \rfloor \geq 1 \Rightarrow \lfloor x \rfloor \geq x/2$ this yields

$$q_k \geq \frac{2^{3p-k}}{2M_{k+2}} \,.$$

Consequently, using the definition of k_0 in the first inequality and (15.100) in the last one,

$$2^k q_k \geq \frac{2^{3p}}{2M_{k+2}} \geq \frac{2^{3p}}{2M_{k_0+2}} \geq \frac{2^{k_0}}{2} \geq \frac{p}{L} \,. \tag{15.105}$$

For τ in B consider $\eta_\tau = \pm 1$, and for $\tau \notin B$ set $\eta_\tau = 0$. Then since $\operatorname{card} B = 2^{s(k)} q_k \leq 2^k q_k$, we have $\sum_{\tau \in G} \eta_\tau^2 \leq 2^k q_k$. We use (15.99) with $V = G$ so that $\operatorname{card} V = 2^{3p}$. It implies, using (15.15) in the second line

$$\begin{aligned} \mathsf{P}\Big(\sum_{\tau \in G} \eta_\tau Y_\tau \geq u\Big) &\leq 2\exp\Big(-\frac{1}{L}\min\Big(u, \frac{u^2 2^{3p}}{N 2^k q_k}\Big)\Big) \\ &\leq 2\exp\Big(-\frac{1}{L}\min\Big(u, \frac{u^2}{m_0 2^k q_k}\Big)\Big) \,. \end{aligned}$$

Consequently, given a parameter $A \geq 1$, and choosing $u = A\sqrt{m_0} q_k 2^k$, we obtain

$$\mathsf{P}\Big(\sum_{\tau \in G} \eta_\tau Y_\tau \geq u\Big) \leq 2\exp\Big(-\frac{A}{L} q_k 2^k\Big) \,. \tag{15.106}$$

Using (15.103) there are

$$\binom{2^{3p-k}}{q_k} \leq \Big(\frac{e 2^{3p-k}}{q_k}\Big)^{q_k} \leq (L M_{k+2})^{q_k} \leq \exp(L 2^k q_k)$$

choices for B. Given B, there are $2^{\operatorname{card} B} \leq 2^{2^k q_k}$ choices for the signs $(\eta_\tau)_{\tau \in B}$. Consequently with probability at least

$$1 - L\exp\Big(\Big(L - \frac{A}{L}\Big) q_k 2^k\Big)$$

we have $\sum_{\tau \in B} \eta_\tau Y_\tau \leq u$ whatever the choice of B and of the signs $(\eta_\tau)_{\tau \in B}$. Taking for A a large enough constant, and using (15.105) the result follows. □

Proof of Proposition 15.5.1. We assume that the events described in Lemmas 15.5.5 and 15.5.4 occur, and we prove (15.96). First, since for $k \geq k_0$ and $C \in \mathcal{C}_k$ we have $\sum_{\tau \in C} |Y_\tau| \leq L 2^{s(k)} \sqrt{m_0}$,

$$\sum_{k \geq k_0} \sum_{C \in \mathcal{C}_k} z(C) \sum_{\tau \in C} |Y_\tau| \leq L\sqrt{m_0} \sum_{k \geq k_0} \sum_{C \in \mathcal{C}_k} 2^{s(k)} z(C) \,. \tag{15.107}$$

Next, consider $k \le k_0$. We claim that

$$\sum_{C\in\mathcal{C}_k} z(C)\sum_{\tau\in C}|Y_\tau| \le L2^k\sqrt{m_0}\sum_{C\in\mathcal{C}_k} z(C) + L\sqrt{m_0}2^{3p}\frac{M_{k+1}}{M_{k+2}}\ . \tag{15.108}$$

Summation of these inequalities for $1 \le k \le k_0$, and combining with (15.107) proves (15.96). To prove (15.108), let us enumerate as $C_1, C_2, \ldots,$ the elements of $\mathcal{C}_k$, in decreasing order of $w(C) := \sum_{\tau\in C}|Y_\tau|$. For clarity we assume that there are at least q_k such elements (since it should be obvious how to proceed when this is not the case). Then (15.104) used for $B = \cup_{\ell\le q_k} C_\ell$ implies

$$\sum_{\ell\le q_k} w(C_\ell) = \sum_{\tau\in B}|Y_\tau| \le Lq_k 2^k\sqrt{m_0}\ ,$$

and since $q_k w(C_{q_k}) \le \sum_{\ell\le q_k} w(C_\ell)$ we deduce that $w(C_{q_k}) \le L2^k\sqrt{m_0}$ and consequently $w(C_\ell) \le L2^k\sqrt{m_0}$ for $\ell \ge q_k$. Therefore, and since $z(C) \le M_{k+1}$ for $C \in \mathcal{C}_k$,

$$\begin{aligned}\sum_{C\in\mathcal{C}_k} z(C)\sum_{\tau\in C}|Y_\tau| &= \sum_{\ell\ge 1} z(C_\ell)w(C_\ell)\\ &= \sum_{\ell\le q_k} z(C_\ell)w(C_\ell) + \sum_{\ell> q_k} z(C_\ell)w(C_\ell)\\ &\le M_{k+1}\sum_{\ell\le q_k} w(C_\ell) + L2^k\sqrt{m_0}\sum_{C\in\mathcal{C}_k} z(C)\ . \end{aligned}\tag{15.109}$$

Since

$$\sum_{\ell\le q_k} w(C_\ell) \le L2^k\sqrt{m_0}q_k \le L\sqrt{m_0}\frac{2^{3p}}{M_{k+2}},$$

we have proved (15.96). □

Let us denote by L_0 the constant in (15.69).

Definition 15.5.6. *Consider $\Delta > 1$. We define the class $\mathcal{S}^*(\Delta)$ of functions $h : G \to \mathbb{R}$ such that we can find a partition $(B_k)_{k\ge1}$ of G where B_k is $\mathcal{P}_k$-measurable, and recalling the definition $\mathcal{C}_k := \{C \in \mathcal{P}_k; C \subset B_k\}$,*

$$\text{for each } k \ge 1\ ,\ h \text{ is constant on each set } C \in \mathcal{C}_k\ , \tag{15.110}$$

and for each $C \in \mathcal{C}_k$, we can find a number $z(C)$ with the following three properties:

$$\sum_{k\ge1}\sum_{C\in\mathcal{C}_k} 2^{s(k)}z(C) \le L_0 2^{3p}\Delta\ , \tag{15.111}$$

$$k \ge 1\ ,\ C \in \mathcal{C}_k \Rightarrow M_k \le z(C) \le M_{k+1}\ , \tag{15.112}$$

if $C \in \mathcal{C}_k$ and if $C' \in \mathcal{P}_k$ is adjacent to C and such that for $k' > k$, we have $C' \not\subset B_{k'}$, then

$$\tau \in C\ ,\ \tau' \in C' \Rightarrow |h(\tau) - h(\tau')| \le z(C)\ . \tag{15.113}$$

Let us stress the difference between (15.113) and (15.71). In (15.113), C'_k is adjacent to C_k as in (15.71) but satisfies the further condition that $C_k \not\subset B_{k'}$ for $k' \geq k$.

In the rest of this chapter we shall prove the following.

Theorem 15.5.7. *Consider an i.i.d sequence of r.v.s* $(U_i)_{i\leq N}$ *distributed like* μ*. Then with probability* $\geq 1 - L\exp(-100p)$ *the following occurs: for* $\Delta \geq 1$*, whenever* $h \in \mathcal{S}^*(\Delta)$*,*

$$\Big|\sum_{i\leq N}\Big(h(U_i) - \int h\mathrm{d}\mu\Big)\Big| \leq L\sqrt{m_0}2^{3p}\Delta \;. \tag{15.114}$$

Let us observe right away the following fundamental identity, which is obvious from the definition of Y_τ:

$$\sum_{i\leq N}\Big(h(U_i) - \int h\mathrm{d}\mu\Big) = \sum_{\tau\in G} h(\tau)Y_\tau \;. \tag{15.115}$$

Proof of Theorem 15.2.1. Consider a function $h \in \mathcal{S}(\Delta)$, the sets B_k and the numbers $z(C)$ as provided by Theorem 15.4.2. Consider the function h^* on G defined as follows: if $C \in \mathcal{C}_k$, then h^* is constant on C, and the value of this constant is the average value of h on C, i.e. $\int_C h\mathrm{d}\mu = \int_C h^*\mathrm{d}\mu$. Then, using (15.71) for $C' = C$ yields that $|h - h^*| \leq z(C)$ on C. Consequently Proposition 15.5.1 implies that with probability $\geq 1 - \exp(-100p)$,

$$\Big|\sum_{\tau\in G} h(\tau)Y_\tau - \sum_{\tau\in G} h^*(\tau)Y_\tau\Big| \leq \sum_{k\geq 1}\sum_{C\in\mathcal{C}_k} z(C)\sum_{\tau\in C}|Y_\tau| \leq L\sqrt{m_0}2^{3p}\Delta \;.$$

Therefore, using Theorem 15.5.7, it suffices to prove that $h^* \in \mathcal{S}^*(\Delta)$. Using the same sets B_k and the same values $z(C)$ for h^* as for h it suffices to prove (15.113). Consider C and C' as in this condition, and $\tau \in C, \tau' \in C'$. Consider k' such that $\tau' \in C'' \in \mathcal{C}_{k'}$. Then we have $k' \leq k$, for otherwise $C' \subset C'' \subset B_{k'}$, and we assume that this is not the case. Thus $C'' \subset C'$, and consequently by (15.71) for $\rho \in C$ and $\rho' \in C''$ we have $|h(\rho) - h(\rho')| \leq z(C)$. Averaging ρ over C and ρ' over C'' proves that $|h^*(\tau) - h^*(\tau')| \leq z(C)$. □

15.6 Haar Basis Expansion

The strategy to prove Theorem 15.5.7 is very simple. We write an expansion $h = \sum_v a_v(h)v$ along the Haar basis, where $a_v(h)$ is a number and v is a function belonging the Haar basis. (See the details in (15.120) below.) We then write

$$\Big|\sum_{i\le N}\Big(h(U_i)-\int h\mathrm{d}\mu\Big)\Big|\le\sum_v|a_v(h)|\Big|\sum_{i\le N}\Big(v(U_i)-\int v\mathrm{d}\mu\Big)\Big|$$
$$=\sum_v|a_v(h)||Y_v|\;, \tag{15.116}$$

where $Y_v=\sum_{i\le N}(v(U_i)-\int v\mathrm{d}\mu)$. In the right hand side we will bound separately the sums corresponding to functions v "of the same size". It should be pointed out that the reason such an approach works at all (while it fails in dimension 2) is that "the obstacles at different scales cannot combine", so that the scales can be separated as above. The first task is to understand the size of the coefficients $a_v(h)$, of course as a function of the coefficients $z(C)$ of Definition 15.5.6. This is done in Proposition 15.6.2 below, and the final work of controlling the sum in (15.116) is the purpose of the next and last section.

For $1\le r\le p+1$ we define the class $\mathcal{H}(r)$ of functions on $\{1,\dots,2^p\}$ as follows.

$$\mathcal{H}(p+1)\text{ consists of the function that is constant equal to }1\;. \tag{15.117}$$

For $1\le r\le p$, $\mathcal{H}(r)$ consists of the 2^{p-r} functions $f_{i,r}$ for $0\le i<2^{p-r}$ that are defined as follows:

$$f_{i,r}(\sigma)=\begin{cases}1 & \text{if } i2^r<\sigma\le i2^r+2^{r-1}\\ -1 & \text{if } i2^r+2^{r-1}<\sigma\le(i+1)2^r\\ 0 & \text{otherwise}\end{cases} \tag{15.118}$$

In this manner we define a total of 2^p functions. These functions are orthogonal in $L^2(\theta)$ where θ is the uniform probability on $\{1,\dots,2^p\}$, and thus form a complete orthogonal basis of this space. Let us note that

$$\int f_{i,r}^2\mathrm{d}\theta=2^{-p+\min(r,p)}\;. \tag{15.119}$$

For $1\le q_1,q_2,q_3\le p+1$, let us denote by $\mathcal{V}(q_1,q_2,q_3)$ the set of functions of the type $v=f_1\otimes f_2\otimes f_3$ where $f_j\in\mathcal{H}(q_j)$ for $j\le 3$. The functions $v\in\mathcal{V}(q_1,q_2,q_3)$ have disjoint supports. As q_1,q_2,q_3 take all possible values, these functions form a complete orthogonal system of $L^2(\nu)$, where

ν denotes the uniform probability on G .

Consequently, given any function h on G, we have the expansion

$$h=\sum_{1\le q_1,q_2,q_3\le p+1}\;\sum_{v\in\mathcal{V}(q_1,q_2,q_3)}a_v(h)v\;, \tag{15.120}$$

where

$$a_v(h) = \frac{\int hv\mathrm{d}\nu}{\int v^2\mathrm{d}\nu} \ . \tag{15.121}$$

Setting $q_j^* = \min(q_j, p)$ and recalling (15.119) we obtain that for $v \in \mathcal{V}(q_1, q_2, q_3)$

$$\int v^2\mathrm{d}\nu = 2^{q_1^*+q_2^*+q_3^*-3p} \geq \frac{1}{L}2^{q_1+q_2+q_3-3p} \ . \tag{15.122}$$

The decomposition (15.120) then implies

$$\Big|\sum_{\tau\in G} h(\tau)Y_\tau\Big| \leq \sum_{1\leq q_1,q_2,q_3\leq p+1}\ \sum_{v\in\mathcal{V}(q_1,q_2,q_3)} |a_v(h)|\Big|\sum_{\tau\in G} v(\tau)Y_\tau\Big| \ . \tag{15.123}$$

This will be our basic tool to prove Theorem 15.5.7, keeping (15.115) in mind. Fixing q_1, q_2, q_3, the main effort will be to find competent bounds for

$$\sum_{v\in\mathcal{V}(q_1,q_2,q_3)} |a_v(h)|\Big|\sum_{\tau\in G} v(\tau)Y_\tau\Big| \ . \tag{15.124}$$

Since we think of q_1, q_2 and q_3 as fixed, we lighten notation by writing

$$V = \mathcal{V}(q_1, q_2, q_3) \ ; \ Y_v = \sum_{\tau\in G} v(\tau)Y_\tau \ . \tag{15.125}$$

We observe that since $q_j \leq p+1$ we have $q_j - 1 \leq q_j^* \leq q_j$ and thus

$$2^{3p-q_1-q_2-q_3} \leq \operatorname{card} V = 2^{3p-q_1^*-q_2^*-q_3^*} \leq L2^{3p-q_1-q_2-q_3} \ . \tag{15.126}$$

Recalling Definition 15.5.2, a first observation is as follows.

Lemma 15.6.1. *The family of r.v.s $(Y_v)_{v\in V}$ belongs to $\mathcal{B}(N)$.*

Proof. This relies on the fact that the functions $v \in V$ have disjoint supports, and is really obvious from the definition:

$$Y_v = \sum_{i\leq N}\Big(v(U_i) - \int v\mathrm{d}\mu\Big) \ . \qquad \square$$

Recall Definition 15.5.6 of the class $\mathcal{S}^*(\Delta)$. The next task is, given a function $h \in \mathcal{S}^*(\Delta)$, to gather information about the coefficients $a_v(h)$. This information will of course depend on the information we have about h, that is the sets B_k and the coefficients $z(C)$. We think of h as fixed, and for $k \geq 0$ we consider the function R_k on G defined as follows:

$$R_k = 0 \text{ outside } B_k \ . \tag{15.127}$$

$$\text{If } C \in \mathcal{C}_k \text{ then } R_k \text{ is constant } = z(C) \text{ on } C \ . \tag{15.128}$$

These functions will be essential for the rest of this chapter. We may think of them as the parameters which governs "the size of h", and we should keep in mind that they depend on h. We observe that since $\nu(C) = 2^{s(k)-3p}$ for $C \in \mathcal{P}_k$,

$$\sum_{C \in \mathcal{C}_k} 2^{s(k)} z(C) = 2^{3p} \int R_k \mathrm{d}\nu , \tag{15.129}$$

and thus from (15.111)

$$\sum_{k \geq 0} \int R_k \mathrm{d}\nu \leq L\Delta . \tag{15.130}$$

Our basic bound is as follows.

Proposition 15.6.2. *Consider $v \in \mathcal{V}(q_1, q_2, q_3)$, and $j \leq 3$ such that $q_j \leq p$. Then, setting $q = q_1 + q_2 + q_3$,*

$$|a_v(h)| \leq L 2^{3p-q+q_j} \sum_{n_j(\ell) < q_j} 2^{-n_j(\ell)} \int |v| R_\ell \mathrm{d}\nu . \tag{15.131}$$

This means than on the right-hand side the summation is taken only over those values of ℓ for which $n_j(\ell) < q_j$. The reason why only these terms appear is closely related to the fact that h is constant on the elements C of $\mathcal{C}_k$ and that $a_v(\mathbf{1}_C) = 0$ for $C \in \mathcal{P}_k$ as soon as $q_j \leq n_j$ and $q_j \leq p$ for some $j \leq 3$. Let us also observe that $|v| \in \{0, 1\}$ so that $\int |v| R_\ell \mathrm{d}\nu$ is simply the integral of R_ℓ on the support of v. A fundamental idea is that Proposition 15.6.2 offers three different bounds (one for each value of j) for $|a_v(h)|$. We will choose properly between these bounds.

In view of (15.121) and (15.122), to prove Proposition 15.6.2 is suffices to show that

$$\left| \int v h \mathrm{d}\nu \right| \leq 2^{q_j} \sum_{n_j(\ell) < q_j} 2^{-n_j(\ell)} \int |v| R_\ell \mathrm{d}\nu . \tag{15.132}$$

The proof ultimately relies on a simple principle, to which we turn now. We say that a subset of $\mathbb{N}^*$ is a *dyadic interval* if it is of the type $\{r 2^q + 1, \ldots, (r+1) 2^q\}$ for some integers $r, q \geq 0$. The essential property is that given two dyadic intervals I and J with $\operatorname{card} J \geq \operatorname{card} I$, we have

$$I \cap J \neq \emptyset \Rightarrow I \subset J .$$

Lemma 15.6.3. *Consider a dyadic interval I and a partition $\mathcal{Q}$ of I into dyadic intervals. Assume that to each $J \in \mathcal{Q}$ is associated a number $z(J)$. Consider a function $g : I \to \mathbb{R}$. Assume that whenever $J, J' \in \mathcal{Q}$ are adjacent and* $\operatorname{card} J \geq \operatorname{card} J'$,

$$\sigma \in J \ , \ \sigma' \in J' \Rightarrow |g(\sigma) - g(\sigma')| \leq z(J) . \tag{15.133}$$

Then for all σ, σ' in I we have

$$|g(\sigma) - g(\sigma')| \leq 2 \sum_{J \in \mathcal{Q}} z(J) . \tag{15.134}$$

Let us insist that (15.133) holds in particular if $J = J' \in \mathcal{Q}$.

Proof. Let us enumerate $\mathcal{Q}$ as $J_1, J_2, \dots$ in a way that $J_{\ell+1}$ is immediately to the right of J_ℓ. Without loss of generality we may assume that $\sigma \leq \sigma'$. If for some $J \in \mathcal{Q}$ we have $\sigma, \sigma' \in J$, then (15.133) implies $|g(\sigma) - g(\sigma')| \leq z(J)$ and hence (15.134). Otherwise we have $\sigma \in J_{\ell_1}$ and $\sigma' \in J_{\ell_2}$ for some $\ell_1 < \ell_2$. For $\ell_1 < \ell < \ell_2$ consider a point $\sigma_\ell \in J_\ell$. Let $\sigma_{\ell_1} = \sigma$ and $\sigma_{\ell_2} = \sigma'$. Then

$$|g(\sigma') - g(\sigma)| = |g(\sigma_{\ell_2}) - g(\sigma_{\ell_1})| \leq \sum_{\ell_1 \leq \ell < \ell_2} |g(\sigma_{\ell+1}) - g(\sigma_\ell)| \,. \tag{15.135}$$

Moreover, it follows from (15.133) (distinguishing whether $\operatorname{card} J_{\ell+1} \geq \operatorname{card} J_\ell$ or the other way around) that

$$|g(\sigma_{\ell+1}) - g(\sigma_\ell)| \leq z(I_\ell) + z(I_{\ell+1}) \,. \tag{15.136}$$

Combining with (15.135) completes the proof. □

Proof of (15.132). Without loss of generality we assume that $j = 1$. By definition of the class $\mathcal{V}(q_1, q_2, q_3)$, v is of the type $f_1 \otimes f_2 \otimes f_3$. Also, $\nu = \nu_1 \otimes \nu_2 \otimes \nu_3$, where ν_j is the uniform probability on $\{1, \dots, 2^p\}$. Therefore

$$\int vh \mathrm{d}\nu = \int \Big(\int f_1 h \mathrm{d}\nu_1 \Big) f_2 f_3 \mathrm{d}\nu_2 \mathrm{d}\nu_3 \,,$$

and consequently

$$\Big| \int vh \mathrm{d}\nu \Big| \leq \int \Big| \int f_1 h \mathrm{d}\nu_1 \Big| |f_2 f_3| \mathrm{d}\nu_2 \mathrm{d}\nu_3 \,. \tag{15.137}$$

Let us fix τ^2 and τ^3 in $\{1, \dots, 2^p\}$. We shall prove that, setting $g(\sigma) = h(\sigma, \tau^2, \tau^3)$,

$$\Big| \sum_\sigma g(\sigma) f_1(\sigma) \Big| \leq 2^{q_1} \sum_{\ell; n_1(\ell) < q_1} 2^{-n_1(\ell)} \sum_{1 \leq \sigma \leq 2^p} |f_1(\sigma)| R_\ell(\sigma, \tau^2, \tau^3) := 2^{q_1} S \,, \tag{15.138}$$

or, equivalently,

$$\begin{aligned} &\Big| \int h(\sigma, \tau^2, \tau^3) f_1(\sigma, \tau^2, \tau^3) \mathrm{d}\nu_1(\sigma) \Big| \\ &\leq 2^{q_1} \sum_{\ell; n_1(\ell) < q_1} 2^{-n_1(\ell)} \int |f_1(\sigma)| R_\ell(\sigma, \tau^2, \tau^3) \mathrm{d}\nu_1(\sigma) \,. \end{aligned}$$

Therefore, using Fubini's theorem,

$$\int \Big| \int f_1 h \mathrm{d}\nu_1 \Big| |f_2 f_3| \mathrm{d}\nu_2 \mathrm{d}\nu_3 \leq 2^{q_1} \sum_{\ell; n_1(\ell) < q_1} 2^{-n_1(\ell)} \int |v| R_\ell \mathrm{d}\nu \,,$$

and combining with (15.137) yields the result.

We turn to the proof of (15.138). Recalling the definition of the class $\mathcal{V}(q_1, q_2, q_3)$, f_1 is of the type $f_{i,r}$ given by (15.118) for $r = q_1 \le p$ and a certain value of i. Let us consider the dyadic interval

$$I = \{i2^r + 1, \dots, (i+1)2^r\} \ ,$$

so that $f_1(\sigma) = 0$ for $\sigma \notin I$, and

$$\Big|\sum_{\sigma} g(\sigma) f_1(\sigma)\Big| = \Big|\sum_{i2^r < \sigma \le i2^r + 2^{r-1}} (g(\sigma) - g(\sigma + 2^{r-1}))\Big| \ . \tag{15.139}$$

(It is here that we use that $q_1 \le p$.) Therefore to prove (15.138), and since $r = q_1$, it suffices to prove that

$$i2^r < \sigma \le i2^r + 2^{r-1} \Rightarrow |g(\sigma) - g(\sigma + 2^{r-1})| \le 2S \ ,$$

or even that

$$\sigma, \sigma' \in I \Rightarrow |g(\sigma) - g(\sigma')| \le 2S \ . \tag{15.140}$$

This we shall deduce from Lemma 15.6.3. Consider the map $\psi : \{1, \dots, 2^p\} \to G$ given by $\psi(\sigma) = (\sigma, \tau^2, \tau^3)$, and let $I^* = \psi(I)$. Assume first

$$\exists \ell \ , \ n_1(\ell) \ge r = q_1 \ , \ \exists C \in \mathcal{C}_\ell \ , \ C \cap I^* \ne \emptyset \ . \tag{15.141}$$

Since $C \in \mathcal{P}_\ell$, $J = \psi^{-1}(C)$ is a dyadic interval with $\operatorname{card} J = 2^{n_1(\ell)} \ge 2^{q_1} = \operatorname{card} I$, and since $I \cap J \ne \emptyset$ because $C \cap I^* \ne \emptyset$ by (15.141), we have $I \subset J$. Now h is constant on C, so g is constant on $J \supset I$ and (15.140) holds true. So we may assume that (15.141) fails, i.e.

$$\forall \ell, C \in \mathcal{C}_\ell \ , \ C \cap I^* \ne \emptyset \Rightarrow n_1(\ell) < q_1 \ . \tag{15.142}$$

We consider the partition $\mathcal{Q}$ of I that consists of the sets of the type $\psi^{-1}(C)$, where, for some $\ell \ge 1$, $C \in \mathcal{C}_\ell$ and $C \cap I^* \ne \emptyset$. When $J = \psi^{-1}(C) \in \mathcal{Q}$ we set $z(J) = z(C)$. We define

$$\begin{aligned} S^* &= \sum \Big\{ z(J) \ ; \ J \in \mathcal{Q} \Big\} \\ &= \sum \Big\{ z(C) \ ; \ C \cap I^* \ne \emptyset \ , \ C \in \bigcup_{\ell \ge 1} \mathcal{C}_\ell \Big\} \ . \end{aligned} \tag{15.143}$$

We now prove that Condition (15.133) follows from (15.113). When $J, J' \in \mathcal{Q}$ are adjacent we have $J = \psi^{-1}(C)$ and $J' = \psi^{-1}(C')$ where C and C' are adjacent and $C \in \mathcal{C}_\ell$, $C' \in \mathcal{C}_{\ell'}$. When moreover $\operatorname{card} J \ge \operatorname{card} J'$ we may assume that $\ell \ge \ell'$. Indeed, when $\operatorname{card} J > \operatorname{card} J'$ we have $\ell > \ell'$, and when $\operatorname{card} J = \operatorname{card} J'$ we can exchange J and J' if necessary. Then $C' \not\subset B_k$ for $k > \ell$, and then by (15.113) we have $|h(\tau) - h(\tau')| \le z(C) = z(I)$ for $\tau \in C$ and $\tau' \in C'$ and this proves (15.133).

Lemma 15.6.3 then implies that $|g(\sigma) - g(\sigma')| \leq 2S^*$ for each $\sigma, \sigma' \in I$. Recalling the quantity S of (15.138) we now prove that $S^* = S$, finishing the proof of (15.140) and of the lemma.

We observe that if $C \in \mathcal{C}_\ell$ is such that $C \cap I^* \neq \emptyset$, then $J = \psi^{-1}(C)$ is a dyadic interval with $J \cap I \neq \emptyset$. Moreover since (15.142) implies $\operatorname{card} J = 2^{n_1(\ell)} \leq \operatorname{card} I = 2^{q_1}$ we have $J \subset I$, so that $\operatorname{card}(C \cap I^*) = 2^{n_1(\ell)}$. Consequently,

$$z(C) = 2^{-n_1(\ell)} \sum_{\sigma \in J} |f_1(\sigma)| R_\ell(\sigma, \tau^2, \tau^3) \tag{15.144}$$

because there are $2^{n_1(\ell)}$ non-zero terms in the summation, and for each of these terms $|f_1(\sigma)| = 1$ and $R_\ell(\sigma, \tau^2, \tau^3) = z(C)$. Summation of the relations (15.144) over $C \in \bigcup_{\ell \geq 1} \mathcal{C}_\ell$ with $C \cap I^* \neq \emptyset$ then proves that $S = S^*$. □

15.7 Probability, II

We go back to the problem of bounding the quantities (15.124):

$$\sum_{v \in \mathcal{V}(q_1, q_2, q_3)} |a_v(h)||Y_v| \,. \tag{15.145}$$

We think of q_1, q_2 and q_3 as fixed, and we write $q = q_1 + q_2 + q_3$, $V = \mathcal{V}(q_1, q_2, q_3)$. The plan is to combine the bound of Proposition 15.6.2 with probabilistic estimates. Computation of $\mathsf{E} Y_v^2$ shows that we should think of $|Y_v|$ as being typically of size about $\sqrt{m_0} 2^{q/2}$. The trouble of course comes from the fact that some of the quantities $|Y_v|$ are much larger than their typical values. Our first goal is to provide a simple argument showing than none of these quantities is larger than about p times its typical value. It will settles the case $q \geq p$.

Lemma 15.7.1. *With probability $\geq 1 - L\exp(-100p)$ the following occurs: for each choice of q_j, $j \leq 3$ we have*

$$\max_{v \in V} |Y_v| \leq Lp\sqrt{m_0} 2^{q/2} \,. \tag{15.146}$$

Proof. Since the family $(Y_v)_{v \in V}$ belongs to $\mathcal{B}(N)$ by Lemma 15.6.1, it follows from (15.96) that for $u > 0$,

$$\mathsf{P}(|Y_v| \geq u) \leq 2\exp\Big(-\frac{1}{L}\min\Big(u, \frac{u^2 \operatorname{card} V}{N}\Big)\Big) \,.$$

Now, using (15.99) and (15.15) we have $\operatorname{card} V/N \geq 2^{-q-1}/m_0$, so that

$$\mathsf{P}(|Y_v| \geq u) \leq 2\exp\Big(-\frac{1}{L}\min\Big(u, \frac{u^2}{m_0 2^q}\Big)\Big) \,, \tag{15.147}$$

and

$$\mathsf{P}(\max_{v\in V}|Y_v|\geq u)\leq 2\,\mathrm{card}\,V\exp\Big(-\frac{1}{L}\min\Big(u,\frac{u^2}{m_0 2^q}\Big)\Big)\ . \tag{15.148}$$

Since $\mathrm{card}\,V\leq 2^{3p}$, the result follows. □

Lemma 15.7.2. *Assume (15.146). Then for $h\in\mathcal{S}^*(\Delta)$ we have*

$$\sum\sum_{v\in\mathcal{V}(q_1,q_2,q_3)}|a_v(h)||Y_v|\leq L2^{3p}\sqrt{m_0}\Delta\ , \tag{15.149}$$

where the first summation is over $1\leq q_1,q_2,q_3\leq p+1, q_1+q_2+q_3\geq p$.

Proof. First, if $q_1=q_2=q_3=p+1$, then $Y_v=\sum_{\tau\in G}Y_\tau=0$ for the unique element v of $\mathcal{V}(q_1,q_2,q_3)$. Next, given $v\in\mathcal{V}(q_1,q_2,q_3)$ with $q=q_1+q_2+q_3<3(p+1)$, we use the bound (15.132). We choose $j\leq 3$ such that $q_j\leq q/3<p+1$. Thus $q_j\leq p$. We use the trivial bound $n_j(\ell)\geq 0$ and we get

$$\Big|\int vh\mathrm{d}\nu\Big|\leq 2^{q/3}\sum_{\ell\geq 1}\int|v|R_\ell\mathrm{d}\nu\ ,$$

and, recalling (15.121) and (15.122),

$$|a_v(h)|\leq L2^{3p-2q/3}\sum_{\ell\geq 1}\int|v|R_\ell\mathrm{d}\nu\ .$$

Thus, since the functions $v\in\mathcal{V}(q_1,q_2,q_3)$ have disjoint support and satisfy $|v|\leq 1$, and using (15.130),

$$\sum_{v\in\mathcal{V}(q_1,q_2,q_3)}|a_v(h)|\leq L2^{3p-2q/3}\Delta\ .$$

Combining with (15.146) yields

$$\sum_{v\in\mathcal{V}(q_1,q_2,q_3)}|a_v(h)||Y_v|\leq Lp2^{3p-q/6}\sqrt{m_0}\Delta\ .$$

Now there are at most q^3 possible choices of q_1,q_2,q_3 for which $q_1+q_2+q_3=q$. To conclude we simply use that $p\sum_{q\geq p}q^3 2^{-q/6}\leq L$. □

The previous argument does a lot better than taking care only of the values of $q\geq p$, but for small values of q we need to be much more sophisticated anyway. From now on we always assume that $q\leq p$. The following lemma is closely connected to Lemmas 15.5.5 and 15.5.4. It controls the possible number of quantities $|Y_v|$ which are larger than $2^{k/2}$ times their typical values. It might help to know right now that the crucial case of this lemma is $k\sim q$.

Lemma 15.7.3. *With probability* $\geq 1 - \exp(-100p)$ *the following occurs. Consider* q_1, q_2, q_3 *with* $q = q_1 + q_2 + q_3 \leq p$. *Consider any* $k \leq q \leq p$ *and* $r := \lfloor \operatorname{card} V / M_{k+2} \rfloor$. *Then, if* $r \geq 1$, *for each subset* W *of* $V = \mathcal{V}(q_1, q_2, q_3)$ *with* $\operatorname{card} W = r$ *we have*

$$\sum_{v \in W} |Y_v| \leq L r \sqrt{m_0} 2^{k/2} 2^{q/2} \,. \tag{15.150}$$

Moreover, if $r = 0$ *then*

$$\max_{v \in V} |Y_v| \leq L \sqrt{m_0} 2^{k/2} 2^{q/2} \,. \tag{15.151}$$

Proof. Assume first that $r \geq 1$, so that $\operatorname{card} V / r \leq 2M_{k+2}$. Let us first prove that $r 2^k \geq p/L$. If $M_{k+2} \leq 2^p$ this holds because $\operatorname{card} V \geq 2^{3p-q} \geq 2^{2p}$, so that $r \geq 2^p$. If $M_{k+2} \geq 2^p$ this holds because then $2^k \geq p/L$.

For $v \in W$ consider $\eta_v = \pm 1$, and let $\eta_v = 0$ if $v \notin W$. Then $\sum_v \eta_v^2 = \operatorname{card} W = r$. Since $\operatorname{card} V = 2^{3p-q}$, (15.99) implies (since $N \leq 2^{3p+1} m_0$)

$$\mathsf{P}\Big(\Big|\sum_{v \in V} \eta_v Y_v\Big| \geq u\Big) \leq 2 \exp\Big(-\frac{1}{L} \min\Big(u, \frac{u^2}{2^q m_0 r}\Big)\Big) \,. \tag{15.152}$$

There are at most

$$\binom{\operatorname{card} V}{r} \leq \Big(\frac{e \operatorname{card} V}{r}\Big)^r \leq 2^{L 2^k r}$$

choices of W, and given W, there are at most 2^r choices for the signs $(\eta_v)_{v \in W}$. Consider a parameter $A \geq 1$. For $u = A r \sqrt{m_0} 2^{k/2} 2^{q/2}$, since $2^{q/2 + k/2} \geq 2^k$, we have $\min(u, u^2 / 2^q m_0 r) \geq A 2^k r$, so with probability $\geq 1 - \exp(2^k r(L - A/L))$ we have $\sum_{v \in W} \eta_v Y_v \leq u$ for each choice of W and (η_v), and hence $\sum_{v \in W} |Y_v| \leq u$ for each choice of W. Therefore, since $r 2^k \geq p/L$ it suffices to take for A a large enough constant to obtain $1 - \exp(2^k r(L - A/L)) \geq 1 - L \exp(-100p)$.

Finally assume that $r = 0$ so that $2^{2p} \leq \operatorname{card} V \leq M_{k+2}$ and thus $p \leq L 2^k$. Moreover since $k \leq q$ for $u = \sqrt{m_0} 2^{k/2} 2^{q/2}$ we have $\min(u, u^2/(m_0 2^q)) \geq 2^k$ and the result then follows from (15.148). □

Lemma 15.7.4. *Assume that the event of Lemma 15.7.3 occurs. Then for each numbers* $(a_v)_{v \in V}$, $a_v > 0$ *and any value of* $k \leq q \leq p$ *we have*

$$\sum_{v \in V} a_v |Y_v| \leq L \sqrt{m_0} 2^{k/2 + q/2} \Big(\sum_{v \in V} a_v + \frac{\operatorname{card} V}{M_{k+2}} \max_v a_v\Big) \,. \tag{15.153}$$

It is correct to think of the last term as an annoying but lower order term.

Proof. Let $r = \lfloor \operatorname{card} V / M_{k+2} \rfloor$. If $r = 0$, the result follows immediately from (15.151), so we may assume that $r \geq 1$. Let us enumerate V as $\{v_1, \ldots\}$ in such a way that the sequence $(|Y_{v_\ell}|)$ is non-increasing. Then (15.150) implies

$$r|Y_{v_r}| \le \sum_{\ell \le r} |Y_{v_\ell}| \le Lr\sqrt{m_0}2^{q/2+k/2} ,$$

and thus $|Y_{v_\ell}| \le L\sqrt{m_0}2^{q/2+k/2}$ for $\ell \ge r$. We then use that

$$\sum_{\ell \le r} a_{v_\ell}|Y_{v_\ell}| \le \max_{v\in V} a_v \sum_{\ell \le r} |Y_{v_\ell}| \le Lr\sqrt{m_0}2^{k/2}2^{q/2} \max_{v\in V} a_v ,$$

and

$$\sum_{\ell > r} a_{v_\ell}|Y_{v_\ell}| \le \max_{\ell > r} |Y_{v_\ell}| \sum_{v\in V} a_v \le L\sqrt{m_0}2^{k/2}2^{q/2} \sum_{v\in V} a_v . \qquad \square$$

Inequality (15.153) is our fundamental tool to control the size of the quantities Y_v. To apply it we will split the sum in (15.145) into several pieces, which we define now. We define $k_0 = \lfloor q/4 \rfloor$, and we define k_1 as the largest integer k such that

$$\forall\, j \le 3 \ , \ n_j(k) < q_j \ . \tag{15.154}$$

Thus

$$k_1 = \sum_{j\le 3} n_j(k_1) \le \sum_{j\le 3}(q_j - 1) \le q - 3 \ . \tag{15.155}$$

In particular, since $q \le p$,

$$k \le k_1 \Rightarrow s(k) = k \ . \tag{15.156}$$

We define

$$V_{k_1} = \Big\{ v \in V \ ; \ \int |v| R_{k_1} \mathrm{d}\nu \ne 0 \Big\} \ . \tag{15.157}$$

For $k > k_0$ we define V_k recursively:

$$V_k = \Big\{ v \in V \ ; \ \int |v| R_k \mathrm{d}\nu \ne 0 \Big\} \setminus V_{k+1} \ . \tag{15.158}$$

Finally we define

$$V_{k_0} = V \setminus V_{k_0+1} \ . \tag{15.159}$$

The following bound is pretty crude but sufficient.

Lemma 15.7.5. *For $k_0 < k \le k_1$ we have*

$$\operatorname{card} V_k \le L\Delta \frac{2^{3p-k}}{M_k} \ . \tag{15.160}$$

Proof. The support of R_k is the union of the sets $C \in \mathcal{C}_k$, so that when $\int |v| R_k \mathrm{d}\nu \ne 0$ there exists $C \in \mathcal{C}_k$ with $\int_C |v| R_k \mathrm{d}\nu \ne 0$. The support of v is a product $I_1 \times I_2 \times I_3$ of dyadic intervals with $\operatorname{card} I_j = 2^{q_j}$, while the sets $C \in \mathcal{P}_k$ are products of dyadic intervals $J_1 \times J_2 \times J_3$ where $\operatorname{card} J_j = 2^{n_j(k)}$. Since $k \le k_1$ we have $n_j(k) < q_j$ for each $j \le 3$, so that if the support of v meets a set $C \in P_k$ it contains it entirely. Consequently $\operatorname{card} V_k \le \operatorname{card} \mathcal{C}_k$. Since $s(k) = k$, combining (15.112) and (15.111) yields that $\operatorname{card} \mathcal{C}_k \le L\Delta 2^{3p-k}/M_k$. $\square$

Let us set

$$\beta_k (= \beta_k(h)) = \int R_k \mathrm{d}\nu \;, \tag{15.161}$$

so that using (15.130) we obtain

$$\sum_{k \ge 1} \beta_k \le L\Delta \;. \tag{15.162}$$

Lemma 15.7.6. *Assume that $k_0 < k \le k_1$, and assume that the event of Lemma 15.7.3 occurs. Then*

$$\sum_{v \in V_k} |a_v(h)||Y_v| \le L\sqrt{m_0}2^{3p}2^{(k-q)/6}(\beta_k + \beta_{k-1}) \\ + L\sqrt{m_0}2^{3p}\Big(\Delta\frac{2^{2q}}{M_{k-1}} + \frac{2^q}{M_{k+1}}\Big) \;. \tag{15.163}$$

In this bound the crucial term is the factor $2^{(k-q)/6}$, that will sum nicely over $q \ge k$. There is plenty of room for the second order quantities represented by the second term.

Proof. We recall the bound (15.131):

$$|a_v(h)| \le L2^{3p-q+q_j} \sum_{n_j(\ell) < q_j} 2^{-n_j(\ell)} \int |v| R_\ell \mathrm{d}\nu \;.$$

When $v \in V_k$ we know by definition of V_k that for $k+1 \le \ell \le k_1$ we have $\int |v| R_\ell \mathrm{d}\nu = 0$. Consequently,

$$|a_v(h)| \le L2^{3p-q+q_j} \sum_{\ell \le k, n_j(\ell) < q_j} 2^{-n_j(\ell)} \int |v| R_\ell \mathrm{d}\nu \;. \tag{15.164}$$

We choose j such that

$$q_j - n_j(k) \le \frac{1}{3}\sum_{j' \le 3}(q_{j'} - n_{j'}(k)) = \frac{1}{3}(q - s(k)) = \frac{1}{3}(q-k) \;.$$

Using that $n_j(k-1) \ge n_j(k) - 1$ we get for $\ell = k$ or $\ell = k-1$ that $2^{q_j - n_j(\ell)} \le L2^{(q-k)/3}$. Using the crude bound $q_j - n_j(\ell) \le q$ for $\ell \le k-2$, (15.164) yields

$$|a_v(h)| \le L2^{3p}2^{-2q/3-k/3} \int |v|(R_k + R_{k-1})\mathrm{d}\nu + L2^{3p} \sum_{\ell \le k-2} \int |v| R_\ell \mathrm{d}\nu \;. \tag{15.165}$$

Since $R_\ell \le M_{\ell+1}$, we have, using that $\int |v|\mathrm{d}\nu = 2^{k-3p}$ for $v \in V_k$,

$$\sum_{\ell \le k-2} \int |v| R_\ell \mathrm{d}\nu \le LM_{k-1} \int |v| \mathrm{d}\nu \le L2^{q-3p}M_{k-1} \;, \tag{15.166}$$

and therefore

$$\sum_{v\in V_k} |a_v(h)| \le L2^{3p}2^{-2q/3-k/3}(\beta_k+\beta_{k-1}) + L2^q M_{k-1}\,\mathrm{card}\, V_k \ .$$

Using (15.160), and since $M_k = M_{k-1}^2$, we obtain

$$\sum_{v\in V_k} |a_v(h)| \le L2^{3p}2^{-2q/3-k/3}(\beta_k+\beta_{k-1}) + L2^{3p}\Delta\frac{2^q}{M_{k-1}} \ . \tag{15.167}$$

Also, as in (15.166) we have $\sum_{\ell\le k}\int |v|R_\ell \mathrm{d}\nu \le L2^{q-3p}M_{k+1}$ and thus

$$|a_v(h)| \le L2^q M_{k+1} \ . \tag{15.168}$$

To conclude we use the bound (15.153) with $a_v = a_v(h)$ for $v\in V_k$ and $a_v = 0$ otherwise. We note that $\mathrm{card}\, V \le 2^{3p-q}$, $M_{k+2} = M_{k+1}^2$, and we use crude bounds such as $2^{k/2+q/2} \le 2^q$ for the secondary terms. □

It remains to take care of the case $k = k_0$, but (because $k_0 \le q/4$) this term is not critical and does not need great care, as the following shows.

Lemma 15.7.7. *Assume that the event of Lemma 15.7.3 occurs, and recall that $k_0 = \lfloor q/4 \rfloor$. Then*

$$\sum_{v\in V_{k_0}} |a_v(h)||Y_v| \le L\sqrt{m_0}2^{3p}\Big(2^{-q/24}\Delta + \frac{2^{2q}}{M_{k_0+1}}\Big) \ . \tag{15.169}$$

Proof. We use the bound (15.131). We choose j such that $q_j \le q/3$ and we observe that for $v\in V_{k_0}$ we have $\int |v|R_\ell \mathrm{d}\nu \ne 0 \Rightarrow \ell \le k_0$ to obtain

$$v\in V_{k_0} \Rightarrow |a_v(h)| \le L2^{3p}2^{-2q/3}\sum_{\ell\le k_0}\int |v|R_\ell \mathrm{d}\nu \ .$$

Consequently, using (15.161) and (15.162),

$$\sum_{v\in V_{k_0}} |a_v(h)| \le L2^{3p}2^{-2q/3}\Delta$$

and, since $R_\ell \le M_{\ell+1}$ and $\int |v|\mathrm{d}\nu = 2^{k_0-3p}$, we have (crudely)

$$\max_{v\in V_{k_0}} |a_v(h)| \le L2^q M_{k_0+1} \ .$$

We now use (15.153) for $k = k_0$ and with $a_v = a_v(h)$ for $v\in V_{k_0}$ and $a_v = 0$ otherwise to obtain

$$\sum_{v\in V_{k_0}} |a_v(h)||Y_v| \le L\sqrt{m_0}2^{3p}\Big(2^{k_0/2-q/6}\Delta + \frac{2^{2q}}{M_{k_0+1}}\Big) \ ,$$

and the result since $k_0 \le q/4$. □

Combining Lemmas 15.7.6 and 15.7.7 we obtain the following.

Proposition 15.7.8. *Assume that the event of Lemma 15.7.3 occurs. Then*

$$\sum_{v\in V}|a_v(h)||Y_v|\le L\sqrt{m_0}2^{3p}\Big(2^{-q/24}\Delta+\sum_{k_0\le k\le k_1}(A(k,q)+B(k,q))\Big)\,, \tag{15.170}$$

where

$$A(k,q):=A(k,q,h)=2^{(k-q)/6}(\beta_k+\beta_{k-1})\,. \tag{15.171}$$

$$B(k,q):=\Delta\frac{2^{2q}}{M_{k-1}}+\frac{2^q}{M_{k+1}}\,. \tag{15.172}$$

Proof of Theorem 15.5.7. We assume that the events of Lemmas 15.7.1 and 15.7.3 occur. We then prove that (15.114) holds. Combining (15.115) and (15.123), we have to prove that

$$\sum_{3\le q\le 3p+3}\ \sum_{q_1+q_2+q_3=q}\ \sum_{v\in\mathcal{V}(q_1,q_2,q_3)}|a_v(h)||Y_v|\le L\sqrt{m_0}2^{3p}\Delta\,. \tag{15.173}$$

Lemma 15.7.2 takes care of the summation over $q\ge p$. Control of the summation for $q\le p$ will be obtained by summing the inequalities (15.170) and interchanging the summation in k and q. Given q there are at most q^3 possible values of (q_1,q_2,q_3) with $q=q_1+q_2+q_3$, and

$$\sum_{q\ge1}q^3 2^{-q/24}\le L\,.$$

Also,

$$\sum_{q\ge1}\sum_{k\ge k_0}q^3B(k,q)\le L(1+\Delta)\le L\Delta\,,$$

because $k_0=\lfloor q/4\rfloor$ and M_k is doubly exponential in k. It remains only to take care of the contribution of the term $A(k,q)$. For this we have to be more sophisticated in counting the number of triples (q_1,q_2,q_3) for which this term occurs. Recalling (15.154) and that $\sum_{j\le3}n_j(k)=k$, we observe that there are at most $(q-k)^3$ such triples. Now,

$$\begin{aligned}\sum_{q\ge1}\sum_{k\le q}(q-k)^3 2^{(k-q)/12}(\beta_k+\beta_{k+1})&=\sum_{k\ge0}(\beta_k+\beta_{k+1})\sum_{q\ge k}(q-k)^3 2^{(k-q)/12}\\&\le L\sum_{k\ge0}(\beta_k+\beta_{k+1})\\&\le L\Delta\,,\end{aligned}$$

using (15.162) in the last line. □

References

1. Shor, P.W., Yukich, J.E.: Minimax grid matching and empirical measures. Ann. Probab. **19**(3), 1338–1348 (1991)
2. Talagrand, M.: The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. Ann. Probab. **22**, 919–959 (1944)

16. Applications to Banach Space Theory

16.1 Cotype of Operators from $C(K)$

We start by recalling some basic definitions. More background can be found in classical books such as [6] or [29].

Given an operator U (i.e. a continuous linear map) from a Banach space X to a Banach space Y and a number $q \geq 2$, we denote by $C_q^g(U)$ its Gaussian cotype-q constant, that is, the smallest number A (possibly infinite) for which, given any integer n, any elements $x_1, \ldots, x_n$ of X, we have

$$\Big(\sum_{i\leq n} \|U(x_i)\|^q\Big)^{1/q} \leq A\mathsf{E}\Big\|\sum_{i\leq n} g_i x_i\Big\| \,.$$

Here, $(g_i)_{i\leq n}$ are i.i.d. standard Gaussian r.v.s, the norm of $U(x_i)$ is in Y and the norm of $\sum_{i\leq n} g_i x_i$ is in X.

The occurrence of the quantity

$$\mathsf{E}\Big\|\sum_{i\leq n} g_i x_i\Big\| = \mathsf{E}\sup\Big\{\sum_{i\leq n} g_i x^*(x_i) \;;\; x^* \in X^* \,,\; \|x^*\| \leq 1\Big\}$$

suggests that results on Gaussian processes will bear on this notion. This is only true to a small extent. Because of the statement "for any elements $x_1, x_2, \ldots, x_n$", geometry dominates. Geometry dominates less in the case when X is a space of continuous functions on a compact space. Then the theory of processes really bears on the notion of cotype. We specialize to this case in the present section.

Given a number $q \geq 2$, we define the Rademacher cotype-q constant $C_q^r(U)$ as the smallest number A (possibly infinite) such that, given any integer n, any elements $(x_i)_{i\leq n}$ of X, we have

$$\Big(\sum_{i\leq n} \|U(x_i)\|^q\Big)^{1/q} \leq A\mathsf{E}\Big\|\sum_{i\leq n} \varepsilon_i x_i\Big\| \,, \tag{16.1}$$

where $(\varepsilon_i)_{i\leq n}$ are i.i.d. Bernoulli r.v.s. The name "Rademacher cotype" stems from the fact that Bernoulli r.v.s are usually (but inappropriately) called Rademacher r.v.s in Banach space theory. Since Bernoulli processes are tricker

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2_16,

than Gaussian processes we expect that Rademacher cotype will be harder to understand than Gaussian cotype. This certainly seems to be the case.

Given $q \geq 1$, we define the $(q,1)$-summing norm $\|U\|_{q,1}$ of U as the smallest number A (possibly infinite) such that, for any integer n, any vectors $x_1, \ldots, x_n$ of X we have

$$\Big(\sum_{i \leq n} \|U(x_i)\|^q\Big)^{1/q} \leq A \sup\Big\{\sum_{i \leq n} |x^*(x_i)| \; ; \; x^* \in X^* \, , \, \|x^*\| \leq 1\Big\} \, . \tag{16.2}$$

For an operator W from Y to another Banach space we have

$$\|W \circ U\|_{q,1} \leq \|W\| \|U\|_{q,1} \, . \tag{16.3}$$

The proof is immediate. These quantities are related as follows.

Proposition 16.1.1. *We have*

$$C_q^g(U) \leq \sqrt{\frac{\pi}{2}} C_q^r(U) \tag{16.4}$$

$$\|U\|_{q,1} \leq C_q^r(U) \, . \tag{16.5}$$

Proof. To prove (16.4) we simply observe that (3.27) implies

$$\mathsf{E}\Big\|\sum_{i \leq n} \varepsilon_i x_i\Big\| \leq \sqrt{\pi/2}\,\mathsf{E}\Big\|\sum_{i \leq n} g_i x_i\Big\| \, .$$

To prove (16.5) we observe that

$$\begin{aligned}\Big\|\sum_{i \leq n} \varepsilon_i x_i\Big\| &= \sup\Big\{\sum_{i \leq n} \varepsilon_i x^*(x_i) \; ; \; x^* \in X^* \, , \, \|x^*\| \leq 1\Big\} \\ &\leq \sup\Big\{\sum_{i \leq n} |x^*(x_i)| \; ; \; x^* \in X^* \, , \, \|x^*\| \leq 1\Big\} \, . \end{aligned}$$ □

In the rest of this section we specialize to the case where X is the space ℓ_N^∞ of sequences $x = (x_j)_{j \leq N}$ provided with the norm

$$\|x\| = \sup_{j \leq N} |x_j| \, .$$

It is possible to show that similar results hold in the case where $X = C(W)$, the space of continuous functions over a compact topological space W. This is deduced from the case $X = \ell_N^\infty$ using a reduction technique which is unrelated to the methods of this book, see [16].

The proof of (16.5) amounts to apply the inequality (5.4). One should then expect that questions about the exact relationship between the quantities $C_q^g(U)$ (related to Gaussian processes) and the one hand and $C_q^r(U)$ and $U_{q,1}(U)$ (related to Bernoulli processes) on the other hand will be related through Theorem 5.1.5.

Theorem 16.1.2. *Given $q \geq 2$ and an operator U from ℓ_N^∞ to a Banach space Y, we have*

$$\sqrt{\frac{2}{\pi}}\max(C_q^g(U), \|U\|_{q,1}) \leq C_q^r(U) \qquad (16.6)$$
$$\leq L\max(C_q^g(U), \|U\|_{q,1}) \; .$$

We observe right away that the left-hand side inequality is a consequence of Proposition 16.1.1. It does not appear to be known whether this result holds for operators between any two Banach spaces.

Proposition 16.1.3. *Consider vectors $(x_i)_{i\leq n}$ in ℓ_N^∞. Then we can find a decomposition $x_i = x_i' + x_i''$ such that*

$$\mathsf{E}\Big\|\sum_{i\leq n} g_i x_i'\Big\| \leq L\mathsf{E}\Big\|\sum_{i\leq n}\varepsilon_i x_i\Big\| \qquad (16.7)$$

and

$$\sup\Big\{\sum_{i\leq n}|x^*(x_i'')| \;\; ; \; x^* \in \ell_N^1 \; , \; \|x^*\|_1 \leq 1\Big\} \leq L\mathsf{E}\Big\|\sum_{i\leq n}\varepsilon_i x_i\Big\| \; . \qquad (16.8)$$

Proof of Theorem 16.1.2. We shall prove that

$$C_q^r(U) \leq L(C_q^g(U) + \|U\|_{q,1}) \; , \qquad (16.9)$$

from which the right-hand side inequality of (16.6) follows.

Let us consider a decomposition $x_i = x_i' + x_i''$ as in Proposition 16.1.3. Then

$$\Big(\sum_{i\leq n}\|U(x_i')\|^q\Big)^{1/q} \leq LC_q^g(U)\mathsf{E}\Big\|\sum_{i\leq n}\varepsilon_i x_i\Big\| \qquad (16.10)$$

$$\Big(\sum_{i\leq n}\|U(x_i'')\|^q\Big)^{1/q} \leq L\|U\|_{q,1}\mathsf{E}\Big\|\sum_{i\leq n}\varepsilon_i x_i\Big\| . \qquad (16.11)$$

Since $\|U(x_i)\| \leq \|U(x_i')\| + \|U(x_i'')\|$, the triangle inequality in ℓ_n^q implies

$$\Big(\sum_{i\leq n}\|U(x_i)\|^q\Big)^{1/q} \leq \Big(\sum_{i\leq n}\|U(x_i')\|^q\Big)^{1/q} + \Big(\sum_{i\leq n}\|U(x_i'')\|^q\Big)^{1/q} ,$$

and combining with (16.10) and (16.11), this proves (16.9). □

Proof of Proposition 16.1.3. Let us write $x_i = (x_{ij})_{1\leq j\leq N}$. For $1 \leq j \leq N$, consider $t_j \in \mathbb{R}^n$ given by $t_j = (x_{ij})_{i\leq n}$. Let $t_0 = 0$ and consider $T = \{t_0, t_1, \ldots, t_N\}$, so that

$$b(T) = \mathsf{E}\max\Big(0, \sup_{1\leq j\leq N}\sum_{i\leq n}\varepsilon_i x_{ij}\Big) \leq \mathsf{E}\Big\|\sum_{i\leq n}\varepsilon_i x_i\Big\| \; . \qquad (16.12)$$

Theorem 5.1.5 provides for $0 \le j \le N$ a decomposition $t_j = t'_j + t''_j$, where $t'_j = (x'_{ij})_{i\le n}, t''_j = (x''_{ij})_{i\le n}$, and

$$\mathsf{E} \sup_{0\le j\le N} \sum_{i\le n} g_i x'_{ij} \le Lb(T) \tag{16.13}$$

$$\forall j \le N \,, \sum_{i\le n} |x''_{ij}| \le Lb(T) \,. \tag{16.14}$$

Since $t_0 = 0 = t'_0 + t''_0$, for each $0 \le j \le N$ we can replace t'_j by $t'_j - t'_0$ and t''_j by $t''_j - t''_0$, so that we may assume that $t'_0 = t''_0 = 0$. For $i \le n$, we consider the elements $x'_i = (x'_{ij})_{j\le N}$ and $x''_i = (x''_{ij})_{j\le N}$ of ℓ^∞_N. Thus $x_i = x'_i + x''_i$.

Next we observe that in the quantity

$$\sup\Big\{\sum_{i\le n} |x^*(x''_i)| \;;\; \|x^*\|_1 \le 1\Big\}\,,$$

by convexity the supremum is attained at an extreme point of the unit ball of $(\ell^\infty_N)^* = \ell^1_N$. These extreme points are the canonical basis vectors, so that (16.14) implies (16.8).

Lemma 2.2.1 implies that when the process $(X_t)_{t\in T}$ is symmetric and $X_s = 0$ for some s, then

$$\mathsf{E}\sup_{t\in T} |X_t| \le \mathsf{E} \sup_{s,t\in T} |X_s - X_t| = 2\mathsf{E}\sup_{t\in T} X_t \,.$$

Using this for $X_t = \sum_{i\le n} g_i x_i$ when $t = (x_i)_{i\le n}$ and $T = \{t'_0, t'_1, \ldots, t'_N\}$ yields (using that $X_{t'_0} = 0$ since $t'_0 = 0$),

$$\mathsf{E}\Big\|\sum_{i\le n} g_i x'_i\Big\| = \mathsf{E} \sup_{0\le j\le N} \Big|\sum_{i\le n} g_i x'_{ij}\Big| \le 2\mathsf{E} \sup_{0\le j\le N} \sum_{i\le n} g_i x'_{ij} \le Lb(T) \,. \qquad \square$$

The following very interesting question was raised by S. Kwapien (private communication).

Research problem 16.1.4. Does there exist a universal constant L with the following property. Given any Banach space E and elements $x_1, \ldots, x_n$ of E, we can write $x_i = x'_i + x''_i$ where

$$\mathsf{E}\Big\|\sum_{i\le n} g_i x'_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\le n} \varepsilon_i x_i\Big\| \;;\; \max_{\eta_i=\pm1}\Big\|\sum_{i\le n} \eta_i x''_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\le n} \varepsilon_i x_i\Big\| \,.$$

When $E = \ell^\infty_N$, Proposition 16.1.3 provides a positive answer when $E = \ell^\infty_N$. A positive answer to Problem 16.1.4 in general would allow to extend Theorem 16.1.2 to the case of operators between any two Banach spaces.

In the rest of this section we turn to another topic, the computation of $C^g_q(U)$ (still in the case where $X = \ell^\infty_N$). S. Montgomery-Smith discovered an effective approach to this computation, using again the theory of Gaussian processes. We denote by $H_q(U)$ the quantity

$$H_q(U) = \sup\Big\{ \big(\sum_{i\le n} \|U(x_i)\|^q\big)^{1/q} \Big\} ,$$

where the supremum is taken over all n and all families $(x_i)_{i\le n}$ with $x_i = \sum_{k\ge 2} a_{ik} u_k$, where $u_k \in \ell_N^\infty$, the elements $(u_k)_{k\ge 2}$ have disjoint supports, $\|u_k\|_\infty \le 1$, and the numbers a_{ik} satisfy

$$\forall k \ge 2\,,\ \sum_{i=1}^n a_{ik}^2 \le \frac{1}{\log k}\ . \tag{16.15}$$

Theorem 16.1.5 (S. Montgomery-Smith [17]). *For all $U : \ell_N^\infty \to Y$ we have*

$$\frac{1}{L} H_q(U) \le C_q^g(U) \le L H_q(U)\,.$$

Proof. Suppose first that for $i \le n$ the elements x_i satisfy $x_i = \sum_{k\ge 2} a_{ik} u_k$, where the elements $(u_k)_{k\ge 2}$ of ℓ_N^∞ have disjoint support, $\|u_k\|_\infty \le 1$, and (16.15) holds. Then, defining $u_k = (u_{kj})_{j\le N}$, we have

$$\Big\|\sum_{i\le n} g_i x_i\Big\| = \sup_{j\le N}\Big|\sum_{i\le n} g_i \sum_{k\ge 2} a_{ik} u_{kj}\Big| = \sup_{j\le N}\Big|\sum_{k\ge 2} X_k u_{kj}\Big| ,$$

where $X_k = \sum_{i\le n} g_i a_{ik}$.

Since the elements u_k have disjoint support and $\|u_k\|_\infty \le 1$, for each j we have $\sum_{k\ge 2} |u_{kj}| \le 1$, and hence $\sup_{j\le N} |\sum_{k\ge 2} X_k u_{kj}| \le \sup_k |X_k|$. Now the r.v.s X_k are Gaussian and by (16.15) we have $\mathsf{E} X_k^2 \le 1/\log k$. Thus $\mathsf{E}\sup_{k\ge 2} |X_k| \le L$ by Proposition 2.4.16, and therefore $\mathsf{E}\|\sum_{i\le n} g_i x_i\| \le L$. Hence

$$\Big(\sum_{i\le n} \|U(x_i)\|^q\Big)^{1/q} \le L C_q^g(U)\,,$$

and thus $H_q(U) \le L C_q^g(U)$.

We now turn to the proof of the converse inequality. Consider for $i \le n$ elements x_i in ℓ_N^∞ , $x_i = (x_{ij})_{j\le N}$ and

$$S := \mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| = \mathsf{E}\sup_{j\le N}\Big|\sum_{i\le n} g_i x_{ij}\Big| \ge \mathsf{E}\sup_T \sum_{i\le n} g_i t_i$$

where

$$T = \{0\} \cup \{t_j = (x_{ij})_{i\le n}\ ;\ 1 \le j \le N\}\ .$$

Since $0 \in T$, by Theorem 2.4.18 we can find a sequence $(a_k)_{k\ge 2}$ of points of ℓ_n^2, with $\|a_k\|_2 \le 1/\sqrt{\log k}$, for which

$$T \subset LS\,\mathrm{conv}(\{a_k\ ;\ k \ge 2\} \cup \{0\})\ .$$

Consequently, for each $j \le N$, we can find numbers $(u_{jk})_{k\ge 2}$ with $t_j = \sum_{k\ge 2} u_{jk} a_k$ and

$$\forall j \le N\,,\; \sum_{k\ge 2} |u_{jk}| \le LS\,.$$

Writing $a_k = (a_{ik})_{i\le n}$, the condition $t_j = \sum_{k\ge 2} u_{jk} a_k$ means

$$\forall j \le N\,,\; \forall i \le n\,,\; x_{ij} = \sum_{k\ge 2} u_{jk} a_{ik}\,,$$

so that

$$\forall i \le n\,,\; x_i = \sum_{k\ge 2} a_{ik} u_k, \tag{16.16}$$

where $u_k = (u_{jk})_{j\le N}$. We observe that

$$\sum_{i\le n} a_{ik}^2 = \|a_k\|_2^2 \le \frac{1}{\log k}\,.$$

When we fix the numbers a_{ik}, the quantity $\sum_{i\le n} \|U(\sum_{k\ge 2} a_{ik} u_k)\|^q$ is a convex function of the numbers $(u_{jk})_{j\le N, k\ge 2}$. On the set

$$\Big\{\forall j \le N\,,\; \sum_{k\ge 2} |u_{jk}| \le LS\Big\}\,,$$

the maximum of this function is attained at an extreme point $(v_{jk})_{j\le N,k\ge 2}$. By extremality, for each j, there is at most one value of k for which $v_{jk} \ne 0$, and of course $|v_{jk}| \le LS$. Thus if we define $v_k = (v_{jk})_{j\le N}$, this means that the elements $(v_k)_{k\ge 2}$ have disjoint supports and satisfy $\|v_k\| \le LS$. Hence

$$\begin{aligned}\Big(\sum_{i\le n} \|U(x_i)\|^q\Big)^{1/q} &= \Big(\sum_{i\le n} \Big\|U\Big(\sum_{k\ge 2} a_{ik} u_k\Big)\Big\|^q\Big)^{1/q} \\ &\le \Big(\sum_{i\le n} \Big\|U\Big(\sum_{k\ge 2} a_{ik} v_k\Big)\Big\|^q\Big)^{1/q} \\ &\le LSH_q(U)\,,\end{aligned}$$

where the first inequality follows from the choice of the numbers $(v_{jk})_{j\le N,k\ge 2}$ and the second inequality from the definition of $H_q(U)$. This completes the proof. □

Theorem 16.1.5 is the starting point of a rather complete theory for the cotype of operators from ℓ_N^∞. We will refer the reader to [25] for a full development. Here we will simply indulge in proving Theorems 16.1.10 and 16.1.11 below to enjoy a nice $\sqrt{\log\log N}$ "exact" term and illustrate how precise and non-trivial matters can get.

For a function f on a space provided with a positive measure μ, let us consider the quantity

$$\|f\|_{q,1} = \int_0^\infty \big(\mu(\{|f| \geq t\})\big)^{1/q} \mathrm{d}t \, . \tag{16.17}$$

It can be shown to be equivalent to a norm, and defines the space $L^{q,1}(\mu)$, $q \geq 1$ (that will occur again in Section 16.6 below).

We start by collection a few simple facts. We first note, using (2.5) in the first equality, and change of variable in the second one,

$$\|f\|_p^p = \int_0^\infty \mu(\{|f|^p \geq t\}) \mathrm{d}t = \int_0^\infty p t^{p-1} \mu(\{|f| \geq t\}) \mathrm{d}t \, .$$

Proposition 16.1.6. *If $p < q$ then*

$$\|f\|_{q,1} \leq K(p,q) \|f\|_\infty^{1-p/q} \|f\|_p^{p/q} \, , \tag{16.18}$$

where $K(p,q)$ depends only on p and q and $K(1,q) = 1$. Moreover, if μ is a probability measure, then

$$q < p \ \Rightarrow \ \|f\|_{q,1} \leq K(p,q) \|f\|_p \tag{16.19}$$

and

$$\|f\|_q \leq K(q) \|f\|_{q,1} \, . \tag{16.20}$$

Proof. To prove (16.18), if $q' = q/(q-1)$ denotes the conjugate exponent of q, by Hölder's inequality we have

$$\begin{aligned} \|f\|_{q,1} &= \int_0^{\|f\|_\infty} \big(\mu(\{|f| \geq t\})\big)^{1/q} \mathrm{d}t \\ &\leq \Big(\int_0^{\|f\|_\infty} t^{(1-p)\frac{q'}{q}} \mathrm{d}t\Big)^{1/q'} \Big(\int_0^\infty t^{p-1} \mu(\{|f| \geq t\}) \mathrm{d}t\Big)^{1/q} \\ &\leq K(p,q) \|f\|_\infty^{1-p/q} \|f\|_p^{p/q} \, , \end{aligned}$$

since $(p-1)q'/q = (p-1)/(q-1) < 1$ and $(1 + (1-p)q'/q)/q' = 1 - p/q$. Obviously when $p = 1$ this holds for $K(q,1) = 1$.

To prove (16.19), assuming without loss of generality that $\|f\|_p = 1$, we have $\mu(\{|f| \geq t\}) \leq \min(1, t^{-p})$ and $\|f\|_{q,1} \leq K(p,q)$ by (16.17).

To prove (16.20), assuming without loss of generality that $\|f\|_{q,1} = 1$, by (16.17) we have

$$t\mu(\{|f| \geq t\})^{1/q} \leq \int_0^t \mu(\{|f| \geq u\}) \mathrm{d}u \leq 1 \, ,$$

so that

$$\begin{aligned} t^{q-1}\mu(\{|f| \geq t\}) &= (t\mu(\{|f| \geq t\})^{1/q})^{q-1}(\mu(\{|f| \geq t\}))^{1/q} \\ &\leq (\mu(\{|f| \geq t\}))^{1/q} , \end{aligned}$$

and thus

$$\|f\|_q^q = \int_0^\infty q t^{q-1}\mu(\{|f| \geq t\})\mathrm{d}t \leq q \int_0^\infty (\mu(\{|f| \geq t\})^{1/q}\mathrm{d}t \leq q ,$$

using (16.17) again. □

In the next few pages, we find it convenient to view an element of ℓ_N^∞ as a function on $\{1, \ldots, N\}$. Thus, for $x = (x_j)_{j \leq N} \in \ell_N^\infty$, $|x|^p$ is the element $(|x_j|^p)_{j \leq N}$ of ℓ_N^∞.

Proposition 16.1.7. *Consider a probability measure μ on $\{1, \ldots, N\}$ and $q \geq 1$. Then the canonical injection $\mathrm{Id} : \ell_N^\infty \hookrightarrow L^{q,1}(\mu)$ satisfies $\|\mathrm{Id}\|_{q,1} \leq 1$.*

Proof. Consider elements $(x_i)_{i \leq n}$ of ℓ_N^∞, $x_i = (x_{ij})_{j \leq N}$, and assume that

$$\forall x^* \in \ell_N^1 = (\ell_N^\infty)^* , \sum_{i \leq n} |x^*(x_i)| \leq \|x^*\| .$$

Therefore,

$$\forall j \leq N , \sum_{i \leq n} |x_{ij}| \leq 1 , \tag{16.21}$$

which, viewing x_i as a function on $\{1, \ldots, N\}$ (which basically means identifying x_i and $\mathrm{Id}(x_i)$) we simply write as

$$\sum_{i \leq n} |x_i| \leq 1 . \tag{16.22}$$

From (16.22) we have $\|x_i\|_\infty \leq 1$, so that, still viewing x_i as a function on $\{1, \ldots, N\}$, (16.18) implies

$$\|x_i\|_{q,1}^q \leq \int |x_i| \mathrm{d}\mu ,$$

and thus $\sum_{i \leq n} \|x_i\|_{q,1}^q \leq 1$ by (16.22) and since μ is a probability. □

The importance of the previous example stems from the fact that it is essentially "generic" as the following factorization theorem, due to G. Pisier, shows.

Theorem 16.1.8. *Given an operator $U : \ell_N^\infty \to Y$, there is a probability measure μ on $\{1, \ldots, N\}$ such that if we denote by V the operator U as seen operating from $L^{q,1}(\mu)$ to Y then*

$$\|V\| \leq L\|U\|_{q,1} . \tag{16.23}$$

We refer the reader to [18] for a proof. This result witnesses the value of $\|U\|_{q,1}$ (within the multiplicative constant L). Indeed, by (16.3) and Proposition 16.1.7 we have

$$\|U\|_{q,1} = \|V \circ \mathrm{Id}\|_{q,1} \le \|V\| \|\mathrm{Id}\|_{q,1} \le \|V\| \; .$$

Here is a simple fact.

Lemma 16.1.9. *If $M \ge 2$, for a positive measure μ on $\{1, \ldots, M\}$ and functions $(x_i)_{i \le n}$ on $\{1, \ldots, M\}$, we have*

$$\sum_{i \le n} |x_i|^2 \le 1 \Rightarrow \sum_{i \le n} \|x_i\|_{2,1}^2 \le L \log M \mu(\{1, \ldots, M\}) \; .$$

Proof. By homogeneity we can and do assume that μ is a probability measure. Consider the probability measure μ' on $\{1, \ldots, M\}$ which is the sum of $\mu/2$ and of point masses $1/(2M)$ a each point of $\{1, \ldots, M\}$. With obvious notation we have $\|x\|_{2,1,\mu} \le \sqrt{2} \|x\|_{2,1,\mu'}$. Thus we may assume without loss of generality that μ gives mass $\ge 1/(2M)$ to each point of $\{1, \ldots, M\}$. We shall prove that this implies

$$\forall x \; , \; \|x\|_{2,1}^2 \le L \log M \|x\|_2^2 \; ,$$

which conclude the proof since $\sum_{i \le n} \|x_i\|_2^2 \le 1$. We set $t_0 = 0$ and for $\ell \ge 1$, we define

$$t_\ell = \sup\{t \; ; \; \mu(\{|x| \ge t\}) \ge 2^{-\ell}\} \, ,$$

so that

$$t_\ell < t < t_{\ell+1} \Rightarrow 2^{-\ell-1} \le \mu(\{|x| \ge t\}) \le 2^{-\ell} \tag{16.24}$$

and thus

$$\|x\|_{2,1} = \int_0^\infty \sqrt{\mu(\{|x| \ge t\})} \, \mathrm{d}t \le \sum_{\ell \ge 0} 2^{-\ell/2} (t_{\ell+1} - t_\ell) \, . \tag{16.25}$$

If ℓ_0 is the smallest integer with $2^{-\ell_0} < 1/2M$, for $\ell \ge \ell_0$ we have $t_\ell = t_{\ell_0} = \|x\|_\infty$, because μ gives mass $\ge 1(2M)$ to each point $1, \ldots, M$. Consequently the sum in (16.25) has in fact at most $(\ell_0 + 1)$ terms. Since $(t_{\ell+1} - t_\ell)^2 \le t_{\ell+1}^2 - t_\ell^2$, using (16.25), the Cauchy-Schwarz inequality and (16.24) we get

$$\begin{aligned} \|x\|_{2,1}^2 &\le (\ell_0 + 1) \sum_{\ell \ge 0} (t_{\ell+1}^2 - t_\ell^2) 2^{-\ell} \\ &\le 4(\ell_0 + 1) \sum_{\ell \ge 0} \int_{t_\ell}^{t_{\ell+1}} t \mu(\{|x| \ge t\}) \, \mathrm{d}t \\ &= 2(\ell_0 + 1) \|x\|_2^2 \; . \end{aligned}$$

□

Theorem 16.1.10. *For an operator U from ℓ_N^∞ to any Banach space Y, we have, for $N \geq 3$*

$$C_q^r(U) \leq L\sqrt{\log\log N}\|U\|_{2,1} \; .$$

Proof. Consider the probability measure provided by Theorem 16.1.8, and the operator V as in that theorem. Then

$$C_q^r(U) = C_q^r(V \circ \mathrm{Id}) \leq \|V\| C_q^r(\mathrm{Id}) \; .$$

Thus it suffices to prove that $C_q^r(U) \leq L\sqrt{\log\log N}$ when $U = \mathrm{Id}$ is the canonical injection from ℓ_N^∞ to $L^{2,1}(\mu)$, where μ is any probability measure on $\{1, \ldots, N\}$. Combining Theorem 16.1.2 and Proposition 16.1.7, it suffices to show that $C_2^g(U) \leq L\sqrt{\log\log N}$, and, using Theorem 16.1.5, that $H_2(U) \leq L\sqrt{\log\log N}$. To prove this consider elements $(u_k)_{k\geq 2}$ of ℓ_N^∞ with disjoint supports, $\|u_k\|_\infty \leq 1$, and numbers $(a_{ik})_{i\leq n, k\geq 2}$ such that

$$\forall k \geq 2 \, , \; \sum_{i\leq n} a_{ik}^2 \leq \frac{1}{\log k} \; . \tag{16.26}$$

Set $x_i = \sum_{k\geq 2} a_{ik} u_k$. We want to prove that

$$\sum_{i\leq n} \|x_i\|_{2,1}^2 \leq L \log\log N \; . \tag{16.27}$$

We observe that there are at most N of the elements u_k that are not zero (since they have disjoint support). By renumbering them, we may assume that $k \geq N+2 \Rightarrow u_k = 0$. For $\ell \geq 0$, we set

$$x_{i,\ell} = \sum_{M_\ell \leq k < M_{\ell+1}} a_{ik} u_k \; , \tag{16.28}$$

where $M_\ell = 2^{2^\ell}$ (so that $M_0 = 2$ and $M_\ell = N_\ell$ for $\ell \geq 1$). Consider the smallest integer ℓ_0 such that $M_{\ell_0} \geq N+2$. Then

$$x_i = \sum_{0\leq \ell \leq \ell_0} x_{i,\ell} \; ,$$

so that, since $\|\cdot\|_{2,1}$ is equivalent to a norm, we have

$$\|x_i\|_{2,1} \leq L \sum_{0\leq \ell\leq \ell_0} \|x_{i,\ell}\|_{2,1} \; ,$$

and, by the Cauchy-Schwarz inequality,

$$\sum_{i\leq n} \|x_i\|_{2,1}^2 \leq L(\ell_0 + 1) \sum_{0\leq \ell\leq \ell_0, i\leq n} \|x_{i,\ell}\|_{2,1}^2 \; .$$

Therefore it suffices to prove that $\sum_{\ell \le \ell_0, i \le n} \|x_{i,\ell}\|_{2,1}^2 \le L$. Denoting by S_ℓ the union of the supports of the vectors u_k for $M_\ell \le k < M_{\ell+1}$, we observe that the sets S_ℓ are disjoint, so that it suffices to prove that

$$\sum_{i \le n} \|x_{i,\ell}\|_{2,1}^2 \le L\mu(S_\ell) . \tag{16.29}$$

First we prove that for each ℓ we have

$$\sum_{i \le n} |x_{i,\ell}|^2 \le L 2^{-\ell} . \tag{16.30}$$

We set $x_{i,\ell} = (x_{i,\ell,j})_{j \le N}$ and $u_k = (u_{kj})_{j \le N}$. Consider $j \le N$. By (16.28), if j does not belong to the support of any u_k, the numbers $x_{i,\ell,j}$ are 0 for each i. Otherwise, since the supports of the elements u_k are disjoint, j belongs to the support of a unique element u_{k_0}. If either $k_0 < M_\ell$ or $k_0 \ge M_{\ell+1}$, by (16.28) the numbers $x_{i,\ell,j}$ are again 0 for each i. If $M_\ell \le k_0 < M_{\ell+1}$ then (16.28) implies that for each $i \le n$ we have $|x_{i,\ell,j}| \le a_{ik_0}$, so that $\sum_{i \le n} x_{i,\ell,j}^2 \le 1/\log k_0 \le L2^{-\ell}$ by (16.26), and since $k_0 \ge M_\ell$. This proves (16.30).

Since $\|u_k\|_\infty \le 1$ we have $|u_{kj}| \le 1$. Since the vectors u_k have disjoint support $\|x_{i,\ell}\|_{1,2} = \| \sum_{M_\ell \le k < M_{\ell+1}} a_{ik} u_k \|_{2,1}$ increases with $|u_{kj}|$, so that to prove (16.29) we may assume without loss of generality that $|u_{kj}| \in \{0,1\}$. The span of the elements $|x_{i,\ell}|$, $i \le n$ in ℓ_N^∞ consists of functions on $\{1, \dots, N\}$ that are constants on the sets $\{|u_k| = 1\}$ for $M_\ell \le k < M_{\ell+1}$, and that are zero outside the union of these sets. If we identify each of these sets $\{|u_k| = 1\}$ to a point, we are in a situation where the underlying measured space has at most $M_{\ell+1}$ points, and since $\log M_{\ell+1} \le 2^{\ell+1}$, Lemma 16.1.9 and (16.30) imply (16.29). □

Theorem 16.1.11. *Let μ be the uniform probability measure on $\{1, \dots, N\}$. Then for $N \ge 3$ we have*

$$C_2^g(U) \ge \frac{1}{L} \sqrt{\log \log N} ,$$

where U is the canonical injection from ℓ_N^∞ into $L^{2,1}(\mu)$.

In summary, we have shown that $\|U\|_{2,1} \le 1$, and that both $C_2^g(U)$ and $C_2^r(U)$ are of order $\sqrt{\log \log N}$. In particular the bound of Theorem 16.1.10 cannot be improved in general.

Proof. To avoid messy details we shall assume that N is of the type $N = (p-3)2^{2^p}$ for some $p \ge 4$. For $2 \le j \le p-2$ we consider disjoint sets S_j with $\operatorname{card} S_j = 2^{2^{j+2}}$, and $S = \bigcup_{2 \le j \le p-2} S_j$.

Consider the probability measure ν on S that gives mass $1/((p-3) \operatorname{card} S_j)$ to each point of S_j. The mass of each point of S is a multiple of N^{-1}, so that

$L^{2,1}(\nu)$ is isometric to a subspace of $L^{2,1}(\mu)$, and it suffices to prove that the canonical injection V from $\ell^\infty(S)$ into $L^{2,1}(\nu)$ satisfies

$$H_2(V) \geq \frac{\sqrt{p}}{L} . \tag{16.31}$$

We consider the family $\mathcal{X}$ consisting of all the elements x of $\ell^\infty(S)$ of the following type. The element x takes only the values 0 and 2^k for $3 \leq k \leq 2^{p-2}$. For each such value of k there is unique $2 \leq j \leq p-2$ for which $2^{j-1} < k \leq 2^j$. Then the set $\{x = 2^k\}$ consists of exactly $2^{-2k-2j}\mathrm{card}\, S_j = 2^{2^{j+2}-2k-2j}$ points of S_j. This is possible because this number is an integer since $2^{j+2} - 2j - 2k \geq 2^{j+2} - 2j - 2^{j+1} \geq 0$. Thus $\nu(\{x = 2^k\}) = 2^{-2k-2j}/(p-3)$, and

$$\begin{aligned}\|x\|_{2,1} &\geq \sum_{3\leq k\leq 2^{p-2}} \int_{2^{k-1}}^{2^k} \sqrt{\nu(\{|x| \geq t\})}\,\mathrm{d}t \geq \sum_{3\leq k\leq 2^{p-2}} 2^{k-1}\sqrt{\nu(\{|x| = 2^k\})} \\ &\geq \sum_{2\leq j\leq p-2}\ \sum_{2^{j-1}<k\leq 2^j} 2^{k-1}\frac{2^{-k-j}}{\sqrt{p-3}} = \frac{p-3}{4\sqrt{p-3}} \geq \frac{\sqrt{p}}{L} .\end{aligned} \tag{16.32}$$

Let us consider the family $\mathcal{F}$ consisting of the elements of $\ell^\infty(S)$ of the type $x/\sqrt{M}$, where $x \in \mathcal{X}$ and $M = \mathrm{card}\mathcal{X}$. Then (16.32) implies

$$\sum_{y\in\mathcal{F}} \|y\|_{2,1}^2 \geq \frac{p}{L} . \tag{16.33}$$

For $x \in \mathcal{X}$, the average value of x^2 on the set S_j is

$$\sum_{2^{j-1}<k\leq 2^j} 2^{2k}2^{-2k-2j} = 2^{-j-1} .$$

For $y \in \mathcal{F}$ let us write $y = (y_k)_{k\in S}$. By symmetry the quantity $\sum_{y\in\mathcal{F}} y_k^2$ is independent of $k \in S_j$, so that

$$\forall k \in S_j\, , \ \sum_{y\in\mathcal{F}} y_k^2 = 2^{-j-1} . \tag{16.34}$$

We can and do assume that the sets S_j are consecutive intervals. In that case, for $k \in S_j$ we have $\log k \leq L2^j$, and (16.34) implies

$$\forall k \in S_j\, , \ \sum_{y\in\mathcal{F}} y_k^2 \leq \frac{L}{\log k} . \tag{16.35}$$

Let us denote by (u_k) the canonical basis of $\ell^\infty(S)$. Then

$$y = \sum_{k\in S} y_k u_k\ ,$$

and the elements u_k have disjoint supports. Combining with (16.33) and (16.35) we have indeed shown that $H_2(V) \geq \sqrt{p}/L$. □

16.2 Computing the Rademacher Cotype-2 Constant

When U is an operator between two finite dimensional Banach spaces X and Y, we recall the definition (16.1) of the Rademacher cotype-2 constant $C_2^r(U)$ of U. One may ask "how many vectors of X are needed in general to compute $C_2^r(U)$" within a constant factor L. That is, how large should n be so that one can find $(x_1, \ldots, x_n)$ in X with

$$\Bigl(\sum_{i\le n}\|U(x_i)\|^2\Bigr)^{1/2} > \frac{1}{L}C_2^r(U)\mathsf{E}\Bigl\|\sum_{i\le n}\varepsilon_i x_i\Bigr\| .$$

This question is motivated in particular by a result of N. Tomczak-Jaegermann [29] who proved that N vectors suffice to compute the *Gaussian* cotype-2 constant of U, where N is the dimension of X. Similar questions in various settings are also investigated e.g. in [10], and we consider only the case $q = 2$ simply because this is the most difficult. We will approach this question through a comparison principle between Gaussian and Rademacher averages which is of interest in its own right.

Consider a Banach space X of dimension $N \ge 3$, and its dual X^*. Consider elements $x_1, \ldots, x_n$ in X and assume without loss of generality that they span X.

Consider the norm $\|\cdot\|_2$ on X such that its unit ball is the set

$$\Bigl\{\sum_{i\le n}\alpha_i x_i\ ;\ \sum_{i\le n}\alpha_i^2 \le 1\Bigr\} .$$

Let us also denote by $\|\cdot\|_2$ the dual of this norm on X^*. It will be clear for the notation in which space we compute the norm, so no confusion will arise. Thus

$$\begin{aligned}\|x^*\|_2 &= \sup\Bigl\{\Bigl|x^*\Bigl(\sum_{i\le n}\alpha_i x_i\Bigr)\Bigr|\ ;\ \sum_{i\le n}\alpha_i^2 \le 1\Bigr\}\\ &= \Bigl(\sum_{i\le n}x^*(x_i)^2\Bigr)^{1/2} .\end{aligned} \tag{16.36}$$

This norm arises from the dot product given by

$$(x^*, y^*) = \sum_{i\le n}x^*(x_i)y^*(x_i) .$$

Consider an orthonormal basis $(e_j^*)_{j\le N}$ of X^* for this dot product. Then

$$x^* = \sum_{j\le N}(x^*, e_j^*)\, e_j^*\ ;\ \|x^*\|_2^2 = \sum_{j\le N}(x^*, e_j^*)^2 .$$

Thus

$$\|x\|_2 = \sup\{|x^*(x)| \; ; \; \|x^*\|_2 \le 1\}$$
$$= \sup\Big\{\Big|\sum_{j\le N} \beta_j e_j^*(x)\Big| \; ; \; \sum_{j\le N} \beta_j^2 \le 1\Big\} = \Big(\sum_{j\le N} e_j^*(x)^2\Big)^{1/2} .$$

We note that

$$\sum_{i\le n} \|x_i\|_2^2 = \sum_{i\le n}\sum_{j\le N} e_j^*(x_i)^2$$
$$= \sum_{j\le N}\sum_{i\le n} e_j^*(x_i)^2 = N \,, \tag{16.37}$$

using (16.36) with $x^* = e_j^*$ in the last equality.

It is of interest to consider a subset T of X as a subset of the Hilbert space $(X, \|\cdot\|_2)$. One can then define the usual quantity $g(T)$, that is concretely given by

$$g(T) = \mathsf{E} \sup_{t\in T} \sum_{j\le N} g_j e_j^*(t) \,, \tag{16.38}$$

where $(g_i)_{j\le N}$ are independent standard normal r.v. (Interestingly, this formula will not be needed in the sequel.)

Lemma 16.2.1. *If $T = \{x_1, \ldots, x_n\}$ then*

$$g(T) \le L\sqrt{\log(N+1)} \,. \tag{16.39}$$

When the sequence $(\|x_i\|_2)_{i\ge 1}$ is non-increasing, and if $M = \lfloor N \log N \rfloor$, the set $T' = \{x_i \; ; \; M \le i \le n\}$ satisfies

$$g(T') \le L \,. \tag{16.40}$$

Proof. Both results are based on the fact that if $T = \{t_k \; ; \; k \ge 1\}$ then

$$g(T) \le L \sup_{k\ge 1}\big(\|t_k\|_2 \sqrt{\log(k+1)}\big) \,,$$

as shown in Proposition 2.4.16. We observe that $\|x_i\|_2 \le 1$ by definition of the unit ball of $\|\cdot\|_2$. Assuming without loss of generality that the sequence $(\|x_i\|_2)_{i\ge 1}$ is non-increasing, we see from (16.37) that $\|x_i\|_2 \le \sqrt{N/i}$. Thus

$$g(T) \le L \sup_{k\ge 1}\Big(\min\Big(1, \sqrt{\frac{N}{k}}\Big)\sqrt{\log(k+1)}\Big) \le L\sqrt{\log N}$$
$$g(T') \le L \sup_{k\ge 1}\Big(\sqrt{\frac{N}{M+k}}\sqrt{\log(k+1)}\Big) \le L\sqrt{\frac{N}{M}\log M} \le L \,. \qquad \square$$

In the next statement, we define $T = \{x_1, \ldots, x_n\}$, and, for a subset I of $\{1, \ldots, n\}$ we define T_I as the collection of vectors x_i for i *outside* I,

$$T_I = \{x_i \; ; \; i \le n \,, \, i \notin I\} \,. \tag{16.41}$$

Theorem 16.2.2. *We have*

$$\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\le n}\varepsilon_i x_i\Big\|(1+g(T))\;. \tag{16.42}$$

More generally, for any subset I of $\{1,\dots,n\}$ we have

$$\mathsf{E}\Big\|\sum_{i\notin I} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\notin I}\varepsilon_i x_i\Big\|\left(1+\frac{\mathsf{E}\|\sum_{i\le n} g_i x_i\|}{\mathsf{E}\|\sum_{i\notin I} g_i x_i\|}g(T_I)\right)\;. \tag{16.43}$$

Of course (16.42) is the special case of (16.43) where $I=\emptyset$. Using (16.39) we see that (16.42) improves the classical inequality

$$\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \le L\sqrt{\log N}\mathsf{E}\Big\|\sum_{i\le n}\varepsilon_i x_i\Big\|\;. \tag{16.44}$$

Corollary 16.2.3. *There exists a subset I of $\{1,\dots,n\}$ such that* $\operatorname{card} I \le N\log(N+1)$ *and that* **either** *of the following holds true*

$$\mathsf{E}\Big\|\sum_{i\notin I} g_i x_i\Big\| \le \frac{1}{2}\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \tag{16.45}$$

or else

$$\mathsf{E}\Big\|\sum_{i\notin I} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\notin I}\varepsilon_i x_i\Big\|\;. \tag{16.46}$$

Proof. By (16.40) we can find a set I with the required cardinality such that $g(T_I)\le L$, so that if (16.45) fails, (16.46) follows from (16.43). □

Corollary 16.2.4. *Consider an operator U from X to Y, and vectors $(x_i)_{i\le n}$ of X such that*

$$A\mathsf{E}\Big\|\sum_{i\le n}\varepsilon_i x_i\Big\| < \Big(\sum_{i\le n}\|U(x_i)\|^2\Big)^{1/2}\;. \tag{16.47}$$

Then we can find vectors $(y_j)_{j\le M}$ of X such that

$$\frac{A}{L}\mathsf{E}\Big\|\sum_{j\le M}\varepsilon_j y_j\Big\| < \Big(\sum_{j\le M}\|U(y_j)\|^2\Big)^{1/2} \tag{16.48}$$

and $M\le N\log N\log\log N$.

For every $A<C_2^r(U)$, there exists vectors $x_1,\dots,x_n$ such that (16.47) is satisfied, and (16.48) means that within the loss of a constant factor one can take $n=M$. In other words, the "Rademacher cotype-2 constant of U can essentially be computed on M vectors".

Of course, one should ask whether it would actually suffice to consider LN vectors.

Proof. The first part of the proof consists of showing that we can find a subset J of $\{1,\ldots,n\}$ with $\operatorname{card} J \le M$ and

$$\mathsf{E}\Big\|\sum_{i\notin J} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\le n} \varepsilon_i x_i\Big\| \,. \tag{16.49}$$

To this aim, consider the largest integer k_0 with $2^{k_0} \le \sqrt{\log N}$, so that $k_0 \le \log\log N$. Using Corollary 16.2.3, by induction over k, for $k \le k_0$ we construct subsets I_k of $\{1,\ldots,n\}$ with $\operatorname{card} I_k \le N\log N$ and either

$$\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_k} g_i x_i\Big\| \le \frac{1}{2}\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_{k-1}} g_i x_i\Big\| \tag{16.50}$$

or else

$$\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_k} g_i x_i\Big\| \le L\mathsf{E}\Big\|\sum_{i\notin I_1\cup\ldots\cup I_{k-1}} \varepsilon_i x_i\Big\| \,. \tag{16.51}$$

If at one step (16.51) holds, we then stop the construction, and we define $J = I_1\cup\ldots\cup I_k$. Thus $\operatorname{card} J \le kN\log N \le M$ and $\mathsf{E}\|\sum_{i\notin J} g_i x_i\| \le L\mathsf{E}\|\sum_{i\notin J}\varepsilon_i x_i\| \le L\mathsf{E}\|\sum_{i\le n}\varepsilon_i x_i\|$, so that (16.49) holds. If, on the other hand, (16.51) never occurs during the construction, we continue this construction until $k = k_0$, and we define now $J = I_1\cup\ldots\cup I_{k_0}$. Thus $\operatorname{card} J \le k_0 N\log N \le M$ and, iterating (16.50),

$$\mathsf{E}\Big\|\sum_{i\notin J} g_i x_i\Big\| \le 2^{-k_0}\mathsf{E}\Big\|\sum_{i\le n} g_i x_i\Big\| \,.$$

Combining with (16.44) this implies

$$\mathsf{E}\Big\|\sum_{i\notin J} g_i x_i\Big\| \le 2^{-k_0}L\sqrt{\log(N+1)}\mathsf{E}\Big\|\sum_{i\le n}\varepsilon_i x_i\Big\| \,,$$

and this proves (16.49) by the choice of k_0.

Now that we have proved (16.49) we consider 2 cases.

Case 1. We have

$$\sum_{i\in J}\|U(x_i)\|^2 \ge \frac{1}{2}\sum_{i\le n}\|U(x_i)\|^2 \,.$$

Then

$$\frac{A}{2}\mathsf{E}\Big\|\sum_{i\in J}\varepsilon_i x_i\Big\| \le \frac{A}{2}\mathsf{E}\Big\|\sum_{i\le n}\varepsilon_i x_i\Big\| < \frac{1}{2}\Big(\sum_{i\le n}\|U(x_i)\|^2\Big)^{1/2} \le \Big(\sum_{i\in J}\|U(x_i)\|^2\Big)^{1/2} ,$$

and this proves (16.48).

Case 2. We have

$$\sum_{i \notin J} \|U(x_i)\|^2 \geq \frac{1}{2} \sum_{i \leq n} \|U(x_i)\|^2 .$$

Then (16.47) yields

$$\frac{A}{2} \mathsf{E} \Big\| \sum_{i \leq n} \varepsilon_i x_i \Big\| < \Big(\sum_{i \notin J} \|U(x_i)\|^2 \Big)^{1/2}$$

and combining with (16.49) we obtain

$$\frac{A}{L} \mathsf{E} \Big\| \sum_{i \notin J} g_i x_i \Big\| < \Big(\sum_{i \notin J} \|U(x_i)\|^2 \Big)^{1/2} , \tag{16.52}$$

which implies that the Gaussian cotype-2 constant of U is $\geq A/L$. To conclude the proof, we use that this Gaussian cotype-2 constant of U "can be computed on N vectors [29]", so that from (16.52) we can find N vectors $y_1 , \ldots , y_N$ of X such that

$$\frac{A}{L} \mathsf{E} \Big\| \sum_{j \leq N} g_j y_j \Big\| \leq \Big(\sum_{j \leq N} \|U(y_j)\|^2 \Big)^{1/2} ,$$

which by (5.3) implies (16.48). □

We turn to the proof of Theorem 16.2.2. It will use the following general principle, where we recall that $N_0 = 1$ and that $N_n = 2^{2^n}$ for $n \geq 1$.

Lemma 16.2.5. *Consider a set T provided with two distances d and d'. Assume that for a certain number S and every $n \geq 0$, every ball $B_d(t,a)$ of T can be covered by N_n sets of d'-diameter at most $aS2^{-n/2}$. Then*

$$\gamma_1(T, d') \leq LS\gamma_2(T, d) .$$

Proof. Consider an admissible sequence $(\mathcal{B}_n)$ of T with

$$\forall t \in T , \sum_{n \geq 0} 2^{n/2} \Delta(\mathcal{B}_n(t), d) \leq 2\gamma_2(T, d) .$$

We construct by induction an increasing sequence of partitions $(\mathcal{C}_n)$ satisfying

$$\operatorname{card} \mathcal{C}_n \leq N_{n+2} \tag{16.53}$$

$$\forall C \in \mathcal{C}_n , \exists B \in \mathcal{B}_n , C \subset B , \Delta(C, d') \leq S2^{-n/2} \Delta(B, d) . \tag{16.54}$$

First, we set $\mathcal{C}_0 = \{T\}$. We note that using the hypothesis for $a = \Delta(T, d)$ and $n = 0$ we have

$$\Delta(T, d') \leq S\Delta(T, d) . \tag{16.55}$$

Thus (16.54) is true for $n = 0$. Assuming that $\mathcal{C}_n$ has been constructed, we split each element C of $\mathcal{C}_n$ as follows. First we split C in the sets $C \cap B$, $B \in \mathcal{B}_{n+1}$. Then we split each set $C \cap B$ in N_{n+1} pieces C' such that

$$\Delta(C', d') \leq S2^{-(n+1)/2} \Delta(C \cap B, d) .$$

This is possible by hypothesis, and this completes the construction of $\mathcal{C}_{n+1}$. Clearly, $\mathcal{C}_{n+1}$ consists of at most $N_{n+2} \cdot N_{n+1}^2 = N_{n+3}$ sets and it is obvious that (16.53) and (16.54) hold for $n + 1$. A consequence of (16.54) is that

$$\forall t\,,\; \Delta(C_n(t), d') \leq S2^{-n/2} \Delta(B_n(t), d)$$

and thus

$$\begin{aligned} \sum_{n \geq 0} 2^n \Delta(C_n(t), d') &\leq S \sum_{n \geq 0} 2^{n/2} \Delta(B_n(t), d) \\ &\leq 2S\gamma_2(T, d) . \end{aligned}$$

Using (16.55) and Lemma 2.3.5 then yields the result. □

Proof of Theorem 16.2.2. We prove (16.43). On X^* consider the norm $\|\cdot\|_I$ given by

$$\|x^*\|_I = \sup_{i \notin I} |x^*(x_i)| .$$

We now appeal to the dual Sudakov minoration inequality (Lemma 8.3.6): for $n \geq 0$ the unit ball of $(X^*, \|\cdot\|_2)$ can be covered by N_n balls for $\|\cdot\|_I$ of radius $Lg(T_I)2^{-n/2}$. Thus, recalling the set T_I of (16.41), Lemma 16.2.5 implies

$$\gamma_1(X_1^*, d_I) \leq Lg(T_I)\gamma_2(X_1^*, \|\cdot\|_2) ,$$

where d_I is the (quasi-) distance associated to the norm $\|\cdot\|_I$ and where X_1^* is the unit ball of X^*. Applying Theorem 2.4.1 to the process given for x^* in X_1^* by $X_{x^*} = \sum_{i \leq n} g_i x^*(x_i)$ yields $\gamma_2(X_1^*, \|\cdot\|_2) \leq L\mathsf{E}\|\sum_{i \leq n} g_i x_i\|$, and consequently

$$\gamma_1(X_1^*, d_I) \leq Lg(T_I)\mathsf{E}\Big\|\sum_{i \leq n} g_i x_i\Big\| . \tag{16.56}$$

Consider now the set

$$T^{\sim} = \{(x^*(x_i))_{i \notin I}\;;\; x^* \in X_1^*\} .$$

It should be obvious that

$$g(T^{\sim}) = \mathsf{E}\Big\|\sum_{i \notin I} g_i x_i\Big\|\;;\; b(T^{\sim}) = \mathsf{E}\Big\|\sum_{i \notin I} \varepsilon_i x_i\Big\|\;;$$

$$\gamma_1(T^{\sim}, d_\infty) = \gamma_1(X_1^*, d_I) . \tag{16.57}$$

We appeal to Theorem 5.4.1 to obtain

$$g(T^\sim) \le L\big(b(T^\sim) + \sqrt{b(T^\sim)\gamma_1(T^\sim, d_\infty)}\big) \ . \tag{16.58}$$

Let us write

$$A = g(T_I)\frac{\mathsf{E}\|\sum_{i\le n} g_i x_i\|}{\mathsf{E}\|\sum_{i\notin I} g_i x_i\|} \ ,$$

so that (16.56) and (16.57) imply $\gamma_1(T^\sim, d_\infty) \le Lg(T^\sim)A$ and combining with (16.58),

$$g(T^\sim) \le L\big(b(T^\sim) + \sqrt{b(T^\sim)g(T^\sim)A}\big) \ .$$

Using the inequality $\sqrt{xy} \le cx + y/c$, we conclude that

$$g(T^\sim) \le Lb(T^\sim) + Lb(T^\sim)A + \frac{1}{2}g(T^\sim) \ ,$$

so that $g(T^\sim) \le L(1+A)b(T^\sim)$. □

16.3 Classifying the Elements of B_1

In this section we prove Proposition 5.2.5. For this we need in particular to cover the case $T = B_1 = \{t \in \ell^1 ; \|t\|_1 \le 1\}$. This relies on an important fact, a kind of classification of the elements of B_1, Theorem 16.3.1 below. This result will also be also used a number of times in the forthcoming sections.

In order to avoid repetition, we consider the situation of a general measured space (Ω, μ), in which case

$$B_1 = \left\{ f \in L^1(\mu) \ ; \ \int |f| \mathrm{d}\mu \le 1 \right\} .$$

The parameter τ in the forthcoming theorem is of secondary importance. In this section we need only the case $\tau = 0$. We use the notation $a \wedge b = \min(a, b)$.

Theorem 16.3.1. *For any integer $\tau \in \mathbb{Z}$ there exists an admissible sequence of partitions $(\mathcal{C}_n)$ of B_1, and for each $C \in \mathcal{C}_n$ an integer $\ell(C) \in \mathbb{Z}$, such that if we set*

$$\ell(f, n) = \ell(C_n(f)) \tag{16.59}$$

we have

$$\forall\, f \in B_1 \ , \ \int (2^{\ell(f,n)} f)^2 \wedge 1 \mathrm{d}\mu \le 2^{n+\tau} \ , \tag{16.60}$$

and

$$\forall f \in B_1 \ , \ \sum_{n\ge 0} 2^{n-\ell(f,n)} \le 10 \cdot 2^{-\tau} \ . \tag{16.61}$$

We note that (16.60) implies

$$\forall f \in B_1 \ , \ \mu\big(\{|f| > 2^{-\ell(f,n)}\}\big) \le 2^{n+\tau} \ . \tag{16.62}$$

A first (and partial) understanding of the meaning of this result is that we classify the functions f of B_1 according to the values of the integers $\ell(f,n)$ for which

$$\mu\big(\{|f| > 2^{-\ell(f,n)}\}\big) \simeq 2^{n+\tau} \ .$$

Lemma 16.3.2. *For any number $a \in \mathbb{R}$ we have*

$$\sum_{k\in\mathbb{Z}} (2^{k+2}a^2) \wedge 2^{-k} \le 8|a| \ . \tag{16.63}$$

Proof. Without loss of generality we assume that $a > 0$. Consider the smallest integer k_0 such that $2^{k_0+1}a \ge 1$, so that $2^{k_0}a \le 1$. Now,

$$\sum_{k\in\mathbb{Z}} (2^{k+2}a^2) \wedge 2^{-k} \le \sum_{k<k_0} 2^{k+2}a^2 + \sum_{k\ge k_0} 2^{-k} = 2^{k_0+2}a^2 + 2^{-k_0+1} \le 8a \ . \quad \square$$

Proof of Theorem 16.3.1. Given $f \in B_1$ and $n \ge 0$ we define $k(n,f)$ as the largest integer in $\mathbb{Z}$ such that

$$\int (2^{k(f,n)}f)^2 \wedge 1 \mathrm{d}\mu \le 2^{n+\tau} \ . \tag{16.64}$$

(If no such integer exists, we leave to the reader to make the small necessary modifications to the proof.) Thus

$$\int (2^{k(f,n)+1}f)^2 \wedge 1 \mathrm{d}\mu \ge 2^{n+\tau} \ ,$$

and therefore

$$\int (2^{k(f,n)+2}f^2) \wedge 2^{-k(n,f)} \mathrm{d}\mu \ge 2^{n-k(f,n)+\tau} \ . \tag{16.65}$$

It is obvious by construction that the sequence $(k(n,f))_n$ is non-decreasing. We observe first from (16.65) that when the set $B_k = \{n \ge 0; k(f,n) = k\}$ is not empty it has a largest element. Let us consider a subset J of $\mathbb{N}$ with the following property. For each $k \in \mathbb{Z}$, if the set $B_k = \{n \ge 0; k(f,n) = k\}$ is not empty, then $J \cap B_k$ consists exactly of the largest element of B_k. All the values $k(f,n)$ for $n \in J$ are different. Then, gathering the terms on the left-hand side according to which set B_k they belong to, we get

$$\sum_{n\ge 0} 2^{n-k(f,n)} \le 2 \sum_{n\in J} 2^{n-k(f,n)} \ .$$

Moreover summing (16.65) over the values of $n \in J$ and using (16.63) we obtain

$$\sum_{n\in J} 2^{n-k(f,n)+\tau} \le \int \sum_{k\in\mathbb{Z}} (2^{k+2} f^2) \wedge 2^{-k} \mathrm{d}\mu \le 8 \int |f| \mathrm{d}\mu \le 8 ,$$

so finally

$$\sum_{n\ge 0} 2^{n-k(f,n)+\tau} \le 8 .$$

Let us then define $\ell(f,n) = \min(k(f,n), \tau + 2n)$ so that

$$\sum_{n\ge 0} 2^{n-\ell(f,n)+\tau} \le 10 . \tag{16.66}$$

We observe the inequality $h^2 \wedge 1 \le |h|$, so that since $f \in B_1$

$$\int (2^{n+\tau} f)^2 \wedge 1 \mathrm{d}\mu \le 2^{n+\tau} \int f \mathrm{d}\mu \le 2^{n+\tau} ,$$

and the definition of $k(n,f)$ implies $k(f,n) \ge n + \tau$ and therefore

$$\tau + n \le \ell(f,n) \le \tau + 2n . \tag{16.67}$$

We define $\mathcal{C}_0 = \{B_1\}$, and $\ell(B_1) = \tau$. Consider the partition $\mathcal{C}_n$ of B_1 induced by the following equivalence relation: f and f' are equivalent if and only if $\ell(f,m) = \ell(f',m)$ for each $m \le n$. The sequence $(\mathcal{C}_n)$ increases. Moreover since $\ell(f,m)$ can take at most $m+1$ values,

$$\operatorname{card} \mathcal{C}_n \le (n+1)! \le N_n , \tag{16.68}$$

so that the sequence $(\mathcal{C}_n)$ is admissible.

By definition of $\mathcal{C}_n$, if $C \in \mathcal{C}_n$ there exists $\ell(C)$ such that $\ell(f,n) = \ell(C)$ whenever $f \in C$. Consequently for $f \in C$, we have $\ell(C_n(f)) = \ell(C) = \ell(f,n)$, so that (16.59) holds. Also (16.60) holds by construction. Finally (16.61) follows from (16.66). □

As the crude inequality (16.68) shows, the use of admissible sequences is not really canonical for a "classification result" such as Theorem 16.3.1 (one could consider sequences of partitions with a much smaller cardinality). This however suffices for the applications, and we have not yet found uses for sharper results.

Proof of Proposition 5.2.5. We will prove it only when $r = 2$. By homogeneity we may assume that $b^*(T) = 1$, so that $T \subset T_1 + B_1$ where $\gamma_2(T_1) \le 1$. Consider an admissible sequence $(\mathcal{B}_n)$ of T_1 such that

$$\sup_{x\in T_1} \sum_{n\ge 0} 2^{n/2} \Delta(B_n(x)) \le 2 , \tag{16.69}$$

and for $B \in \mathcal{B}_n$ let us define $k_n(B)$ as the largest integer k for which $\Delta(B) \le 2^{n/2-k}$. Consider an admissible sequence $(\mathcal{C}_n)$ of partitions of B_1 and the numbers $\ell(C)$ as provided by Theorem 16.3.1 for $\tau = 0$. For each x in T let us choose $\varphi(x) \in T_1$ and $\psi(x) \in B_1$ such that $x = \varphi(x) + \psi(x)$. For $B \in \mathcal{B}_n$ and $C \in \mathcal{C}_n$ let us define

$$B * C = \{x \in T \,;\, \varphi(x) \in B \,,\; \psi(x) \in C\} \,.$$

There are at most N_n^2 such sets and they form a partition of $\mathcal{D}_n$ of T. The sequence $(\mathcal{D}_n)$ increases. For $D = B * C$ in $\mathcal{D}_n$ let us set $j_n^*(D) = \min(k_n(B), \ell(C))$. Therefore $j_n^*(D_n(x)) = \min(k_n(B_n(\varphi(x))), \ell(C_n(\psi(x))))$ satisfies

$$\sum_{n\ge0} 2^{n-j_n^*(D_n(x))} \le \sum_{n\ge0} 2^{n-k_n(B_n(\varphi(x)))} + \sum_{n\ge0} 2^{n-\ell(C_n(\psi(x)))}. \tag{16.70}$$

By definition of $k_n(B)$ we have $2^{n/2-k_n(B)} \le 2\Delta(B)$, so that $2^{n-k_n(B)} \le 2^{n/2+1}\Delta(B)$, and (16.62), (16.70) and (16.69) prove that

$$\sum_{n\ge0} 2^{n-j_n^*(D_n(x))} \le L \,. \tag{16.71}$$

Since $x = \varphi(x) + \psi(x)$, for any $x, y \in D = B * C$ we have

$$\begin{aligned}(2^{j_n^*(D)}|x-y|) \wedge 1 \le (2^{j_n^*(D)}|\varphi(x)-\varphi(y)|) \wedge 1 + (2^{j_n^*(D)}|\psi(x)|) \wedge 1 \\ + (2^{j_n^*(D)}|\psi(y)|) \wedge 1 \,,\end{aligned} \tag{16.72}$$

and using the triangle inequality in ℓ^2 we obtain, using also (16.60) in the last inequality,

$$\begin{aligned}\|(2^{j_n^*(D)}|x-y|) \wedge 1\|_2 &\le \|(2^{j_n^*(D)}|\varphi(x)-\varphi(y)|) \wedge 1\|_2 + \|(2^{j_n^*(D)}|\psi(x)|) \wedge 1\|_2 \\ &\quad + \|(2^{j_n^*(D)}|\psi(y)|) \wedge 1\|_2 \\ &\le 2^{k(B)}\Delta(B) + \|(2^{\ell(C)}|\psi(x)|) \wedge 1\|_2 \\ &\quad + \|(2^{\ell(C)}|\psi(y)|) \wedge 1\|_2 \\ &\le 3 \cdot 2^{n/2} \le 2^{(n+4)/2} \,.\end{aligned} \tag{16.73}$$

We now define $\mathcal{A}_n = \mathcal{D}_{n-4}$ for $n \ge 4$ and $\mathcal{A}_n = \{T\}$ if $n \le 3$. Since we assume $0 \in T$ we have $\|t\|_\infty \le L$ for each $t \in T$. We then consider the smallest integer j^* such that $2^{-j^*-1} > L$, and we set $j_n(T) = j^*$ for $n \le 3$ to conclude the proof. □

16.4 1-Unconditional Bases and Gaussian Measures

Consider a Banach space E with a basis $(e_i)_{i\ge1}$, and independent standard normal r.v.s g_i. (There is little loss of generality to assume that E is finite dimensional, in which case no knowledge of basis theory is required.) How do we "compute" $\mathsf{E}\|\sum_{i\ge 1} g_i e_i\|$? We first deduce a possible answer from Theorem 2.4.18. We denote by x^* an element of the dual E^* of E.

To lighten notation in this section we write

$$a_n := \frac{1}{\sqrt{\log(n+1)}} \; . \tag{16.74}$$

Theorem 16.4.1. *Consider a Banach space E with a basis $(e_i)_{i\ge 1}$ and set $S = \mathsf{E}\|\sum_{i\ge 1} g_i e_i\|$. Assume that $S < \infty$. Then we can find a sequence $x_n^* \in E^*$ such that for each n we have*

$$\Big(\sum_{i\ge 1} x_n^*(e_i)^2\Big)^{1/2} \le a_n \tag{16.75}$$

and

$$\|x\| \le LS\mathcal{N}(x)$$

where $\mathcal{N}(x) := \sup_n |x_n^*(x)|$.

The point is that $\|x\| \le LS\mathcal{N}(x)$ while (using Proposition 2.4.16)

$$\mathsf{E}S\mathcal{N}\Big(\sum_{i\ge 1} g_i e_i\Big) \le LS = L\mathsf{E}\Big\|\sum_{i\ge 1} g_i e_i\Big\| \; .$$

In words, given a norm $\|\cdot\|$, if we are only interested in the quantity $S = \mathsf{E}\|\sum_{i\ge 1} g_i e_i\|$, our norm is in a sense equivalent to a norm of the type $S\mathcal{N}$, where $\mathcal{N}(x) = \sup_n |x_n^*(x)|$ for a sequence (x_n^*) that satisfies (16.75).

Proof of Theorem 16.4.1. Consider the set of sequences

$$T = \{(x^*(e_i))_{i\ge 1} \; ; \; \|x^*\| \le 1\} \; .$$

We observe that $\mathsf{E}|\sum_i g_i x^*(e_i)| = \mathsf{E}|x^*(\sum_i g_i e_i)| \le \mathsf{E}\|\sum_i g_i e_i\| \le S$, so that $T \subset \ell^2$. As usual, for a sequence $t = (t_i)_{i\ge 1} \in \ell^2$ we write $X_t = \sum_{i\ge 1} t_i g_i$. Keeping in mind the duality formula

$$\|x\| = \sup\{x^*(x) \; ; \; \|x^*\| \le 1\} \; , \tag{16.76}$$

we get $S = \mathsf{E}\|\sum_{i\ge 1} g_i e_i\| = \mathsf{E}\sup_{\|x^*\|\le 1} \sum_{i\ge 1} x^*(x_i) g_i = \mathsf{E}\sup_{t\in T} X_t$. To conclude, we apply Theorem 2.4.18 to a dense countable set T' of T to obtain a sequence $y_n = (y_{n,i})_{i\ge 1}$ with $\|y_n\|_2 \le a_n$ and $T' \subset LS\,\mathrm{conv}\{y_n; n \ge 1\}$, and where y_n is moreover a multiple of the difference of two elements of T. This last condition implies that there exists x_n^* in E^* with $y_n = (x_n^*(e_i))_{i\ge 1}$, i.e. $y_{n,i} = x_n^*(e_i)$. Thus (16.75) follows from $\|y\|_2 \le a_n$. Moreover, when $x = \sum_{i\ge 1} x_i e_i$ we obtain from (16.76) that

$$\begin{aligned}\|x\| &= \sup\Big\{\sum_{i\geq 1} x^*(e_i)x_i \; ; \; \|x^*\| \leq 1\Big\} = \sup\Big\{\sum_{i\geq 1} t_i x_i \; ; \; (t_i) \in T\Big\} \\ &= \sup\Big\{\sum_{i\geq 1} t_i x_i \; ; \; (t_i) \in T'\Big\} \leq LS \sup\Big\{\sum_{i\geq 1} y_{n,i} x_i \; ; \; n \geq 1\Big\} \\ &= LS \sup\Big\{\sum_{i\geq 1} x_n^*(e_i) x_i \; ; \; n \geq 1\Big\} = LS \sup_{n\geq 1} x_n^*(x) \; .\end{aligned}$$

This finishes the proof. □

Suppose now that the basis $(e_i)_{i\geq 1}$ is 1-unconditional for the norm $\|\cdot\|$, in the sense that the quantity $\|\sum_{i\geq 1} x_i e_i\|$ is invariant under changes of signs of any number of coordinates x_i. (We shall abuse terminology and simply describe this situation by saying that "the norm is 1-unconditional".) Then Theorem 16.4.1 is still true, but is not satisfactory because the norm $\mathcal{N}$ it produces is not 1-unconditional. In the present section we provide a version of Theorem 16.4.1 which is adapted to the case where the norm $\|\cdot\|$ is 1-unconditional.

Theorem 16.4.2. *Consider a Banach space E with a 1-unconditional basis $(e_i)_{i\geq 1}$. Assume that $\mathsf{E}\|\sum_{i\geq 1} g_i e_i\| = S < \infty$. Then we can find a sequence (I_n) of subsets of $\mathbb{N}^*$ with the following properties:*

$$\forall n \geq 1 \; , \; \operatorname{card} I_n \leq \log(n+1) \; , \tag{16.77}$$

and

$$\forall x \in E \, , \; x = \sum_{i\geq 1} x_i e_i \, , \; \|x\| \leq LS \sup_{n\geq 1} a_n \Big(\sum_{i\in I_n} x_i^2\Big)^{1/2} \; . \tag{16.78}$$

To explain this result, when $x = \sum_{i\geq 1} x_i e_i$, let us define

$$\mathcal{N}(x) = \sup_{n\geq 1} a_n \Big(\sum_{i\in I_n} x_i^2\Big)^{1/2} \; .$$

This norm is 1-unconditional and (16.78) implies $\|x\| \leq LS\mathcal{N}(x)$. Moreover, as will follow from Lemma 16.4.4 below, we have $\mathsf{E}\mathcal{N}(\sum_{i\geq 1} g_i e_i) \leq L$. In words, given a 1-unconditional norm $\|\cdot\|$, if we are only interested in the quantity $S = \mathsf{E}\|\sum_{i\geq 1} g_i e_i\|$, our norm is in a sense equivalent to a 1-unconditional norm of the type $S\mathcal{N}$.

Exercise 16.4.3. In the statement of Theorem 16.4.2 prove that one may instead request $\operatorname{card} I_n \geq \log(1+n)$ and replace (16.78) by

$$\|x\| \leq LS \sup_{n\geq 1} \Big(\frac{1}{\operatorname{card} I_n} \sum_{i\in I_n} x_i^2\Big)^{1/2} \; .$$

(Hint: one is permitted to add more sets I_n!)

Lemma 16.4.4. *Assume that the sets I_n satisfy (16.77). Then if (g_i) are independent standard Gaussian r.v.s we have*

$$\mathsf{E}\sup_{n\geq 1} a_n \Big(\sum_{i\in I_n} g_i^2\Big)^{1/2} \leq L\;. \tag{16.79}$$

Proof. For each i we have $\mathsf{E}\exp(g_i^2/4)\leq 2$, so that for any set I,

$$\mathsf{E}\exp\Big(\frac{1}{4}\sum_{i\in I} g_i^2\Big) \leq 2^{\operatorname{card} I}\;,$$

and, for $v\geq 8\operatorname{card} I$,

$$\mathsf{P}\Big(\sum_{i\in I} g_i^2 \geq v\Big) \leq 2^{\operatorname{card} I}\exp\Big(-\frac{v}{4}\Big) \leq \exp\Big(-\frac{v}{8}\Big)\;.$$

Now, (16.77) implies that for $w^2\geq 8$, we have, using of course the value (16.74) of a_n,

$$\begin{aligned}\mathsf{P}\Big(\sup_{n\geq 1} a_n\Big(\sum_{i\in I_n} g_i^2\Big)^{1/2}\geq w\Big) &\leq \sum_{n\geq 1}\mathsf{P}\Big(\sum_{i\in I_n} g_i^2\geq w^2\log(n+1)\Big)\\ &\leq \sum_{n\geq 1}\exp\Big(-\frac{w^2\log(n+1)}{8}\Big)\;,\end{aligned}$$

and the last sum is $\leq L\exp(-w^2/L)$ for w large enough. □

Exercise 16.4.5. If the sets I_n satisfy $\operatorname{card} I_n\geq \log(n+1)$, prove that

$$\mathsf{E}\sup_{n\geq 1}\Big(\frac{1}{\operatorname{card} I_n}\sum_{i\in I_n} x_i^2\Big)^{1/2}\leq L\;.$$

Exercise 16.4.6. Given a finite subset I of $\mathbb{N}^*$, and a number $a>0$, let us define $B_2(I,a)$ as the set of elements with support in I and with ℓ^2 norm $\leq a$, i.e.

$$B_2(I,a) = \Big\{x\in\ell^2\;;\; i\notin I\Rightarrow x_i=0\;;\;\sum_{i\in I}x_i^2\leq a^2\Big\}\;. \tag{16.80}$$

Find an other proof of Lemma 16.4.4 by constructing a sequence (u_k) with $\|u_k\|\leq La_k$ and

$$\bigcup_{n\geq 1} B_2(I_n,a_n)\subset \operatorname{conv}\{u_k\;;\;k\geq 1\}\;.$$

(Hint: the Euclidean unit ball of $\mathbb{R}^d$ is contained in the convex hull of a set of 5^d vectors of length ≤ 2. This follows from (2.41) with $\epsilon = 1/2$. Use this for each ball $B_2(I,a)$ in the left-hand side above.)

We recall that for $T \subset \ell^2$ we write

$$g(T) = \mathsf{E}\sup_{t\in T} X_t = \mathsf{E}\sup_{t\in T}\sum_{i\geq 1} t_i g_i \ .$$

It does not hurt to present another proof of Lemma 16.4.4 based on a general principle that we spell out now.

Proposition 16.4.7. *Consider subsets T_n of ℓ^2, and assume that for certain numbers b_n we have $\|x\| \leq b_n$ for $x \in T_n$. Then*

$$g\Big(\bigcup_{n\geq 1} T_n\Big) \leq \sup_n \Big(g(T_n) + 2b_n\sqrt{\log(n+1)}\Big) + L\sup_n b_n \ . \tag{16.81}$$

This is a generalization of Proposition 2.4.16 which we recover when the sets T_n consist of a single point.

Proof. We may assume each set T_n to be finite. Lemma 2.4.7 (i.e. the concentration inequality for the supremum of a Gaussian process) implies

$$\mathsf{P}\Big(\sup_{T_n} X_t \geq g(T_n) + u\Big) \leq 2\exp\Big(-\frac{u^2}{2b_n^2}\Big)$$

so that

$$\mathsf{P}\Big(\sup_{T_n} X_t \geq g(T_n) + 2b_n\sqrt{\log(n+1)} + ub_n\Big) \leq 2\exp\Big(-\frac{u^2}{2} - 2\log(n+1)\Big)$$

and thus

$$\mathsf{P}\bigg(\sup_{\bigcup T_n} X_t \geq \sup_n \Big(g(T_n) + 2b_n\sqrt{\log(n+1)}\Big) + u\sup_n b_n\bigg) \leq L\exp\Big(-\frac{u^2}{2}\Big) \ ,$$

and in turn

$$g\Big(\bigcup_n T_n\Big) \leq \sup_n \Big(g(T_n) + 2b_n\sqrt{\log(n+1)}\Big) + L\sup_n b_n \ . \qquad \square$$

Second proof of Lemma 16.4.4. Since for $x \in B_2(I,a)$ we have

$$\sum_{i\geq 1} x_i g_i \leq a\Big(\sum_{i\in I} g_i^2\Big)^{1/2} \ ,$$

using the Cauchy-Schwarz inequality in the second inequality below we obtain

$$g(B_2(I,a)) \leq a\mathsf{E}\Big(\sum_{i\in I} g_i^2\Big)^{1/2} \leq a\sqrt{\operatorname{card} I} \ . \tag{16.82}$$

We recall the notation $a_n = 1/\sqrt{\log(n+1)}$. Let $T_n = B_2(I_n, a_n)$, so that $g(T_n) \le a_n \sqrt{\operatorname{card} I_n} \le 1$ by (16.82) and (16.77). Thus (16.81) implies

$$g\Big(\bigcup_n B_2(I_n, a_n)\Big) = g(\cup_n T_n) \le L \, . \qquad \square$$

We start the proof of Theorem 16.4.2 with a simple observation.

Lemma 16.4.8. *Assume that the norm* $\|\cdot\|$ *is* 1*-unconditional and let* $S = \mathsf{E}\|\sum_{i\ge1} g_i x_i\|$*. Then the set*

$$T = \big\{(x^*(e_i))_{i\ge1} \; ; \; x^* \in E^* \, , \; \|x^*\| \le 1\big\} \tag{16.83}$$

satisfies

$$\forall\, y \in T \; , \; \sum_{i\ge1} |y_i| \le 2S \, . \tag{16.84}$$

Proof. Denote by η_i the sign of $g_i x^*(e_i)$, so that

$$\begin{aligned}\sum_{i\ge1} |g_i||x^*(e_i)| &= \sum_{i\ge1} |x^*(g_i e_i)| = \sum_{i\ge1} \eta_i x^*(g_i e_i) \\ &= x^*\Big(\sum_{i\ge1} \eta_i g_i e_i\Big) \le \Big\|\sum_{i\ge1} \eta_i g_i e_i\Big\| = \Big\|\sum_{i\ge1} g_i e_i\Big\| \, .\end{aligned}$$

Taking expectation completes the proof since $\mathsf{E}|g_i| = \sqrt{2/\pi} \ge 1/2$. $\square$

Let us now recall the notation $B_1 = \{y \in \ell^2; \sum_{i\ge1} |y_i| \le 1\}$ and let us state the main step in the proof of Theorem 16.4.2, where we use the notation (16.80).

Theorem 16.4.9. *Consider a subset* T *of* ℓ^2*. Assume that for a certain number* S *we have* $\gamma_2(T, d_2) \le S$ *and* $T \subset SB_1$*. Then there exist sets* I_n *such that* $\operatorname{card} I_n \le \log(n+1)$ *with*

$$T \subset LS \, \overline{\operatorname{conv}} \bigcup_{n\ge1} B_2(I_n, a_n) \, , \tag{16.85}$$

where $\overline{\operatorname{conv}}\, Z$ *denotes the closed convex hull of* Z*.*

Proof of Theorem 16.4.2. We recall the set T of (16.83). Lemma 16.4.8 implies that $T \subset 2SB_1$. Moreover Theorem 2.4.1 implies that $\gamma_2(T) \le Lg(T)$, whereas

$$g(T) = \mathsf{E} \sup_{\|x^*\|\le1} x^*\Big(\sum_{i\ge1} g_i e_i\Big) = \mathsf{E}\Big\|\sum_{i\ge1} g_i e_i\Big\| = S \, .$$

Theorem 16.4.9 provides sets I_n that satisfy (16.77) and $T \subset LST_1$, where

$$T_1 = \overline{\mathrm{conv}} \bigcup_{n \ge 1} B_2(I_n, a_n) \, .$$

Thus, by duality, if $x = \sum_{i \ge 1} x_i e_i \in E$, we have for any k

$$\Big\| \sum_{i \le k} x_i e_i \Big\| \le LS \sup_{t \in T_1} \sum_{i \le k} t_i x_i \le LS \sup_{n \ge 1} a_n \Big(\sum_{i \in I_n} x_i^2 \Big)^{1/2}$$

and this proves (16.78) since $\|x\| = \sup_k \| \sum_{i \le k} x_i e_i \|$ (a property which is part of the definition that (e_i) is a basis). □

Proof of Theorem 16.4.9. By homogeneity we may assume that $S = 1$. We recall that we denote by $\Delta_2(A)$ the diameter of A for the distance induced by ℓ^2. We consider an admissible sequence $(\mathcal{B}_n)$ with

$$\sup_{t \in T} \sum_{n \ge 0} 2^{n/2} \Delta_2(B_n(t)) \le 2 \, . \tag{16.86}$$

We consider the admissible sequence $(\mathcal{C}_n)$ provided by Theorem 16.3.1 when $\tau = 0$, $\Omega = \mathbb{N}^*$ and μ is the counting measure. We consider the increasing sequence of partitions $(\mathcal{A}_n)_{n \ge 0}$ where $\mathcal{A}_n$ is generated by $\mathcal{B}_n$ and $\mathcal{C}_n$, so $\mathrm{card}\, \mathcal{A}_n \le N_{n+1}$. The numbers $\ell(t,n)$ of (16.59) depend only on $A_n(t)$. Therefore

$$s \in A_n(t) \Rightarrow \ell(s,n) = \ell(t,n) \, . \tag{16.87}$$

For every $A \in \mathcal{A}_n$, we pick an arbitrary element $x(A) = (x_i(A))_{i \ge 1}$ of A, and we set

$$J_n(A) = \big\{ i \in \mathbb{N}^* \; ; \; |x_i(A)| > 2^{-\ell(x(A),n)} \big\} \, ,$$

so that $\mathrm{card}\, J_n(A) \le 2^n$ by (16.62). For $n \ge 1$ and $A \in \mathcal{A}_n$, consider the unique element $B \in \mathcal{A}_{n-1}$ such that $A \subset B$, and set

$$I_n(A) = J_n(A) \backslash J_{n-1}(B)$$

so that $\mathrm{card}\, I_n(A) \le 2^n$ and

$$i \in I_n(A) \Rightarrow |x_i(B)| \le 2^{-\ell(x(B),n-1)} \, . \tag{16.88}$$

We define $I_0(T) = J_0(T)$ and $\mathcal{F}$ as the family of pairs $(I_n(A), 2^{-n/2})$ for $A \in \mathcal{A}_n$ and $n \ge 0$. The heart of the argument is to prove that

$$T \subset L \, \overline{\mathrm{conv}} \bigcup_{\mathcal{F}} B_2(I,a) \, . \tag{16.89}$$

So let us fix $t \in T$ and for $n \ge 1$ define $I_n(t) = I_n(A_n(t))$. We observe that since $x(A_{n-1}(t)) \in A_{n-1}(t)$, it follows from (16.87) (used for $s = x(A_{n-1}(t))$ and $n-1$ rather than n) that

$$\ell(x(A_{n-1}(t)), n-1) = \ell(t, n-1) \, .$$

Using (16.88) for $B = A_{n-1}(t)$ implies

$$\|x(A_{n-1}(t))\mathbf{1}_{I_n(t)}\|_\infty \le 2^{-\ell(t,n-1)}\,, \tag{16.90}$$

and thus, since $\operatorname{card} I_n(t) \le 2^n$,

$$\|x(A_{n-1}(t))\mathbf{1}_{I_n(t)}\|_2 \le 2^{n/2-\ell(t,n-1)}\,.$$

Since $t, x(A_n(t)) \in A_{n-1}(t)$ we have $\|t - x(A_{n-1}(t))\|_2 \le \Delta_2(A_{n-1}(t))$ and thus

$$\|t\mathbf{1}_{I_n(t)}\|_2 \le c(t,n) := \Delta_2(A_{n-1}(t)) + 2^{n/2-\ell(t,n-1)}\,, \tag{16.91}$$

and hence

$$t\mathbf{1}_{I_n(t)} \in 2^{n/2}c(t,n)B_2(I_n(t), 2^{-n/2})\,. \tag{16.92}$$

For each $t \in T$ we define $c(t,0) = 1$. Since $t \in T \subset B_1$ and $\operatorname{card} J_0(T) = \operatorname{card} I_0(t) \le 2^0 = 1$, (16.92) also holds for $n = 0$. We claim now that

$$t = \sum_{n\ge 0} t\mathbf{1}_{I_n(t)}\,. \tag{16.93}$$

Since by construction the sets $(I_n(t))_{n\ge 0}$ are disjoint, it suffices to show that the support of t is contained in the union of the supports of the t_i's, i.e. that

$$|t_i| > 0 \Rightarrow i \in \bigcup_{n\ge 0} I_n(t) = \bigcup_{n\ge 0} I_n(A_n(t)) = \bigcup_{n\ge 0} J_n(A_n(t))\,. \tag{16.94}$$

To prove this, consider i with $|t_i| > 0$ and n large enough so that $\Delta_2(A_n(t)) < |t_i|/2$. Then for all $x \in A_n(t)$ we have $|x_i - t_i| \le |t_i|/2$ and hence $|x_i| > |t_i|/2$. Since $\ell(x,n) \ge n-3$ by (16.61), if n is large enough, for all $x \in A_n(t)$ we have $2^{-\ell(x,n)} < |x_i|$. This holds in particular for $x = x(A_n(t))$. Thus, by definition of $J_n(A)$ this shows that $i \in J_n(A_n(t))$. This proves (16.94) and hence (16.93).

Thus we have written $t = \sum_{n\ge 0} t_n$ where

$$t_n := t\mathbf{1}_{I_n(t)} \in b(t,n)B_2(I_n(t), 2^{-n/2})\,, \tag{16.95}$$

for $b(t,n) = 2^{n/2}c(t,n)$. By (16.86) and (16.61) we have $\sum_{n\ge 0} b(t,n) \le L$. Let us write (16.95) as $t_n = b(t,n)u_n$ where $u_n \in B_2(I_n(t), 2^{-n/2})$. Then

$$t = \sum_n t_n = \sum_n b(t,n)u_n = A\sum_n \alpha_n u_n\,,$$

where $A = \sum_n b(t,n)$ and $\alpha_n = b(t,n)/A$, so that $\sum_n \alpha_n = 1$. This completes the proof of (16.89).

We now enumerate all the sets of the type $I_n(A)$ for $n \ge 0$ and $A \in A_n$ as a single sequence $(I_k)_{k\ge 1}$, in a special way, as follows. Let $M_n = \sum_{0\le k\le n} N_k$, so that $M_{n+1} - M_n = N_{n+1} \ge \operatorname{card} \mathcal{A}_n$. We enumerate the elements of the type $I_n(A)$ for $A \in \mathcal{A}_n$ using the integers $M_{n+1} < k \le M_{n+2}$, and, allowing repetition if necessary, using all these integers. Thus if $I_n(A) = I_k$ then

$$N_{n+1} < M_{n+1} < k \le M_{n+2} = N_0 + \cdots + N_{n+2} \le N_{n+3} - 1$$

and thus, since $2^n \le \log N_{n+1} = 2^{n+1} \log 2 \le 2^{n+1}$,

$$2^n \le \log(k+1) \le 2^{n+3} .$$

In particular,

$$a_k = \frac{1}{\sqrt{\log(k+1)}} \ge 2^{-n/2-2}$$

and $B(I_n(A), 2^{-n/2}) \subset 4B(I_k, a_k)$, so that (16.89) proves (16.85). Finally we have $\operatorname{card} I_k = \operatorname{card} I_n(A) \le 2^n \le \log(k+1)$. □

Consider a set T, and the following properties: $\gamma_2(T, d_2) \le 1$; $T \subset B_1$; $\gamma_1(T, d_\infty) \le 1$ (where d_∞ denotes the distance associated with the supremum norm). Numerous relations exist between these properties, a theme that we started exploring in Chapter 5. We pursue its investigation in the rest of this section, a circle of ideas closely connected to the investigations of Section 16.5 below. The essence of Theorem 16.4.10 below is that the conditions $T \subset B_1$ and $\gamma_1(T, d_\infty) \le 1$ taken together are very restrictive.

For $I \subset \mathbb{N}^*$ and $a > 0$, we define $B_\infty(I, a)$ as the set of elements of support in I and of ℓ^∞ norm $\le a$, i.e.

$$B_\infty(I, a) = \bigl\{x = (x_i)_{i \ge 1} \; ; \; i \notin I \Rightarrow x_i = 0 \, ; \, i \in I \Rightarrow |x_i| \le a\bigr\} . \tag{16.96}$$

We have

$$x \in B_\infty(I, a) \Rightarrow \sum_{i \ge 1} x_i^2 \le a^2 \operatorname{card} I$$

and thus, recalling the sets $B_2(I, a)$ of (16.80), this implies

$$B_\infty(I, a) \subset B_2(I, a\sqrt{\operatorname{card} I}) . \tag{16.97}$$

Theorem 16.4.10. *Consider a set $T \subset SB_1$, and assume that $\gamma_1(T, d_\infty) \le S$. Then we can find subsets I_n of $\mathbb{N}^*$ with $\operatorname{card} I_n \le \log(n+1)$, for which*

$$T \subset LS \, \overline{\operatorname{conv}} \bigcup_{n \ge 1} B_\infty\Bigl(I_n, \frac{1}{\log(n+1)}\Bigr) . \tag{16.98}$$

Let us observe that (16.97) and (16.98) imply that

$$T \subset LS \, \overline{\operatorname{conv}} \bigcup_{n \ge 1} B_2(I_n, a_n) ,$$

As Lemma 16.4.4 shows, this implies that $\gamma_2(T, d_2) \le LS$. The information provided by (16.98) is however very much stronger than this. In Theorem 16.4.12 below we provide a direct and more general proof that $\gamma_2(T, d_2) \le L$.

Proof. By homogeneity we may assume that $S = 1$. We proceed as in the proof of Theorem 16.4.9, but we may now assume

$$\forall t \in T\,,\ \sum_{n\ge 0} 2^n \Delta_\infty(A_n(t)) \le 2\,.$$

Using (16.90) rather than (16.91) we get

$$\|t\mathbf{1}_{I_n(t)}\|_\infty \le c(t,n) := \Delta_\infty(A_{n-1}(t)) + 2^{-\ell(t,n-1)}$$

so that

$$t\mathbf{1}_{I_n(t)} \in 2^n c(t,n) B_\infty(I_n(t), 2^{-n})$$

and the proof is finished exactly as before. □

Exercise 16.4.11. Assume that the r.v.s Y_i are independent, symmetric and satisfy $\mathsf{P}(|Y_i| \ge x) = \exp(-x)$. Consider a Banach space E with a 1-unconditional basis (e_i), and assume that $S = \mathsf{E}\|\sum_{i\ge 1} Y_i e_i\| < \infty$. Then we can find a sequence (I_n) of subsets of $\mathbb{N}^*$ with $\operatorname{card} I_n \le \log(n+1)$ and

$$\forall\, x \in E\ ,\ x = \sum_{i\ge 1} x_i e_i\ ,\ \|x\| \le LS \sup_{n\ge 1} \frac{1}{\log(n+1)} \sum_{i\in I_n} |x_i|\ .$$

(Hint: copy the proof of Theorem 16.4.2 using Theorems 10.2.8 and 16.4.10.)

As we have explained, under the conditions of Theorem 16.4.10, we have $\gamma_2(T, d_2) \le LS$. It is of course of interest to give a more direct (and more general) proof of this fact. This is the purpose of the next result, where we go back to the general setting of a measured space (Ω, μ).

Theorem 16.4.12. *Consider a measured space (Ω, μ) and $T \subset SB_1$, such that $\gamma_1(T, d_\infty) < \infty$. Then*

$$\gamma_2(T, d_2) \le L\sqrt{S\gamma_1(T, d_\infty)}\ . \tag{16.99}$$

Proof. Consider $\tau \in \mathbb{Z}$ to be chosen later, and consider the admissible sequence $(\mathcal{C}_n)$ as provided by Theorem 16.3.1. As T is a class of functions we denote its generic element by f. By homogeneity we may assume that $S = 1$.

Consider an admissible sequence of partitions $(\mathcal{B}_n)$ such that

$$\sup_{f\in T} \sum_{n\ge 0} 2^n \Delta_\infty(B_n(f)) \le 2\gamma_1(T, d_\infty)\ .$$

Consider the partition $\mathcal{A}_n$ generated by $\mathcal{B}_n$ and $\mathcal{C}_n$. As in the proof of Theorem 16.4.9 it satisfies $\operatorname{card} \mathcal{A}_n \le N_{n+1}$, and

$$\sup_{f\in T} \sum_{n\ge 0} 2^n \Delta_\infty(A_n(f)) \le 2\gamma_1(T, d_\infty)\,. \tag{16.100}$$

Moreover, the numbers $\ell(f,n)$ depend only on $A_n(f)$.

Consider $g, h \in A_n(f)$, and set $\Delta = \Delta_\infty(A_n(f))$, so that $\|g-h\|_\infty \le \Delta$. Thus

$$\begin{aligned}\|g-h\|_2 &\le \|\min(|g-h|, \Delta)\|_2 \\ &\le \|\min(|g|, \Delta)\|_2 + \|\min(|h|, \Delta)\|_2 \,. \end{aligned} \tag{16.101}$$

Now

$$\min(|f|, \Delta) \le \Delta \mathbf{1}_{\{|f| \ge 2^{-\ell(f,n)}\}} + |f| \mathbf{1}_{\{|f| \le 2^{-\ell(f,n)}\}}$$

and, using (16.62) and (16.60) in the second line we get,

$$\begin{aligned}\|\min(|f|, \Delta)\|_2 &\le \Delta \sqrt{\mu(\{|f| \ge 2^{-\ell(f,n)}\})} + \||f| \wedge 2^{-\ell(f,n)}\|_2 \\ &\le L\Delta 2^{n/2+\tau/2} + L 2^{n/2-\ell(f,n)+\tau/2} \,. \end{aligned} \tag{16.102}$$

Therefore, combining with (16.101),

$$\Delta_2(A_n(f)) \le L\big(2^{n/2+\tau/2}\Delta_\infty(A_n(f)) + 2^{n/2-\ell(f,n)+\tau/2}\big) \,.$$

Combining (16.100) and (16.61) yields

$$\sum_{n \ge 0} 2^{n/2} \Delta_2(A_n(f)) \le L(2^{\tau/2}\gamma_1(T, d_\infty) + 2^{-\tau/2}) \,.$$

Since $\Delta_2(T) \le L$, appealing to Lemma 2.3.5 and choosing τ appropriately finishes the proof. □

16.5 Restriction of Operators

Consider $q > 1$, the space ℓ_N^q, and its canonical basis $(e_i)_{i \le N}$. Consider a Banach space X and an operator $U : \ell_N^q \to X$. We will give (surprisingly mild) conditions under which there are large subsets J of $\{1, \ldots, N\}$ such that the norm $\|U_J\|$ of the restriction U_J to the span of the vectors $(e_i)_{i \in J}$ is much smaller than the norm of U. We first compute this norm. We denote by X_1^* the unit ball of the dual of X, by p the conjugate exponent of q. Setting $x_i = U(e_i)$, we have

$$\begin{aligned}\|U_J\| &= \sup\Big\{\sum_{i \in J} \alpha_i x^*(x_i) \; ; \; \sum_{i \in J} |\alpha_i|^q \le 1 \,, \, x^* \in X_1^*\Big\} \\ &= \sup\Big\{\Big(\sum_{i \in J} |x^*(x_i)|^p\Big)^{1/p} \; ; \; x^* \in X_1^*\Big\} \,. \end{aligned} \tag{16.103}$$

The set J will be constructed by a random choice. Specifically, given a number $0 < \delta < 1$, we consider (as in Section 12.2) i.i.d. r.v.s $(\delta_i)_{i \le N}$ with

$$\mathsf{P}(\delta_i = 1) = \delta \ ; \ \mathsf{P}(\delta_i = 0) = 1 - \delta \,, \tag{16.104}$$

and we set $J = \{i \le N ; \delta_i = 1\}$. Thus (16.103) implies

$$\|U_J\|^p = \sup_{t \in T} \sum_{i \le N} \delta_i |t_i|^p \,, \tag{16.105}$$

where

$$T = \{(x^*(x_i))_{i \le N} \ ; \ x^* \in X_1^*\} \,. \tag{16.106}$$

For a subset T of $\mathbb{R}^N$, we set

$$|T|^p = \{(|t_i|^p)_{i \le N} \ ; \ t \in T\} \,.$$

Thus, if T is the set (16.106) we may rewrite (16.105) as

$$\|U_J\|^p = \sup_{t \in |T|^p} \sum_{i \le N} \delta_i t_i \,. \tag{16.107}$$

This brings forward the essential point: to control $\mathsf{E}\|U_J\|^p$ we need information on the set $|T|^p$. On the other hand, information we might gather from the properties of X as a Banach space is likely to bear on T rather than $|T|^p$. The link between the properties of T and $|T|^p$ is provided in Theorem 16.5.1 below, which transfers a certain "smallness" property of T into an appropriate smallness property of $|T|^p$.

Before we state this theorem we make a simple observation: interchanging the supremum and the expectation, we have

$$\mathsf{E}\|U_J\|^p \ge \sup_{t \in T} \mathsf{E}\Big(\sum_{i \le N} \delta_i |t_i|^p\Big) = \delta \sup_{t \in T} \sum_{i \le N} |t_i|^p \,. \tag{16.108}$$

This demonstrates the relevance of the quantity

$$\sup_{t \in T} \sum_{i \le N} |t_i|^p = \sup_{\|x^*\| \le 1} \sum_{i \le N} |x^*(e_i)|^p \,, \tag{16.109}$$

and why we need to control it from above if we want to control $\mathsf{E}\|U_J\|^p$ from above.

We recall from (16.96) that for a subset I of $\{1, \ldots, N\}$ and for $a > 0$ we write

$$B_\infty(I, a) = \big\{(t_i)_{i \le N} \ ; \ i \notin I \Rightarrow t_i = 0 \,, \forall i \in I \,, |t_i| \le a\big\} \,.$$

The following is closely related to Theorem 16.4.10.

Theorem 16.5.1. *Consider a subset T of $\mathbb{R}^N$ with $0 \in T$. Assume that there exists an admissible sequence $(\mathcal{B}_n)$ of T such that*

$$\forall t \in T \,, \sum_{n \ge 0} 2^n \Delta^p(B_n(t), d_\infty) \le A \tag{16.110}$$

and let

$$B = \max\Big(A, \sup_{t \in T} \sum_{i \le N} |t_i|^p\Big) . \tag{16.111}$$

Then we can find a sequence $(I_n)_{n \ge 1}$ *of subsets of* $\{1, \ldots, N\}$ *with*

$$\operatorname{card} I_n \le \frac{LB}{A} \log(n+1) , \tag{16.112}$$

and

$$|T|^p \subset K(p)A \ \operatorname{conv} \bigcup_{n \ge 1} B_\infty\Big(I_n, \frac{1}{\log(n+1)}\Big) . \tag{16.113}$$

Proof. The proof resembles that of Theorems 16.4.9 and 16.4.10 (but may be read independently). Consider the largest integer τ for which $2^\tau \le B/A$. Since $B \ge A$, we have $\tau \ge 0$, and $2^{-\tau} \le 2A/B$.

The set $|T|^p$ is a subset of the ball W of $L^1(\mu)$ of center 0 and radius B, where μ is the counting measure on $\{1, \ldots, N\}$, and the first step of the proof is to take full advantage of this. Homogeneity and Theorem 16.3.1 provide us with an admissible sequence of partitions $(\mathcal{D}_n)$ of $|T|^p$ and for each $D \in \mathcal{D}_n$ an integer $\ell^*(D) \in \mathbb{Z}$, such that if for $t \in |T|^p$ we set

$$\ell^*(t,n) = \ell^*(D_n(t)) \tag{16.114}$$

then

$$\forall t \in |T|^p , \operatorname{card} \{i \le N \, ; \, t_i \ge 2^{-\ell^*(t,n)}\} \le 2^{n+\tau} \le \frac{2^n B}{A} \tag{16.115}$$

$$\forall t \in |T|^p \ , \ \sum_{n \ge 0} 2^{n-\ell^*(t,n)} \le 12 \cdot 2^{-\tau} B \le LA . \tag{16.116}$$

Consider the canonical map $\varphi : T \to |T|^p$ given by $\varphi((t_i)_{i \le N}) = (|t_i|^p)_{i \le N}$. We consider on T the admissible sequence of partitions $(\mathcal{C}_n)$ where $\mathcal{C}_n$ consists of the sets $\varphi^{-1}(D)$ where $D \in \mathcal{D}_n$. For $t \in T$ we define $\ell(t,n) = \ell^*(\varphi(t), n)$, and this number depends only on $C_n(t)$. Moreover, we deduce from (16.115) and (16.116) respectively that

$$\forall t \in T , \operatorname{card} \{i \le N \, ; \, |t_i|^p \ge 2^{-\ell(t,n)}\} \le \frac{2^n B}{A} \tag{16.117}$$

$$\forall t \in T \ , \ \sum_{n \ge 0} 2^{n-\ell(t,n)} \le LA . \tag{16.118}$$

The sequence of partitions $\mathcal{A}_n$ generated by $\mathcal{B}_n$ and $\mathcal{C}_n$ is increasing and $\operatorname{card} \mathcal{A}_n \le N_{n+1}$. Moreover since $A_n(t) \subset B_n(t)$, (16.110) implies

$$\forall t \in T , \ \sum_{n \ge 0} 2^n \Delta^p(A_n(t), d_\infty) \le A , \tag{16.119}$$

and furthermore the integer $\ell(t,n)$ depends only on $A_n(t)$.

After these preparations we start the main construction. For $A \in \mathcal{A}_n$, $n \geq 0$, let us choose in an arbitrary manner $u_n(A) \in A$, and set $\pi_n(t) = u_n(A_n(t))$. We write $\pi_n(t) = (\pi_{n,i}(t))_{i \leq N}$ and we define

$$I_0(t) = \{i \leq N \; ; \; |\pi_{0,i}(t)|^p \geq 2^{-\ell(t,0)}\} . \tag{16.120}$$

For $n \geq 1$ we further define

$$I_n(t) = \{i \leq N \; ; \; |\pi_{n,i}(t)|^p \geq 2^{-\ell(t,n)} \, , \, |\pi_{n-1,i}(t)|^p < 2^{-\ell(t,n-1)}\} \, .$$

It is good to observe right now that $I_n(t)$ depends only on $A_n(t)$, so that there are at most $\operatorname{card} \mathcal{A}_n \leq N_{n+1}$ sets of this type. Next, since $|t_i - \pi_{n,i}(t)| \leq \Delta(B_n(t), d_\infty)$ we have $\lim_{n\to\infty} |t_i - \pi_{n,i}(t)| = 0$ and thus

$$\{i \leq N \; ; \; |t_i| \neq 0\} \subset \bigcup_{n \geq 0} I_n(t) \; . \tag{16.121}$$

Finally we note from (16.117) that

$$\operatorname{card} I_n(t) \leq \frac{2^n B}{A} \; . \tag{16.122}$$

The definition of $I_n(t)$ show that, for $n \geq 1$ and $i \in I_n(t)$, we have

$$\begin{aligned} |t_i| &\leq |t_i - \pi_{n-1,i}(t)| + |\pi_{n-1,i}(t)| \\ &\leq \Delta(A_{n-1}(t), d_\infty) + 2^{-\ell(t,n-1)/p} \end{aligned}$$

and hence

$$|t_i|^p \leq K(p)(\Delta(A_{n-1}(t), d_\infty)^p + 2^{-\ell(t,n-1)}) := c(t,n) \; . \tag{16.123}$$

Since $0 \in T$, this remains true for $n = 0$ if we define $c(t,0) = \Delta(T, d_\infty)^p$, so that, finally,

$$n \geq 0 \; , \; i \in I_n(t) \Rightarrow |t_i|^p \leq c(t,n) \; . \tag{16.124}$$

Moreover (16.119) and (16.118) imply

$$\forall t \in T \, , \sum_{n \geq 0} 2^n c(t,n) \leq K(p) A \; . \tag{16.125}$$

We consider the family $\mathcal{F}$ of all pairs $(I_n(t), 2^{-n})$ for $t \in T$ and $n \geq 0$, and we prove that

$$|T|^p \subset K(p) A \operatorname{conv} \bigcup_{(I,a) \in \mathcal{F}} B_\infty(I,a) \; . \tag{16.126}$$

For this we simply write, using (16.121)

$$|t|^p = \sum_{n \geq 0} 2^n c(n,t) u_n \ , \tag{16.127}$$

where, using (16.124),

$$u_n = \frac{1}{2^n c(n,t)} |t|^p \mathbf{1}_{I_n(t)} \in B_\infty(I_n(t), 2^{-n}) \ .$$

Then (16.127) and (16.125) prove that $|t|^p \in K(p) A \operatorname{conv} \cup_{(I,a)\in\mathcal{F}} B_\infty(I,a)$ and (16.126).

Let us now enumerate all the sets of the type $I_n(t)$ for $n \geq 0$ and $t \in T$ as a single sequence $(I_k)_{k\geq 1}$, in a way that if $n < n'$ the sets of the type $I_n(t)$ are enumerated before the sets of the type $I_{n'}(t')$. Then if $I_k = I_n(t)$, since $I_n(t)$ depends only on $A_n(t)$, and since $\operatorname{card} \mathcal{A}_n \leq N_{n+1}$, we have

$$k \leq N_1 + \cdots + N_{n+1} \leq N_{n+2} - 1$$

and thus $k + 1 \leq N_{n+2} = 2^{2^{n+2}}$ so that

$$\frac{1}{\log(k+1)} \geq 2^{-n-2}$$

and thus $B_\infty(I_n(t), 2^{-n}) \subset 4B_\infty(I_k, 1/\log(k+1))$, so that (16.126) implies (16.113).

Furthermore, by labeling several times if necessary certain elements $I_n(t)$ we may also assume that when $I_n(t) = I_k$ then $k \geq N_n$, and (16.122) implies then (16.112). □

The smallness criterion provided by (16.113) is perfectly adapted to the control of $\mathsf{E}\|U_J\|^p$.

Theorem 16.5.2. *Consider the set* $T = \{(x^*(e_i)); x^* \in X_1^*\}$ *of (16.106), and assume that (16.110) and (16.111) hold. Consider* $\epsilon > 0$ *and* $\delta \leq 1$ *such that*

$$\delta \leq \frac{A}{B\epsilon N^\epsilon \log N} \ . \tag{16.128}$$

Then if the r.v.s $(\delta_i)_{i\leq N}$ *are as in (16.104) and* $J = \{i \leq N; \delta_i = 1\}$*, for* $v > 0$ *we have*

$$\mathsf{P}\left(\|U_J\|^p \geq v K(p) \frac{A}{\epsilon \log N} \right) \leq L \exp\left(-\frac{v}{L}\right)$$

and in particular

$$\mathsf{E}\|U_J\|^p \leq K(p) \frac{A}{\epsilon \log N} \ . \tag{16.129}$$

Proof. The magic is that

$$\sup_{t\in B_\infty(I,a)}\sum_{i\le N}\delta_i t_i \le a\sum_{i\in I}\delta_i \ ,$$

so that (16.113) implies

$$\sup_{t\in|T|^p}\sum_{i\le N}\delta_i t_i \le K(p)A\sup_{n\ge 1}\frac{1}{\log(n+1)}\sum_{i\in I_n}\delta_i \ , \tag{16.130}$$

and that the right-hand side can be controlled by using the union bound. For $n\ge 1$ we have $\operatorname{card} I_n \le L_0\log(n+1)B/A$ so that

$$\delta\operatorname{card} I_n \le \frac{L_0\log(n+1)}{\epsilon N^\epsilon\log N} \ .$$

We recall the inequality (12.6): If $u\ge 6\delta\operatorname{card} I$,

$$P\Big(\sum_{i\in I}\delta_i\ge u\Big)\le\exp\Big(-\frac{u}{2}\log\frac{u}{2\delta\operatorname{card} I}\Big) \ .$$

Considering $v\ge 6$, we use this inequality for $u=L_0 v\log(n+1)/(\epsilon\log N)\ge 6\delta N^\epsilon\operatorname{card} I_n\ge 6\delta\operatorname{card} I_n$ to obtain

$$\begin{aligned}\mathsf{P}\Big(\sum_{i\in I_n}\delta_i\ge\frac{L_0 v\log(n+1)}{\epsilon\log N}\Big)&\le\exp\Big(-\frac{L_0 v\log(n+1)}{2\epsilon\log N}\log(N^\epsilon)\Big)\\&=\exp\Big(-\frac{L_0 v\log(n+1)}{2}\Big) \ .\end{aligned} \tag{16.131}$$

Thus, if we define the event

$$\Omega(v)\ :\ \forall n\ge 1\ ,\ \sum_{i\in I_n}\delta_i\le\frac{L_0 v\log(n+1)}{\epsilon\log N}\ ,$$

we obtain from (16.131) that $\mathsf{P}(\Omega(v)^c)\le L\exp(-v/L)$. When $\Omega(v)$ occurs, for $n\ge 1$ we have

$$\frac{1}{\log(n+1)}\sum_{i\in I_n}\delta_i\le\frac{v}{\epsilon\log N} \ .$$

Then (16.130) and (16.107) imply $\|U_J\|^p\le K(p)vA/(\epsilon\log N)$. □

A drawback of Theorem 16.5.2 is that the quantity (16.109), namely $\sup_{t\in T}\sum_{i\le N}|t_i|^p$ (and hence B) might be too large. It might be sometimes to our advantage to change the norm (as little as we can) to decrease this quantity. For this, given a number $C>0$, we denote by $\|\cdot\|_C$ the norm on X such that the unit ball of the dual norm is

$$\Big\{x^* \in X^* \; ; \; \|x^*\| \le 1 \, , \, \sum_{i \le N} |x^*(x_i)|^p \le C\Big\} \, . \tag{16.132}$$

This definition of course ensures that for the norm $\|\cdot\|_C$ the quantity (16.109) is now $\le C$. Another very nice feature is that the set T_C of (16.106) corresponding to the new norm is *smaller* than the set T corresponding to the original norm, so that we will be able to prove that T_C is small in the sense of (16.110) simply because T is already small in this sense. This will be done by using the geometric properties of the original norm, and we shall *not* have to be concerned with the geometric properties of the norm $\|\cdot\|_C$.

Theorem 16.5.3. *Consider* $1 < q \le 2$ *and its conjugate exponent* $p \ge 2$. *Consider a Banach space* X *such that* X^* *is* p*-convex (see Definition 4.1.2). Consider vectors* $x_1, \ldots, x_N$ *of* X, *and* $S = \max_{i \le N} \|x_i\|$. *Denote by* U *the operator* $\ell_N^q \to X$ *such that* $U(e_i) = x_i$. *Then, for a number* $K(\eta, p)$ *depending only on* p *and on the constant* η *in Definition 4.1.2, if* $B = \max(K(\eta, p) S^p \log N, C)$, *and if for some* $\epsilon > 0$,

$$\delta \le \frac{S^p}{B \epsilon N^\epsilon} \le 1 \, , \tag{16.133}$$

we have

$$\mathsf{E}\|U_J\|_C^p \le K(\eta, p) \frac{S^p}{\epsilon} \, , \tag{16.134}$$

where J *is as in Theorem 16.5.2.*

In the situations of interest, S will be much smaller than $\|U\|_C$, so that (16.134) brings information. It is remarkable that the right-hand side of (16.134) does not depend on $\|U\|_C$ but only on $S = \max_{i \le n} \|U(e_i)\|$.

Lemma 16.5.4. *Consider the (quasi) distance* d_∞ *on* X_1^* *defined by*

$$d_\infty(x^*, y^*) = \max_{i \le N} |x^*(x_i) - y^*(x_i)| \, .$$

Then

$$e_k(X_1^*, d_\infty) \le K(p, \eta) S 2^{-k/p} (\log N)^{1/p} \tag{16.135}$$

or, equivalently, for $\epsilon > 0$,

$$\log N(X_1^*, d_\infty, \epsilon) \le K(p, \eta) \Big(\frac{S}{\epsilon}\Big)^p \log N \, . \tag{16.136}$$

Here X_1^* is the unit ball of X^*, $N(X_1^*, d_\infty, \epsilon)$ is the smallest number of balls for d_∞ of radius ϵ needed to cover X_1^* and e_k is defined in (2.34).

It would be nice to have a simple proof of this statement. The only proof we know is somewhat indirect. It involves geometric ideas. First, one proves a "duality" result, namely that if W denotes the convex hull of the points $(\pm x_i)_{i \le N}$, to prove (16.136) it suffices to show that

$$\log N(W, \|\cdot\|, \epsilon) \leq K(p,\eta)\Big(\frac{S}{\epsilon}\Big)^p \log N \ . \tag{16.137}$$

This duality result is proved in [3], Proposition 2, *(ii)*. We do not reproduce the simple and very nice argument, which is not related to the ideas of this work.

The proof of (16.137) involves more geometrical ideas. Briefly, since X^* is p-convex, it is classical that "X is of type p, with a type p constant depending only on p and η" as proved in [14], and then the conclusion follows from a beautiful probabilistic argument of Maurey, which is reproduced e.g. in [27], Lemma 3.2. □

Proof of Theorem 16.5.3. We combine (16.135) with Theorem 4.1.4 (used for $\alpha = p$) to obtain

$$\gamma_{p,p}(X_1^*, d_\infty) \leq K(p,\eta) S (\log N)^{1/p} \ ,$$

i.e. there exists an admissible sequence $(\mathcal{B}_n)$ on X_1^* for which

$$\forall t \in X_1^* \ , \ \sum_{n\geq 0} 2^n \Delta^p(B_n(t), d_\infty) \leq K(p,\eta) S^p \log N := A \ . \tag{16.138}$$

The set T_C corresponding to the norm (16.132) is

$$T_C = \Big\{ (x^*(x_i))_{i\leq N} \ ; \ \|x^*\| \leq 1 \ , \ \sum_{i\leq N} |x^*(x_i)|^p \leq C \Big\} \ .$$

It follows from (16.138) that this set satisfies (16.110), and since $\sum_{i\leq N} |t_i|^p \leq C$ for $t \in T_C$, it also satisfies (16.111) for $B = \max(A, C)$. We then conclude with Theorem 16.5.2. □

To conclude this section, we describe an example showing that Theorem 16.5.3 is very close to being optimal in certain situations. Consider two integers r, m and $N = rm$. We divide $\{1, \ldots, N\}$ into m disjoint subsets $I_1, \ldots, I_m$ of cardinality r. We consider $1 < q \leq 2$ and the operator $U : \ell_N^q \to \ell_m^q = X$ such that $U(e_i) = e_j$ for $i \in I_j$, where $(e_i)_{i\leq N}, (e_j)_{j\leq m}$ are the canonical bases of ℓ_N^q and ℓ_m^q respectively. Thus $S = 1$. It is classical [14] that $X^* = \ell_m^p$ is p-convex. Consider δ with $\delta^r = 1/m$. Then

$$\mathsf{P}(\exists j \leq m \ ; \ \forall i \in I_j \ , \ \delta_j = 1) = 1 - \Big(1 - \frac{1}{m}\Big)^m \geq \frac{1}{L} \ ,$$

and when this event occurs we have $\|U_J\| \geq r^{1/p}$, since $\|\sum_{i\in I_j} e_i\| = r^{1/q}$ and $\|U_J(\sum_{i\in I_j} e_i)\| = r\|e_j\| = r$. Thus

$$\mathsf{E}\|U_J\|^p \geq \frac{r}{L} \ . \tag{16.139}$$

On the other hand, let us try to apply Theorem 16.5.3 to this situation, so that $x_i = e_j$ for $i \in I_j$. Then we must take C large enough that $\|\cdot\|_C = \|\cdot\|$. Since

$$\sum_{i \le N} |x^*(x_i)|^p = r \sum_{j \le m} |x^*(e_j)|^p$$

can be as large as r for $\|x^*\| \le 1$, one has to take $C = r$. Then $B = r$ whenever $K(q) \log N \le r$. Let us choose $\epsilon = 1/(2r)$ so that for large m

$$\delta = \frac{1}{m^{1/r}} \le \frac{S^p}{B\epsilon N^\epsilon} = \frac{1}{r\epsilon N^\epsilon} = \frac{1}{r\epsilon m^\epsilon r^\epsilon} .$$

Thus (16.139) shows that (16.134) gives the exact order of $\|U_J\|$ in this case.

16.6 The $\Lambda(p)$-Problem

We denote by λ the uniform measure on $[0,1]$. Consider functions $(x_i)_{i\le N}$ on $[0,1]$ such that

$$\forall i \le N ,\ \|x_i\|_\infty \le 1 \tag{16.140}$$

$$\text{the sequence } (x_i)_{i\le N} \text{ is orthogonal in } L^2 = L^2(\lambda) . \tag{16.141}$$

Consider a number $p > 2$. J. Bourgain [2] proved the remarkable fact that there exists a subset J of $\{1,\dots,N\}$ with card $J = N^{2/p}$, for which we have an estimate

$$\forall (\alpha_i)_{i\in J} ,\ \Big\|\sum_{i\in J} \alpha_i x_i\Big\|_p \le K(p)\Big(\sum_{i\in J} \alpha_i^2\Big)^{1/2} , \tag{16.142}$$

where $\|\cdot\|_p$ denotes the norm in $L^p(\lambda)$. The most interesting case of application of this theorem is the case of the trigonometric system. Even in that case, no simpler proof is known. Bourgain's argument is probabilistic, showing in fact that a random choice of J works with positive probability.

We will give a sharpened version of (16.142), proving that instead of (16.142) one may even require

$$\forall (\alpha_i)_{i\in J} ,\ \Big\|\sum_{i\in J} \alpha_i x_i\Big\|_{p,1} \le K(p)\Big(\sum_{i\in J} \alpha_i^2\Big)^{1/2} , \tag{16.143}$$

where the norm in the left-hand side is in the space $L^{p,1}(\lambda)$ and is defined in (16.17). This is stronger than (16.142) by (16.20).

We consider r.v.s δ_i as in (16.104) with $\delta = N^{2/p-1}$, and $J = \{i \le N\ ;\ \delta_i = 1\}$.

Theorem 16.6.1. *Consider $p < p_1 < \infty$ and $p < p' < 2p$. Then there is a r.v. $W \ge 0$ with $\mathsf{E}W \le K$ such that for any numbers $(\alpha_i)_{i\in J}$ with $\sum_{i\in J} \alpha_i^2 \le 1$ we can write*

$$f := \sum_{i \in J} \alpha_i x_i = f_1 + f_2 + f_3 \tag{16.144}$$

where

$$\|f_1\|_{p_1} \le W \tag{16.145}$$

$$\|f_2\|_2 \le W\sqrt{\log N} N^{1/p-1/2} \; ; \; \|f_2\|_\infty \le W N^{1/p'} \tag{16.146}$$

$$\|f_3\|_2 \le W N^{1/p-1/2} \; ; \; \|f_3\|_\infty \le W N^{1/p} \, . \tag{16.147}$$

Here, as well as in the rest of this section, K denotes a number depending only on p, p' and p_1, that need not be the same on each occurrence. To understand Theorem 16.6.1, it helps to keep in mind that by (16.18), for any function h we have

$$\|h\|_{p,1} \le K(p)\|h\|_2^{2/p}\|h\|_\infty^{1-2/p} \, . \tag{16.148}$$

Thus (16.147) implies $\|f_3\|_{p,1} \le KW$ while (16.146) implies $\|f_2\|_{p,1} \le KWN^{-1/K}$. Since $\|f_1\|_{p,1} \le K\|f_1\|_{p_1}$ by (16.19), (16.144) implies $\|f\|_{p,1} \le KW$, so that we have the estimate

$$\forall (\alpha_i)_{i \in J} \; , \; \Big\|\sum_{i \in J} \alpha_i x_i\Big\|_{p,1} \le KW\Big(\sum_{i \in J} \alpha_i^2\Big)^{1/2} \, . \tag{16.149}$$

Moreover, since $\mathsf{P}(\operatorname{card} J \ge N^{2/p}) \ge 1/L$, with positive probability we have both $\operatorname{card} J \ge N^{2/p}$ and $W \le K$ and in this case we obtain (16.143).

Furthermore Theorem 16.6.1 lets us guess the exact reason why we cannot increase p in (16.142). This can be due only to the fact that we may have near equality for both parts of condition (16.144), i.e. the function f of (16.144) might take a value about $N^{1/p}$ on a set of measure about $1/N$. We believe (but cannot prove) that the lower order term f_2 is not needed in (16.144).

We consider the operator $U : \ell_N^2 \to L^p$ given by $U(e_i) = x_i$, and we denote by U_J its restriction to ℓ_J^2.

We choose once and for all $p_2 > p_1$. (This might be the time to mention that there is some room in the proof, and that some of the choices we make are simply convenient and in no way canonical.) We consider on L^{p_2} the norms $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ such that the unit ball of the dual norm is given respectively by

$$\Big\{x^* \in L^{q_2} \; ; \; \|x^*\|_{q_2} \le 1 \, , \; \sum_{i \le N} x^*(x_i)^2 \le N^{1/2-1/p}\Big\} \, . \tag{16.150}$$

$$\Big\{x^* \in L^{q_2} \; ; \; \|x^*\|_{q_2} \le 1 \, , \; \sum_{i \le N} x^*(x_i)^2 \le N^{1-2/p}\Big\} \, , \tag{16.151}$$

where q_2 is the conjugate exponent of p_2. These are sets of the type (16.132) in the case $X = L^{p_2}$.

Lemma 16.6.2. *We have*

$$\mathsf{E}\|U_J\|_{(1)} \le K \; ; \; \mathsf{E}\|U_J\|_{(2)} \le K\sqrt{\log N} \, .$$

Proof. We recall the classical fact that L^{q_2} is 2-convex [14], so that we may apply Theorem 16.5.3 to the case $X = L^{p_2}$, where the value of p in that theorem is taken equal to 2. We observe that it suffices to prove the result for N large enough. We choose $S = 1$, $C = N^{1/2-1/p}$ and $\epsilon = 1/2 - 1/p$. In this case $S^2/\epsilon \le K$, and for N large enough $B = \max(C, K(\eta)\log N) = C$, so that

$$\delta = N^{2/p-1} \le \frac{1}{B\epsilon N^{\epsilon}} \, , \tag{16.152}$$

and (16.134) proves that $\mathsf{E}\|U_J\|^2_{(1)} \le K$

Next, we choose $S = 1$ and $C = N^{1-2/p}$ and $\epsilon = 1/\log N$. In this case $S^2/\epsilon \le L\log N$, and, for N large enough, $B = \max(C, K(\eta,p)\log N) = C$, so that (16.152) holds again, and (16.134) proves now that $\mathsf{E}\|U_J\|^2_{(2)} \le K\log N$. □

We recall the norm of (9.59):

$$\|f\|_{\psi_2} = \inf\left\{c > 0 \; ; \; \int \exp\left(\frac{f^2}{c^2}\right) \mathrm{d}\lambda \le 2\right\} , \tag{9.59}$$

and we denote by $\|\cdot\|^*_{\psi_2}$ the dual norm.

We consider $a = N^{-1/p'}$, $b = N^{1/p'-1/p}$, and the norm $\|\cdot\|_{(3)}$ on L^p such that the unit ball of the dual norm is the set

$$Z = \left\{x^* \in L^q(\lambda) \; ; \; \|x^*\|_1 \le a \, , \; \|x^*\|^*_{\psi_2} \le b \; ; \; \sum_{i \le N} x^*(x_i)^2 \le N^{1-2/p}\right\} , \tag{16.153}$$

where q is the conjugate exponent of p.

Lemma 16.6.3. *We have* $\mathsf{E}\|U_J\|_{(3)} \le L$.

This uses arguments really different from those of Lemma 16.6.2, and the proof will be given at the end of this section.

Lemma 16.6.4. *Consider a bounded measurable function f. Assume that $\|f\|_{(1)} \le 1$, $\|f\|_{(2)} \le \sqrt{\log N}$, $\|f\|_{(3)} \le 1$. Then we may write $f = f_1 + f_2 + f_3$, where*

$$\|f_1\|_{p_1} \le K \tag{16.154}$$

$$\|f_2\|_2 \le K\sqrt{\log N}N^{1/p-1/2} \; ; \; \|f_2\|_\infty \le KN^{1/p'} \tag{16.155}$$

$$\|f_3\|_2 \le KN^{1/p-1/2} \; ; \; \|f_3\|_\infty \le KN^{1/p} \, . \tag{16.156}$$

Proof of Theorem 16.6.1. This is an obvious consequence of the previous three lemmas, with

$$W = \|U_J\|_{(1)} + \frac{1}{\sqrt{\log N}}\|U_J\|_{(2)} + \|U_J\|_{(3)} \ .$$

Indeed, for any numbers $(\alpha_i)_{i\in J}$ with $\sum_{i\in J}\alpha_j^2 \le 1$, the function $f = W^{-1}\sum_{i\in J}\alpha_i x_i$ satisfies the hypotheses of Lemma 16.6.4, since $\sum_{i\in J}\alpha_i x_i$ is of the type $U_J(y)$ for $y \in \ell_N^2$ and $\|y\|_2 \le 1$. □

Before we prove Lemma 16.6.4, we need to understand how we will use the information that $\|f\|_{(j)} \le 1$. This is through duality. Assume that we are given norms $\mathcal{N}_\ell$ for $\ell \le 3$ (say) on a finite dimensional space E, and denote by B_ℓ the corresponding unit ball. Denote by $\mathcal{N}_\ell^*$ the dual norms on $E^*$and by B_ℓ^* the corresponding unit ball. Consider the norm $\mathcal{N}^*$ on E^* whose unit ball is the set $\cap_{\ell\le 3}B_\ell^*$. Then the unit ball of dual norm $\mathcal{N}$ of $\mathcal{N}^*$ on E is the convex hull of the set $\cup_{\ell\le 3}B_\ell$. This is immediate from the Hahn-Banach theorem. In particular we can write any $x\in E$ as $x = \sum_{\ell\le 3}x_\ell$ with $\sum_{\ell\le 3}\mathcal{N}_\ell(x_\ell) \le \mathcal{N}(x)$, and in particular $\mathcal{N}(x_\ell) \le \mathcal{N}(x)$ for each ℓ. (If we are given the sets B_ℓ^* rather than the norm $\mathcal{N}_\ell$, we simply use the information $\mathcal{N}_\ell(x_\ell) \le A$ on the form $|x^*(x)| \le A$ for $x^* \in B_\ell^*$.) In the following proof we use this fact for E the span of the vectors $(x_i)_{i\le N}$ and $\mathcal{N}$ each of the norms $\|\cdot\|_{(j)}$, $j \le 3$.

Proof of Lemma 16.6.4. The intuitive idea is that f_1 corresponds to the small values of f, f_2 to the intermediate values and f_3 to the large values. This idea will be carried out by studying the three functions

$$f\mathbf{1}_{\{|f|\le c_1\}} \ ; \ f\mathbf{1}_{\{c_1\le|f|\le c_2\}} \ ; \ f\mathbf{1}_{\{|f|\ge c_2\}}$$

for appropriate values of c_1 and c_2. Namely we will prove that

$$\|f\mathbf{1}_{\{|f|\le c_1\}}\|_{p_1} \le K \ , \tag{16.157}$$

$$f\mathbf{1}_{\{c_1\le|f|\le c_2\}} = g_1 + g_2 \tag{16.158}$$

where $\|g_1\|_{p_1} \le K$, $\|g_2\|_\infty \le 6N^{1/p'}$ and $\|g_2\|_2 \le 2\sqrt{\log N}N^{1/p-1/2}$, and finally

$$f\mathbf{1}_{\{|f|\ge c_2\}} = g_3 + g_4 \ , \tag{16.159}$$

where $\|g_3\|_2 \le KN^{1/p-1/2}$, $\|g_3\|_\infty \le KN^{1/p}$, and $\|g_4\|_{p_1} \le K$. Combining (16.159), (16.157) and (16.158) finishes the proof.

Relation (16.157) will be a consequence of the hypothesis that $\|f\|_{(1)} \le 1$. Relation (16.158) will be a consequence of the hypothesis that $\|f\|_{(2)} \le \log N$, and relation (16.159) will be a consequence of the hypothesis that $\|f\|_{(3)} \le 1$. Unfortunately the details of the proof are messy and tedious.

Since $\|f\|_{(1)} \le 1$, by duality we may write $f = u_1 + u_2$ where $\|u_1\|_{p_2} \le 1$ and $u_2 = \sum_{i\le N}\beta_i x_i$ with $\sum_{i\le N}\beta_i^2 \le N^{1/p-1/2}$. By (16.140) and (16.141) we have $\|u_2\|_2^2 \le N^{1/p-1/2}$. The obvious inequality

$$\lambda(\{|f| \geq t\}) \leq \lambda\Big(\Big\{|u_1| \geq \frac{t}{2}\Big\}\Big) + \lambda\Big(\Big\{|u_2| \geq \frac{t}{2}\Big\}\Big) \tag{16.160}$$

implies

$$\lambda(\{|f| \geq t\}) \leq K(t^{-p_2} + t^{-2}N^{1/p-1/2}) \leq Kt^{-p_2}$$

for $t \leq c_1 = N^\alpha$, where $\alpha(p_2 - 2) = 1/2 - 1/p$. This proves (16.157).

Since $\|f\|_{(2)} \leq \sqrt{\log N}$, by duality we may write $f = v_1 + v_2$, where $\|v_1\|_{p_2} \leq \sqrt{\log N}$ and $v_2 = \sum_{i \leq N} \beta_i x_i$, with $\sum_{i \leq N} \beta_i^2 \leq (\log N) N^{2/p-1}$, so that $\|v_2\|_2 \leq \sqrt{\log N} N^{1/p-1/2}$. Let $c_2 = 3N^{1/p'}$. We next show that

$$\begin{aligned} |f|\mathbf{1}_{\{c_1 \leq |f| \leq c_2\}} &\leq 2|v_1|\mathbf{1}_{\{|v_1| \geq c_1/2\}} + 2|v_2|\mathbf{1}_{\{|v_2| \leq 2c_2\}} \\ &:= h_1 + h_2 \ . \end{aligned} \tag{16.161}$$

To see this, we may assume that $c_1 \leq c_2$, for there is nothing to prove otherwise. Assume first that $|v_2| > 2c_2$. Then if $|f| = |v_1 + v_2| \leq c_2$, we have $|v_1| > c_2$ so that since $|f| \leq c_2$, then $|f| \leq c_2 \leq |v_1|\mathbf{1}_{\{|v_1| \geq c_2\}}$ and hence (16.161) holds true since $c_2 \geq c_1$. Hence to prove (16.161) we may assume that $|v_2| \leq 2c_2$. Since $|f| \leq |v_1| + |v_2|$, we are done if $|v_1| \leq |v_2|$, since then $|f| \leq 2|v_2|$ and $|f| \leq 2|v_2|\mathbf{1}_{\{|v_2| \leq 2c_2\}}$. If, on the other hand, $|v_1| \geq |v_2|$, then $|f| \leq 2|v_1|$, so $|f|\mathbf{1}_{\{c_1 \leq |f|\}} \leq 2|v_1|\mathbf{1}_{\{|v_1| \geq c_1/2\}}$, finishing the proof of (16.161).

Since $\|v_1\|_{p_2} \leq \sqrt{\log N}$, we have

$$\lambda(\{|v_1| \geq t\}) \leq (\log N)^{p_2/2} t^{-p_2} \ .$$

Since $p_2 > p_1$, a straightforward computation based on the formula

$$\|h\|_{p_1}^{p_1} = \int p_1 t^{p_1 - 1} \lambda(\{|h| \geq t\}) \mathrm{d}t \tag{16.162}$$

yields $\|h_1\|_{p_1} \leq K$ (and there is lot of room since in fact $\|h_1\|_{p_1} \leq KN^{-\gamma}$ for some $\gamma > 0$). Next we observe that $\|h_2\|_\infty \leq 2c_2 \leq 6N^{1/p'}$ and $\|h_2\|_2 \leq 2\|v_2\|_2 \leq 2\sqrt{\log N} N^{1/p-1/2}$. Then (16.161) implies (16.158).

Since $\|f\|_{(3)} \leq 1$, by duality we may write $f = w_1 + w_2 + w_3$ with $\|w_1\|_\infty \leq a^{-1} = N^{1/p'}$, $\|w_2\|_{\psi_2} \leq b^{-1} = N^{1/p-1/p'}$ and $w_3 = \sum_{i \leq N} \beta_i x_i$ with $\sum_{i \leq N} \beta_i^2 \leq N^{2/p-1}$.

Thus

$$\|w_3\|_2 \leq N^{1/p-1/2} \tag{16.163}$$

and, using (16.140)

$$\|w_3\|_\infty \leq \sum_{i \leq N} |\beta_i| \leq N^{1/2} \Big(\sum_{i \leq N} \beta_i^2\Big)^{1/2} \leq N^{1/p} \ .$$

We note that

$$|f|\mathbf{1}_{\{|f| \geq c_2\}} \leq 3|w_3| + 2|w_2|\mathbf{1}_{\{|w_2| \geq c_2/3\}} \ . \tag{16.164}$$

To see this, we first observe that this is obvious if $|w_2| > c_2/3$, because then $|w_1| \le N^{1/p'} = c_2/3 \le |w_2|$, so $|f| \le |w_3| + 2|w_2|$. If now $|w_2| \le c_2/3$, since $|w_1| \le c_2/3$, when $|f| \ge c_2$, we must have $|w_3| \ge c_2/3$ and hence $|f| \le |w_1|+|w_2|+|w_3| \le 2c_2/3+|w_3| \le 3|w_3|$, finishing the proof of (16.164).

By definition of $\|\cdot\|_{\psi_2}$, and since $\|w_2\|_{\psi_2} \le b^{-1}$, we have

$$\int \exp(w_2^2 b^2) \mathrm{d}\lambda \le 2$$

so that

$$\lambda(\{|w_2| \ge t\}) \le 2\exp(-t^2 b^2)\,. \tag{16.165}$$

Since $p' < 2p$ we have $1/p - 1/p' < 1/p'$ and recalling the values of b and c_2 one checks from (16.165) and (16.162), with huge room to spare, that $h_3 = 2|w_2|\mathbf{1}_{\{|w_2| \ge c_2/3\}}$ satisfies $\|h_3\|_{p_1} \le K$. Thus from (16.164) we obtain (16.159). □

We turn to the proof of Lemma 16.6.3.

Lemma 16.6.5. *When $(g_i)_{i \le N}$ are independent standard Gaussian r.v.s, given a set I we have*

$$\mathsf{E}\Big\|\sum_{i\in I} g_i x_i\Big\|_{\psi_2} \le L\sqrt{\operatorname{card} I}\,.$$

Proof. We have

$$\mathsf{E}\int \exp\frac{(\sum_{i\in I} g_i x_i)^2}{3\operatorname{card} I}\mathrm{d}\lambda = \int \mathsf{E}\exp\frac{(\sum_{i\in I} g_i x_i)^2}{3\operatorname{card} I}\mathrm{d}\lambda \le L \tag{16.166}$$

because for each $t \in [0,1]$, $g = \sum_{i\in I} g_i x_i(t)$ is a Gaussian r.v. with $\mathsf{E}g^2 \le \operatorname{card} I$.

For $u \ge 1$, we have $e^{f^2/u} \le 1 + e^{f^2}/u$. We use this for $u = \int \exp f^2 d\lambda$ and integrate to get $\int \exp(f^2/u)\mathrm{d}\mu \le 2$, so that by definition $\|f\|_{\psi_2} \le \sqrt{u} \le u$, i.e.

$$\|f\|_{\psi_2} \le \int \exp f^2 \mathrm{d}\lambda\,.$$

Taking $f = (3\operatorname{card} I)^{-1/2}\sum_{i\in I} g_i x_i$ and combining with (16.166) yields the result. □

Proof of Lemma 16.6.3. The beginning of the proof uses arguments similar to the Giné-Zinn Theorem (Theorem 9.1.10). We recall the set Z of (16.153), and we set

$$T = \{(x^*(x_i)^2)_{i\le N}\,;\ x^* \in Z\}$$

so that $\sum_{i\le N} t_i \le N^{1-2/p}$ for $t \in T$, and hence $\delta\sum_{i\le N} t_i \le 1$. Thus

$$\begin{aligned}\mathsf{E}\|U_J\|_{(3)}^2 &= \mathsf{E}\sup_{t\in T}\sum_{i\le N}\delta_i t_i \\ &\le 1 + \mathsf{E}\sup_{t\in T}\sum_{i\le N}(\delta_i-\delta)t_i \;.\end{aligned}$$

Consider an independent sequence $(\delta'_i)_{i\le N}$ distributed like $(\delta_i)_{i\le N}$. Then, by Jensen's inequality we have

$$\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}(\delta_i-\delta)t_i\Big| \le \mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}(\delta_i-\delta'_i)t_i\Big| \;.$$

Consider independent Bernoulli r.v.s $(\varepsilon_i)_{i\le N}$, independent of the r.v. δ_i and δ'_i. Since the sequences $(\delta_i-\delta'_i)_{i\le N}$ and $(\varepsilon_i(\delta_i-\delta'_i))_{i\le N}$ have the same distribution,

$$\begin{aligned}\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}(\delta_i-\delta'_i)t_i\Big| &= \mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}\varepsilon_i(\delta_i-\delta'_i)t_i\Big| \\ &\le 2\mathsf{E}\sup_{t\in T}\Big|\sum_{i\le N}\varepsilon_i\delta_i t_i\Big| \\ &= 2\mathsf{E}\sup_{t\in T}\Big|\sum_{i\in J}\varepsilon_i t_i\Big| \\ &\le \sqrt{2\pi}\mathsf{E}\sup_{t\in T}\Big|\sum_{i\in J}g_i t_i\Big| \\ &\le L\mathsf{E}\sup_{t\in T}\sum_{i\in J}g_i t_i \;,\end{aligned}$$

using Proposition 5.1.1 and Lemma 2.2.1, and since $0\in T$. Since E card $J = N^{2/p}$ and $ab = N^{-1/p}$ it suffices to show that given a set I we have

$$\mathsf{E}\sup_{t\in T}\sum_{i\in I}g_i t_i = \mathsf{E}\sup_{x^*\in Z}\sum_{i\in I}g_i x^*(x_i)^2 \le Lab\sqrt{\operatorname{card} I}\;. \tag{16.167}$$

We have $|x^*(x_i)|\le a$ since $\|x^*\|_1\le a$ and $\|x_i\|_\infty\le 1$, so that for a fixed set I we have

$$\begin{aligned}d_1(x^*,y^*) &:= \Big(\sum_{i\in I}\big(x^*(x_i)^2-y^*(x_i)^2\big)^2\Big)^{1/2} \\ &\le 2a\Big(\sum_{i\in I}\big(x^*(x_i)-y^*(x_i)\big)^2\Big)^{1/2} := 2ad_2(x^*,y^*)\;.\end{aligned}$$

Thus

$$\gamma_2(Z,d_1)\le La\gamma_2(Z,d_2)$$

and, by Theorem 2.2.18 and Theorem 2.4.1, we have

$$\mathsf{E} \sup_{x^* \in Z} \sum_{i \in I} g_i x^*(x_i)^2 \le La \mathsf{E} \sup_{x^* \in Z} x^* \Big(\sum_{i \in I} g_i x_i \Big) . \tag{16.168}$$

Now, by definition of Z, for $x^* \in Z$ we have $\|x^*\|^*_{\psi_2} \le b$, where $\|\cdot\|^*_{\psi_2}$ is the dual norm of the $\|\cdot\|_{\psi_2}$ norm, and thus

$$\sup_{x^* \in Z} x^* \Big(\sum_{i \in I} g_i x_i \Big) \le b \Big\| \sum_{i \in I} g_i x_i \Big\|_{\psi_2} .$$

Combining with Lemma 16.6.5 this proves (16.167) and hence Lemma 16.6.3.
□

Remark. One can also deduce (16.168) from the classical comparison theorems for Gaussian r.v.s, see [12].

16.7 Proportional Subsets of Bounded Orthogonal Systems

As in the previous section, we consider a sequence $(x_i)_{i \le N}$ of functions on a probability space (e.g. $[0,1]$ with Lebesgue's measure) that satisfies (16.140) and (16.141), i.e. $(x_i)_{i \le N}$ is an orthogonal sequence with $\|x_i\|_\infty \le 1$. We are now interested in finding a large subset J of $\{1, \dots, N\}$ such that on the span of $(x_i)_{i \in J}$ we know how to bound the L^2 norm by a suitable multiple of the L^1 norm.

Theorem 16.7.1 ([8]). *Assume that, for some number $\tau > 0$*

$$\forall\, i \le N \;,\; \|x_i\|_2 \ge \tau \; . \tag{16.169}$$

Then, given any integer $1 < k < N$ there exists a set $J \subset \{1, \dots, N\}$ with $\operatorname{card} J \ge N - k$ *such that for every $(\alpha_i)_{i \in J}$,*

$$\Big\| \sum_{i \in J} \alpha_i x_i \Big\|_2 \le L\kappa (\log \kappa)^{5/2} \Big\| \sum_{i \in J} \alpha_i x_i \Big\|_1 \;, \tag{16.170}$$

where $\kappa = \tau^{-1} \sqrt{N/k} \sqrt{\log k}$.

To gain a first understanding of this inequality, let us observe that since $\|x_i\|_\infty \le 1$, given numbers $(\alpha_i)_{i \le N}$ the function $f = \sum_{i \le N} \alpha_i x_i$ satisfies $\|f\|_\infty \le \sqrt{N} (\sum_{i \le N} \alpha_i^2)^{1/2}$. Moreover from (16.169) it satisfies $\|f\|_2 \ge \tau (\sum_{i \le N} \alpha_i^2)^{1/2}$. The easy inequality $\|f\|_2^2 \le \|f\|_\infty \|f\|_1$ then proves that

$$\Big\| \sum_{i \le N} \alpha_i x_i \Big\|_2 \le \tau^{-1} \sqrt{N} \Big\| \sum_{i \le N} \alpha_i x_i \Big\|_1 \;, \tag{16.171}$$

which is (better than) the "case $k = 0$" of (16.170). Let us now consider an example proving that (16.171) is essentially optimal. Assume that the underlying measure space is the group $\{-1,1\}^n$ provided with the uniform measure. For $i \le n$ let us denote by ε_i the n-th coordinate function, and for a subset I of $1,\ldots,n$ let $\varepsilon_I = \prod_{i\in I}\varepsilon_i$. The $N = 2^n$ functions ε_I form an orthonormal system. Moreover $\sum_I \varepsilon_I = \prod_{i\le n}(1+\varepsilon_i)$ is zero everywhere except on the set of measure $2^{-n} = 1/N$ where it takes the value N. Thus we cannot have a better constant than $\sqrt{N}$ in (16.171). It is quite remarkable that even a rather small value of k allows a considerable improvement on this estimate. (Indeed (16.170) starts to improves upon (16.171) when k is about $(\log N)^5 \log\log N$. One might of course wonder what happens for smaller values of k.)

A particular case of (16.170) is when $\operatorname{card} J$ is a proportion of N, i.e. when $\operatorname{card} J = cN$. In that case, the coefficient in the right-hand side is basically (with a coefficient depending on c but not on N) $\sqrt{\log N}(\log\log N)^{5/2}$. When c is small enough, it is proved in [28] that the better estimate $\sqrt{\log N}\sqrt{\log\log N}$ holds true. The method of [28] does not seem to extend to the case where c is not small. It is known (and explained in [8]) that the factor $\sqrt{\log N}$ is necessary. The term in $\sqrt{\log\log N}$ is probably parasitic, and of course it is debatable how interesting it is to reduce such terms to the smallest possible power when one does not know how to eliminate them altogether.

The basis for the approach to Theorem 16.7.1 is to try first to control the norm L^p rather than the norm L^1.

Theorem 16.7.2. *Under the hypothesis of Theorem 16.7.1, for every* $1 < k < N$ *there exists a subset* J *of* $\{1,\ldots,N\}$ *with* $\operatorname{card} J = k$ *such that for every* $(\alpha_i)_{i\in J}$ *one has*

$$\Big\|\sum_{i\in J}\alpha_i x_i\Big\|_2 \le \frac{L\sqrt{N/k}\sqrt{\log k}}{\tau(p-1)^{5/2}}\Big\|\sum_{i\in J}\alpha_i x_i\Big\|_p \,. \tag{16.172}$$

The idea to deduce Theorem 16.7.1 is simply to take p close enough to 1 that on the span of the vectors $(x_i)_{i\in J}$ the L^1 and L^p norm are within a factor 2 of each other. The point of using the L^p norm in Theorem 16.7.2 is that we will be able to use its convexity properties. The use of the convexity property of the L^p norm in the present approach explains why the estimate of [28] is better than the present estimate in its range (i.e. $\operatorname{card} J$ a small proportion of N): rather than using an auxiliary L^p norm, the method of [28] is to exhibit some "directions of smoothness" in certain spaces where the norm itself is not smooth, and extra information is gained this way.

To prove Theorem 16.7.2 it turns out that, rather than assuming (16.140) and (16.141), it is more convenient instead to assume

$$\|x_i\|_2 = 1 \;;\; \|x_i\|_\infty \le \frac{1}{\tau} \,, \tag{16.173}$$

which changes nothing, replacing x_i by $x_i/\|x_i\|_2$.

Let us first explain the basic geometrical principle on which the approach to Theorem 16.7.1 is based. We denote

$$E \text{ the span of } x_1, \ldots, x_N \,. \tag{16.174}$$

Given a parameter $\rho > 0$ we consider the following norm on E:

$$\|y\|_\rho^\sim := (\|y\|_p^2 + \rho^{-2}\|y\|_2^2)^{1/2} \,. \tag{16.175}$$

We denote by $B(\rho)$ the ball of E for $\|\cdot\|_\rho^\sim$ centered at 0 and of radius 1. We denote by $\langle x, y\rangle$ the dot product in L^2 of two vectors in E, and we observe that since the sequence (x_i) is orthonormal from (16.173),

$$\forall\, y \in E \;;\; \sum_{i\le N} \langle x_i, y\rangle^2 = \|y\|_2^2 \,. \tag{16.176}$$

Our first lemma provides a criteria under which we can compare the L^2 and the L^p norms as in (16.170).

Lemma 16.7.3. *Consider $k \le N$ and $i_1, \ldots, i_k \in \{1, \ldots, N\}$. (These indexes need not be distinct.) Assume that*

$$\sup_{y\in B(\rho)} \left|\sum_{j=1}^k \left(\langle x_{i_j}, y\rangle^2 - \frac{\|y\|_2^2}{N}\right)\right| \le \frac{k\rho^2}{2N} \,. \tag{16.177}$$

Let

$$J = \{1, \ldots, N\} \setminus \{i_1, \ldots, i_k\} \,. \tag{16.178}$$

Then

$$\forall\, (\alpha_i)_{i\in J} \;;\; \Big\|\sum_{i\in J} \alpha_i x_i\Big\|_2 \le \rho \Big\|\sum_{i\in J} \alpha_i x_i\Big\|_p \,. \tag{16.179}$$

Proof. Consider $y = \sum_{i\in J} \alpha_i x_i$. Then $\langle x_{i_j}, y\rangle = 0$ for $j \le k$, so that

$$\left|\sum_{j=1}^k \left(\langle x_{i_j}, y\rangle^2 - \frac{\|y\|_2^2}{N}\right)\right| = \frac{k\|y\|_2^2}{N} \,.$$

Consequently (16.177) implies that for such a y

$$y \in B(\rho) \Rightarrow \|y\|_2^2 \le \frac{\rho^2}{2} \,,$$

and, by homogeneity,

$$\|y\|_2^2 \le \frac{\rho^2}{2}\|y\|_\rho^{\sim 2} = \frac{\rho^2}{2}\|y\|_p^2 + \frac{1}{2}\|y\|_2^2 \,. \qquad \square$$

We shall prove that for a suitable choice of ρ the relation (16.177) holds with probability $\geq 1/2$ when the indexes $i_1, \ldots, i_k$ are chosen independently at random, uniformly over $\{1, \ldots, N\}$. The main estimate, Proposition 16.7.4 below, will involve a 2-convex Banach space E, that is with the property that

$$\|x\|, \|y\| \leq 1 \Rightarrow \left\|\frac{x+y}{2}\right\| \leq 1 - \eta\|x-y\|^2 , \tag{16.180}$$

for some number $\eta > 0$. We denote by $T_2(E^*)$ the type-2 constant of the dual E^* of E, that is the smallest constant for which the inequality

$$\mathsf{E}\Big\|\sum_{i\leq m} g_i y_i\Big\| \leq T_2(E^*)\Big(\sum_{i\leq m}\|y_i\|^2\Big)^{1/2} \tag{16.181}$$

for any vectors $y_1, \ldots, y_m$ of E^*, where of course $g_1, \ldots, g_m$ are independent standard Gaussian r.v.s. (For convenience we use the same notation for the norms of E and E^* since this does not create ambiguities.) We denote as usual by ε_i independent Bernoulli r.v.s.

Proposition 16.7.4. *Consider a Banach space E such that (16.180) and (16.181) hold. Consider $k \geq 2$ and vectors $z_1, \ldots, z_k$ in E^* with $\|z_j\| \leq 1$. Then*

$$\mathsf{E}\sup_{y\in E, \|y\|\leq 1}\Big|\sum_{j\leq k}\varepsilon_j z_j(y)^2\Big| \leq L\eta^{-2}T_2(E^*)\sqrt{\log k}\sup_{\|y\|\leq 1}\Big(\sum_{j\leq k} z_j(y)^2\Big)^{1/2} . \tag{16.182}$$

It is important in (16.182) to obtain an explicit and sharp dependence in η and $T_2(E^*)$. We shall use (16.182) when the space E is endowed with the norm $\|\cdot\|_\rho^\sim$ for an appropriate value of ρ. (This norm is 2-convex because L^p is 2-convex so that the norm $\|\cdot\|_\rho^\sim$ is a "Hilbert sum" of 2-convex norms, see [14]). In that case η and $T_2(E^*)$ depend on p, and the sharp dependence of (16.182) in these quantities yields a sharp dependence of the estimate (16.182) in p. This is essential to obtain the factor $(p-1)^{-5/2}$ in Theorem 16.7.2, and in turn to obtain Theorem 16.7.1 through the correct choice of p.

Until the end of the proof of Proposition 16.7.4 we fix the vectors $(z_j)_{j\leq k}$. Without loss of generality we may assume that

$$\max_{j\leq k}\|z_j\| = 1 ,$$

so that

$$S := \sup_{\|y\|\leq 1}\Big(\sum_{j\leq k} z_j(y)^2\Big)^{1/2} \geq 1 . \tag{16.183}$$

We denote by T the unit ball of E and we consider on T the distance given by

$$\widehat{d}(y, y')^2 = \sum_{j\leq k}(z_j(y)^2 - z_j(y')^2)^2 . \tag{16.184}$$

We shall prove that

$$\gamma_2(T, \widehat{d}\,) \le L\eta^{-2} T_2(E^*) S \sqrt{\log k} \; , \tag{16.185}$$

from which (16.182) follows, using the subgaussian inequality (3.2.2) and the generic chaining bound of Theorem 2.2.18.

The proof of (16.185) is a small variation of the proof of the Ellipsoid Theorem, Theorem 4.1.10. A technical difficulty is however that the balls for $\widehat{d}$ are not convex sets. To bypass this problem, we shall replace $\widehat{d}$ by a larger, but better behaved distance. The definition of this new distance is the purpose of the next lemma.

Lemma 16.7.5. *Consider on T the quantity $\overline{d}(y,y')$ given by*

$$\overline{d}(y,y')^2 = \sum_{j\le k} (z_j(y) - z_j(y'))^2 (z_j(y)^2 + z_j(y')^2) \; , \tag{16.186}$$

and the distance d on T given by

$$d(y,y') = \inf \sum_{1\le r\le n} \overline{d}(y_r, y_{r+1}) \; ,$$

where the infimum is taken over all n and sequences $y = y_1, \ldots, y_{n+1} = y'$ in T. Then

$$\widehat{d}(y,y') \le 2d(y,y') \le 2\overline{d}(y,y') \; , \tag{16.187}$$

and

$$\overline{d}(y,y') \le 2d(y,y') \; . \tag{16.188}$$

It is elegant to use the distance d, but the proof of (16.188) is somewhat tricky. It is possible to dispense with the use of d and to work directly with $\overline{d}$, which can be shown to satisfy the following substitute of the triangle inequality: $\overline{d}(x,y) \le 4(\overline{d}(x,z) + \overline{d}(x,y))$ (and moreover the balls for $\overline{d}$ are sufficiently convex in the sense of (16.196) below), see [21] for the details and Section 16.10 below for more comments. The drawback of this approach is that one has to check that Theorem 2.3.14 is still true in this setting, which however offers no difficulty.

Proof. The proof will use many times the relation

$$\sqrt{a^2+b^2} \le |a| + |b| \le \sqrt{2}\sqrt{a^2+b^2} \; . \tag{16.189}$$

Consider the function $f(a,b) = |a-b|\sqrt{a^2+b^2}$, so that

$$\overline{d}(y,y')^2 = \sum_{j\le k} f(z_j(y), z_j(y'))^2 \; . \tag{16.190}$$

First we observe that

$$|a^2 - b^2| \le |a-b|(|a|+|b|) \le \sqrt{2}|a-b|\sqrt{a^2+b^2} = \sqrt{2} f(a,b) \; , \qquad (16.191)$$

and using (16.190) this proves that $\widehat{d}(y,y') \le \sqrt{2}\,\overline{d}(y,y')$ and, since $\widehat{d}$ is a distance, that $\widehat{d}(y,y') \le \sqrt{2} d(y,y')$. Also, it is obvious by definition that $d(y,y') \le \overline{d}(y,y')$ so only (16.188) remains to be proved. Given numbers $a_1, \ldots, a_{n+1}$ we prove that

$$f(a_1, a_{n+1}) \le 2 \sum_{1 \le r \le n} f(a_r, a_{r+1}) \; . \qquad (16.192)$$

First, if $ab \ge 0$, then

$$f(a,b) = ||a|-|b||\sqrt{a^2+b^2} \le ||a|-|b|| \cdot ||a|+|b|| = |a^2-b^2| \; , \qquad (16.193)$$

while if $ab < 0$ then, using the right-hand side of (16.189) in the left-hand side inequality,

$$f(a,b) = (|a|+|b|)\sqrt{a^2+b^2} \le \sqrt{2}(a^2+b^2) \le \sqrt{2} f(a,b) \; . \qquad (16.194)$$

Thus when $a_1 a_{n+1} \ge 0$ then (16.192) follows from (16.193), (16.191) and the triangle inequality $|a_1^2 - a_{n+1}^2| \le \sum_{1 \le r \le n} |a_r^2 - a_{r+1}^2|$ in $\mathbb{R}$. When $a_1 a_{n+1} < 0$ there must exist $1 \le m \le n$ with $a_m a_{m+1} < 0$, and then

$$a_1^2 + a_{n+1}^2 \le a_m^2 + a_{m+1}^2 + \sum_{r \ne m} |a_r^2 - a_{r+1}^2| \; ,$$

so that (16.192) follows from (16.191) and (16.194).

Using (16.192) and (16.190) proves that for $y_1, \ldots, y_{n+1} \in T$,

$$\overline{d}(y_1, y_{n+1}) \le 2 \sum_{1 \le r \le n} \overline{d}(y_r, y_{r+1}) \; ,$$

and this implies (16.188). □

Rather than (16.185) we shall prove that

$$\gamma_2(T,d) \le L\eta^{-2} T_2(E^*) S \sqrt{\log k} \; , \qquad (16.195)$$

which suffices since $\widehat{d} \le 2d$. The great advantage of d over $\widehat{d}$ is that the balls for d are almost convex, as the next lemma shows.

Lemma 16.7.6. *For every $y \in T$ and $\epsilon > 0$ we have*

$$\operatorname{conv} B_d(y,\epsilon) \subset B_d(y, 2\epsilon) \; . \qquad (16.196)$$

Proof. In view of (16.187) and (16.188) we have

$$B_d(y,\epsilon) \subset \{y \in T \; ; \; \overline{d}(y,y') \le 2\epsilon\} \subset B_d(y, 2\epsilon) \; ,$$

so that it suffices to prove that given $y' \in T$ the set

$$\{y \in T \; ; \; \overline{d}(y,y')^2 \le 4\epsilon^2\}$$

is convex. For this it suffices to prove that for any a the function $f(x) = (x-a)^2(x^2+a^2)$ is convex. Now,

$$f(x) = (x^2 - 2ax + a^2)(x^2+a^2) = x^4 - 2ax^3 + 2a^2x^2 - 2a^3x + a^4 \; ,$$

so that

$$f(x)'' = 12x^2 - 12ax + 4a^2 = 12(x-a/2)^2 + a^2 \ge 0 \; . \qquad \square$$

On E we consider the norm

$$\|y\|_\infty = \max_{j \le k} |z_j(y)| \; . \tag{16.197}$$

Given an element w of E we also consider on E the weighted ℓ^2 norm $\|\cdot\|_w$ given by

$$\|y\|_w^2 = \sum_{j \le k} z_j(y)^2 z_j(w)^2 \; . \tag{16.198}$$

We recall the quantity S of (16.183).

Lemma 16.7.7. *For y, y' in T we have*

$$d(y,y')^2 \le 4\|y-y'\|_w^2 + 2S^2\|y-y'\|_\infty^2(\|y-w\|^2 + \|y'-w\|^2) \; . \tag{16.199}$$

Proof. We use the inequalities $z_j(y)^2 \le 2z_j(y-w)^2 + 2z_j(w)^2$ and

$$\begin{aligned}\sum_{j\le k}(z_j(y) - z_j(y'))^2 z_j(y-w)^2 &\le \|y-y'\|_\infty^2 \sum_{j \le k} z_j(y-w)^2 \\ &\le S^2\|y-y'\|_\infty^2 \|y-w\|^2\end{aligned}$$

to obtain

$$\overline{d}(y,y')^2 \le 4\|y-y'\|_w^2 + 2S^2\|y-y'\|_\infty^2(\|y-w\|^2 + \|y'-w\|^2) \; ,$$

and the conclusion since $d \le \overline{d}$ by (16.187). $\square$

We recall the entropy numbers defined in (2.34). We denote by d_w and d_∞ the distances on T associated to the norms $\|\cdot\|_w$ and $\|\cdot\|_\infty$ respectively.

Lemma 16.7.8. *For $w \in T$ we have*

$$e_n(T, d_w) \le L2^{-n/2} T_2(E^*) S \; . \tag{16.200}$$

Proof. Consider independent standard r.v.s $(g_j)_{j\le k}$. Then d_w is the canonical distance associated to the Gaussian process on E^* defined by $X_y = \sum_{j\le k} g_j z_j(w) z_j(y)$. Using successively the definition of $T_2(E^*)$, the fact that $\|z_j\| \le 1$ and the definition (16.183) of S we obtain

$$\begin{aligned}\mathsf{E}\sup_{\|y\|\le 1} X_y &= \mathsf{E}\Big\|\sum_{j\le k} g_j z_j(w) z_j\Big\| \\ &\le T_2(E^*)\Big(\sum_{j\le k} z_j(w)^2 \|z_j\|^2\Big)^{1/2} \\ &\le T_2(E^*)\Big(\sum_{j\le k} z_j(w)^2\Big)^{1/2} \\ &\le T_2(E^*) S \ . \end{aligned} \tag{16.201}$$

The conclusion then follows from Sudakov's minoration (2.82). (Alternatively, one may combine (2.44) with Theorem 2.4.1.) □

Lemma 16.7.9. *We have*

$$e_n(T, d_\infty) \le L 2^{-n/2} \eta^{-1} T_2(E^*) \sqrt{\log k} \ . \tag{16.202}$$

This is basically Lemma 16.5.4, the difference being that the dependence in the modulus of convexity has been carried out explicitly (and also the dependence on $T_2(E^*)$ but this part is immediate). Figuring this out requires to go into the details of Proposition 2 of [3]. These details are provided in [8] on page 1079.

Exercise 16.7.10. (a) Using Lemma 16.7.5 prove that $\widehat{d} \le LSd_\infty$. Using (16.202) prove that

$$e_n(T, \widehat{d}) \le L 2^{-n/2} \eta^{-1} T_2(E^*) S \sqrt{\log k} \ .$$

(b) What is the parasitic factor you expect when using Dudley's integral to bound $\gamma_2(T, \widehat{d})$ using the previous inequality? How do you expect to reduce this factor using convexity as in Theorem 4.1.4? (Hint: one expects an extra factor $\log k$, which one should reduce to $\sqrt{\log k}$ using convexity. The remarkable feature of (16.185) is that this extra factor $\sqrt{\log k}$ itself has disappeared. The remaining factor $\sqrt{\log k}$ already occurs in the entropy estimate.)

Proof of (16.195). We set $r = 8$. We consider the largest integer n_0 such that $2^{2^{n_0}} \le (1+2r)^k$. We consider the functionals

$$F_n(A) = 1 - \inf\{\|y\| \ ; \ y \in \operatorname{conv} A\} + \eta \frac{\max(n_0 + 1 - n, 0)}{\log k} \ . \tag{16.203}$$

We shall prove that they satisfy the growth condition of Definition 2.3.10 with

$$c^* = \frac{\eta^2}{LST_2(E^*)\sqrt{\log k}} \,. \tag{16.204}$$

Since $n_0 \le L \log k$ and $\eta \le 1$ we have $F_0(T) \le L$ and the conclusion then follows from Theorem 2.3.16.

To prove the growth condition, it follows from Exercise 2.3.13 that it suffices to consider the case where $n \le n_0$. We copy the proof of Theorem 4.1.10. We consider $0 \le n \le n_0, m = N_{n+1}$, and points $(t_\ell)_{\ell \le m}$ in T, such that $d(t_\ell, t_{\ell'}) \ge a$ whenever $\ell \ne \ell'$. Consider also sets $H_\ell \subset T \cap B_d(t_\ell, a/r)$, where the index d emphasizes that the ball is for the distance d rather than for the norm $\|\cdot\|$ of E. Set

$$u = \inf\Bigl\{\|v\| \; ; \; v \in \operatorname{conv} \bigcup_{\ell \le m} H_\ell\Bigr\} \,, \tag{16.205}$$

and consider u' such that

$$u' > \max_{\ell \le m} \inf\{\|v\| \; ; \; v \in \operatorname{conv} H_\ell\} \,. \tag{16.206}$$

Let us define $u'' := \min(u', 1)$. For $\ell \le m$ consider $v_\ell \in \operatorname{conv} H_\ell$ with $\|v_\ell\| \le u''$. It follows from (16.180) that for $\ell, \ell' \le m$,

$$\Bigl\|\frac{v_\ell + v_{\ell'}}{2u''}\Bigr\| \le 1 - \eta \Bigl\|\frac{v_\ell - v_{\ell'}}{u''}\Bigr\|^2 \,. \tag{16.207}$$

Moreover, since $(v_\ell + v_{\ell'})/2 \in \operatorname{conv} \bigcup_{\ell \le m} H_\ell$, we have $u \le \|v_\ell + v_{\ell'}\|/2$, and (16.207) implies

$$\frac{u}{u''} \le 1 - \eta \Bigl\|\frac{v_\ell - v_{\ell'}}{u''}\Bigr\|^2 \,,$$

so that, using that $u'' \le 1$ in the second inequality,

$$\|v_\ell - v_{\ell'}\| \le u'' \Bigl(\frac{u'' - u}{\eta u''}\Bigr)^{1/2} \le R := \Bigl(\frac{u'' - u}{\eta}\Bigr)^{1/2} \,,$$

and hence the points v_ℓ belong to $T' := RT + v_1$.

To lighten notation let

$$A = T_2(E^*)S \; ; \; B = \eta^{-1} T_2(E^*)\sqrt{\log k} \,.$$

Since the distances d_∞ and d_{v_1} both arise from a norm, we deduce from Lemmas 16.7.8 and 16.7.9 that we can cover T' with N_n sets that are of diameter $\le L2^{-n/2}AR$ for d_{v_1} and of diameter $\le L2^{-n/2}BR$ for d_∞.

Now, since $H_\ell \subset B_d(t_\ell, a/r)$ it follows from (16.196) that $v_\ell \in B_d(t_\ell, 2a/r)$. Since $r = 8$, we have $2a/r \le a/4$ so that since $d(t_\ell, t'_\ell) \ge a$ for $\ell \ne \ell'$ we have $d(v_\ell, v_{\ell'}) \ge a/2$. There are N_{n+1} such points v_ℓ, so two of them must fall inside one of the N_n sets we constructed in the previous paragraph. Using (16.199) for y, y' equal to these two points, and $w = v_1$ we obtain the relation

$$\frac{a^2}{4} \le L(R^2 2^{-n} A^2 + S^2 2^{-n} B^2 R^4) \le L \max(R^2 2^{-n} A^2, S^2 2^{-n} B^2 R^4) ,$$

so that

$$R^2 \ge \frac{1}{L} \min\Big(\frac{a^2 2^n}{A^2}, \frac{a 2^{n/2}}{SB}\Big) .$$

Since

$$\frac{a^2 2^n}{A^2} + \frac{1}{\log k} \ge \frac{a 2^{n/2}}{A\sqrt{\log k}}$$

we obtain

$$R^2 + \frac{1}{\log k} \ge \frac{1}{L} \min\Big(\frac{a 2^{n/2}}{A\sqrt{\log k}}, \frac{a 2^{n/2}}{SB}\Big) \ge \frac{a 2^{n/2}}{LSB} , \tag{16.208}$$

because $A\sqrt{\log k} \le SB$. Since $u' - u \ge u'' - u = \eta R^2$ it follows that

$$u' + \frac{\eta}{\log k} \ge u + \eta \frac{a 2^{n/2}}{LSB} ,$$

and hence

$$1 - u + \frac{\eta}{\log k} \ge 1 - u' + \eta \frac{a 2^{n/2}}{LSB} .$$

Since u' is arbitrary in (16.206) we deduce using also (16.205) that

$$1 - \inf\Big\{\|v\| \; ; \; v \in \operatorname{conv} \bigcup_{\ell \le m} H_\ell\Big\} + \frac{\eta}{\log k}$$

$$\ge 1 - \max_{\ell \le m} \inf\{\|v\| \; ; \; v \in \operatorname{conv} H_\ell\} + \eta \frac{a 2^{n/2}}{LSB}$$

i.e.

$$F_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge \min_{\ell \le m} F_{n+1}(H_\ell) + \eta \frac{a 2^{n/2}}{LSB} .$$

This completes the proof of the growth condition (2.147) for the value c^* of (16.204). We have proved (16.185) and hence Proposition 16.7.4. □

We now complete the proof of Theorem 16.7.2 along well understood lines.

Proposition 16.7.11. *Consider a Banach space E such that (16.180) and (16.181) hold. Consider a random vector Z valued in E^*, with $\|Z\| \le 1$, and i.i.d. copies $(Z_i)_{i \le k}$ of Z. Let*

$$D = \eta^{-2} T_2(E^*) \sqrt{\log k} , \tag{16.209}$$

and

$$\sigma^2 = \sup_{y \in E, \|y\| \le 1} \mathsf{E} Z(y)^2 . \tag{16.210}$$

Then

$$\mathsf{E} \sup_{y \in E, \|y\| \le 1} \Big| \sum_{j \le k} (Z_j(y)^2 - \mathsf{E} Z(y)^2) \Big| \le L(D^2 + D\sigma\sqrt{k}) . \tag{16.211}$$

Proof. We use (9.21) (of course with $R_i = 1$ if $i \le k$ and $R_i = 0$ if $i > k$) and we take expectation in the r.v.s ε_i using (16.182) to obtain

$$\begin{aligned} U := \mathsf{E} \sup_{y \in E, \|y\| \le 1} \Big| \sum_{j \le k} (Z_j(y)^2 - \mathsf{E} Z(y)^2) \Big| \\ \le 2\mathsf{E} \sup_{\|y\| \le 1} \Big| \sum_{j \le k} \varepsilon_j Z_j(y)^2 \Big| \\ \le LD\mathsf{E}\Big(\sup_{\|y\| \le 1} \sum_{j \le k} Z_j(y)^2 \Big)^{1/2} . \end{aligned} \tag{16.212}$$

Now

$$\sup_{\|y\| \le 1} \sum_{j \le k} Z_j(y)^2 \le k\sigma^2 + \sup_{\|y\| \le 1} \Big| \sum_{j \le k} (Z_j(y)^2 - \mathsf{E} Z(y)^2) \Big| ,$$

so that using the Cauchy-Schwarz inequality to put the expectation inside the square root in the right-hand side of (16.212) yields

$$U \le LD(U + k\sigma^2)^{1/2} \le LD\sqrt{U} + LD\sigma\sqrt{k} \le \frac{U}{2} + \frac{(LD)^2}{2} + LD\sigma\sqrt{k} . \quad \square$$

Proof of Theorem 16.7.2. We use Proposition 16.7.11 for the space E of (16.174) endowed with the norm (16.175), where the parameter ρ will be chosen later.

It is proved in [4] that for two functions f, g in L^p,

$$\Big\| \frac{f+g}{2} \Big\|_p^2 + \frac{p(p-1)}{8} \Big\| \frac{f-g}{2} \Big\|_p^2 \le \frac{1}{2} (\|f\|_p^2 + \|g\|_p^2) ,$$

and this implies that the norm (16.175) satisfies (16.180) with $\eta \ge (p-1)/8$. Also, $T_2(E^*) \le L\sqrt{q} = L\sqrt{p/(p-1)}$, where q is the conjugate exponent of p, because this bound is already true for L^p. Consequently the quantity D of (16.209) satisfies

$$D \le L(p-1)^{-5/2} \sqrt{\log k} . \tag{16.213}$$

We apply Proposition 16.7.11 to the random vector Z such that for $i \le N$, $\mathsf{P}(Z = x_i) = 1/N$. We think of Z as being valued in E^*. Then (16.176) implies $\mathsf{E} Z(y)^2 = \|y\|_2^2 / N$, and the quantity σ of (16.210) satisfies

$$\sigma^2 \le \sup_{\|y\| \le 1} \frac{\|y\|_2^2}{N} \le \frac{\rho^2}{N} ,$$

because $\|y\|_2^2 \le \rho^2$ when $\|y\| = \|y\|_\rho^\sim \le 1$ by (16.175). Thus $\sigma \le \rho/\sqrt{N}$. Since $\|Z\| \le \tau^{-1}$ by (16.173), (16.211) implies by homogeneity

$$\mathsf{E} \sup_{y \in E, \|y\| \le 1} \Big| \sum_{j \le k} (Z_j(y)^2 - \mathsf{E} Z(y)^2) \Big| \le L(\tau^{-2} D^2 + \tau^{-1} D\rho\sqrt{k/N}) . \tag{16.214}$$

Let us consider a parameter C and $\rho = C\sqrt{N/k}D/\tau$. We substitute $D\tau^{-1} = \rho\sqrt{k/N}/C$ in the right-hand side of (16.214) to see that we may chose

$$\rho = \frac{L}{\tau}\sqrt{\frac{N}{k}}D \le \frac{L\sqrt{N/k}\sqrt{\log k}}{\tau(p-1)^{5/2}} ,$$

so that this right-hand side of (16.214) is $\le k\rho^2/4N$. With this choice, we can find (through a realization of the r.v. Z) indexes $i_1, \ldots, i_k$ as in Lemma 16.7.3, and this concludes the proof. □

Proof of Theorem 16.7.1. It remains simply to choose p appropriately. to prove that (16.170) follows from (16.172). We leave to the reader to deduce from Hölder's inequality that for $1 < p < 2$ and $\theta = (2-p)/p$, then for any function f,

$$\|f\|_p \le \|f\|_1^{\theta}\|f\|_2^{1-\theta} .$$

Consequently, if $\|f\|_2 \le a\|f\|_p$ then $\|f\|_2 \le a\|f\|_1^{\theta}\|f\|_2^{1-\theta}$ and hence $\|f\|_2 \le a^{1/\theta}\|f\|_1$. Assume now that $a = a(p)$ is of the type $a = \kappa(p-1)^{-5/2}$. Then, since for $p \le 3/2$ we have $1/\theta = 1 + 2(p-1)/(2-p) \le 1 + 4(p-1)$, we get then

$$a^{1/\theta} = (p-1)^{-5/(2\theta)}\kappa^{1/\theta} \le (p-1)^{-5/2}\kappa^{1+4(p-1)} .$$

Assuming that $\kappa \ge 3$, let us then take $p = 1 + 1/(2\log\kappa)$ to obtain $a^{1/\theta} \le L\kappa(\log\kappa)^{5/2}$, so that $\|f\|_2 \le L\kappa(\log\kappa)^{5/2}\|f\|_1$. Using this for $\kappa = L\sqrt{N/k}\sqrt{\log k}/\tau$ proves Theorem 16.7.1. □

16.8 Embedding Subspaces of L^p into ℓ^p_N

Given a k-dimensional subset X of $L^p = L^p([0,1])$, what is the smallest value of N such that X is 2-isomorphic to a subset of ℓ^p_N? Here we shall study the case $1 < p < 2$, which turns out to the most difficult one. We recall that for two Banach spaces E and F (say, of the same finite dimension) we define their Banach-Mazur distance as $\inf\{\|T\|\|T^{-1}\|\}$ where the infimum is taken over all choices of the isomorphism T between E and F. We shall prove the following.

Theorem 16.8.1 ([26]). *A k-dimensional subspace of L^p is within Banach-Mazur distance 2 of a subspace of ℓ^p_N where*

$$N \le Kk\log k(\log\log k)^2 . \tag{16.215}$$

Here, as well as in the rest of the section, K denotes a number depending on p only. We shall not attempt to track the dependence of K on p, but let us say that the dependence obtained through our methods is unsatisfactory, since the value of K blows up as $p \to 2$.

How sharp is Theorem 16.8.1? It does not seem to be known if one could in fact take N a multiple of k. An educated guess is that the factor $\log k$ cannot be removed when one uses a random method of the type we shall use. (There are no reasons to believe that this factor $\log k$ is necessary, but for the time being, no other method seems to exist, except in very special cases, see [23] for the case p even integer.) On the other hand it seems most likely that the factor $\log\log k$ is parasitic. The lame excuse we make up for presenting such an imperfect result is that it has stood unimproved for quite a while, and that the proof we present is instructive (and hopefully easier to follow than the original proof).

Theorem 16.8.1 is one of the most difficult (and pretty) results of this work. The reader may find in Theorem 15.3 of [12] a proof of the weaker estimate $N \le Kk(\log k)^3$ rather than (16.215) using only Dudley's bound rather than the generic chaining. The key ingredient to the proof of Theorem 16.8.1 is the following.

Theorem 16.8.2. *Consider a probability measure μ on $\{1,\dots,M\}$ and assume that*

$$\forall\, i \le M \ ,\ \mu_i := \mu(\{i\}) \le \frac{2}{M}\ . \tag{16.216}$$

Consider a k-dimensional subspace G of $L^p(\mu)$, and assume that it admits a basis $(\psi_j)_{j\le k}$ orthogonal in $L^2(\mu)$, such that if we write $\psi_j = (\psi_{j,i})_{i\le M}$ then for each $i \le M$ one has $\sum_{j\le k}\psi_{j,i}^2 = 1$ and moreover $\|\psi_j\|_2^2 = 1/k$ for each $j \le k$. Then

$$\mathsf{E}\sup_{x\in G,\|x\|_p\le 1}\Big|\sum_{i\le M}\mu_i\varepsilon_i|x_i|^p\Big| \le K\sqrt{\frac{k\log M}{M}}\Big(\log\log M + \log\frac{M}{k}\Big)\ . \tag{16.217}$$

Here of course $x = (x_i)_{i\le M}$, and $(\varepsilon_i)_{i\le M}$ is an independent Bernoulli sequence. The relevance of this statement to the problem at hand is that it will allow us to nicely embed G into a space of dimension about $M/2$ simply by dropping each coordinate at random with probability $1/2$. Iteration of this procedure will then yield Theorem 16.8.1. We have nothing new to say on this part of the proof, which we will recall at the end of the section for completeness.

Consider

$$T = \{(|x_i|^p)_{i\le M}; x\in G, \|x\|_p \le 1\}$$

and the distance d' on T given by

$$d'(s,t)^2 = \sum_{i\le M}\mu_i^2(s_i - t_i)^2\ . \tag{16.218}$$

The plan is to bound $\gamma_2(T,d')$ and to use Theorem 2.2.22. Let us define

$$\nu_i = \max\Big(\mu_i, \frac{1}{M(\log M)^3}\Big)\ . \tag{16.219}$$

As a technical device, we consider instead of d' the distance d given by

$$d(s,t)^2 = \sum_{i\le M} \mu_i \nu_i (s_i - t_i)^2 \ , \tag{16.220}$$

where ν_i is given in (16.219). Of course, the reason for this, and for many other technical devices we shall use will become apparent only in due time (the fact that $\nu_i \ge 1/M(\log M)^3$ is used crucially in (16.245) below). Since $d' \le d$, we have $\gamma_2(T,d') \le \gamma_2(T,d)$, and we shall bound $\gamma_2(T,d)$.

Let us start with a simple observation.

Lemma 16.8.3. *If G is as in Theorem 16.8.2 then for $x \in G$ we have*

$$\forall\, i \le M \ , \ |x_i| \le k^{1/p} \|x\|_p \ . \tag{16.221}$$

Proof. We recall that $(k^{1/2}\psi_j)_{j\le k}$ is an orthonormal basis of G. If $x = \sum_{j\le k} \alpha_j \psi_j$, then $\|x\|_2 = k^{-1/2}(\sum_{j\le k} \alpha_j^2)^{1/2}$, so that, since $x_i = \sum_{j\le k} \alpha_j \psi_{j,i}$ for each $i \le M$, the Cauchy-Schwarz inequality implies

$$|x_i| \le \Big(\sum_{j\le k} \alpha_j^2\Big)^{1/2} \Big(\sum_{j\le k} \psi_{j,i}^2\Big)^{1/2} = k^{1/2}\|x\|_2 \ , \tag{16.222}$$

since we assume that $\sum_{j\le k} \psi_{j,i}^2 = 1$ for each i. Now, using (16.222) in the second inequality, and since $p \le 2$,

$$\|x\|_2^2 \le \max_{i\le M} |x_i|^{2-p} \|x\|_p^p \le k^{1-p/2} \|x\|_2^{2-p} \|x\|_p^p \ ,$$

so that $\|x\|_2 \le k^{1/p-1/2}\|x\|_p$, and combining with (16.222) completes the proof. □

We observe from (16.221) and since $\nu_i \le 2/M$ that, when $\|x\|_p \le 1$,

$$\sum_{i\le M} \mu_i \nu_i x_i^{2p} \le \frac{2k}{M} \sum_{i\le M} \mu_i |x_i|^p \le \frac{2k}{M} \ ,$$

so that $d(t,0) \le \sqrt{2k/M}$ for each $t \in T$ and thus

$$\Delta(T,d) \le K\sqrt{k/M} \le K \ . \tag{16.223}$$

For $n \in \mathbb{Z}$ we define

$$c_n = \frac{2^n}{M(\log M)^4} \ . \tag{16.224}$$

For a subset A of T we define

$$F'_n(A) = 1 - \sup_{\mu(I)\le c_n} \inf_{t\in A} \sum_{i\in I} \mu_i t_i \ , \tag{16.225}$$

where the supremum is taken over all the subsets I of $\{1,\dots,M\}$ with $\mu(I) \le c_n$. This is an increasing function of A, and obviously $F'_{n+1}(A) \le F'_n(A)$. We consider an integer n_1 and we further define

$$F_n(A) = F'_n(A) + \frac{1}{n_1}\max(n_1 - n, 0)\ . \tag{16.226}$$

In this manner we define an decreasing sequence of functionals on T, and $F_0(T) \le 2$. The objective is to prove the following.

Proposition 16.8.4. *We can find an integer n_1, an integer s such that*

$$s \le K\Big(\log\log M + \log\frac{M}{k}\Big)\ , \tag{16.227}$$

and a number c^* with

$$c^* = \frac{1}{K}\sqrt{\frac{M}{k\log M}} \tag{16.228}$$

such that the sequence (F_n) of functionals satisfies the following growth condition. Consider $n \ge 0$ and $m = N_{n+1}$. Then, whenever the subsets $(H_\ell)_{\ell\le m}$ of T are far apart in the sense that there exist $s_0, t_1, \dots, t_m \in T$ such that

$$\forall \ell \le m\,,\ t_\ell \in B(s_0, 4a)\ ;\ \forall \ell\,,\ \ell' \le m\,,\ \ell \ne \ell' \Rightarrow d(t_\ell, t_{\ell'}) \ge a\,, \tag{16.229}$$

and

$$\forall \ell \le m\,,\ H_\ell \subset B(t_\ell, a4^{-s})\ , \tag{16.230}$$

we have

$$F_n\Big(\bigcup_{\ell\le m} H_\ell\Big) \ge c^* a 2^{n/2} + \min_{\ell\le m} F_{n+s}(H_\ell)\ . \tag{16.231}$$

The non-standard growth condition (16.231) is adapted to Theorem 2.7.6. This theorem implies (when applied to the functionals $(c^*)^{-1}F_n$ and using also (16.223))

$$\gamma_2(T,d) \le Ks(c^*)^{-1} \le Ks\sqrt{\frac{k\log M}{M}}\ , \tag{16.232}$$

and this proves Theorem 16.8.2. We fix once and for all s as the smallest for which

$$2^{-s} \le \frac{k}{M(\log M)^{6+4/(2-p)}}\ , \tag{16.233}$$

so that (16.227) holds.

We will deduce Proposition 16.8.4 from the following.

Proposition 16.8.5. *Consider $n \ge 0$ and $a > 0$ such that*

$$a^2 2^n \ge \frac{k}{M\log M}\ , \tag{16.234}$$

and define $\tau > 0$ *by*

$$\tau := \frac{2}{c_{n+s}} = 2^{-n-s+1} M (\log M)^4 \,. \tag{16.235}$$

Consider points $(t_\ell)_{\ell \le m}$ *of* T *which satisfy (16.229). For each* ℓ *we write* $t_\ell = (t_{\ell,i})_{i \le M}$*. Consider a subset* I *of* $\{1, \ldots, M\}$ *with*

$$\mu(I) \le c_n = \frac{2^n}{M (\log M)^4} \,. \tag{16.236}$$

and define

$$S := \max_{\ell \le m} \sum_{i \notin I} \mu_i t_{\ell,i} \mathbf{1}_{\{t_{\ell,i} \ge \tau\}} \,. \tag{16.237}$$

Then

$$S^2 \ge \frac{M a^2 2^n}{K k \log M} \,. \tag{16.238}$$

Proof of Proposition 16.8.4. Since (T, d) is a subset of the M-dimensional normed space $L^2(\mu)$, the ball $B(s_0, 4a)$ can be covered by L^M balls of radius $a/4$, and therefore there may exist points $t_1, \ldots, t_m$ as in (16.229) only when $N_{n+1} \le L^M$, i.e. $n \le L \log M$. We can then choose n_1 such that $n_1 \le L \log M$ such that it suffices to prove (16.231) in the case $n < n_1$. In that situation the last term of (16.226) forces that $F_{n+1}(A) \le F_n(A) - 1/n_1$, and (16.231) is automatically satisfied unless

$$c^* a 2^{n/2} \ge \frac{1}{n_1} \ge \frac{1}{L \log M} \,. \tag{16.239}$$

If the constant K of (16.228) is large enough, (16.239) implies (16.234). Therefore to prove (16.231) it suffices to prove that when (16.234) occurs then

$$F'_n\Big(\bigcup_{\ell \le m} H_\ell\Big) \ge c^* a 2^{n/2} + \min_{\ell \le m} F'_{n+s}(H_\ell) \,. \tag{16.240}$$

Consider a subset I of $\{1, \ldots, M\}$ as in (16.236). Combining (16.237) and (16.238) we may then fix $\ell \le m$ for which

$$\sum_{i \in J} \mu_i t_{\ell,i} \ge a 2^{n/2} \frac{1}{K_0} \sqrt{\frac{M}{k \log M}} \,, \tag{16.241}$$

where

$$J = \{ i \notin I \,;\, t_{\ell,i} \ge \tau \} \tag{16.242}$$

is disjoint from I. Since $\sum_{i \le M} \mu_i t_{\ell,i} \le 1$ we have

$$\mu(J) \le 1/\tau = \frac{2^{n+s-1}}{M (\log M)^4} \,, \tag{16.243}$$

so that, with lots of room,

$$\mu(I \cup J) \le c_{n+s} = \frac{2^{n+s}}{M(\log M)^4} \,. \tag{16.244}$$

Consider $t \in H_\ell$. The Cauchy-Schwarz inequality implies

$$\sum_{i \in J} \mu_i |t_{\ell,i} - t_i| \le \sqrt{\sum_{i \in J} \frac{\mu_i}{\nu_i}} \sqrt{\sum_{i \le M} \mu_i \nu_i (t_{\ell,i} - t_i)^2} \le \sqrt{\sum_{i \in J} \frac{\mu_i}{\nu_i}} d(t_\ell, t) \,.$$

Now, since $\nu_i \ge 1/(M(\log M)^3)$,

$$\sqrt{\sum_{i \in J} \frac{\mu_i}{\nu_i}} \le \sqrt{\sum_{i \in J} \mu_i} \sqrt{M(\log M)^3} = \sqrt{\mu(J)} \sqrt{M(\log M)^3} \,.$$

Finally, since $\mu(J) \le 1/\tau$ and $d(t_\ell, t) \le a 4^{-s}$ by (16.230),

$$\begin{aligned} \sum_{i \in J} \mu_i |t_{\ell,i} - t_i| &\le \sqrt{1/\tau} \sqrt{M(\log M)^3} d(t_\ell, t) \\ &\le 2^{n/2} 2^{s/2} a 4^{-s} \le a 2^{n/2} 2^{-s} \,. \end{aligned} \tag{16.245}$$

Recalling the constant K_0 of (16.241) and using (16.233), since $k \le M$ we have with plenty of room that for $M \ge K$

$$2^{-s} \le \frac{k}{M(\log M)^{6+4/(2-p)}} \le \frac{1}{2K_0} \sqrt{\frac{M}{k \log M}} \,,$$

and consequently from (16.245),

$$\sum_{i \in J} \mu_i |t_{\ell,i} - t_i| \le \frac{a 2^{n/2}}{2K_0} \sqrt{\frac{M}{k \log M}} \,.$$

Since this occurs for each $t \in H_\ell$, recalling (16.241) we obtain

$$\inf_{t \in H_\ell} \sum_{i \in J} \mu_i t_i \ge \frac{a 2^{n/2}}{2K_0} \sqrt{\frac{M}{k \log M}} \,.$$

Therefore

$$\begin{aligned} \inf_{t \in H_\ell} \sum_{i \in I \cup J} \mu_i t_i &\ge \frac{a 2^{n/2}}{K} \sqrt{\frac{M}{k \log M}} + \inf_{t \in H_\ell} \sum_{i \in I} \mu_i t_i \\ &\ge \frac{a 2^{n/2}}{K} \sqrt{\frac{M}{k \log M}} + \inf_{t \in \bigcup_{\ell' \le m} H_{\ell'}} \sum_{i \in I} \mu_i t_i \,, \end{aligned} \tag{16.246}$$

and since I is arbitrary, using (16.244) this proves (16.240) and completes the proof. □

The overall strategy to prove Proposition 16.8.5 is clear. We have to show that if S is small, there is not "enough room" in G so that one can find all these points t_ℓ in T which satisfy $d(t_\ell, t_{\ell'}) \geq a$ for $\ell \neq \ell'$. The program is to prove some "smallness" of $B_{G,p}$ for various distances in the form of entropy estimates, and then to relate these distances with the distance d on T.

We start this program by proving suitable entropy estimates. Given $r > 1$ we define

$$B_{G,r} = \Big\{ x \in G \, ; \, \sum_{i \leq M} \mu_i |x_i|^r \leq 1 \Big\} \,. \tag{16.247}$$

Given a subset W of $\{1, \ldots, M\}$, we consider the distance $d_{W,r}$ on G induced by the semi-norm

$$\alpha(x) := \alpha_{W,r}(x) := \Big(\sum_{i \in W} \mu_i |x_i|^r \Big)^{1/r} \,. \tag{16.248}$$

When $W = \{1, \ldots, M\}$ we will write d_r for the distance $d_{W,r}$, which is simply the distance induced by $L^r(\mu)$, of unit ball (16.247). There is a simple reason why one should consider the distances $d_{W,r}$. In order to say anything at all about the quantity S of (16.238), one must prove that the separation property (16.229) does not occur entirely from the coordinates in I, so it will be helpful to control what happens on sets of coordinates that are not too large.

The following estimates will be crucial in the proof of Proposition 16.8.5.

Proposition 16.8.6. *We have*

$$e_n(B_{G,p}, d_2) \leq K(2^{-n} k \log k)^{1/p - 1/2} \,. \tag{16.249}$$

Moreover, recalling the distance $d_{W,r}$ induced by the semi-norm (16.248), we have

$$e_n(B_{G,p}, d_{W,p}) \leq K(2^{-n} k \log k \mu(W))^{1/p} \tag{16.250}$$

and

$$e_n(B_{G,p}, d_r) \leq K 2^{-n/p} (k \log k)^{1/p - 1/2} \sqrt{kr} \,. \tag{16.251}$$

The meaning of these estimates is not very intuitive. In the range $2^{-n} k \geq 1$ we do not know how to do better. When $2^{-n} k \leq 1$ one could get much better estimates by using that G is of dimension k and Exercise 2.2.14, (e). These better estimates will not be needed.

The proof of Proposition 16.8.6 takes the next three pages or so. It will be an opportunity to learn some of the most useful methods to bound entropy numbers. These methods will be further used in the proof of Proposition 16.8.5.

Lemma 16.8.7. *Consider a subset W of $\{1,\ldots,M\}$ and recall the distance $d_{W,r}$ on G induced by the semi-norm (16.248). Then*

$$e_n(B_{G,2},d_{W,r}) \le K2^{-n/2}\sqrt{kr}\mu(W)^{1/r}\,. \tag{16.252}$$

Proof. The proof is based on the dual Sudakov inequality (8.3.6). Since $(k^{1/2}\psi_j)_{j\le k}$ is an orthonormal basis of $L^2(\mu)$, when $(g_j)_{j\le k}$ are independent standard Gaussian r.v.s the vector $\sum_{j\le k}k^{1/2}g_j\psi_j$ is a standard Gaussian vector on G (seen as a subspace of $L^2(\mu)$). Now,

$$\begin{aligned}\mathsf{E}\alpha\Big(\sum_{j\le k}k^{1/2}g_j\psi_j\Big) &= k^{1/2}\mathsf{E}\Big(\sum_{i\in W}\mu_i\Big|\sum_{j\le k}g_j\psi_{j,i}\Big|^r\Big)^{1/r}\\ &\le k^{1/2}\Big(\sum_{i\in W}\mu_i\mathsf{E}\Big|\sum_{j\le k}g_j\psi_{j,i}\Big|^r\Big)^{1/r}\,. \end{aligned}\tag{16.253}$$

Since we assume that $\sum_{j\le k}\psi_{j,i}^2=1$, the r.v. $\sum_{j\le k}g_j\psi_{j,i}$ is standard Gaussian so that

$$\mathsf{E}\Big|\sum_{j\le k}g_j\psi_{j,i}\Big|^r \le (Lr)^{r/2},$$

and substitution in (16.253) yields that the left-hand side is $\le L\sqrt{kr}\mu(W)^{1/r}$, and use of (8.3.6) completes the proof. □

Lemma 16.8.8. *Consider $r>2$. Then*

$$e_n(B_{G,2},d_r) \le K(r)(2^{-n}k\log k)^{1/2-1/r}\,. \tag{16.254}$$

This has a tendency to improves on (16.252) in the case where $2^{-n}k\ge 1$ (which is the important one) because $1/2-1/r<1/2$.

Proof. Consider $h>r$ and define θ, $0\le\theta\le 1$ by

$$\frac{1}{r}=\frac{1-\theta}{2}+\frac{\theta}{h}\,.$$

It follows from Hölder's inequality that $\|x\|_r\le\|x\|_2^{1-\theta}\|x\|_h^\theta$. Consequently for $x,y\in B_{G,2}$ we have $d_r(x,y)\le 2d_h(x,y)^\theta$ and using (16.252) in the second inequality implies

$$e_n(B_{G,2},d_r)\le 2e_n(B_{G,2},d_h)^\theta \le 2(K2^{-n/2}\sqrt{kh})^\theta\,.$$

Now

$$\beta := \frac12-\frac1r=\frac{\theta}{2}\Big(1-\frac{2}{h}\Big)\,,$$

so that if $h\ge 4$, and since $1/(1-x)\le 1+2x$ for $0\le x\le 1/2$,

$$2\beta\le\theta=2\beta\frac{1}{1-2/h}\le 2\beta+\frac{8\beta}{h}\,.$$

Thus if $2^{-n/2}\sqrt{kh} \le 1$ we have $(2^{-n/2}\sqrt{kh})^\theta \le (2^{-n/2}\sqrt{kh})^{2\beta}$, while if $2^{-n/2}\sqrt{kh} \ge 1$ we have

$$(2^{-n/2}\sqrt{kh})^\theta \le (2^{-n/2}\sqrt{kh})^{2\beta}(hk)^{4\beta/h} ,$$

and the choice $h = r\log k$ completes the proof since then $(hk)^{4\beta/h} \le (hk)^{4/h} \le K(r)$ because $\beta \le 1$ and also $K^\theta \le K(r)$. □

The following classical property of the entropy numbers is fundamental.

Lemma 16.8.9. *Consider two distances d_1 and d_2 on $\mathbb{R}^m$ that arise from semi-norms, of unit balls U_1 and U_2 respectively. Then for any set $T \subset \mathbb{R}^m$ one has*

$$e_{n+1}(T, d_2) \le 2e_n(T, d_1)e_n(U_1, d_2) . \tag{16.255}$$

Proof. Consider $a > e_n(T, d_1)$ so that we can find points $(t_\ell)_{\ell \le N_n}$ of T such $T \subset \cup_{\ell \le N_n}(t_\ell + aU_1)$. Consider $b > e_n(U_1, d_2)$, so that we can find points $(u_\ell)_{\ell \le N_n}$ for which $U_1 \subset \cup_{\ell \le N_n}(u_\ell + bU_2)$. Then

$$T \subset \bigcup_{\ell,\ell' \le N_n} (t_\ell + au_{\ell'} + abU_2) .$$

Let

$$I = \{(\ell,\ell') \; ; \; \ell,\ell' \le N_n \; , \; (t_\ell + au_{\ell'} + abU_2) \cap T \ne \emptyset\} ,$$

so that $\operatorname{card} I \le N_n^2 = N_{n+1}$. For $(\ell,\ell') \in I$ let $v_{\ell,\ell'} \in (t_\ell + au_{\ell'} + abU_2) \cap T$. Then

$$T \subset \bigcup_{(\ell,\ell') \in I} (v_{\ell,\ell'} + 2abU_2) ,$$

so that $e_{n+1}(T, d_2) \le 2ab$. □

The following very useful result of N. Tomczak-Jaegermann is weaker than the duality results of [3] which we used in the previous section. Since the proof takes only a few lines, we give it for convenience.

Lemma 16.8.10. *Consider on $\mathbb{R}^m$ a distance d_V induced by a norm of unit ball V, and let V° be the polar set of V, i.e. the unit ball of the dual norm under the canonical duality of $\mathbb{R}^m$ with itself. Denote by B_2 the Euclidean ball of $\mathbb{R}^m$ and by d_2 the Euclidean distance. Assume that for some numbers $\alpha \ge 1$, A and n^* we have*

$$0 \le n \le n^* \Rightarrow e_n(B_2, d_V) \le 2^{-n/\alpha}A . \tag{16.256}$$

Then

$$0 \le n \le n^* \Rightarrow e_n(V^\circ, d_2) \le 16 \cdot 2^{-n/\alpha}A . \tag{16.257}$$

Proof. Consider $n \le n^*$. Using (16.255) in the first inequality and (16.256) in the second one we obtain

$$e_{n+1}(V^\circ, d_V) \le 2e_n(V^\circ, d_2)e_n(B_2, d_V) \le 2^{-n/\alpha+1} A e_n(V^\circ, d_2) \ . \qquad (16.258)$$

Let us now denote by $\langle \cdot, \cdot \rangle$ the canonical duality of $\mathbb{R}^m$ with itself, so that if $y \in V$ and $z \in V^\circ$ we have $\langle y, z \rangle \le 1$. Consider $x, t \in V^\circ$, and $a = d_V(s,t)$. Then $x - t \in 2V^\circ$ and $x - t \in aV$, so that

$$\|x - t\|_2^2 = \langle x - t, x - t \rangle \le 2a \ ,$$

and thus $d_2(x,t)^2 \le 2d_V(s,t)$. Consequently $e_{n+1}(V^\circ, d_2)^2 \le 2e_{n+1}(V^\circ, d_V)$. Combining with (16.258),

$$e_{n+1}(V^\circ, d_2)^2 \le 2^{-n/\alpha+2} A e_n(V^\circ, d_2) \ ,$$

from which (16.257) follows by induction over n. □

Proof of Proposition 16.8.6. We use (16.254) for $r = p^*$, the conjugate exponent of p, so that $1/2 - 1/p^* = 1/p - 1/2$ and we then use Lemma 16.8.10 to obtain (16.249). Next we use (16.255) to obtain

$$e_{n+1}(B_{G,p}, d_{W,r}) \le 2e_n(B_{G,p}, d_2)e_n(B_{G,2}, d_{W,r}) \ . \qquad (16.259)$$

We then use (16.249) and (16.252) to obtain

$$e_n(B_{G,p}, d_{W,r}) \le K 2^{-n/p}(k \log k)^{1/p-1/2}\sqrt{kr}(\mu(W))^{1/r} \ .$$

Taking $r = p$ yields (16.250) and taking $W = \{1, \dots, M\}$ yields (16.251). □

This finishes for the time being our consideration of entropy estimates. We now define p_0 as the conjugate exponent of $\log M$ and we observe that since $k \le M$,

$$0 \le u \le k \Rightarrow u^{p_0} \le Lu \ . \qquad (16.260)$$

(It is of course this property which motivates the definition of p_0.) We further define

$$q = p \log M \ . \qquad (16.261)$$

We need the following elementary fact.

Lemma 16.8.11. *Consider $\eta > 0$, a subset W of $\{1, \dots, M\}$ and $x, y \in B_{G,p}$. Consider a number A with*

$$A \ge \sum_{i \in W} \mu_i |x_i|^p \mathbf{1}_{\{|x_i| \ge \eta\}} \ ; \ A \ge \sum_{i \in W} \mu_i |y_i|^p \mathbf{1}_{\{|y_i| \ge \eta\}} \ . \qquad (16.262)$$

Then

$$\sum_{i \in W} \mu_i \nu_i (|x_i|^p - |y_i|^p)^2 \le K \frac{\eta^p}{M} + K(\max_{i \in W} \nu_i)\|x - y\|_q^p A^{1/p_0} \ . \qquad (16.263)$$

In particular, taking $\eta = 0$,

$$\sum_{i \in W} \mu_i \nu_i (|x_i|^p - |y_i|^p)^2 \le K(\max_{i \in W} \nu_i) \|x - y\|_q^p \, . \tag{16.264}$$

This lemma performs the second part of our program. If $t = (|x_i|^p)_{i \le M}$ and $t' = (|y_i|^p)_{i \le M}$ then the left-hand side of (16.263) is relevant to the estimation of $d(t,t')^2 = \sum_{i \le M} \mu_i \nu_i (|x_i|^p - |y_i|^p)^2$, so that one might say that this inequality relates the distance d on T with the norm $\| \cdot \|_q$ on $B_{G,q}$.

Proof. Let $W_1 = \{i \in W ; |x_i| \le \eta, |y_i| \le \eta\}$ and $W_2 = W \setminus W_1$. For $i \in W_1$ we have $(|x_i|^p - |y_i|^p)^2 \le (|x_i|^p + |y_i|^p)^2 \le 2\eta^p(|x_i|^p + |y_i|^p)$ so that since $\nu_i \le 2/M$,

$$\sum_{i \in W_1} \mu_i \nu_i (|x_i|^p - |y_i|^p)^2 \le \frac{4\eta^p}{M} \sum_{i \in W_1} \mu_i (|x_i|^p + |y_i|^p) \le \frac{8\eta^p}{M} \, . \tag{16.265}$$

Next, we observe that for $u, v \in \mathbb{R}$ one has

$$(|u|^p - |v|^p)^2 \le K|u - v|^p (|u|^p + |v|^p) \, .$$

To prove this we may assume that $0 \le v \le u$ and we use that $u^p \le v^p + p(u - v)u^{p-1}$ and that $(u-v)^2 u^{2p-2} \le |u - v|^p u^p$ since $p \le 2$. Consequently, using in the second line Hölder's inequality with exponents $\log M$ and p_0 (and since $q = p \log M$)

$$\begin{aligned} \sum_{i \in W_2} \mu_i \nu_i (|x_i|^p - |y_i|^p)^2 &\le K(\max_{i \in W} \nu_i) \sum_{i \in W_2} \mu_i |x_i - y_i|^p (|x_i|^p + |y_i|^p) \\ &\le K(\max_{i \in W} \nu_i) \|x - y\|_q^p \Big(\sum_{i \in W_2} \mu_i (|x_i|^{pp_0} + |y_i|^{pp_0}) \Big)^{1/p_0} \, . \end{aligned} \tag{16.266}$$

Next we observe that for $i \in W_2$ we have

$$|x_i|^p + |y_i|^p \le 2(|x_i|^p \mathbf{1}_{\{|x_i| \ge \eta\}} + |y_i|^p \mathbf{1}_{\{|y_i| \ge \eta\}}) \, ,$$

since one of the terms on the right-hand side in not zero. Consequently, using (16.221) and (16.260),

$$\sum_{i \in W_2} \mu_i (|x_i|^{pp_0} + |y_i|^{pp_0}) \le KA \, ,$$

and combining with (16.266) completes the proof. □

We now begin the proof of Proposition 16.8.4. Consider the points $(t_\ell)_{\ell \le m}$ as in (16.229), and for $\ell \le m$ let $x^\ell \in B_{G,p}$ such that $t_\ell = (|x_i^\ell|^p)_{i \le M}$. This notation will be used until the end of the proof.

Lemma 16.8.12. *If $M \geq K$ we can find a subset V of $\{1, \dots, m\}$, with* $\operatorname{card} V \geq N_{n-2}$ *and the following properties:*

$$\ell, \ell' \in V \ , \ \ell \neq \ell' \Rightarrow \frac{a^2}{2} \leq \sum_{i \notin I} \mu_i \nu_i (t_{\ell,i} - t_{\ell',i})^2 \ . \tag{16.267}$$

$$\ell, \ell' \in V \Rightarrow \|x^{\ell} - x^{\ell'}\|_q \leq K(2^{-n} k \log M)^{1/p} \ . \tag{16.268}$$

$$\ell, \ell' \in V \Rightarrow \sum_{i \in I} \mu_i |x_i^{\ell} - x_i^{\ell'}|^p \leq \frac{Kk \log M}{M(\log M)^4} \ . \tag{16.269}$$

In particular (16.267) shows that the separation property (16.229) basically occurs outside I. The role of the other two conditions will became apparent soon.

Proof. Let

$$I_0 = \left\{ i \in I \ ; \ \mu_i \geq \frac{1}{M(\log M)^3} \right\} ,$$

so that (16.236) implies that $\operatorname{card} I_0 \leq 2^n / \log M$. Next we claim that for $M \geq K$ we can partition $\{1, \dots, m\}$ into N_n sets V which satisfy (16.268) and (16.269) as well as

$$\ell, \ell' \in V \Rightarrow \sum_{i \in I_0} \mu_i \nu_i (t_{\ell,i} - t_{\ell',i})^2 \leq \frac{a^2}{4} \ . \tag{16.270}$$

To see this, for each of the properties (16.268) to (16.270) we prove that we can partition $\{1, \dots, m\}$ into $\leq N_{n-2}$ sets V which satisfy this property. Since $N_{n-2}^3 \leq N_n$ this proves the claim.

We first observe that (16.251) used for $r = q = p \log M$ implies

$$e_n(B_{G,p}, d_q) \leq K(2^{-n} k \log M)^{1/p} \ ,$$

and using this for $n-2$ rather than n shows that we can partition $\{1, \dots, m\}$ into N_{n-2} sets V that satisfy (16.268).

Next, using (16.250) for $W = I$ yields

$$e_n(B_{G,p}, d_{I,p}) \leq K(2^{-n} k \log k \mu(I))^{1/p} \leq K \left(\frac{k \log k}{M(\log M)^4} \right)^{1/p} ,$$

and using this for $n-2$ rather than n proves that we can partition $\{1, \dots, m\}$ into N_{n-2} sets V that satisfy (16.269).

To prove that we can partition $\{1, \dots, m\}$ into N_{n-2} sets which satisfy (16.270) we simply use (16.229) and a dimensionality argument. Namely, we consider the space E of sequences $(y_i)_{i \in I_0}$, provided with the norm $\|y\|^2 = \sum_{i \in I_0} \mu_i \nu_i y_i^2$. For $t = (t_i)_{i \leq M} \in T$ let us write $\pi(t) = (t_i)_{i \in I_0}$. Consider points s_0 and t_ℓ as in (16.229). In E the ball $B(\pi(s_0), 4a)$ can be covered by $L^{\operatorname{card} I_0}$

balls of radius $a/4$, so in particular this ball can be partitioned in $\leq L^{\operatorname{card} I_0}$ sets of diameter $\leq a/2$. This proves that we can partition $\{1,\ldots,m\}$ into $\leq L^{\operatorname{card} I_0}$ sets which satisfy (16.270). Since $\operatorname{card} I_0 \leq 2^n/\log M$, we have $L^{\operatorname{card} I_0} \leq N_{n-2}$ for $M \geq K$.

This completes the construction of the sets V. Since $m = N_{n+1} = N_n^2$ and there are at most N_n sets V, we can find such a set V with $\operatorname{card} V \geq N_n$. This set satisfies (16.268) and (16.269) and we have to show that it satisfies (16.267).

Let us define $I' = I \setminus I_0$. Using (16.264) for $W = I'$, and since $\nu_i \leq 1/M(\log M)^3$ for $i \in W$ we obtain, using (16.268) in the second inequality and (16.234) in the third inequality that for $\ell, \ell' \in V$ we have

$$\sum_{i\in I'} \mu_i\nu_i(t_{\ell,i} - t_{\ell',i})^2 \leq \frac{K\|x^\ell - x^{\ell'}\|_q^p}{M(\log M)^3} \leq \frac{K2^{-n}k\log M}{M(\log M)^3} \leq \frac{Ka^2}{\log M} .$$

Combining this inequality with (16.270) we then obtain that for $M \geq K$ we have

$$\ell, \ell' \in V \Rightarrow \sum_{i\in I} \mu_i\nu_i(t_{\ell,i} - t_{\ell',i})^2 \leq \frac{a^2}{2} . \tag{16.271}$$

Since by (16.229) one has $\sum_{i\leq M} \mu_i\nu_i(t_{\ell,i} - t_{\ell',i})^2 \geq a^2$ for $\ell \neq \ell'$ we have proved (16.267). □

We recall the quantity τ of (16.235) and we define $\eta > 0$ by

$$\eta^p = \tau .$$

Recalling the quantity S of (16.237) we then have

$$S = \max_{\ell\leq M} \sum_{i\notin I} \mu_i t_{\ell,i}\mathbf{1}_{\{t_{\ell,i}\geq\tau\}} = \max_{\ell\leq M} \sum_{i\notin I} \mu_i |x_i^\ell|^p \mathbf{1}_{\{|x_i^\ell|\geq\eta\}} . \tag{16.272}$$

Lemma 16.8.13. *If $M \geq K$ have*

$$\ell, \ell' \in V \ , \ \ell \neq \ell' \Rightarrow \frac{a^2}{4} \leq \frac{K}{M}\|x^\ell - x^{\ell'}\|_q^p S . \tag{16.273}$$

This is a significant progress because now the separation pertains to points of G, for a simple norm.

Proof. Using (16.263) we obtain

$$\ell, \ell' \in V \Rightarrow \sum_{i\notin I} \mu_i\nu_i(t_{\ell,i} - t_{\ell',i})^2 \leq K\frac{\eta^p}{M} + \frac{K}{M}\|x^\ell - x^{\ell'}\|_q^p S^{1/p_0} . \tag{16.274}$$

Then, using (16.235) in the first inequality, (16.233) in the second one, and (16.234) in the last one,

$$K\frac{\eta^p}{M} = K\frac{\tau}{M} \le K2^{-n-s}(\log M)^4 \le K2^{-n}\frac{k}{M(\log M)^2} \le \frac{K}{\log M}a^2 \; .$$

In particular for $M \ge K$ we have $K\eta^p/M \le a^2/4$. Combining with (16.267) and (16.274) we have shown that for $M \ge K$ we have

$$\ell, \ell' \in V \; , \; \ell \ne \ell' \Rightarrow \frac{a^2}{4} \le \frac{K}{M}\|x^\ell - x^{\ell'}\|_q^p S^{1/p_0} \; , \tag{16.275}$$

and, consequently using (16.268),

$$\frac{a^2}{4} \le \frac{K}{M}2^{-n}k\log M S^{1/p_0} \; . \tag{16.276}$$

Using (16.234) again we obtain $S^{1/p_0} \ge 1/K(\log M)^2$. Since p_0 is the conjugate exponent of $\log M$, for $M \ge K$ we have $p_0 \le 1 + 2/\log M$. Since $(\log M)^{1/\log M} \le K$ this implies

$$S \ge \frac{1}{(K\log M)^{2p_0}} \ge \frac{1}{K(\log M)^2} \tag{16.277}$$

and

$$S^{1/p_0} = S \times S^{-1/\log M} \le KS \; . \tag{16.278}$$

Combining with (16.275), this proves (16.273). □

Let us fix $\ell_0 \in V$ and for $\ell \in V$ define $y^\ell = x^\ell - x^{\ell_0}$. These points are far apart from (16.273). We are now going to show that they belong to a rather small special subset of G. For $r \ge 1$ let us denote by B_r the unit ball of $L^r(\mu)$. Considering $0 \le \xi \le 1$, let us define

$$U(\xi, \eta) = \xi B_p + (B_p \cap \eta B_\infty) \; . \tag{16.279}$$

Lemma 16.8.14. *We have $y^\ell \in 2U(\xi, \eta)$ where $\xi = (3S)^{1/p}$.*

Proof. We deduce from (16.269), using (16.277) in the second inequality, that for $M \ge K$

$$\ell \in V \Rightarrow \sum_{i \in I}\mu_i|y_i^\ell|^p \le \frac{kK\log M}{M(\log M)^4} \le S \; . \tag{16.280}$$

Next, we observe the inequality

$$|u - v|^p \mathbf{1}_{\{|u-v| \ge 2\eta\}} \le 2^p(|u|^p\mathbf{1}_{\{|u| \ge \eta\}} + |v|^p\mathbf{1}_{\{|v| \ge \eta\}}) \; ,$$

that follows simply from the fact that $|u - v| \le 2\max(|u|, |v|)$. Using this for $u = x_i^\ell$ and $v = x_i^{\ell_0}$ and combining with the definition of S this implies that for $\ell \in V$,

$$\sum_{i \notin I}\mu_i\Big|\frac{y_i^\ell}{2}\Big|^p \mathbf{1}_{\{|y_i^\ell| \ge 2\eta\}} \le 2S \; . \tag{16.281}$$

Let us then define $u = (u_i)_{i \le M}$ by $u_i = y_i^\ell \mathbf{1}_{\{|y_i^\ell| < 2\eta\}}$ if $i \notin I$ and $u_i = 0$ otherwise. Thus $u/2 \in \eta B_\infty \cap B_p$, and combining (16.280) with (16.281) shows that $\|y^\ell/2 - u/2\|_p^p \le 3S$. This proves indeed that $y^\ell \in 2U(\xi, \eta)$ for $\xi = (3S)^{1/p}$. □

To take advantage of the fact that the points $(y_\ell)_{\ell \in V}$ are far from each other for d_q from (16.273), we need information about the covering numbers of the sets $G \cap U(\xi, \eta)$ for this distance. This is the purpose of the next (and crucial) estimate.

Proposition 16.8.15. *Consider $n \ge 0$ such that*

$$\eta \le \xi^{2/(2-p)} (2^{-n} k \log M)^{1/p} . \tag{16.282}$$

Then

$$e_n(G \cap U(\xi, \eta), d_q) \le K\xi (2^{-n} k \log M)^{1/p} . \tag{16.283}$$

Let us stress the point here. An element $x \in G \cap U(\xi, \eta)$ can be written as $x = x_1 + x_2$ where $x_1 \in \xi B_p$ and $x_2 \in B_p \cap \eta B_\infty$. There is a priori no reason why one should have such a decomposition with x_1, x_2 in G.

Proof. The interpolation formula $\|f\|_2 \le \|f\|_\infty^{1-p/2} \|f\|_p^{p/2}$ implies

$$B_p \cap \eta B_\infty \subset \eta^{1-p/2} B_2 . \tag{16.284}$$

We consider $L^2(\mu)$ in canonical duality with itself, and the polar V of $U(\xi, \eta)$. Let p^* be the conjugate exponent of p. Consider the distance d_V induced by the norm of unit ball V, so that, using (16.279) and (16.284),

$$d_V(x, y) = \sup\{\langle x - y, z\rangle \; ; \; z \in U(\xi, \eta)\} \le \xi d_{p^*}(x, y) + \eta^{1-p/2} d_2(x, y) , \tag{16.285}$$

and therefore

$$e_n(B_{G,2}, d_V) \le \xi e_n(B_{G,2}, d_{p^*}) + 2\eta^{1-p/2} .$$

Using (16.254) for $r = p^*$ yields

$$e_n(B_{G,2}, d_{p^*}) \le K(2^{-n} k \log M)^{1/p - 1/2} ,$$

and consequently,

$$e_n(B_{G,2}, d_V) \le K\xi (2^{-n} k \log M)^{1/p-1/2} + 2\eta^{1-p/2} ,$$

so that, when (16.282) holds,

$$e_n(B_{G,2}, d_V) \le K\xi (2^{-n} k \log M)^{1/p-1/2} .$$

Now (16.282) holds for $n \le n^*$ for a certain integer n^* and (16.257) of Lemma 16.8.10 implies that for these values of n,

$$e_n(G \cap U(\xi, \eta), d_2) \le K\xi(2^{-n} k \log M)^{1/p - 1/2} . \tag{16.286}$$

Finally we deduce from (16.255) that

$$e_{n+1}(G \cap U(\xi, \eta), d_q) \le 2e_n(G \cap U(\xi, \eta), d_2) e_n(B_{G,2}, d_q) ,$$

and we use (16.286) for the first term and (16.252) for $r = q$ and $W = \{1, \ldots, M\}$ for the second term. □

Proof of Proposition 16.8.4. We rewrite (16.282) as

$$\tau = \eta^p \le (\xi^p)^{2/(2-p)} 2^{-n} k \log M ,$$

and, recalling the value of τ from (16.235), and since $\xi^p = 3S$, this holds provided

$$2^{-n-s+1} M (\log M)^4 \le (3S)^{2/(2-p)} 2^{-n} k \log M . \tag{16.287}$$

It then follows from (16.277) and (16.233) that this relation is satisfied when $M \ge K$. Consequently, we know from (16.283) that

$$e_{n-3}(G \cap U(\xi, \eta), d_q) \le K\xi(2^{-n} k \log M)^{1/p} .$$

Since $\operatorname{card} V = N_{n-2}$ we can find ℓ and ℓ' in V, $\ell \ne \ell'$ such that

$$\|x^\ell - x^{\ell'}\|_q \le K\xi(2^{-n} k \log M)^{1/p} .$$

Consequently, using (16.273) in the first inequality and since $\xi^p = 3S$,

$$a^2 \le \frac{K}{M} \|x^\ell - x^{\ell'}\|_q^p S \le \frac{K}{M} S^2 2^{-n} k \log M ,$$

so that

$$S^2 \ge \frac{M a^2 2^n}{K k \log M} ,$$

which is the minoration of S we have been looking for. □

Let us end this section with a sketch of the proof of Theorem 16.8.1. Denoting by Λ the left-hand side of (16.217) we can find a choice of the sequence (ε_i) for which $\operatorname{card}\{i \le M; \varepsilon_i = 1\} \le M/2$ and

$$\sup_{x \in B_{G,p}} \Big| \sum_{i \le M} \mu_i \varepsilon_i |x_i|^p \Big| \le \Lambda . \tag{16.288}$$

Consequently

$$\sup_{x \in B_{G,p}} \Big| \sum_{i \le M} \mu_i (1 + \varepsilon_i) |x_i|^p - \|x\|_p^p \Big| \le \Lambda , \tag{16.289}$$

and if we define the norm $\mathcal{N}(x)$ by $\mathcal{N}(x)^p = \sum_{i \le N} (1 + \varepsilon_i) |x_i|^p$, we have

$$|\mathcal{N}(x)^p - \|x\|_p^p| \le \Lambda \|x\|_p^p ,$$

and

$$(1-\Lambda)^{1/p} \le \mathcal{N}(x) \le \|x\|_p (1+\Lambda)^{1/p} .$$

In particular for $\Lambda \le 1/2$, the Banach-Mazur distance between the spaces $(G, \|\cdot\|_p)$ and $(G, \mathcal{N}(\cdot))$ is $\le K\Lambda^2$. The space $(G, \mathcal{N}(\cdot))$ is isomorphic to a subspace of a space $L^p(\mu')$ where μ' is supported by a set of cardinality $M' \le M/2$. A theorem of D. Lewis [13] asserts that one can make a change of density in the space $L^p(\mu')$ to ensure the existence of a sequence ψ_j as in Theorem 16.8.2. By splitting the atoms of mass $\ge 2/M'$ one can finally find a subspace G' of $L^p(\mu'')$ within Banach-Mazur distance $K\Lambda^2$ of G, which satisfies the conditions of Theorem 16.8.2 for $M'' \le 3M/4$. This is the basic step of "dimension reduction". Iteration of this step yields Theorem 16.8.1. We refer to [12] for the details (which are straightforward).

16.9 Gordon's Embedding Theorem

One should think that there are many other potential applications of the material presented in this book to Banach Spaces, but new ideas take a long time to percolate. The following can be obtained as a particularly elegant application of Theorem 2.4.1. For any integer n, we denote by $\|\cdot\|_2$ the Euclidean norm on $\mathbb{R}^n$.

Theorem 16.9.1 (Gordon's embedding theorem [7]). *Consider two integers n and m. Denote by S^{m-1} the unit sphere of $\mathbb{R}^m$. Consider a norm $\|\cdot\|$ on $\mathbb{R}^n$, and assume that $\|\cdot\| \le \|\cdot\|_2$. Denote by X_t the canonical Gaussian process on $\mathbb{R}^m$, and by $(e_j)_{j\le n}$ the canonical basis of $\mathbb{R}^n$. Then for every subset T of S^{m-1} there is a linear operator $U : \mathbb{R}^m \to \mathbb{R}^n$ such that*

$$\forall t \in T\ ,\ 1 - L\epsilon \le \|U(t)\| \le 1 + L\epsilon$$

where

$$\epsilon = \frac{\mathsf{E} \sup_{t\in T} X_t}{\mathsf{E}\|\sum_{j\le n} g_j e_j\|} .$$

Gordon's original proof does not use Theorem 2.4.1. Theorem 16.9.1 was rediscovered by G. Schechtman [22], whose approach we follow here. Discussing all the remarkable consequences of this statement in Banach Space Theory goes beyond the purpose of this work, and we refer to [22] and references therein for this.

Proof. Consider independent standard Gaussian random variables $(g_i)_{i\ge 1}$, $(g_{ij})_{i,j\ge 1}$, and for $t = (t_1, \ldots, t_m) \in \mathbb{R}^m$ define $C_t \in \mathbb{R}^n$ by

$$C_t = \sum_{i\le m, j\le n} t_i g_{ij} e_j = \sum_{j\le n} e_j \Big(\sum_{i\le m} g_{ij} t_i\Big) ,$$

so that the law of C_t in $\mathbb{R}^n$ is the same for all $t \in S^{m-1}$, because in that case the sequence $(\sum_{i\le m} g_{ij}t_i)_{j\le m}$ is an independent sequence of standard normal r.v. Moreover, for the same reason, when $t \in S^{m-1}$ we have

$$\mathsf{E}\|C_t\| = \mathsf{E}\Big\|\sum_{j\le n} g_j e_j\Big\| \,. \tag{16.290}$$

We fix $t_0 \in T$, and for $t \in S^{m-1}$ we define

$$Y_t = \|C_t\| - \|C_{t_0}\| \,,$$

so that $\mathsf{E}Y_t = 0$. The key of the proof is to establish the inequality

$$\forall u > 0\,,\ \forall s,t \in S^{m-1}\,,\ \mathsf{P}(|Y_s - Y_t| \ge u) \le 2\exp\Big(-\frac{u^2}{L\|s-t\|_2^2}\Big)\,. \tag{16.291}$$

Once this is proved, we proceed as follows. Since $Y_{t_0} = 0$, it follows from Theorem 2.4.12 that

$$\mathsf{E}\sup_{t\in T}|Y_t| \le L\mathsf{E}\sup_{t\in T} X_t \,. \tag{16.292}$$

It follows from Lemma 8.1.11 that

$$\mathsf{P}\Big(\|C_{t_0}\| \ge \frac{1}{2}\mathsf{E}\|C_{t_0}\|\Big) \ge \frac{1}{L}\,,$$

and combining with (16.290) and (16.292), we see that we find a realization of the r.v. such that

$$\sup_{t\in T}|\|C_t\| - \|C_{t_0}\|| \le L\mathsf{E}\sup_{t\in T} X_t$$

$$\|C_{t_0}\| \ge \frac{1}{2}\mathsf{E}\|C_{t_0}\| = \frac{1}{2}\mathsf{E}\Big\|\sum_{j\le n} g_j e_j\Big\| \,.$$

The operator U given by $U(t) = C_t/\|C_{t_0}\|$ then satisfies our requirements.

The proof of (16.291) given by G. Schetchman is very beautiful. First, we note that for any $x \in \mathbb{R}^n$ and any $b \in \mathbb{R}^m$ the r.v.s $\|x + C_b\|$ and $\|x - C_b\|$ have the same law because the distribution of C_b is symmetric, and thus

$$\mathsf{E}\|x + C_b\| = \mathsf{E}\|x - C_b\| \,,$$

and also

$$\begin{aligned} &\mathsf{P}\big(\big|\|x + C_b\| - \|x - C_b\|\big| \ge u\big) \qquad (16.293)\\ &\le \mathsf{P}\Big(\big|\|x + C_b\| - \mathsf{E}\|x + C_b\|\big| \ge \frac{u}{2}\Big) + \mathsf{P}\Big(\big|\|x - C_b\| - \mathsf{E}\|x - C_b\|\big| \ge \frac{u}{2}\Big)\\ &= 2\mathsf{P}\Big(\big|\|x + C_b\| - \mathsf{E}\|x + C_b\|\big| \ge \frac{u}{2}\Big)\,. \end{aligned}$$

Now,

$$\|x+C_b\| = \sup\bigl\{x^*(x+C_b)\ ;\ x^*\in W\bigr\} = \sup\bigl\{Z_{x^*}\ ;\ x^*\in W\bigr\}\ ,$$

where W is the unit ball of the dual of the Banach space $(\mathbb{R}^N, \|\cdot\|)$, and where $Z_{x^*} = x^*(x+C_b)$. The crucial fact now is that Lemma 2.4.7 remains true when the Gaussian process Z_t is not necessarily centered, provided one replaces the condition $\mathsf{E}Z_t^2 \le \sigma^2$ by the condition $\mathsf{E}(Z_t - \mathsf{E}Z_t)^2 \le \sigma^2$. Such a property is formally stated in the case of Bernoulli processes in (5.56). In the Gaussian case, as in Lemma 2.4.7, it simply takes its roots in the remarkable behavior of the canonical Gaussian measure on $\mathbb{R}^k$ with respect to Lipschitz functions [11]. We have

$$\mathsf{E}(Z_{x^*} - \mathsf{E}Z_{x^*})^2 = \mathsf{E}(x^*(C_b))^2 = \sum_{i\le m, j\le n} x^*(e_j)^2 b_i^2\ .$$

Since we assume that $\|\cdot\| \le \|\cdot\|_2$, for $x^*\in W$, we have $|x^*(e_j)|\le 1$, and thus $\mathsf{E}(Z_{x^*} - \mathsf{E}Z_{x^*})^2 \le \|b\|_2^2$. We can then deduce from the extension of Lemma 2.4.7 mentioned above that

$$\mathsf{P}\Bigl(\bigl|\|x+C_b\| - \mathsf{E}\|x+C_b\|\bigr| \ge \frac{u}{2}\Bigr) \le 2\exp\Bigl(-\frac{u^2}{8\|b\|_2^2}\Bigr)\ ,$$

and combining with (16.293) yields

$$\mathsf{P}\bigl(\bigl|\|x+C_b\| - \|x-C_b\|\bigr| \ge u\bigr) \le 4\exp\Bigl(-\frac{u^2}{8\|b\|_2^2}\Bigr)\ . \tag{16.294}$$

Consider finally s and t in S^{m-1}. Writing $a=(s+t)/2$ and $b=(s-t)/2$ we notice that

$$C_s = C_a + C_b\ ;\ C_t = C_a - C_b\ .$$

Most importantly, since $\|s\| = \|t\|$, the vectors a and b are orthogonal, so that by the rotational invariance property of Gaussian measures the random vectors C_a and C_b are independent, and (16.291) follows using (16.294) for $x = C_a$ conditionally on C_a. □

The one objection one might raise against the previous proof is that (16.291) is a rather immediate consequence of an equally beautiful, but much older argument that G. Pisier [19] used to prove the very form of concentration of measure which is the cornerstone of the previous proof. Pisier's argument goes as follows. Consider $a, b\in S^{m-1}$ and assume that they are orthogonal. Consider a number v and the function

$$\varphi(\theta) = \exp v(\|C_{a\cos\theta + b\sin\theta}\| - \|C_a\|)\ .$$

Differentiation and Gaussian integration by parts (using that $\|\cdot\|\le\|\cdot\|_2$) yield (after a few lines of computation) the inequality

$$\varphi'(\theta) \le Lv^2 |\sin\theta| \varphi(\theta)$$

and since $\varphi(0) = 1$ by integration

$$\varphi(\theta) \le \exp L\theta^2 v^2 \ . \tag{16.295}$$

Now, given $s, t \in S^{m-1}$ we can find $b \in S^{m-1}$ orthogonal to t with $s = t\cos\theta + b\sin\theta$, and then (with the proper choice of θ modulo 2π) $\theta^2 \le \|s-t\|^2$ so that from (16.295)

$$\exp v(\|C_s\| - \|C_t\|) \le \exp Lv^2 \|s-t\|^2 \ ,$$

from which (16.291) follows.

16.10 Notes and Comments

The idea of using convexity through the functionals (4.18) (or previous incarnations of the same idea) has been successful in a number of situations. First of course is the Ellipsoid theorem, Theorem 4.1.4, which clarifies the proof of several matching theorems as in explained in Chapter 4. The second occurrence was in Rudelson's paper [21]. Rudelson discovered and proved the result of [21] alone, but his original argument for the proof of the crucial probabilistic estimate was significantly more complicated than the proof which appears in [21]. The argument which is published in [21] was produced by this author after seeing the original proof. It is essentially the proof of Proposition 16.7.4 in the case where the norm is the Euclidean norm. (It is written of course in [21] using "pre-generic chaining techniques".) Despite Rudelson's very clear statement about this origin of his argument, it has been referred in several subsequent papers as "Rudelson's argument". It has been used in particular under this name in [9], [8] and [5]. Unfortunately the arguments of [5] use a false claim about the convexity of certain balls and are incorrect as written. See also [30].

In the case of the Euclidean norm, Proposition 16.7.4 has been essentially superseded by the non-commutative Khinchin's inequalities [15] and further results in the same direction, see e.g. [24] for a review. In the simplest form, the non-commutative Khinchin's inequality asserts that for self-adjoint operators (A_i) on $\mathbb{R}^d$, and for $p \ge 2$,

$$\mathsf{E}\,\mathrm{tr}\Big(\sum_{i\ge 1} g_i A_i\Big)^{2p} \le (Lp)^p \mathrm{tr}\Big(\sum_{i\ge 1} A_i^2\Big)^p \ , \tag{16.296}$$

where tr denotes the trace. (This version with the correct growth of the constants goes back to G. Pisier's work [20].) Now, since the operators A_i are self-adjoint, so is $\sum_i g_i A_i$ and thus $\|\sum_i g_i A_i\|^{2p} \le \mathrm{tr}(\sum_i g_i A_i)^{2p}$. Moreover, $\mathrm{tr}(\sum_{i\ge 1} A_i^2)^p \le d\sigma^{2p}$ where $\sigma^2 = \|\sum_{i\ge 1} A_i^2\|$. Thus (16.296) yields

$$\mathsf{E}\Big\|\sum_i g_i A_i\Big\|^{2p} \le (Lp)^p d\sigma^{2p} ,$$

and in particular $\mathsf{E}\|\sum_i g_i A_i\| \le L\sqrt{p} d^{1/2p}\sigma$. Taking p about $\log d$ yields

$$\mathsf{E}\Big\|\sum_{i\ge1} g_i A_i\Big\| \le L\sigma\sqrt{\log d} \,. \tag{16.297}$$

In particular, if we denote by $\langle\cdot,\cdot\rangle$ the dot product in $\mathbb{R}^d$, and if

$$T = \{(\langle x, A_i(x)\rangle)_{i\ge1} \; ; \; \|x\| \le 1\} ,$$

then $\gamma_2(T, d_2) \le L\sigma\sqrt{\log d}$. It would be quite interesting to find a direct "geometrical" proof of this statement.

When (z_i) are vectors in $\mathbb{R}^d$ with $\|z_i\| \le 1$, and $A_i(x) = z_i\langle z_i, x\rangle$, then $A_i^2(x) = \|z_i\|^2 z_i\langle z_i, x\rangle$ and $\sigma \le (\sup_{\|x\|\le1}\langle z_i, x\rangle^2)^{1/2}$, so that (16.297) yields

$$\mathsf{E}\sup_{x\le1} g_i \sum_i \langle x, z_i\rangle^2 \le L\sqrt{\log d}\Big(\sup_{\|x\|\le1}\langle z_i, x_i\rangle^2\Big)^{1/2},$$

to be compared with Proposition 16.7.4. Also relevant are the very strong deterministic results of [1].

References

1. Batson, J., Spielman, D., Srivastava, N.: Twice-Ramanujan sparsifiers. In: Proceedings of the 41st Annual Symposium on the Theory of Computing, pp. 47–55 (2009)
2. Bourgain, J.: Bounded orthogonal sets and the Λ_p-problem. Acta Math. **162**, 227–246 (1989)
3. Bourgain, J., Pajor, A., Szarek, S.J., Tomczak-Jaegerman, N.: On the duality problem for entropy numbers. In: Geometric Aspects of Functional Analysis. Lecture Notes in Math, vol. 1376, pp. 53–63 (1989)
4. Clarkson, J.: Uniformly convex spaces. Trans. Am. Math. Soc. **40**(3), 39–414 (1936)
5. Candès, E., Plan, Y.: A probabilistic and RIPless theory of compressed sensing. IEEE Trans. Inf. Theory **5**(11), 7235–7254 (2011)
6. Diestel, J., Jarchow, H., Tonge, A.: Absolutely Summing Operators. Cambridge Studies in Advanced Mathematics, vol. 43. Cambridge University Press, Cambridge (1955)
7. Gordon, Y.: A note on an observation of G. Schechtman. In: Geometric Aspects of Functional Analysis. Springer Lecture Notes in Math., vol. 1910, pp. 127–132 (2010)
8. Guédon, O., Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.: Majorizing measures and proportional subsets of bounded orthonormal systems. Rev. Mat. Iberoam. **24**(3), 1071–1095 (2008)
9. Guédon, O., Rudelson, M.: L_p-moments of random vectors via majorizing measures. Adv. Math. **208**(2), 798–823 (2007)

10. König, H., Tzafriri, L.: Some estimates for type and cotype constants. Math. Ann. **256**(1), 85–94 (1981)
11. Ledoux, M.: The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001). x+181 pp. ISBN 0-8218-2864-9
12. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
13. Lewis, D.: Finite dimensional subspaces of L_p. Stud. Math. **63**(2), 207–212 (1978)
14. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces, II. Function Spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 97. Springer, Berlin (1979). x+243 pp. ISBN: 3-540-08888-1
15. Lust-Piquard, F.: Inégalités de Khintchine dans C_p. C. R. Acad. Sci. Paris Sér. I Math. **303**(7), 289–292 (1986)
16. Maurey, B.: Type et cotype dans les espaces munis de structures locales inconditionelles. In: Séminaire Maurey-Schwartz 1973-74, Exposé 12. Ecole Polytechnique, Paris (1974)
17. Montgomery-Smith, S.J.: The cotype of operators form $C(K)$. Ph.D. thesis, Cambridge (1988)
18. Pisier, G.: Factorization of operators through $L_{p,\infty}$ or $L_{p,1}$ and noncommutative generalizations. Math. Ann. **276**, 105–136 (1986)
19. Pisier, G.: Probabilistic methods in the geometry of Banach spaces. In: Probability and Analysis, Varenna, 1985. Lecture Notes in Math., vol. 1206, pp. 167–241. Springer, Berlin (1986)
20. Pisier, G.: Noncommutative vector valued L_p-spaces and completely p-summing maps. Astérisque **247**, 1–131 (1998)
21. Rudelson, M.: Almost orthogonal submatrices of an orthogonal matrix. Isr. J. Math. **111**, 143–155 (1999)
22. Schechtman, G.: Two observations regarding embedding subsets of Euclidean spaces in normed spaces. Adv. Math. **200**(1), 125–135 (2006)
23. Schechtman, G.: Tight embedding of subspaces of L^p in ℓ^p_n for even p. Proc. Am. Math. Soc. **139**(12), 4419–4421 (2001)
24. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Found. Comput. Math. **12**(4), 389–434 (2012)
25. Talagrand, M.: Cotype of operators from $C(K)$. Invent. Math. **107**, 1–40 (1992)
26. Talagrand, M.: Embedding subspaces of L_p into ℓ^n_p. In: Operator Theory: Advances and Applications, vol. 77, pp. 311–326. Birkhäuser, Basel (1995)
27. Talagrand, M.: Sections of smooth convex bodies via majorizing measures. Acta Math. **175**, 273–306 (1995)
28. Talagrand, M.: Selecting a proportion of characters. Isr. J. Math. **108**, 275–284 (1999)
29. Tomczak-Jaegermann, N.: Banach-Mazur Distances and Finite-Dimensional Operator Ideals. Pitman Monographs and Surveys in Pure and Applied Mathematics, vol. 38. Wiley, New York (1989). xii+395 pp. ISBN: 0-582-01374-7
30. Zvavitch, A.: More on embedding subspaces of L_p into ℓ^N_p, $0 < p < 1$. In: Geometric Aspects of Functional Analysis. Lecture Notes in Math., vol. 1745, pp. 269–280. Springer, Berlin (2000)

A. Appendix: What This Book Is *Really* About

A.1 Introduction

The present book is designed to be read with little prior knowledge, but it might also help to discuss a few classical results, in the perspective of what is done in this book. This appendix is self-contained, even though this entails some repetition from the main body of work.

A.2 The Kolmogorov Conditions

Since Kolmogorov invented chaining, it is appropriate to start with processes that satisfy the so-called Kolmogorov conditions (1.2), that is processes $(X_t)_{t\in T}$ where $T = [0,1]^m$, for which

$$\forall\, s,t \in [0,1]^m\ ,\ \mathsf{E}|X_s - X_t|^p \le d(s,t)^\alpha\ . \tag{1.2}$$

where $d(s,t)$ denotes the Euclidean distance and $p > 0, \alpha > m$. Let us try to prove that such processes are continuous (i.e. have a continuous version).

Chaining uses successive approximations $\pi_n(t)$ of the points t of T. When $T = [0,1]^m$ and we use the Euclidean distance, it is natural to assume that $\pi_n(t) \in G_n$, where G_n is the set of points x in $[0,1]^m$ such that the coordinates of $2^n x$ are integers $\ne 0$. Thus $\operatorname{card} G_n = 2^{nm}$. For $n \ge 0$, let us define

$$U_n = \{(s,t)\ ;\ s \in G_n\ ,\ t \in G_{n+1}\ ,\ d(s,t) \le 3\sqrt{m}2^{-n}\}\ , \tag{A.1}$$

so that we have the crucial property

$$\operatorname{card} U_n \le K(m)2^{nm}\ , \tag{A.2}$$

where $K(m)$ denotes a number depending only on m, which need not be the same on each occurrence. Consider then the r.v.

$$Y_n = \max\{|X_s - X_t|\ ;\ (s,t) \in U_n\}\ , \tag{A.3}$$

so that, using that for a finite family of numbers $V_i \ge 0$, it holds

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2,

$$(\max_i V_i)^p \le \sum_i V_i^p \; , \tag{A.4}$$

we get

$$\mathsf{E} Y_n^p \le \mathsf{E} \sum_{(s,t)\in U_n} |X_s - X_t|^p \le K(m,\alpha) 2^{n(m-\alpha)} \; ,$$

since $\mathsf{E}|X_s - X_t|^p \le K(m,\alpha)2^{-n\alpha}$ for $(s,t) \in U_n$ and using (A.2). To proceed one needs to distinguish whether or not $p \ge 1$. For specificity we assume $p \ge 1$. Since, as we just proved, $\|Y_n\|_p := (\mathsf{E}|Y|^p)^{1/p} \le K(m,p)2^{n(m-\alpha)/p}$ the triangle inequality in L^p yields

$$\Big\| \sum_{n\ge k} Y_n \Big\|_p \le K(m,p,\alpha) 2^{k(m-\alpha)/p} \; . \tag{A.5}$$

To avoid having to explain what is "a version of the process", and since we care only about inequalities, let $G = \bigcup_{n\ge 0} G_n$. We claim that

$$\sup_{s,t\in G; d(s,t)\le 2^{-k}} |X_s - X_t| \le 3 \sum_{n\ge k} Y_n \; . \tag{A.6}$$

Combining with (A.5) we then obtain

$$\Big\| \sup_{s,t\in G; d(s,t)\le 2^{-k}} |X_s - X_t| \Big\|_p \le K(m,p,\alpha) 2^{k(m-\alpha)/p} \; , \tag{A.7}$$

a sharp inequality from which it is then simple to prove (with some loss of sharpness) results such as the fact that for $0 < \beta < \alpha - m$ one has

$$\mathsf{E} \sup_{s,t\in G} \frac{|X_s - X_t|^p}{d(s,t)^\beta} < \infty \; . \tag{A.8}$$

To prove (A.6), for each n and each $u \in T = [0,1]^m$ denote by $\pi_n(u)$ a point of G_n which is as close to u as possible, so that $d(u, \pi_n(u)) \le \sqrt{m}2^{-n}$, and in particular $d(\pi_n(u), \pi_{n+1}(u)) \le 2\sqrt{m}2^{-n}$, so that $(\pi_n(u), \pi_{n+1}(u)) \in U_n$ and

$$|X_{\pi_n(u)} - X_{\pi_{n+1}(u)}| \le Y_n \; .$$

Also, when $d(s,t) \le 2^{-k}$,

$$d(\pi_k(s), \pi_k(t)) \le d(s, \pi_k(s)) + d(s,t) + d(t, \pi_k(t)) \le 3\sqrt{m}2^{-k} \; ,$$

so that $(\pi_k(s), \pi_k(t)) \in U_k$ and

$$|X_{\pi_k(s)} - X_{\pi_k(t)}| \le Y_k \; .$$

To prove (A.6) we may assume without loss of generality that $s, t \in G_m$ for some $m \ge k$, so that $s = \pi_m(s)$ and $t = \pi_m(t)$. We then use the previous inequalities and the identities

$$X_s - X_t = X_s - X_{\pi_k(s)} + X_{\pi_k(s)} - X_{\pi_k(t)} + X_{\pi_k(t)} - X_t \ ,$$

and, for $u \in \{s, t\}$,

$$X_u - X_{\pi_k(u)} = X_{\pi_m(u)} - X_{\pi_k(u)} = \sum_{k \le n < m} X_{\pi_{n+1}(u)} - X_{\pi_n(u)} \ .$$

Thus, chaining not only proves that the process (X_t) has a continuous version, it also provides the very good estimate (A.7). One reason for which everything is so easy in this case is that not only we bound the size of the terms $X_{\pi_{n+1}(u)} - X_{\pi_n(u)}$ independently of u, but also that the size of these terms decreases like a geometric series.

A.3 More Chaining in $\mathbb{R}^m$

One may also consider situations more general than (1.2), for example situations such as

$$\forall n \ge 0 \ , \ \forall s, t \in T \ , \ d(s,t) \le 3\sqrt{m}2^{-n} \Rightarrow \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{c_n}\Big) \le d_n \ , \tag{A.9}$$

where φ is a convex function ≥ 0 with $\varphi(0) = 0$, and c_n, d_n are numbers. Of course the factor $3\sqrt{m}$ is to simplify the statement of the forthcoming inequality (A.13) and is not important. Equivalently, one may consider conditions such as

$$\forall s, t \in T \ , \ \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{\psi(d(s,t))}\Big) \le \theta(d(s,t)) \ . \tag{A.10}$$

where ψ and θ are functions. We follow exactly the same method as previously, but instead of (A.4) we use now that for r.v.s $V_i \ge 0$ we have $\varphi(\max_i V_i) \le \sum_i \varphi(V_i)$, so that

$$\varphi(\mathsf{E}\max_i V_i) \le \mathsf{E}\varphi(\max_i V_i) \le \sum_i \mathsf{E}\varphi(V_i)$$

and hence

$$\mathsf{E}\max_i V_i \le \varphi^{-1}\Big(\sum_i \mathsf{E}\varphi(V_i)\Big) \ . \tag{A.11}$$

Therefore the r.v. Y_n of (A.3) satisfies

$$\mathsf{E}Y_n \le c_n \varphi^{-1}(2^{nm} d_n) \ , \tag{A.12}$$

and combining with (A.6),

$$\mathsf{E} \sup_{s,t \in G, d(s,t) \le 2^{-n}} |X_s - X_t| \le \sum_{k \ge n} c_k \varphi^{-1}(2^{km} d_k) \ . \tag{A.13}$$

Of course in the case (A.10) one may write the right-hand side as an integral. The key again to this result is (A.2), which expresses the nice homogeneity of T when provided with the Euclidean distance. However here the series in (A.13) has no reason to converge like a geometric series, so we already are being more sophisticated than in the case of the Kolmogorov conditions.

In the left-hand side of (A.13) we would like to do better than controlling the expectation, but one really needs some regularity of the function φ for this. It suffices here to say that when $\varphi(x) = |x|^p$ for $p \geq 1$ we may replace the expectation by the norm of L^p, proceeding exactly as we did in the case of the Kolmogorov conditions.

A.4 The Garsia-Rodemich-Rumsey Lemma

We all like to have at our disposal robust tools which will handle most situations in a mechanical manner, but the comfort they provide should not make us blind to their shortcomings. It may help to discuss one popular such tool here, the Garsia-Rodemich-Rumsey (GRR) lemma [35]. (The reader who has not used this lemma before may skip this discussion.) This lemma had some fundamental historical importance. It directly influenced Fernique's discovery of the "majorizing measures" approach to regularity of Gaussian processes, and this entire book in turn. Our discussion will be *very* informal. Let us quote a form of this lemma as in [106] p. 60. We use again the notation $T = [0,1]^m$, and we are again in the setting with "homogeneity". The point we are going to make is that (even in this simplest setting where there is "homogeneity of T") the method based on this form of the lemma, while effective in certain situations, does not even do as well as the most basic chaining method explained above.

Lemma A.4.1. *Consider two non-decreasing functions φ, ψ on $\mathbb{R}^+$ with $\varphi(0) = \psi(0) = 0$. Consider a continuous function u on T. Assume that*

$$J := \int_{T\times T} \varphi\Big(\frac{|u(x)-u(y)|}{\psi(d(x,y))}\Big)\mathrm{d}x\mathrm{d}y < \infty \,. \tag{A.14}$$

Then, for a constant K depending on m only and each $s, t \in T$,

$$|u(s)-u(t)| \leq K \int_0^{2d(s,t)} \varphi^{-1}\Big(\frac{KJ}{x^{2m}}\Big)\psi(\mathrm{d}x) \,. \tag{A.15}$$

The way this is used is by considering the function $u(x) = X_x$ for a given realization of the process, and the finiteness of J is obtained by assuming that

$$\mathsf{E}J = \int_{T\times T} \mathsf{E}\,\varphi\Big(\frac{|X_x - X_y|}{\psi(d(x,y))}\Big)\mathrm{d}x\mathrm{d}y < \infty \,.$$

The main drawback of Lemma A.4.1 is the power $2m$ in the denominator in the right-hand side. When studying Gaussian processes, where often φ^{-1} behaves like $\sqrt{\log}$, this does not matter (but in that case the GRR lemma does not give better results than chaining in the most straightforward fashion as we performed above). It matters when φ has a polynomial behavior. For example, assume that the process (X_t) satisfies

$$\forall s, t \in T \ , \ \mathsf{E}\,\varphi\Big(\frac{|X_s - X_t|}{\psi(d(s,t))}\Big) \le 1 \ , \tag{A.16}$$

so that we are in the situation (A.9) with $d_n = 1$ and $c_n = \psi(K(m)2^{-n})$. The process $(X_t)_{t\in T}$ is sample-continuous under the condition $\sum_n c_n \varphi^{-1}(2^{nm}) < \infty$, which is equivalent to

$$\int_0^1 \varphi^{-1}\Big(\frac{K(m)}{x^m}\Big)\psi(\mathrm{d}x) < \infty \ , \tag{A.17}$$

where x appears at the power m (rather than $2m$) in the denominator. We do not see how this could be deduced from Lemma A.4.1.

This being said, the principle behind the use of the GRR inequality is of fundamental importance (and is implicitly the key to the results of Section 13.5). In the situation here, to prove sample-boundedness of a process, the method is to find a suitable probability measure θ on $T \times T$ such that the finiteness of

$$\int_{T\times T} \varphi\Big(\frac{|u(x) - u(y)|}{\psi(d(x,y))}\Big)\mathrm{d}\theta(x,y) \tag{A.18}$$

implies the boundedness of $\sup_{x,y\in T} |u(x)-u(y)|$. It is explained in particular in [114] why this method is in a sense optimal. The point however is that to make full use of the method one must consider probabilities θ which are quite different from the uniform probability used in GRR. It is precisely because the special form of the conditions (1.2) allows just this that the GRR lemma can be used to prove (A.8) for these processes, as is done in [10], but this success seems accidental.

A.5 Chaining in a Metric Space

Suppose now that we want to study the uniform convergence on $[0,1]$ of a random Fourier series $X_t = \sum_{k\ge 1} a_k g_k \exp(2\pi i k t)$ where (g_k) are independent standard Gaussian r.v.s. Then the natural distance on $[0,1]$ is of little use, it is much more relevant to consider the distance given by

$$d(s,t)^2 = \mathsf{E}|X_s - X_t|^2 = \sum_k |a_k|^2 |\exp(2i\pi ks) - \exp(2i\pi kt)|^2 \ . \tag{A.19}$$

The space $(T = [0,1], d)$ is in a sense homogeneous, because d is translation invariant, but there is no reason that it looks in any way "finite dimensional".

Let us consider therefore an abstract situation, that of metric space (T, d) and of a process $(X_t)_{t\in T}$ which satisfies

$$\forall\, s,t \in T\ ,\ \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{d(s,t)}\Big) \le 1\ , \tag{A.20}$$

where φ is convex function with $\varphi(0) = 0, \varphi \ge 0$. This situation is canonical because we may define

$$d(s,t) = \inf\Big\{a > 0\ ;\ \mathsf{E}\varphi\Big(\frac{|X_s - X_t|}{a}\Big) \le 1\Big\}\ . \tag{A.21}$$

To find a substitute for the sets G_n of Section A.2 we use the *covering numbers*. For $\epsilon > 0$ we define the covering number $N(T, d, \epsilon)$ as the smallest integer N such that T can be covered by N balls of radius ϵ, or equivalently, such that there exists a set $V \subset T$ with $\operatorname{card} V \le N$ and such that each point of T is within distance ϵ of V. Let us denote by $\Delta(T)$ the diameter of T, defined as $\Delta(T) = \sup_{s,t\in T} d(s,t)$, and observe that $N(T, d, \Delta(T)) = 1$. Consider the largest integer n_0 with $\Delta(T) \le 2^{-n_0}$. For $n \ge n_0$ consider a set $T_n \subset T$ with $\operatorname{card} T_n = N(T, d, 2^{-n})$ such that each point of T is within distance 2^{-n} of a point of T_n. For each $x \in T$ and each $n \ge 0$ we consider a point $\pi_n(x) \in T_n$ with $d(x, \pi_n(x)) \le 2^{-n}$ and we consider

$$U_n = \{(s,t)\ ;\ s \in T_n\ ,\ t \in T_{n+1}\ ,\ d(s,t) \le 3\cdot 2^{-n}\}\ ,$$

so that

$$\operatorname{card} U_n \le \operatorname{card} T_n \operatorname{card} T_{n+1} \le \operatorname{card} T_{n+1}^2 = N(T, d, 2^{-n-1})^2\ .$$

This bound is crude, but we cannot do much better in general. It should be compared to (A.2). Using (A.11) the r.v.

$$Y_n = \max\{|X_s - X_t|\ ;\ (s,t) \in U_n\}$$

satisfies

$$\mathsf{E}Y_n \le 3\cdot 2^{-n}\varphi^{-1}(N(T, d, 2^{-n-1})^2)\ ,$$

and exactly as in the case of the Kolmogorov conditions we obtain

$$\mathsf{E}\sup_{d(s,t)\le 2^{-n}} |X_s - X_t| \le L\sum_{k\ge n} 2^{-k}\varphi^{-1}(N(T, d, 2^{-k-1})^2)\ ,$$

where L is a number, which one usually writes in the integral form

$$\mathsf{E}\sup_{d(s,t)\le\delta} |X_s - X_t| \le L\int_0^\delta \varphi^{-1}(N(T, d, \epsilon)^2)\mathrm{d}\epsilon\ . \tag{A.22}$$

This simple and general bound is also very handy. It is worth to state its consequences for Gaussian processes, that is, when the family $(X_t)_{t\in T}$ is jointly Gaussian centered. In that case one uses canonical distance d on the index set given by $d(s,t) = (\mathsf{E}(X_s - X_t)^2)^{1/2}$. The tail properties of Gaussian r.v.s then imply that (A.20) holds for the function $\varphi(x) = \exp(x^2/L) - 1$, in which case $\varphi^{-1}(x) = L\sqrt{\log(1+x)}$. Inequality (A.22) is then easily shown to be equivalent to the following more elegant formulation (Dudley's bound):

$$\mathsf{E} \sup_{d(s,t)\le\delta} |X_s - X_t| \le L \int_0^\delta \sqrt{\log N(T,d,\epsilon)}\mathrm{d}\epsilon \ . \tag{A.23}$$

This *very general* inequality is effective even in simple situations.

Exercise A.5.1. Prove that the previous bound gives the correct modulus of continuity for Brownian motion on $[0,1]$.

One drawback of (A.22) is that one would like to have an integrand $\varphi^{-1}(N(T,d,\epsilon))$ rather than $\varphi^{-1}(N(T,d,\epsilon)^2)$. Interestingly, it does not appear to be known whether this is true in general or not. This is known for $\delta = \Delta(T)$, a result which requires simply to define $\pi_n(t)$ in a smarter way. This is explained in the self-contained Section B.2, which we advise the reader to study next. Theorem B.2.4 there provides a partial result for smaller values of δ. The arguments of Section B.2 are more sophisticated than those of the present section but are still simple.

The main issue of the generic chaining is however in a *different direction.* Namely the covering numbers do not provide a fine enough description of a metric space, because they do not take in account the "local irregularities". Let us try to explain this in terms of the fundamental chaining identity:

$$X_t - X_{\pi_0(t)} = \sum_{n\ge 0} X_{\pi_{n+1}(t)} - X_{\pi_n(t)} \ . \tag{A.24}$$

When performing the chaining using covering numbers, one basically bounds the size of the term $X_{\pi_{n+1}(t)} - X_{\pi_n(t)}$ *independently of* t. The generic chaining bounds allows the size of this term *to depend on* t, a subtle but sometimes crucial difference.

A.6 Two Classical Inequalities

Consider standard Brownian motion $(B_t)_{t\ge 0}$ on $\mathbb{R}^+$ and let

$$B_t^* = \sup_{0\le s\le t} |B_s| \ . \tag{A.25}$$

Consider a r.v. X, which may or may not be a stopping time. Classical inequalities estimate from above the moments $\|B_X^*\|_p$ for $p \ge 1$, and we would like to briefly discuss such inequalities from our abstract point of view, as an illustration of what it might lead to. Let us hurry to say that the results of this investigation are not spectacular. In some sense there is "considerable room" in these classical inequalities, and the power of the modern methods is certainly better demonstrated in tighter situations. Still, the "abstract" approach does bring a different light, for better or for worse.

When one tries to understand what makes a result about Brownian motion "really work" there are usually multiple answers, because Brownian motion has so many facets. Our discussion largely follows the paper [10]. Our emphasis here is to show that obtaining very precise inequalities does not require undue work (provided one stays away from the GRR inequality), so we assume more than in [10], namely we assume

$$\forall\, p \ge 1\;,\; \forall\, s,t\;,\; 0 \le s \le t\;,\; \|B_t - B_s\|_p \le \sqrt{t-s}\sqrt{p}\;. \tag{A.26}$$

A basic fact is then as follows.

Lemma A.6.1. *If the process $(B_t)_{t\ge 0}$ satisfies (A.26) and $B_0 = 0$ then for each $p \ge 1$ and each $a > 0$ we have, using the notation (A.25), and denoting by L a universal constant,*

$$\|B_a^*\|_p \le L\sqrt{p}\sqrt{a}\;. \tag{A.27}$$

Proof. This is chaining of the most brutish type, which actually copies the case of the Kolmogorov conditions. We write

$$B_a^* \le \sum_{n\ge 0} Y_n\;, \tag{A.28}$$

where

$$Y_n = \sup_{0\le k<2^n} |B_{k2^{-n}a} - B_{(k+1)2^{-n}a}|\;,$$

and

$$\|B_{k2^{-n}a} - B_{(k+1)2^{-n}a}\|_p \le L\sqrt{p}\sqrt{a}2^{-n/2}\;,$$

from (A.26). Using (A.4), we obtain $\|Y_n\|_p \le L\sqrt{p}\sqrt{a}2^{n/p-n/2}$. We then use (A.28) and the triangle inequality in L^p to obtain (A.27) when $p \ge 4$, from which the general case follows. □

Theorem A.6.2. *Consider a non-decreasing process (B_t^*) on $\mathbb{R}^+$ and assume that for each $p \ge 1$,*

$$\forall\, t > 0\;,\; \|B_t^*\|_p \le \sqrt{tp}\;. \tag{A.29}$$

Consider a r.v $X \ge 0$, a scaling factor $a \ge 0$ and the sets

$$A_0 = \{X \le a\}\;;\; \text{for } n \ge 1\;,\; A_n = \{a2^{n-1} < X \le a2^n\}\;. \tag{A.30}$$

Then for each $p \ge 1$ we have

$$\|B_X^*\|_p \le L\sqrt{pa}\Big(\sum_{n\ge 0} 2^{np/2}\mathsf{P}(A_n)\Big(\log\frac{e}{\mathsf{P}(A_n)}\Big)^{p/2}\Big)^{1/p}\;. \tag{A.31}$$

Combining with Lemma A.6.1 this yields precise information on the process (A.25). There is of course no simple expression of the right-hand side in general, but let us give some consequences of this bound. It is natural to assume that a has been chosen such that $\mathsf{P}(A_0) \ge 1/2$. Let us write $I = \{n \ge 1; \mathsf{P}(A_n) \ge e^{-np}\}$, so that

$$\sum_{n\ge 1} 2^{np/2}\mathsf{P}(A_n)\Big(\log \frac{e}{\mathsf{P}(A_n)}\Big)^{p/2} = \sum_{n\in I} + \sum_{n\notin I} := \mathrm{I} + \mathrm{II}\,.$$

For $n \in I$ we have $\log(e/\mathsf{P}(A_n)) \le Lnp$, so that

$$\mathrm{I} \le L^p p^{p/2} \sum_{n\ge 1} 2^{np/2} n^{p/2} \mathsf{P}(A_n)\,.$$

Also, the function $x \mapsto x(\log(e/x))^p$ increases for $x \le e^{-p}$ so that

$$\mathrm{II} \le \sum_{n\ge 1} e^{-np}(Lnp)^{p/2} \le (Lp)^{p/2}\,.$$

We then obtain the inequality

$$\|B_X^*\|_p \le L\sqrt{ap} + Lp\sqrt{a}\Big(\sum_{n\ge 1} 2^{np/2} n^{p/2}\mathsf{P}(A_n)\Big)^{1/p}\,,$$

and thus, using the notation $\log^+ x = \max(\log x, 0)$,

$$\|B_X^*\|_p \le L\sqrt{ap} + Lp\sqrt{a}\Big\|\Big(\frac{X}{a}\log^+\frac{X}{a}\Big)^{1/2}\Big\|_p\,.$$

Lemma A.6.3. *If a r.v. Y satisfies $\|Y\|_p \le \sqrt{p}$ for each $p \ge 2$, then for each event A we have*

$$\|Y\mathbf{1}_A\|_p \le L\sqrt{p}\,\mathsf{P}(A)^{1/p}\sqrt{\log\frac{e}{\mathsf{P}(A)}}\,. \tag{A.32}$$

Proof. Consider conjugate exponents a, b. Then Hölder's inequality yields

$$\mathsf{E}|Y|^p\mathbf{1}_A \le \||Y|^p\|_a\|\mathbf{1}_A\|_b = \|Y\|_{ap}^p\|\mathbf{1}_A\|_b\,,$$

so that

$$\|Y\mathbf{1}_A\|_p \le \|Y\|_{ap}\mathsf{P}(A)^{1/pb} \le \sqrt{ap}\mathsf{P}(A)^{1/pb}\,.$$

Now

$$\mathsf{P}(A)^{1/pb} = \mathsf{P}(A)^{1/p}\mathsf{P}(A)^{-1/pa}\,,$$

and the second term is $\le L$ for $a = \log(e/\mathsf{P}(A)) \ge 1$. □

Proof of Theorem A.6.2. The key point is the inequality

$$B^*_X \le \sum_{n\ge 0} B^*_{a2^n}\mathbf{1}_{A_n} ,$$

an obvious consequence of the fact that the process (B^*_t) increases. Since the sets A_n are disjoint,

$$\| \sum_{n\ge 0} B^*_{a2^n}\mathbf{1}_{A_n}\|_p^p = \sum_{n\ge 0} \|B^*_{a2^n}\mathbf{1}_{A_n}\|_p^p .$$

Bounding the terms in the right-hand side using the hypothesis (A.29) and (A.32) then concludes the proof. □

The methods we used are very general, and there is no doubt that they apply to more general "functionals of Brownian motion" than the supremum.

Assume now that X is a stopping time, and that (B^*_t) is as in (A.25), where (B_t) is Brownian motion. In that case the Burkholder-Davis-Gundy inequality states that

$$\|B^*_X\|_p \le L\sqrt{p}\|X\|_p . \tag{A.33}$$

The usual proofs of this inequality heavily rely on the martingale property of Brownian motion. However, as we show now, (A.33) also holds for processes that satisfy a certain condition on their increments, *irrespective* of the martingale property.

Theorem A.6.4. *Consider a process* $(Y_t)_{t\in\mathbb{R}^+}$ *with* $Y_0 = 0$*, and a r.v.* $X \ge 0$*. Assume that for each* $s > 0$ *the following two properties hold:*

$$\mathsf{P}(\exists t \ge s, Y_t \ne Y_s) \le \mathsf{P}(X \ge s) , \tag{A.34}$$

$$\forall\, u > 0 \,,\ \forall\, t > s \,,\ \mathsf{P}(|Y_t - Y_s| \ge u) \le 2\mathsf{P}(X \ge s)\exp\Big(-\frac{u^2}{2(t-s)}\Big) . \tag{A.35}$$

Then, for each $p \ge 1$ *we have*

$$\Big\|\sup_{t\ge 0} |Y_t|\Big\|_p \le L\sqrt{p}\|X\|_p . \tag{A.36}$$

The relevance of this result to (A.33) is that, if (B_t) denotes standard Gaussian motion and X is a stopping time, the process $Y_t = B_{\min(t,X)}$ satisfies the hypotheses of Theorem A.6.4. (Unfortunately, one does not expect to find many other natural examples of processes which satisfy these hypotheses.) Theorem A.6.4 is a simple consequence of the following fact, interesting in its own right.

Lemma A.6.5. *Consider for* $n \ge 0$ *r.v.s* $Z_n \ge 0$*, and assume that for numbers* $b_n \ge 0$ *and each* $u > 0$ *we have*

$$\mathsf{P}(Z_n \ge u) \le b_n \exp\Big(-\frac{u^2}{2^n}\Big) . \tag{A.37}$$

Then for each $p \ge 1$ *we have*

$$\Big\|\sum_{n\ge 0} Z_n\Big\|_p \le L\sqrt{p}\Big(1 + \sum_{n\ge 1} b_n 2^{np/2}\Big)^{1/p} . \tag{A.38}$$

Proof of Theorem A.6.4. Consider a scaling factor $a > 0$ and for $n \ge 0$ the r.v.s

$$W_0 = \sup_{t\le a} |Y_t| \ ; \text{ for } n \ge 1 \ , \ W_n = \sup_{a2^{n-1}\le t\le a2^n} |Y_t - Y_{a2^{n-1}}| \ , \tag{A.39}$$

and observe that obviously

$$\sup_{t\ge 0} |Y_t| \le \sum_{n\ge 0} W_n . \tag{A.40}$$

The key property of the r.v.s W_n is

$$\forall\, u > 0 \ , \ \mathsf{P}(W_n \ge u) \le b_n \exp\Big(-\frac{u^2}{La2^n}\Big) , \tag{A.41}$$

where $b_0 = 2$ and $b_n = 2\mathsf{P}(X \ge a2^{n-1})$ for $n \ge 1$. This is proved from (A.34) and (A.35) using a chaining argument about as straightforward as that of Lemma A.6.1. As this type of chaining argument is developed right at the beginning of Chapter 2 we do not reproduce the details here. Use of (A.38) and (A.40) then yields

$$\Big\|\sup_{t\ge 0} |Y_t|\Big\|_p \le L\sqrt{p}\sqrt{a}\Big(1 + \sum_{n\ge 0} b_n 2^{np/2}\Big)^{1/p} .$$

Now

$$\sum_{n\ge 0} b_n 2^{np/2} \le 2\sum_{n\ge 0} 2^{np/2}\mathsf{P}(X \ge a2^{n-1}) \le L^p\mathsf{E}(X/a)^{p/2} .$$

To prove (A.36) we choose a small enough that $1 \le 2\mathsf{E}(X/a)^{p/2}$. □

Proof of Lemma A.6.5. The plan is to bound $\mathsf{E}(\sum_n Z_n)^p$ from above. First, we observe that for a r.v $Z \ge 0$,

$$\mathsf{E}Z^p = p\int_0^\infty t^{p-1}\mathsf{P}(Z \ge t)\mathrm{d}t \le 1 + \sum_{k\ge 1} 2^{(k+1)p}\mathsf{P}(Z \ge 2^k) \ , \tag{A.42}$$

simply by writing that $\mathsf{P}(Z \ge t) \le \mathsf{P}(Z \ge 2^k)$ for $2^k \le t \le 2^{k+1}$. Next, consider numbers $\alpha(n,k) \ge 0$ and assume that

$$\forall\, k \geq 1\ , \ \sum_{n\geq 0} \alpha(n,k) \leq 1\ , \tag{A.43}$$

so that, using (A.37) in the second inequality,

$$\begin{aligned} \mathsf{P}\Big(\sum_{n\geq 0} Z_n \geq 2^k\Big) &\leq \sum_{n\geq 0} \mathsf{P}(Z_n \geq \alpha(n,k)2^k) \\ &\leq \sum_{n\geq 0} b_n \exp\Big(-\frac{\alpha(n,k)^2 2^{2k}}{2^n}\Big)\ . \end{aligned} \tag{A.44}$$

Combining with (A.42) and exchanging the order of summation we obtain

$$\mathsf{E}\Big(\sum_{n\geq 0} Z_n\Big)^p \leq 1 + \sum_{n\geq 0} b_n \sum_{k\geq 1} 2^{(k+1)p} \exp\Big(-\frac{\alpha(n,k)^2 2^{2k}}{2^n}\Big)\ . \tag{A.45}$$

Now we choose the numbers $\alpha(n,k)$ in order to control these latter sums. Define c_n by $2^{2c_n} = p2^n$, and define

$$\alpha(n,k) = 0 \text{ if } k \leq c_n + 100\ ;\ \alpha(n,k) = 2^{c_n - k}\sqrt{k - c_n} \text{ if } k > c_n + 100\ .$$

We observe that $c_{n-\ell} = c_n - \ell/2$, so that given k, and denoting by n^* the largest integer n for which $c_n + 100 \leq k$, we get

$$\sum_{n\geq 0} \alpha(n,k) \leq \sum_{\ell\geq 0} 2^{c_{n^*} - k - \ell/2}\sqrt{k - c_{n^*} + \ell/2} \leq 1\ ,$$

and this proves (A.43). Moreover, recalling that $2^{2c_n} = p2^n$,

$$\begin{aligned} &\sum_{k > c_n + 100} 2^{(k+1)p} \exp\Big(-\frac{\alpha(n,k)^2 2^{2k}}{2^n}\Big) \\ &= \sum_{k > c_n + 100} 2^{(k+1)p} \exp(-p(k - c_n)) \\ &= 2^{(c_n+1)p} \sum_{k > c_n + 100} 2^{p(k-c_n)} \exp(-p(k - c_n)) \\ &\leq L 2^{p(c_n+1)}\ . \end{aligned} \tag{A.46}$$

Also,

$$\sum_{k \leq c_n + 100} 2^{(k+1)p} \exp\Big(-\frac{\alpha(n,k)^2 2^{2k}}{2^n}\Big) \leq \sum_{k \leq c_n + 100} 2^{(k+1)p} \leq 2 \cdot 2^{p(c_n + 100)}\ ,$$

and using (A.45) and since $2^{2c_n} = p2^n$ we have proved that

$$\mathsf{E}\Big(\sum_{n\geq 0} Z_n\Big)^p \leq 1 + \sum_{n\geq 0} b_n (Lp2^n)^{p/2}\ . \qquad \square$$

B. Appendix: Continuity

B.1 Introduction

When trying to prove "regularity" of a stochastic process, the most difficult task is to prove boundedness. For this reason the main body of the book contains no results about continuity. Mainstream probabilists are however adamant about the need of studying continuity of stochastic processes. We shall give a few typical results in this direction. We do not strive for maximum generality or sharpness. Rather, we show what can be done with our methods, and we try to state results which can conceivably be useful as such. The next section requires only Section A.5 as a prerequisite, while Section B.3 requires the main results of Chapter 2.

B.2 Continuity Under Metric Entropy Conditions

Consider a Young function φ, that is, a convex function with $\varphi(0) = 0$, $\varphi(x) = \varphi(-x) \geq 0$. Given a metric space (T, d) we are interested in the continuity of processes $(X_t)_{t\in T}$ that satisfy condition (A.20), i.e.

$$\forall\, s, t \in T\ ,\ \mathsf{E}\varphi\Big(\frac{X_s - X_t}{d(s,t)}\Big) \leq 1\ . \tag{A.20}$$

Recalling (A.21) we may also write this condition as $\|X_s - X_t\|_\varphi \leq d(s,t)$. The case of Lemma B.3.1 below is essentially the case where $\varphi(x) = \exp(x^2) - 1$. This theorem generalizes easily to the case where instead $\varphi(x) = \exp|x|^\alpha - 1$, or to more general functions "with exponential growth." What we have in mind in the present section is the more difficult case of "polynomial growth", e.g. $\varphi(x) = |x|^p$ for $p > 1$. For the convenience of the reader this section uses only Section A.5 as a prerequisite, but let us point out that considerably more elaborate arguments (in a somewhat different direction) are given in Chapter 13. Since we know only very few natural situations where these sophisticated arguments are needed, we shall only consider here "metric entropy conditions". Let us repeat the bound (A.22) for convenience:

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*,
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of
Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2,

$$\mathsf{E} \sup_{d(s,t)\le\delta} |X_s - X_t| \le L \int_0^\delta \varphi^{-1}(N(T,d,\epsilon)^2)\mathrm{d}\epsilon \ . \tag{A.22}$$

The problem with this bound is the occurrence of $N(T,d,\epsilon)^2$ rather than $N(T,d,\epsilon)$.

Research problem B.2.1. Is it true in full generality that

$$\mathsf{E} \sup_{d(s,t)\le\delta} |X_s - X_t| \le L \int_0^\delta \varphi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon \ ? \tag{B.1}$$

And if this fails, what is the worst possible value of the left-hand side given φ and the numbers $N(T,d,\epsilon)$?

It seems very unlikely that (B.1) holds in general. In particular, it one further assumes that (say) $d(s,t) \ge \delta/10$ for $s,t \in T$, $s \ne t$, then (B.1) would imply that $\mathsf{E}\sup_{d(s,t)\le\delta} |X_s - X_t| \le L\delta\varphi^{-1}(N(T,d,\delta/10))$. This does not seem to be true, but examples are hard to construct. The bound (B.10) below, which holds under mild regularity conditions on φ is not very far from (B.1). It is even conceivable that this bound is optimal in some sense.

The chaining argument that leads to (A.22) (and which is given in Section A.5) is really brutal, and our first goal is to improve it. Without loss of generality we assume that T is finite. We consider the largest integer n_0 such that $\Delta(T) \le 2^{-n_0}$ and for $n \ge n_0$ we consider a subset T_n of T with $\operatorname{card} T_n = N(T,d,2^{-n})$ such that each point of T is within distance $\le 2^{-n}$ of a point of T_n. (So that T_{n_0} consists of a unique point t_0.) For $n \ge n_0$ we consider a map $\theta_n : T_{n+1} \to T_n$ such that $d(\theta_n(t), t) \le 2^{-n}$ for each $t \in T_n$. Since we assume that T is finite, we have $T = T_m$ when m is large enough. We fix such an m, and we define $\pi_n(t) = t$ for each $t \in T$ and each $n \ge m$. Starting with $n = m$ we then define recursively $\pi_n(t) = \theta_n(\pi_{n+1}(t))$ for $n \ge n_0$. The point of this construction is that $\pi_{n+1}(t)$ determines $\pi_n(t)$ so that there are at most $N(T,d,2^{-n-1})$ pairs $(\pi_{n+1}(t), \pi_n(t))$, and the bound (A.11) implies

$$\mathsf{E}\sup_{t\in T} |X_{\pi_{n+1}(t)} - X_{\pi_n(t)}| \le 2^{-n}\varphi^{-1}(N(T,d,2^{-n-1})) \ . \tag{B.2}$$

Using the chaining identity

$$X_t - X_{\pi_n(t)} = \sum_{k\ge n} X_{\pi_{k+1}(t)} - X_{\pi_k(t)} \ ,$$

we have proved the following.

Lemma B.2.2. *We have*

$$\mathsf{E}\sup_{t\in T} |X_t - X_{\pi_n(t)}| \le \sum_{k\ge n} 2^{-k}\varphi^{-1}(N(T,d,2^{-k-1})) \ . \tag{B.3}$$

Taking $n = n_0$ this yields the following:

Theorem B.2.3. *We have*

$$\mathsf{E}\sup_{s,t\in T}|X_s - X_t| \le L\int_0^{\Delta(T)} \varphi^{-1}(N(t,d,\epsilon))\mathrm{d}\epsilon \;. \tag{B.4}$$

Since $d(\pi_{n+1}(t),\pi_n(t)) \le 2^{-n}$ by construction, it follows by decreasing induction over n that $d(t,\pi_n(t)) \le 2^{-n+1}$. Therefore

$$d(\pi_n(s),\pi_n(t)) \le 2^{-n+2} + d(s,t) \;,$$

and consequently

$$\mathsf{E}\sup_{d(s,t)\le 2^{-n+2}}|X_s - X_t| \le \mathsf{E}\sup_{s,t\in T_n, d(s,t)\le 2^{-n+3}}|X_s - X_t| + 2\mathsf{E}\sup_{t\in T}|X_t - X_{\pi_n(t)}| \;. \tag{B.5}$$

We know from (B.3) how to control the second term, but the first one is challenging. One may crudely apply (A.11) to this term, and (choosing of course for n the largest with $2^{-n+2}\ge\delta$, so that $2^{-n}\ge\delta/4$) we obtain that for any $\delta>0$,

$$\mathsf{E}\sup_{d(s,t)\le\delta}|X_s - X_t| \le L\delta\varphi^{-1}(N(T,d,\delta/4)^2) + L\int_0^{\delta}\varphi^{-1}(N(t,d,\epsilon))\mathrm{d}\epsilon \;. \tag{B.6}$$

The problem with the bound (B.6) is of course that there is no reason why the term $\delta\varphi^{-1}(N(t,d,\delta/4)^2)$ should be small, even if the integral in the right-hand side converges. We shall prove the following improvement of (B.6).

Theorem B.2.4. *For each $\delta>0$, each number $2\le A\le N(T,d,\delta/4)+1$ and each process $(X_t)_{t\in T}$ which satisfies (13.125) one has*

$$\mathsf{E}\sup_{d(s,t)\le\delta}|X_s - X_t| \le L\delta\frac{\log(N(T,d,\delta/4))}{\log A}\varphi^{-1}(AN(T,d,\delta/4)) + L\int_0^{\delta}\varphi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon \;. \tag{B.7}$$

It is known in full generality that processes satisfying (A.20) are automatically sample-continuous when the entropy integral is finite in (B.4). The known proof is unsavory, and we refer to [58] for it. In the usual cases, (B.7) provides a cleaner approach to this result.

Corollary B.2.5 (Continuity of processes under the metric entropy condition). *Let us assume that the function φ as in Theorem B.2.4 satisfies the following regularity condition for N large enough, where C is a number,*

$$\inf_{2\le A\le N}\frac{\log N}{\log A}\varphi^{-1}(AN) \le C\varphi^{-1}(N)\log\varphi^{-1}(N) \;. \tag{B.8}$$

Then for each process $(X_t)_{t\in T}$ *on a metric space, which satisfies (A.20) and for each* $\alpha > 0$ *there exists a number* δ *for which*

$$\mathsf{E} \sup_{d(s,t)\le\delta} |X_s - X_t| \le \alpha \ . \tag{B.9}$$

For example, (B.8) is satisfied when $\varphi(x) = |x|^p$, as is shown by the choice $A = 2$. In Proposition B.2.6 below we prove that actually (B.8) always hold under minimum regularity conditions.

Proof. Combining (B.8) and (B.7) we get

$$\begin{aligned}\mathsf{E} \sup_{d(s,t)\le\delta} |X_s - X_t| \le L\delta\varphi^{-1}(N(T,d,\delta/4)) \log\big(\varphi^{-1}(N(T,d,\delta/4))\big) \\ + L\int_0^\delta \varphi^{-1}(N(T,d,\epsilon))\mathrm{d}\epsilon \ . \end{aligned} \tag{B.10}$$

Let $f(\epsilon) = \varphi^{-1}(N(T,d,\epsilon))$. The integral $\int_0^{\Delta(T)} f(\epsilon)\mathrm{d}\epsilon$ converges, so that given $0 < a < 1$ there are arbitrarily small values of ϵ for which $f(\epsilon) \le a/(\epsilon\log(1/\epsilon))$. Then $f(\epsilon) \le 1/\epsilon$ so that $\epsilon f(\epsilon)\log(f(\epsilon)) \le a$ is arbitrarily small. Thus there exist arbitrarily small values of δ for which the first term on the right-hand side of (B.10) is arbitrarily small. □

Proposition B.2.6. *Assume that the function*

$$f(x) := \log\varphi^{-1}(\exp x)$$

is concave. Then (B.8) holds.

Since it is reasonable to assume that $\varphi(x) \ge x$ and hence that $f(x) \le x$, the hypothesis that f is concave is most reasonable.

Proof. Letting $N = \exp U$ and $A = \exp B$ we have to prove that for U large enough we have

$$\inf_{1\le B\le U} \frac{U}{B} \exp f(U+B) \le Cf(U)\exp f(U) \ . \tag{B.11}$$

Now, since f is concave, we have $f(U+B) \le f(U) + f'(U)B$ and $Uf'(U) \le f(U) - f(0)$, so that for large U we have $f(U+B) \le f(U) + 2Bf(U)/U$ and the left-hand side of (B.11) is at most

$$\inf_{1\le B\le U} \frac{U}{B} \exp(f(U) + 2Bf(U)/U) \ ,$$

so that (B.11) follows by taking $B = \max(1, U/f(U))$. □

We turn to the proof of Theorem B.2.4.

Lemma B.2.7. *Consider a number* $A \geq 1$ *and an integer* p *with* $A^p \geq \operatorname{card} T_n$. *Given* n *and* $c > 0$ *there exists a set* $U_n \subset T_n^2$ *with the following properties*

$$\operatorname{card} U_n \leq A \operatorname{card} T_n \ , \tag{B.12}$$

$$(s,t) \in U_n \Rightarrow d(s,t) \leq Lcp \ . \tag{B.13}$$

$$\sup_{s,t \in T_n, d(s,t) \leq c} |X_s - X_t| \leq 2 \sup_{(s,t) \in U_n} |X_s - X_t| \ . \tag{B.14}$$

Proof. We perform the following construction of a decreasing sequence $(V_\ell)_{\ell \geq 1}$ of subsets of T_n and of a sequence $(r_\ell)_{\ell \geq 1}$ of integers, $1 \leq r_\ell \leq p$. We first set $V_1 = T_n$ and choose any $t_1 \in V_1$. Since $\operatorname{card} V_1 \leq \operatorname{card} T_n \leq A^p$ there exists an integer $r \leq p$ for which

$$\operatorname{card}\{s \in V_1 \ ; \ d(t_1, s) \leq rc\} \leq A^r \ , \tag{B.15}$$

because $r = p$ satisfies this conditions. We consider the smallest integer $r_1 \geq 1$ which satisfies (B.15) and we define

$$B_1 = \{s \in T_n \ ; \ d(s, t_1) \leq (r_1 - 1)c\} \ .$$

We observe that

$$\operatorname{card} B_1 \geq A^{r_1 - 1} \ .$$

If $r_1 > 1$ this follows from the definition of r_1, while if $r_1 = 1$ this is obvious because $\operatorname{card} B_1 \geq 1$ since $t_1 \in B_1$. We then set $V_2 = T_n \setminus B_1 = V_1 \setminus B_1$, and we continue this procedure. We pick any point $t_2 \in V_2$ and we consider the smallest integer $1 \leq r_2 \leq p$ for which

$$\operatorname{card}\{s \in V_2 \ ; \ d(t_2, s) \leq r_2 c\} \leq A^{r_2} \ ,$$

so that as before

$$\operatorname{card} B_2 \geq A^{r_2 - 1} \ .$$

We then set

$$B_2 = \{s \in V_2 \ , \ d(s, t_2) \leq (r_2 - 1)c\} \ . \tag{B.16}$$

We define $V_3 = V_2 \setminus B_2$ and we continue in this manner until T_n is exhausted. Thus, for $\ell \geq 1$ we have

$$\operatorname{card} B_\ell \geq A^{r_\ell - 1} \tag{B.17}$$

and

$$\operatorname{card}\{s \in V_\ell \ ; \ d(t_\ell, s) \leq r_\ell c\} \leq A^{r_\ell} \ . \tag{B.18}$$

Since the sets B_ℓ are disjoint, (B.17) implies

$$\sum_{\ell \geq 1} A^{r_\ell - 1} \leq \sum_{\ell \geq 1} \operatorname{card} B_\ell \leq \operatorname{card} T_n \ ,$$

and thus
$$\sum_{\ell \geq 1} A^{r_\ell} \leq A \operatorname{card} T_n \ . \tag{B.19}$$

We define U_n as the set of pairs (t_ℓ, s) for $\ell \geq 1$ with $s \in V_\ell$ and $d(t_\ell, s) \leq c r_\ell$. Thus (B.18) and (B.19) imply (B.12), while (B.13) holds since $r_\ell \leq p$. It remains only to prove (B.14). Consider $s, t \in T_n$ with $d(s,t) \leq c$ and let ℓ be the largest integer such that both s and t belong to V_ℓ. For definitiveness, assume that $s \notin V_{\ell+1}$ so that since $V_{\ell+1} = V_\ell \setminus B_\ell$ we have $s \in B_\ell$ i.e. $d(t_\ell, s) \leq (r_\ell - 1)c$, and since $d(s,t) \leq c$ we have $d(t_\ell, t) \leq r_\ell c$ and therefore by definition of U_n we have $(t_\ell, s) \in U_n$ and $(t_\ell, t) \in U_n$. Moreover, $|X_s - X_t| \leq |X_s - X_{t_\ell}| + |X_t - X_{t_\ell}|$. This proves (B.14) and completes the proof. □

Let us also observe that the smallest integer p with $A^p \leq \operatorname{card} T_n$ satisfies $A^{p-1} \geq \operatorname{card} T_n$ so that, when $A \leq 1 + \operatorname{card} T_n$ and $\operatorname{card} T_n \geq 2$,
$$p \leq 1 + \frac{\log \operatorname{card} T_n}{\log A} \leq L \frac{\log \operatorname{card} T_n}{\log A} \ . \tag{B.20}$$

Proof of Theorem B.2.4. We may assume that $N(T, d, \delta/2) \geq 2$, for otherwise $\Delta(T) \leq \delta$ and the result follows (B.4). Consider the largest integer n with $\delta \leq 2^{-n+2}$, so that $\delta/4 \leq 2^{-n} \leq \delta/2$, and in particular $2 \leq \operatorname{card} T_n = N(T, d, 2^{-n}) \leq N(T, d, \delta/4)$. We use Lemma B.2.7 for this value of n and $c = 2^{-n+3}$, and the smallest possible value of p, which satisfies (B.20) since $\operatorname{card} T_n \geq N(T, d, \delta/2) \geq 2$. We use (A.11) to obtain, using also (B.12) and (B.13),
$$\mathsf{E} \sup_{(s,t) \in U_n} |X_s - X_t| \leq L 2^{-n} p \varphi^{-1}(A N(T, d, 2^{-n})) \ .$$

Since $\delta/4 \leq 2^{-n} \leq \delta/2$, it follows from (B.14) that
$$\mathsf{E} \sup_{s,t \in T, d(s,t) \leq 2^{-n+3}} |X_s - X_t| \leq L \delta p \varphi^{-1}(A N(T, d, \delta/4)) \ .$$

On the other hand, from (B.3),
$$\mathsf{E} \sup_{t \in T} |X_t - X_{\pi_n(t)}| \leq \sum_{k \geq n} 2^{-k} \varphi^{-1}(N(T, d, 2^{-k-1})) \leq L \int_0^\delta \varphi^{-1}(N(T, d, \epsilon)) \mathrm{d}\epsilon \ .$$

We combine with (B.5) and (B.20) to conclude the proof. □

The second part of the following exercise assumes that you know well the material of Chapter 13.

Exercise B.2.8. When $\varphi(x) = x^p$ for some $p > 1$, improve (B.7) by replacing the left-hand side by
$$\Big(\mathsf{E} \sup_{s,t \in T, d(s,t) \leq c_k} |X_s - X_t|^p \Big)^{1/p} \ .$$

Generalize this fact in the spirit of Proposition 13.5.16.

B.3 Continuity of Gaussian Processes

It is worth insisting that *by far* the most important result concerning continuity of Gaussian processes is Dudley's bound (A.23). However since the finiteness of the right hand side of (A.23) is not necessary for the Gaussian process to be continuous, there are situations where this bound is not appropriate. The purpose of the present section is to show that a suitable form of the generic chaining allows to capture the exact modulus of continuity of a Gaussian process with respect to the canonical distance in full generality. Not surprisingly, the modulus of continuity is closely related to the rate at which the series $\sum_n 2^{n/2}\Delta(A_n(t))$ converges uniformly on T for a suitable admissible sequence $(\mathcal{A}_n)$. Our first result shows how to obtain a modulus of continuity using the generic chaining. The idea of the proof is simply to use (B.5), and to evaluate the last term using the generic chaining rather than plain chaining.

Lemma B.3.1. *Consider a metric space (T,d) and a process $(X_t)_{t\in T}$ which satisfies the increment condition (1.4):*

$$\forall u>0\ ,\ \mathsf{P}(|X_s-X_t|\geq u)\leq 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right)\ . \tag{1.4}$$

Assume that there exists a sequence (T_n) of subsets of T with $\operatorname{card} T_n\leq N_n$ such that for certain integer m, and a certain number B one has

$$\sup_{t\in T}\sum_{n\geq m}2^{n/2}d(t,T_n)\leq B\ . \tag{B.21}$$

Consider $\delta>0$. Then, for a number $u\geq 1$, with probability $\geq 1-\exp(-u^2 2^m)$ we have

$$\forall\, s,t\in T\ ,\ d(s,t)\leq\delta\Rightarrow |X_s-X_t|\leq Lu(2^{m/2}\delta+B)\ . \tag{B.22}$$

Proof. We assume T finite for simplicity. For $n\geq m$ and $t\in T$ denote by $\pi_n(t)$ an element of T_n such that $d(t,\pi_n(t))=d(t,T_n)$. Consider the event $\Omega(u)$ defined by

$$\forall\, n\geq m+1\ ,\ \forall\, t\in T_n\ ,\ |X_{\pi_{n-1}(t)}-X_{\pi_n(t)}|\leq Lu2^{n/2}d(\pi_{n-1}(t),\pi_n(t))\ ,$$

and

$$\forall\, s',t'\in T_m\ ,\ |X_{s'}-X_{t'}|\leq Lud(s',t')2^{m/2}\ . \tag{B.23}$$

Then, as usual, we have $\mathsf{P}(\Omega(u))\geq 1-\exp(-u^2 2^m)$. Now, when $\Omega(u)$ occurs, for any $t\in T$ and any $k\geq 0$, using chaining as usual and (B.21) we get

$$|X_t-X_{\pi_m(t)}|\leq LuB\ . \tag{B.24}$$

Moreover, using (B.21) again,

$$d(t, \pi_m(t)) \le d(t, T_m) \le B2^{-m/2} ,$$

so that, using (B.24),

$$\begin{aligned} d(s,t) \le \delta \Rightarrow d(\pi_m(s), \pi_m(t)) &\le \delta + 2B2^{-m/2} \\ \Rightarrow |X_{\pi_m(s)} - X_{\pi_m(t)}| &\le Lu(\delta 2^{m/2} + B) . \end{aligned}$$

Combining with (B.24) proves that $|X_s - X_t| \le Lu(\delta 2^{m/2} + B)$ and completes the proof. □

Exercise B.3.2. Deduce Dudley's bound (A.23) from Lemma B.3.1

We now turn to our main result, which allows to exactly describe the modulus of continuity of a Gaussian process in term of certain admissible sequences. It implies in particular the remarkable fact (discovered by X. Fernique) that for Gaussian processes the "local modulus of continuity" (as in (B.25)) is also "global".

Theorem B.3.3. *Consider a Gaussian process $(X_t)_{t\in T}$, with canonical associated distance d given by (0.1). Assume that $S = \mathsf{E}\sup_t X_t < \infty$. For $k \ge 1$ consider $\delta_k > 0$ and assume that*

$$\forall\, t \in T\ ;\ \mathsf{E} \sup_{\{s\in T; d(s,t)\le \delta_k\}} |X_s - X_t| \le 2^{-k} S . \tag{B.25}$$

Let $n_0 = 0$ and for $k \ge 1$ consider an integer n_k for which

$$L_1 S 2^{-n_k/2-k} \le \delta_k . \tag{B.26}$$

Then we can find an admissible sequence $(\mathcal{A}_n)$ of partitions of T such that

$$\forall\, k \ge 0\ ;\ \sup_{t\in T} \sum_{n\ge n_k} 2^{n/2} \Delta(A_n(t)) \le LS2^{-k} . \tag{B.27}$$

Conversely, given an admissible sequence $(\mathcal{A}_n)$ as in (B.27), and defining now $\delta_k^* = S2^{-n_k/2-k}$, with probability $\ge 1 - \exp(-u^2)$ we have

$$\sup_{\{s,t\in T; d(s,t)\le \delta_k^*\}} |X_s - X_t| \le Lu2^{-k} S . \tag{B.28}$$

The abstract formulation here might make it hard at first to grab the power of the statement. The numbers δ_k describe the (uniform) modulus of continuity of the process. The numbers n_k describe the uniform convergence (over t) of the series $\sum_{n\ge 0} 2^{n/2}\Delta(A_n(t))$. Both are related by the relation $\delta_k \sim S2^{-n_k/2-k}$. The first part of the theorem assumes only the "local" modulus of continuity (B.25), while the converse provides a uniform modulus of continuity (B.28).

Proof. Let us set $L_1 = 2L_0$ where L_0 is the constant of (2.80). By induction over k we construct an admissible sequence $(\mathcal{A}_n)_{n\le n_k}$ such that

$$1 \le p \le k \Rightarrow \sup_{t\in T} \sum_{n_{p-1}<n\le n_p} 2^{n/2}\Delta(A_n(t)) \le 2L_0 S 2^{-p} . \tag{B.29}$$

For $k = 1$ the existence of the sequence $(\mathcal{A}_n)_{n<n_1}$ follows from the left-hand side of (2.80), so we turn to the induction step from k to $k+1$. Using (B.29) for $p = k$ we deduce that for each $t \in T$, $2^{n_k/2}\Delta(A_{n_k}(t)) \le 2L_0 S 2^{-k} = L_1 S 2^{-k}$, so that, using (B.26), $\Delta(A_{n_k}(t)) \le L_1 S 2^{-n_k/2-k} \le \delta_k$. Consequently, for any element B of $\mathcal{A}_{n_k}$ we have $\Delta(B) \le \delta_k$, so that considering any element t of B we have

$$\mathsf{E}\sup_{s\in B} X_s = \mathsf{E}\sup_{s\in B}(X_s - X_t) \le \mathsf{E} \sup_{\{s\in T; d(s,t)\le\delta_k\}} |X_s - X_t| \le S2^{-k} .$$

Using again (2.80) we obtain for each $B \in \mathcal{A}_{n_k}$ an admissible sequence $(\mathcal{A}_{B,n})_{n\ge 0}$ for which

$$\forall t \in B , \sum_{n\ge 0} 2^{n/2}\Delta(A_{B,n}(t)) \le L_0 S 2^{-k} . \tag{B.30}$$

For $n_k < n \le n_{k+1}$ we simply define $\mathcal{A}_n$ as the collection of all sets in one of the partitions $A_{B,n-1}$ where $B \in \mathcal{A}_{n_k}$, so that $\operatorname{card}\mathcal{A}_n \le N_{n-1}\operatorname{card}\mathcal{A}_{n_k} \le N_{n-1}^2 \le N_n$, and since $A_n(t) \subset A_{B,n-1}(t)$ it follows from (B.30) that

$$\sup_{t\in T} \sum_{n_k<n\le n_{k+1}} 2^{n/2}\Delta(A_n(t)) \le \sum_{n\ge n_k} 2^{n/2}\Delta(A_{B,n-1}(t)) \le 2L_0 S 2^{-k} .$$

This completes the induction and the construction of the sequence $(\mathcal{A}_n)$ since (B.29) implies (B.27).

It remains to prove the "conversely" part. For this for each $n \ge 0$ we simply consider a subset T_n of T such that

$$\forall A \in \mathcal{A}_n , \operatorname{card}(T_n \cap A) = 1 .$$

We then use Lemma B.3.1 for $m = n_k$ and $B = S2^{-k}$. □

References

1. Adamczak, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. J. Am. Math. Soc. **23**(2), 535–561 (2010)
2. Adamczak, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Sharp bounds on the rate of convergence of the empirical covariance matrix. C. R. Math. Acad. Sci. Paris **349**(3–4), 195–200 (2011)
3. Adamczak, R., Latała, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. Ann. Inst. Henri Poincaré Probab. Stat. **48**(4), 1103–1136 (2012)
4. Adamczak, R., Latała, R., Litvak, A., Pajor, A., Tomczak-Jaegermann, N.: Chevet type inequality and norm of submatrices. Stud. Math. **210**(1), 35–56 (2012)
5. Ajtai, M., Komlós, J., Tusnády, G.: On optimal matchings. Combinatorica **4**(4), 259–264 (1984)
6. Arcones, M.: The class of Gaussian chaos of order two is closed by taking limits in distribution. In: Advances in Stochastic Inequalities, Atlanta, GA, 1997. Contemp. Math., vol. 234, pp. 13–19. Amer. Math. Soc., Providence (1999)
7. Arcones, M., Giné, E.: On decoupling, series expansions, and tail behavior of chaos processes. J. Theor. Probab. **6**, 101–122 (1993)
8. Artstein, S.: The change in the diameter of a convex body under a random sign-projection. In: Geometric Aspects of Functional Analysis. Springer Lecture Notes in Math., vol. 1850, pp. 31–39 (2004)
9. Bollobas, B.: Combinatorics. Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability. Cambridge University Press, Cambridge (1986)
10. Barlow, M.T., Yor, M.: Semimartingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. J. Funct. Anal. **49**(2), 198–229 (1982)
11. Barthe, F., Guédon, O., Mendelson, S., Naor, A.: A probabilistic approach to the geometry of the ℓ_p^n-ball. Ann. Probab. **33**(2), 480–513 (2005)
12. Batson, J., Spielman, D., Srivastava, N.: Twice-Ramanujan sparsifiers. In: Proceedings of the 41st Annual Symposium on the Theory of Computing, pp. 47–55 (2009)
13. Bednorz, W.: On a Sobolev type inequality and its applications. Stud. Math. **176**(2), 113–137 (2006)
14. Bednorz, W.: A theorem on majorizing measures. Ann. Probab. **34**(5), 1771–1781 (2006)
15. Bednorz, W.: Hölder continuity of random processes. J. Theor. Probab. **20**(4), 917–934 (2007)

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2,

16. Bednorz, W.: On Talagrand's admissible net approach to majorizing measures and boundedness of stochastic processes. Bull. Pol. Acad. Sci., Math. **56**(1), 83–91 (2008)
17. Bednorz, W.: Majorizing measures on metric spaces. C. R. Math. Acad. Sci. Paris **348**(1–2), 75–78 (2006)
18. Bednorz, W.: Majorizing measures and ultrametric spaces. Bull. Pol. Acad. Sci., Math. **60**(1), 91–100 (2012)
19. Bednorz, W.: The complete characterization of a.s. convergence of orthogonal series. Ann. Probab. **41**(2), 1055–1071 (2013)
20. Bednorz, W., Latała, R.: On the suprema of Bernoulli processes. C. R. Math. Acad. Sci. Paris **351**(3–4), 131–134 (2013)
21. Bismut, J.-M., Yor, M.: An inequality for processes which satisfy Kolmogorov's continuity criterion. Application to continuous martingales. J. Funct. Anal. **51**(2), 166–173 (1983)
22. Bobkov, S.G., Nazarov, F.L.: On convex bodies and log-concave probability measures with unconditional basis. In: Geometric Aspects of Functional Analysis. Lecture Notes in Math., vol. 1807, pp. 53–69. Springer, Berlin (2003)
23. Bourgain, J.: Bounded orthogonal sets and the Λ_p-problem. Acta Math. **162**, 227–246 (1989)
24. Bourgain, J.: Random points in isotropic convex sets. In: Convex Geometric Analysis, Berkeley, CA, 1996. Math. Sci. Res. Inst. Publ., vol. 34, pp. 53–58. Cambridge Univ. Press, Cambridge (1999)
25. Bourgain, J., Pajor, A., Szarek, S.J., Tomczak-Jaegerman, N.: On the duality problem for entropy numbers. In: Geometric Aspects of Functional Analysis. Lecture Notes in Math, vol. 1376, pp. 53–63 (1989)
26. Clarkson, J.: Uniformly convex spaces. Trans. Am. Math. Soc. **40**(3), 39–414 (1936)
27. Candès, E., Plan, Y.: A probabilistic and RIPless theory of compressed sensing. IEEE Trans. Inf. Theory **5**(11), 7235–7254 (2011)
28. Coffman, E.G. Jr., Shor, P.W.: A simple proof of the $O(\sqrt{n}\log^{3/4} n)$ upright matching bound. SIAM J. Discrete Math. **4**(1), 48–57 (1991)
29. Dudley, R.M.: The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J. Funct. Anal. **1**, 290–330 (1967)
30. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics, vol. 63. Cambridge University Press, Cambridge (1999)
31. Diestel, J., Jarchow, H., Tonge, A.: Absolutely Summing Operators. Cambridge Studies in Advanced Mathematics, vol. 43. Cambridge University Press, Cambridge (1955)
32. Fernique, X.: Regularité des trajectoires des fonctions aléatoires gaussiennes. In: Ecole d'été de Probabilités de Saint-Flour, IV-1974. Lecture Notes in Math., vol. 480, pp. 1–96. Springer, Berlin (1975)
33. Fernique, X.: Eléments d'analyse des chaos gaussiens. Probab. Math. Stat. **15**, 291–300 (1995)
34. Fernique, X.: Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens. Université de Montréal, Centre de Recherches Mathématiques, Montreal (1977). iv+217 pp
35. Garsia, A.M., Rodemich, E., Rumsey, H.: A real variable lemma and the continuity of path of some Gaussian processes. Indiana Univ. Math. J. **20**, 565–578 (1970/1971)
36. Giné, E., Zinn, J.: Some limit theorems for empirical processes. Ann. Probab. **12**, 929–989 (1984)

37. Gluskin, E.D., Kwapien, S.: Tail and moment estimates for sums of independent random variables with logarithmically concave tails. Stud. Math. **114**(3), 303–309 (1995)
38. Gordon, Y.: Some inequalities for Gaussian processes and applications. Isr. J. Math. **50**(4), 265–289 (1985)
39. Gordon, Y.: On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. In: Geometric Aspects of Functional Analysis. Springer Lecture Notes in Math., vol. 1317, pp. 84–106 (1988)
40. Gordon, Y.: A note on an observation of G. Schechtman. In: Geometric Aspects of Functional Analysis. Springer Lecture Notes in Math., vol. 1910, pp. 127–132 (2010)
41. Guédon, O., Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.: Majorizing measures and proportional subsets of bounded orthonormal systems. Rev. Mat. Iberoam. **24**(3), 1071–1095 (2008)
42. Guédon, O., Rudelson, M.: L_p-moments of random vectors via majorizing measures. Adv. Math. **208**(2), 798–823 (2007)
43. Heinrich, P.: The size of diagonal two-order Gaussian chaoses. Stat. Probab. Lett. **51**(1), 35–39 (2001)
44. Hanson, D.L., Wright, F.T.: A bound on tail probabilities for quadratic forms of independent random variables. Ann. Math. Stat. **42**(6), 1079–1083 (1971)
45. Kahane, J.-P.: Some Random Series of Functions, 2nd edn. Cambridge Studies in Advanced Mathematics, vol. 5. Cambridge University Press, Cambridge (1985). xiv+305 pp
46. Kashin, B.S., Saakyan, A.A.: Orthogonal Series. Translations of Mathematical Monographs, vol. 75. Am. Math. Soc., Providence (1989)
47. Klartag, B., Mendelson, S.: Empirical processes and random projections. J. Funct. Anal. **225**(1), 229–245 (2005)
48. Kramher, F., Mendelson, S., Rauhut, H.: Suprema of chaos processes and the restricted isometry property. Manuscript (2012)
49. König, H., Tzafriri, L.: Some estimates for type and cotype constants. Math. Ann. **256**(1), 85–94 (1981)
50. Kôno, N.: Sample path properties of stochastic processes. J. Math. Kyoto Univ. **20**, 295–313 (1980)
51. Kwapien, S., Rosinski, J.: Sample Hölder continuity of stochastic processes and majorizing measures. In: Seminar on Stochastic Analysis, Random Fields and Applications IV. Progr. Probab., vol. 58, pp. 155–163. Birkhäuser, Basel (2006)
52. Latała, R.: Sudakov minoration principle and supremum of some processes. Geom. Funct. Anal. **7**, 936–953 (1997)
53. Latała, R.: Estimates of moments and tails of Gaussian chaoses. Ann. Probab. **34**(6), 2315–2331 (2006)
54. Latała, R.: On the boundedness of Bernoulli processes on thin sets. Electron. Commun. Probab. **13**, 175–186 (2008)
55. Latała, R.: On weak tail domination of random vectors. Bull. Pol. Acad. Sci., Math. **57**(1), 75–80 (2009)
56. Latała, R.: Weak and strong moments of random vectors. In: Marcinkiewicz Centenary Volume. Banach Center Publ., vol. 95, pp. 115–121. Polish Acad. Sci. Inst. Math, Warsaw (2011)
57. Ledoux, M.: The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001). x+181 pp. ISBN 0-8218-2864-9

58. Ledoux, M., Talagrand, M.: Probability in a Banach Space: Isoperimetry and Processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 23. Springer, Berlin (1991). xii+480 pp. ISBN: 3-540-52013-9
59. Lehec, J.: Moments of the Gaussian chaos. In: Séminaire de Probabilités XLIII. Lecture Notes in Math., vol. 2006, pp. 327–340. Springer, Berlin (2011)
60. Leighton, T., Shor, P.: Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. Combinatorica **9**(2), 161–187 (1989)
61. Lewis, D.: Finite dimensional subspaces of L_p. Stud. Math. **63**(2), 207–212 (1978)
62. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces, II. Function Spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 97. Springer, Berlin (1979). x+243 pp. ISBN: 3-540-08888-1
63. Lifshits, M., Weber, M.: On the supremum of random Dirichlet polynomials. Stud. Math. **182**(1), 41–65 (2007)
64. Lust-Piquard, F.: Inégalités de Khintchine dans C_p. C. R. Acad. Sci. Paris Sér. I Math. **303**(7), 289–292 (1986)
65. Marcus, M.B.: ξ-Radial Processes and Random Fourier Series. Mem. Amer. Math. Soc., vol. 68(368) (1987). viii+181 pp
66. Marcus, M.B., Pisier, G.: Random Fourier Series with Applications to Harmonic Analysis. Annals of Mathematics Studies, vol. 101. Princeton University Press, Princeton; University of Tokyo Press, Tokyo (1981). v+151 pp. ISBN: 0-691-08289-8; 0-691-08292-8
67. Marcus, M.B., Pisier, G.: Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. Acta Math. **152**(3–4), 245–301 (1984)
68. Marcus, M.B., Rosinski, J.: Sufficient conditions for boundedness of moving averages. In: Stochastic Inequalities and Applications. Progr. Probab., vol. 56, pp. 113–128. Birkhäuser, Basel (2003)
69. Marcus, M.B., Rosinski, J.: Continuity and boundedness of infinitely divisible processes: a Poisson point process approach. J. Theor. Probab. **18**, 109–160 (2005)
70. Marcus, M., Talagrand, M.: Continuity conditions for a class of Gaussian chaoses related to Wick squares. J. Theor. Probab. **11**, 157–179 (1998)
71. Maurey, B.: Type et cotype dans les espaces munis de structures locales inconditionelles. In: Séminaire Maurey-Schwartz 1973-74, Exposé 12. Ecole Polytechnique, Paris (1974)
72. Mendel, M., Naor, A.: Ultrametric subsets with large Hausdorff dimension. Invent. Math. **192**(1), 1–54 (2013)
73. Menchov, D.E.: Sur les séries de fonctions orthogonales. Fundam. Math. **4**, 82–105 (1923)
74. Mendelson, S.: On weakly bounded empirical processes. Math. Ann. **340**(2), 293–314 (2008)
75. Mendelson, S.: Empirical processes with a bounded ψ_1-diameter. Geom. Funct. Anal. **20**(4), 988–1027 (2010)
76. Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.: Reconstruction and subgaussian operators in asymptotic geometric analysis. Geom. Funct. Anal. **17**(4), 1248–1282 (2007)
77. Mendelson, S., Paouris, G.: On generic chaining and the smallest singular value of random matrices with heavy tails. J. Funct. Anal. **262**(9), 3775–3811 (2012)
78. Mendelson, S., Vershynin, R.: Entropy and the combinatorial dimension. Invent. Math. **152**(3), 37–55 (2003)

79. Montgomery-Smith, S.J.: The cotype of operators form $C(K)$. Ph.D. thesis, Cambridge (1988)
80. Montgomery-Smith, S.J.: The Gaussian cotype of operators from $C(K)$. Isr. J. Math. **68**(1), 123–128 (1989)
81. Montgomery-Smith, S.J., Talagrand, M.: The Rademacher cotype of operators from ℓ_∞^N. Isr. J. Math. **68**, 123–128 (1990)
82. Olejnik, J.: On Paszkiewicz-type criterion for a.e. continuity of processes in Lp-spaces. In: Stability in Probability, Banach Center Publ., vol. 90, pp. 103–110. Polish Acad. Sci. Inst. Math, Warsaw (2010)
83. Olejnik, J.: On a construction of majorizing measures on subsets of $\mathbb{R}^n$ with special metrics. Stud. Math. **197**(1), 1–12 (2010)
84. Olevskii, A.M.: Fourier Series with Respect to General Orthogonal Systems. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 86. Springer, New York (1975). viii+136 pp
85. Ossiander, M.: A central limit theorem under metric entropy with L_2 bracketing. Ann. Probab. **15**(3), 897–919 (1987)
86. Paley, R., Zygmund, A.: On some series of functions (1). Proc. Camb. Philos. Soc. **26**, 337–357 (1930)
87. Paley, R., Zygmund, A.: On some series of functions (2). Proc. Camb. Philos. Soc. **26**, 458–474 (1930)
88. Paley, R., Zygmund, A.: On some series of functions (3). Proc. Camb. Philos. Soc. **28**, 190–205 (1932)
89. Paszkiewicz, A.: A complete characterization of coefficients of a.e. convergent orthogonal series and majorizing measures. Invent. Math. **180**(1), 50–110 (2010)
90. Pisier, G.: Conditons d'entropie assurant la continuité de certains processus et applications à l'analyse harmonique. In: Séminaire d'analyse fonctionelle 1979–1980. Ecole Polytechnique, Paris (1980)
91. Pisier, G.: Factorization of operators through $L_{p,\infty}$ or $L_{p,1}$ and noncommutative generalizations. Math. Ann. **276**, 105–136 (1986)
92. Pisier, G.: Probabilistic methods in the geometry of Banach spaces. In: Probability and Analysis, Varenna, 1985. Lecture Notes in Math., vol. 1206, pp. 167–241. Springer, Berlin (1986)
93. Pisier, G.: Noncommutative vector valued L_p-spaces and completely p-summing maps. Astérisque **247**, 1–131 (1998)
94. Radmacher, H.: Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. Math. Ann. **87**(1–2), 112–138 (1922)
95. Rhee, W., Talagrand, M.: Exact bounds for the stochastic upward matching problem. Trans. Am. Math. Soc. **109**, 109–126 (1988)
96. Rosinski, J.: On series representations of infinitely divisible random vectors. Ann. Probab. **18**, 405–430 (1990)
97. Rosinski, J.: Series representation of Lévy processes from the perspective of point processes. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I. (eds.) Lévy Processes, Theory and Applications, pp. 401–415. Birkhäuser, Basel (2001)
98. Rosinski, J.: Representations of Poissonian processes and applications. Manuscript in preparation (2012)
99. Rudelson, M.: Almost orthogonal submatrices of an orthogonal matrix. Isr. J. Math. **111**, 143–155 (1999)
100. Salem, R., Zygmund, A.: Some properties of trigonometric series whose terms have random signs. Acta Math. **91**, 245–301 (1954)
101. Schechtman, G.: Two observations regarding embedding subsets of Euclidean spaces in normed spaces. Adv. Math. **200**(1), 125–135 (2006)

102. Schechtman, G.: Tight embedding of subspaces of L^p in ℓ_n^p for even p. Proc. Am. Math. Soc. **139**(12), 4419–4421 (2001)
103. Shor, P.W.: The average-case analysis of some on-line algorithms for bin packing. Theory of computing (Singer Island, Fla., 1984). Combinatorica **6**(2), 179–200 (1986)
104. Shor, P.W.: How to pack better than best fit: tight bounds for average-case on-line bin packing. In: Proc. 32nd Annual Symposium on Foundations of Computer Science, pp. 752–759 (1991)
105. Shor, P.W., Yukich, J.E.: Minimax grid matching and empirical measures. Ann. Probab. **19**(3), 1338–1348 (1991)
106. Stroock, D.W., Varhadan, S.R.S.: Multidimensional Diffusion Processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 233. Springer, Berlin (1979). xii+338 pp. ISBN 3-540-90353-4
107. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Found. Comput. Math. **12**(4), 389–434 (2012)
108. Talagrand, M.: Regularity of Gaussian processes. Acta Math. **159**, 99–149 (1987)
109. Talagrand, M.: The Glivenko-Cantelli problem. Ann. Probab. **15**, 837–870 (1987)
110. Talagrand, M.: The structure of sign-invariant GB-sets, and of certain Gaussian measures. Ann. Probab. **16**, 172–179 (1988)
111. Talagrand, M.: Necessary conditions for sample boundedness of p-stable processes. Ann. Probab. **16**, 1584–1595 (1988)
112. Talagrand, M.: Donsker classes of sets. Probab. Theory Relat. Fields **78**, 179–191 (1988)
113. Talagrand, M.: On subsets of L^p and p-stable processes. Ann. Inst. Henri Poincaré **25**, 153–166 (1989)
114. Talagrand, M.: Sample boundedness of stochastic processes under increment conditions. Ann. Probab. **18**, 1–49 (1990)
115. Talagrand, M.: Characterization of almost surely continuous 1-stable random Fourier series and strongly stationary processes. Ann. Probab. **18**, 85–91 (1990)
116. Talagrand, M.: A new isoperimetric inequality for product measure and the concentration of measure phenomenon. In: Israel Seminar (GAFA). Springer Lecture Notes in Math., vol. 1469, pp. 94–124 (1991)
117. Talagrand, M.: Cotype of operators from $C(K)$. Invent. Math. **107**, 1–40 (1992)
118. Talagrand, M.: Necessary and sufficient conditions for sample continuity of random Fourier series and of harmonic infinitely divisible processes. Ann. Probab. **20**, 1–28 (1992)
119. Talagrand, M.: Sudakov-type minoration for Gaussian chaos. Isr. J. Math **79**, 207–224 (1992)
120. Talagrand, M.: The Ajtai-Komlos-Tusnady matching theorem for general measures. In: Probability in a Banach Space. Progress in Probability, vol. 8, pp. 39–54. Birkhäuser, Basel (1992)
121. Talagrand, M.: A simple proof of the majorizing measure theorem. Geom. Funct. Anal. **2**, 119–125 (1992)
122. Talagrand, M.: Regularity of infinitely divisible processes. Ann. Probab. **21**, 362–432 (1993)
123. Talagrand, M.: The supremum of certain canonical processes. Am. J. Math. **116**, 283–325 (1994)

124. Talagrand, M.: Matching theorems and discrepancy computations using majorizing measures. J. Am. Math. Soc. **7**, 455–537 (1994)
125. Talagrand, M.: Constructions of majorizing measures Bernoulli processes and cotype. Geom. Funct. Anal. **4**, 660–717 (1994)
126. Talagrand, M.: The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. Ann. Probab. **22**, 919–959 (1944)
127. Talagrand, M.: A remark on Sudakov minoration for chaos. In: Probability in Banach Spaces, vol. 9, pp. 341–346. Birkhäuser, Basel (1994)
128. Talagrand, M.: A borderline random Fourier series. Ann. Probab. **23**, 776–785 (1995)
129. Talagrand, M.: Embedding subspaces of L_p into ℓ_p^n. In: Operator Theory: Advances and Applications, vol. 77, pp. 311–326. Birkhäuser, Basel (1995)
130. Talagrand, M.: Sections of smooth convex bodies via majorizing measures. Acta Math. **175**, 273–306 (1995)
131. Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. IHÉS **81**, 73–205 (1995)
132. Talagrand, M.: Majorizing measures: the generic chaining. Ann. Probab. **24**, 1049–1103 (1996)
133. Talagrand, M.: The Glivenko-Cantelli problem, ten years later. J. Theor. Probab. **9**, 371–384 (1996)
134. Talagrand, M.: A general lower bound for Gaussian chaos. J. Theor. Probab. **11**, 157–179 (1998)
135. Talagrand, M.: Selecting a proportion of characters. Isr. J. Math. **108**, 275–284 (1999)
136. Talagrand, M.: Majorizing measures without measures. Ann. Probab. **29**, 411–417 (2001)
137. Talagrand, M.: Gaussian averages, Bernoulli averages, and Gibbs' measures. Random structures and algorithms (Poznan, 2001). Random Struct. Algorithms **21**(3–4), 197–204 (2002)
138. Talagrand, M.: Vapnik-Chervonenkis type conditions and uniform Donsker classes of functions. Ann. Probab. **31**(2), 1565–1582 (2003)
139. Talagrand, M.: Selector processes on classes of sets. Probab. Theory Relat. Fields **135**(4), 471–486 (2006)
140. Talagrand, M.: The Generic Chaining. Springer Monographs in Mathematics. Springer, Berlin (2005). viii+222 pp. ISBN: 3-540-24518-9
141. Talagrand, M.: Are many small sets explicitly small? In: STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, pp. 13–35. ACM, New York (2010). Available at http://research.microsoft.com/apps/video/dl.aspx?id=137091
142. Talagrand, M., Yukich, J.: The integrability of the square exponential transportation cost. Ann. Appl. Probab. **3**, 1100–1111 (1993)
143. Tandori, K.: Über die divergenz der Orthogonalreihen. Publ. Math. Drecen **8**, 291–307 (1961)
144. Tandori, K.: Über die Konvergenz der Orthogonalreihen. Acta Sci. Math. (Szeged) **24**, 131–151 (1963)
145. Tomczak-Jaegermann, N.: Banach-Mazur Distances and Finite-Dimensional Operator Ideals. Pitman Monographs and Surveys in Pure and Applied Mathematics, vol. 38. Wiley, New York (1989). xii+395 pp. ISBN: 0-582-01374-7
146. Weber, M.: Analyse infiniésimale de fonctions aléatoires. In: Eleventh Saint Flour Probability Summer School, Saint Flour, 1981. Lecture Notes in Math., vol. 976, pp. 383–465. Springer, Berlin (1981)
147. Weber, M.: Some theorems related to almost sure convergence of orthogonal series. Indag. Math. **11**(2), 293–311 (2000)

148. Weber, M.: Estimating random polynomials by means of metric entropy methods. Math. Inequal. Appl. **3**(3), 443–457 (2000)
149. Weber, M.: Sur le caractère gaussien de la convergence presque partout. Fundam. Math. **167**(1), 23–54 (2001)
150. Weber, M.: Uniform bounds under increment conditions. Trans. Am. Math. Soc. **358**(2), 911–936 (2006)
151. Weber, M.: Dynamical Systems and Processes. IRMA Lectures in Mathematics and Theoretical Physics, vol. 14. European Mathematical Society (EMS), Zürich (2009). xii+761 pp. ISBN: 978-3-03719-046-3
152. Yukich, J.: Some generalizations of the Euclidean two-sample matching problem. In: Probability in Banach Spaces, vol. 8, Brunswick, ME, 1991. Progr. Probab., vol. 30, pp. 55–66. Birkhäuser, Boston (1992)
153. Zvavitch, A.: More on embedding subspaces of L_p into ℓ_p^N, $0 < p < 1$. In: Geometric Aspects of Functional Analysis. Lecture Notes in Math., vol. 1745, pp. 269–280. Springer, Berlin (2000)

Index

M. Talagrand, *Upper and Lower Bounds for Stochastic Processes*,
Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of
Modern Surveys in Mathematics 60, DOI 10.1007/978-3-642-54075-2,