Systems & Control: Foundations & Applications

Agamirza E. Bashirov

Partially Observable Linear Systems Under Dependent Noises

Springer Basel AG

Author:

Agamirza E. Bashirov
Department of Mathematics
Eastern Mediterranean University
Gazimagusa
Mersin 10
Turkey
e-mail: agamirza.bashirov@emu.edu.tr

2000 Mathematics Subject Classification 47N70, 60G35, 93C05, 93Exx

A CIP catalogue record for this book is available from the Library of Congress, Washington D.C., USA

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>

ISBN 978-3-0348-9407-4 ISBN 978-3-0348-8022-0 (eBook)

DOI 10.1007/978-3-0348-8022-0

Originally published by Birkhäuser Verlag, Basel – Boston – Berlin in 2003
Softcover reprint of the hardcover 1st edition 2003
Printed on acid-free paper produced of chlorine-free pulp. TCF ∞

ISBN 978-3-0348-9407-4

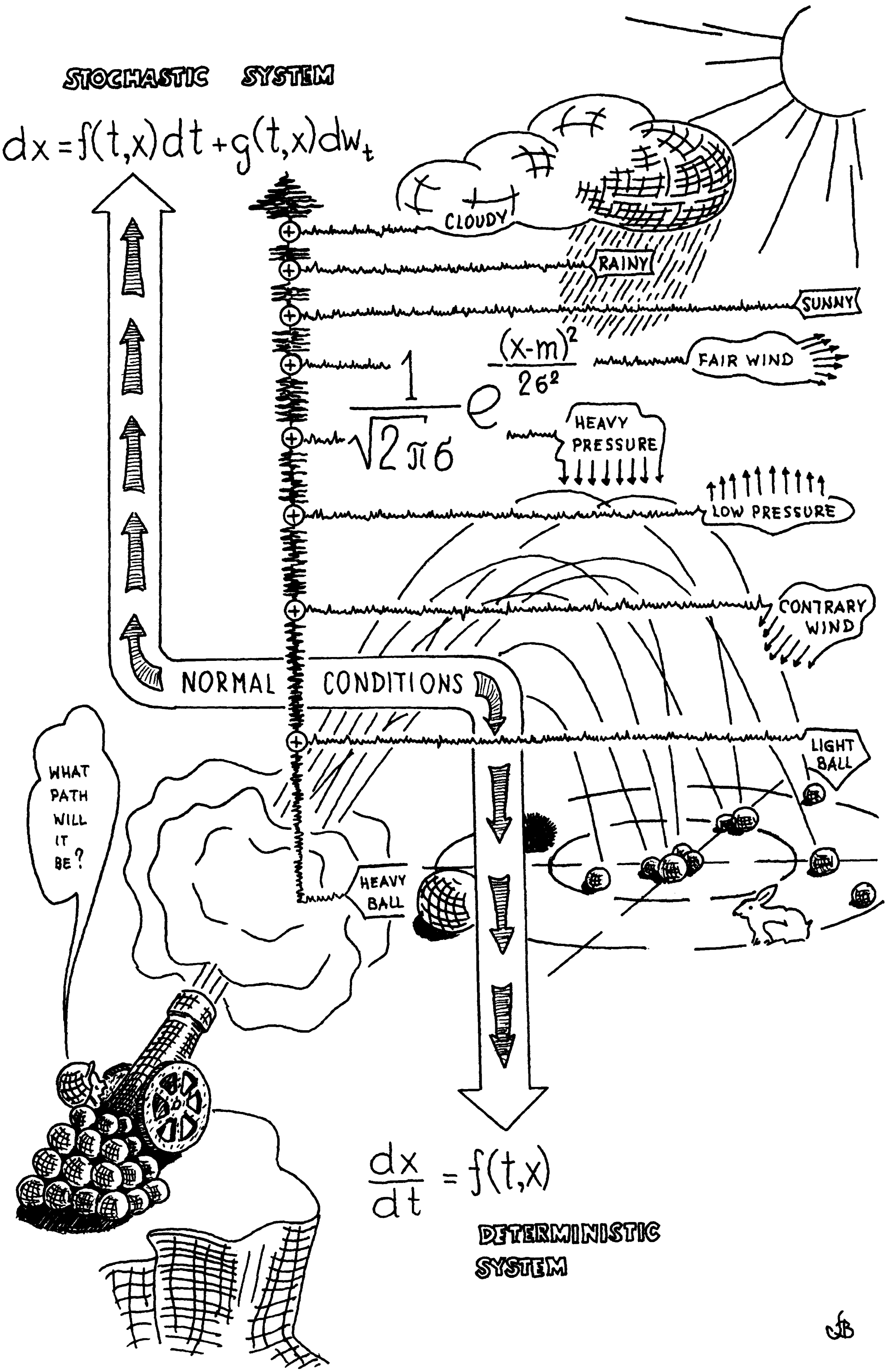
9 8 7 6 5 4 3 2 1

To my parents:

Enver Bashirov

and

Sharqiya Bashirova


STOCHASTIC SYSTEM
$dx = f(t,x)dt + g(t,x)dW_t$
CLOUDY
RAINY
SUNNY
$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$
FAIR WIND
HEAVY PRESSURE
LOW PRESSURE
CONTRARY WIND
NORMAL CONDITIONS
LIGHT BALL
WHAT PATH WILL IT BE?
HEAVY BALL
$\frac{dx}{dt} = f(t,x)$
DETERMINISTIC SYSTEM

Contents

Preface

Modern control and estimation theory originated in the 1960s in the fundamental works of Pontryagin [82], Bellman [28] and Kalman [61, 62, 60]. Subsequent scholarship noted the importance of a global approach to systems of a different nature such as lumped, distributed, delayed etc. This prompted the study of control and estimation problems in spaces of infinite dimension. The pioneering books of Balakrishnan [4], Bensoussan [29] and Curtain and Pritchard [40] advocated this approach, the two latter providing an extensive list of early references concerning the theory.

Nowadays a great number of mathematicians and engineers are promoting the development of control and estimation theory. This is reflected in numerous papers published in such popular journals as *SIAM Journal on Control and Optimization*, *Stochastics and Stochastics Reports*, *IEEE Transactions on Automatic Control*, etc. Many books and papers devoted to different issues in control and estimation theory have been written (see the bibliography in Yong and Zhou [93] and in Grewal and Andrews [52] for theoretical issues and for applications).

Nevertheless, while multiple early and recent findings on the subject have been obtained and challenging problems remain to be solved there is an aspect of control and estimation theory that has not been examined adequately. This aspect concerns partially observable systems under the action of dependent noise processes on the state (or signal) and the observations.

Generally speaking, noise is a rich concept playing an underlying role in human activity. Consideration of the noise phenomenon in arts and sciences, respectively, makes the distinction between both domains more obvious. Artists create "deliberate noise"; the best masterpieces of literature, music, modern fine art etc. are those where a clear idea, traditionally related to such concepts as love, is presented under a skilful veil of "deliberate noise". On the contrary, scientists fight against noise; a scientific discovery is a law of nature extracted from a noisy medium and refined.

Noise in sciences is an unwanted signal generated by processes of nature or by human intelligence. Mathematically, this useless noise is modelled by a random process. Noisy signal, on the other hand, is also modelled by a random process. To distinguish them, the first one is referred to as a noise process or, briefly, a noise. One outstanding result, mathematically proved and stated in the central limit theorem, is that a noise process that is the total of the effects of a large number of independent contributing factors is approximately Gaussian in its behavior. Therefore, throughout this book we assume that the noise processes under discussion are Gaussian. The meaning of this statement will be most clear by our exposition.

From the physical point of view, two noise processes are independent if they are generated by distinct independent sources. Noise processes coming essentially from the same source are dependent. How can dependent noise processes be exposed mathematically? One of the ways is to consider noise processes without any restriction on the relation between them. This general framework includes the case of independent noise processes as well, and so it is reasonable to refer to such noise processes as arbitrarily dependent. Other ways of analysis are to set specific relations between them. One such particular framework, concerning correlated white noise processes, is very popular in the existing literature.

A noise process is white if there is no correlation between its values at different times. A white noise is purely a mathematical concept, as its paths are functions in a generalized sense. Noise in a practical system is at best nearly white and may be far from being white. To improve a noise model, a solution of a linear stochastic differential equation disturbed by a white noise is used and is called a colored noise. Colored noises having the same white noise as a source are evidently dependent, providing another particular specific relation for exposing dependent noise processes.

More complicated dependent noise processes can be exposed if one of two correlated white noises is shifted in time. We regard such noises as shifted white noises. Here are some illustrations of processes that can form shifted noises.

Mapping the ocean floor. Getting a correct map of the ocean floor is important for installation of fixed mobile drilling platforms, locating pipelines in the ocean etc. This job is assisted by a device called a sonar. A sound signal radiates into the water through the sonar transducer that normally is mounted near the keel of a surface ship. Echoes are reflected from the ocean bottom to the sonar, which detects them and determines water depth. However, the ocean waves affect the calculated water depth. If ε is the difference of the detecting and radiating times of the sound signal and y is the actual water depth (corresponding to the ocean level), then the water depth is

$$z = y + w'(t)$$

at the detecting time moment t of the sound signal and it is

$$x = y + w'(t - \varepsilon)$$

at the radiating time moment $t - \varepsilon$ of the same signal. Here $w'(t)$ is the displacement in a surface wave at the time t and can be characterized as the sum of wind-generated waves at previous times over a large area in conjunction with the Earth's gravity. Considering w' as a white noise, we see that x and z are random perturbations of y by shifted white noises.

Space navigation and guidance. In the previous illustration, ε is negligible. For instance, if a sound propagates in water at a speed of about 1500 m/s, for an ordinary water depth of 750 m, one can calculate $\varepsilon = 1\ s$. The change

$$w'(t) - w'(t - \varepsilon)$$

of the ocean-wave height for the time of 1 s is much smaller than the depth of 750 m. But, the previous illustration exposes well a mechanism that forms shifted white noises. Should the ocean bottom be replaced with a spacecraft, a sonar with a ground radar, a sound signal with an electromagnetic signal and a white noise caused by ocean waves with a white noise caused by atmospheric propagation, then the corresponding value of ε would be significant. It is nearly constant for Earth orbiting satellites and time dependent for space probes having interplanetary missions.

To understand the nature of the shift arising in space navigation, fix some time moment t and let ε be the time needed for electromagnetic signals to run the distance from the ground radar to the spacecraft and then to turn back. Assume that the control action u changes the position x of the spacecraft in accordance with the linear equation

$$x' = Ax + Bu$$

if noise effects and the distance to the spacecraft are neglected. Then at the time t the ground radar detects the signal

$$z(t) = x(t - \varepsilon/2) + w'(t)$$

consisting of the useful information $x(t-\varepsilon/2)$ about the position of the spacecraft at $t - \varepsilon/2$, corrupted by white noise $w'(t)$ caused by atmospheric propagation. Furthermore, the position of the spacecraft at $t - \varepsilon/2$ is changed by the control action $u(t - \varepsilon)$ that is sent by the ground radar at the time moment $t - \varepsilon$. This control passing through the atmosphere is corrupted by the noise $w'(t-\varepsilon)$. Hence, the equation for the position of the spacecraft must be written as

$$x'(t - \varepsilon/2) = Ax(t - \varepsilon/2) + B(u(t - \varepsilon) + w'(t - \varepsilon)).$$

Substituting

$$\tilde{x}(t) = x(t - \varepsilon/2) \quad \text{and} \quad \tilde{u}(t) = u(t - \varepsilon),$$

we obtain the partially observable system

$$\begin{cases} \tilde{x}'(t) = A\tilde{x}(t) + B\tilde{u}(t) + Bw'(t - \varepsilon), \\ z(t) = \tilde{x}(t) + w'(t), \end{cases}$$

disturbed by shifted white noises with the state noise delaying the observation noise.

Here ε is a function of time t in general. Since Earth orbiting satellites have nearly constant distance from the Earth, it is reasonable to take ε as a constant for them. But for space probes flying away from the Earth $\varepsilon = ct$ with $0 < c < 1$, since their distance from the Earth increases with nearly constant rate of change. One can deduce that for space probes flying toward the Earth, $\varepsilon = a - ct$, where $a > 0$ and $0 < c < 1$ so that $a - ct > 0$.

The way of forming shifted white noises suggests another interesting relation that, in particular, leads to the important concept of wide band noise. Indeed, consider a distributed shift (instead of pointwise) of a white noise, i.e., define the random process

$$\varphi(t) = \int_{\max(0,t-\varepsilon)}^{t} \Phi(\theta - t) w'(t)\, dt, \tag{1}$$

where $\varepsilon > 0$, Φ is a deterministic function that will be labelled a relaxing function, and w' is a white noise. Then the resulting noise process φ becomes wide band. Fleming and Rishel [48], p. 126, express their views of these noise processes as follows:

"*Wide band noise.* Suppose that some physical process, if unaffected by random disturbances, can be described by a (vector) ordinary differential equation $d\xi = b(t, \xi(t))dt$. If, however, such disturbances enter the system in an additive way, then one might take as a model

$$d\xi = b(t, \xi(t))dt + \varphi(t)dt, \tag{2}$$

where φ is some stationary process with mean 0 and known autocovariance matrix $R(r)$:

$$R_{ij}(r) = \mathbf{E}(\varphi_i(t)\varphi_j(t + r)), \ i, j = 1, \ldots, n.$$

If $R(r)$ is nearly 0 except in a small interval near $r = 0$, then φ is called wide band noise. White noise corresponds to the ideal case when R_{ij} is a constant a_{ij} times a Dirac delta function. Then $\varphi(t)dt$ is replaced by $\sigma w'(t)dt$, where σ is a constant matrix such that $\sigma\sigma^* = a$, $a = (a_{ij})$. (Here σ^* is the transpose of σ and w' is a white noise.) The corresponding diffusion is then an approximation to the solution to (2)."

Additionally, note that in many fields such replacement of wide band noise by white noise gives rise to tangible distortions. Therefore, it is very important to develop the methods of control and estimation for wide band noise driven systems. Evidently, two wide band noise processes in the form (1) having the same source white noise w' are dependent.

For a long time the author together with his colleagues has been working on partially observable systems under dependent noises. The history of this research begins in 1976 when J. E. Allahverdiev, after his visit to the Control Theory Centre at the University of Warwick, initiated the study in the field of stochastic control at the Institute of Cybernetics of the Azerbaijan Academy of Sciences (Baku). By that time, research in the mathematical community of Baku on mathematical problems of optimal control, though of deterministic nature, was prompted by Pontryagin's maximum principle. At the Baku State University such research had been initiated earlier by K. T. Ahmedov. Another strong research group was in the Institute of Cybernetics (Baku), mainly concentrated in Lab. no. 1 of the institute led by J. E. Allahverdiev. The author of this book as a young Ph.D. student was the first to be involved in the study of stochastic control problems. Soon a small research group in Lab. no. 1 was organized also including R. R. Hajiyev, N. I. Mahmudov and, later, L. R. Mishne and others. A significant event for this group of researchers was the International Conference IFIP on Stochastic Differential Systems held in Baku in 1984 by the initiation of A. V. Balakrishnan and J. E. Allahverdiev. After this conference, research on stochastic control in the Institute of Cybernetics became part of the global research on estimation and control of stochastic processes in the former USSR. The founder of the eminent school of Probability and Mathematical Statistics in the USSR was the late A. N. Kolmogorov. His disciples have been successfully continuing the traditions of this school. In particular, research on stochastic processes in connection with control and estimation are coordinated through the seminar "Statistics and Control of Stochastic Processes" organized by A. N. Shiryaev (Steklov Mathematical Institute, Moscow) in collaboration with N. V. Krylov and R. S. Liptser and the seminar "Theory of Random Processes" organized by A. V. Skorohod (Institute of Mathematics of the Ukrainian Academy of Sciences, Kiev). Many other research groups in Vilnius, Tbilisi, Tashkent, Baku etc. were involved in this global research. Conferences, workshops, seminars, and professional contacts undoubtedly played a significant role in consolidating the research group in Baku. After 1990, many researchers involved in this global research went to different institutions around the world. In particular, the author of this book has since 1992 been working and continuing related research at the Eastern Mediterranean University, Famagusta.

This book is intended to discuss in a systematic way some results on partially observable systems under dependent noises. The discussion is given for infinite dimensional systems, since some of the noise processes used in the book are easily described if the state space is enlarged up to infinite dimension. This is very similar to differential delay equations. Hence, it becomes convenient to take underlying systems to be infinite dimensional. The main objective in the book is to establish the specific features following dependent noises. Best of all this can be done within the linear quadratic framework and for continuous and finite time horizon. Hence, the systems under consideration are linear, the functionals are quadratic, time is continuous running in a finite horizon. We are concerned with four basic problems of systems theory, namely,

(a) optimal control,

(b) estimation,

(c) duality and

(d) controllability

for which valuable progress has already been achieved.

Dealing with control and estimation problems in infinite dimensional spaces requires using concepts from functional analysis. In order to make this book self-contained, we discuss these concepts in the first three chapters. The related background can be found in more detail in multiple sources. Therefore, we often avoid complete explanations and keep our discussion at a level that is adequate to read the other chapters. Though separable Hilbert spaces are used as underlying in this book, the spaces over them, such as spaces of bounded linear operators, spaces of continuous functions etc., are not Hilbert spaces. Therefore, sometimes we have to be more general than it may appear considering the concepts of functional analysis in Banach and, occasionally, in metric spaces.

Chapter 1 includes concepts such as sets, functions, abstract spaces, linear operators and different kinds of convergence. The reader can use any textbook on functional analysis to study this chapter in greater detail.

In Chapter 2 we consider the concepts of continuity, differentiability, measurability and integrability for functions with values in infinite dimensional spaces and complete it with two basic classes of linear operators on functional spaces, with integral and differential operators. For more detail, the recommended books are Hille and Phillips [54], Dunford and Schwartz [45], Warga [89], Yosida [94], Kato [63], Balakrishnan [4] etc.

In Chapter 3 we discuss two basic classes of evolution operators, namely, semigroups of bounded linear operators, and mild evolution operators, and related transformations. Theory of semigroups is presented in a number of books including Balakrishnan [4], Bensoussan *et al.* [31], Curtain and Pritchard [40] etc. The concept of mild evolution operator was introduced in Curtain and Pritchard [39]. In this chapter, Riccati equations in operator form are studied as well.

Chapter 4 starts with Hilbert space-valued random variables and processes. For details, we recommend the books of Curtain and Prichard [40], Metivier [77] and Rozovskii [84]. A recommended book on Gaussian systems is Shiryaev [86]. Then in Section 4.2 we discuss Brownian motion and derive stochastic differential equations. The recommended books are Davis [43], Hida [53] and Gihman and Skorohod [50].

Section 4.3 deals with the stochastic integrals with respect to Hilbert space-valued square integrable martingales. There are a number of sources on stochastic integration (see, for example, Liptser and Shiryaev [70, 72], Gihman and Skorohod [50, 51], Kallianpur [59], Elliot [46]). We follow Metivier [77] with some supplements from Rozovskii [84]. The set $\tilde{\Lambda}(0,T;X,Z)$ is introduced in Metivier [77]. To make it a Hilbert space, we consider its quotient set. Perhaps, the results of this section might be found too general for the purposes of control and estimation theory. Our aim in this section is to present a general formulation and a complete proof of the stochastic analogue of Fubini's theorem, which is useful in dealing with stochastic control and estimation problems as well as in other applications of stochastic calculus. The reader may prefer to omit this section without any loss and assume that all stochastic integrals used in this book are integrals of nonrandom functions.

In Section 4.4 we discuss the solution concepts for linear stochastic differential equations and introduce linear stochastic evolution systems and partially observable linear systems, the second of them being the main object of study in this book. There are a number of sources on stochastic differential equations, for example, Liptser and Shiryaev [70, 72], Gihman and Skorohod [51], Ikeda and Watanabe [58] etc. In infinite dimensional spaces this subject is studied in Rozovskii [84], Da Prato and Zabczyk [42] for nonlinear and in Curtain and Pritchard [40] for linear cases. In Section 4.5 we follow Curtain and Pritchard [40] and use Shiryaev [86] to present basic estimation in Hilbert spaces. Finally, Chapter 4 is completed in Section 4.6 with a discussion of white, colored and wide band noise processes. There are different approaches to the concept of wide band noise. For example, Kushner [69] uses an approximative approach. The integral representation of wide band noise, used in this section, is introduced in Bashirov [9].

Chapters 5–10 deal with optimal control and estimation problems. There are a lot of sources about these problems, especially, in linear quadratic case. Two approaches to optimal control problems are basic. One of them concerns necessary conditions of optimality and it is called Pontryagin's maximum principle [82]. The other one, giving sufficient conditions of optimality, is Bellman's dynamic programming [28]. We recommend the books by Krylov [66], Fleming and Rishel [48], Fleming and Soner [49] for a discussion of the dynamic programming approach to stochastic control problems. The recent book by Yong and Zhou [93] discusses the maximum principle for stochastic systems in general form, which essentially differs from Pontryagin's maximum principle, that covers both controlled drift and controlled diffusion. Also, a comparison of these two basic approaches, many other issues as well, are considered in [93]. This seems to be an appropriate place to note that the general stochastic maximum principle was obtained independently by Mahmudov [73] and Peng [81]. The international control community has traditionally referred only to the paper [81] that was published in 1990 based on results obtained in 1988 (see comments in [93]). Due note is taken here that the work [73] was reported in 1987 in the workshop "Statistics and Control of Random Processes", organized by Steklov Mathematical Institute (Moscow), Institute of Mathematics (Kiev) and Institute of Mathematics and Cybernetics (Vilnius) held in Preila and it was published in Russian in 1989.

For linear quadratic optimal control problems under partial observations, both these approaches lead to the same result, called the separation principle. In the continuous time case the separation principle was first stated and studied by Wonham [92]. This result in Hilbert spaces was considered in a number of works, for example, Bensoussan and Viot [33], Curtain and Ichikawa [38] etc.

The first estimation problems were studied independently by Kolmogorov [64] and Wiener [91] who used the spectral expansion of stationary random processes. A significant stage in the development of estimation theory was the famous works of Kalman [61] and Kalman and Bucy [62]. For complete discussion of estimation problems see Liptser and Shiryaev [71, 72], Kallianpur [59], Elliot [46] etc. In infinite dimensional spaces linear estimation problems are studied in Curtain [37] (see also Curtain and Pritchard [40]). In Chapters 5–9 our principal aim consists of discussing optimal control and estimation problems when noise processes of the state (or signal) and observation systems are dependent.

In Chapter 5 the separation principle, which is essential to studying linear quadratic optimal control problems under partial observations, is extended to arbitrarily dependent noise processes. We make a distinction between the two forms of the separation principle and consider the separation principle that holds for independent noise processes to be classical. It is found that this form remains valid for dependent noise processes as well, while the past of observations and the future of the state noise are independent. Otherwise, the separation principle changes its form with appearance of some additional terms in a representation of the optimal control. This form of the separation principle is considered to be extended. In this chapter we also discuss some consequences from the extended separation principle, including its generalization to a game problem, construction of a minimizing sequence, the existence of an optimal control, and presentation of a linear regulator problem. In Section 5.1 we use the ideas from Bensoussan and Viot [33] and Curtain and Ichikawa [38] in setting a linear quadratic optimal control problem under partial observations. Extension of the separation principle to noise processes, acting dependently on state and observations, is studied in Bashirov [7], which we follow up in Section 5.2. Section 5.3 is written on the basis of Bashirov [14]. The idea of minimizing sequence considered in Section 5.4 comes from Bensoussan and Viot [33]. In infinite dimensional spaces the linear regulator problem has been studied in a number of works (see, for example, Curtain and Prichard [39]). In Section 5.5 this problem is considered as a degenerate case of the separation principle. Generally, the results on existence of optimal control are based on weak convergence and weak compactness. In the linear quadratic case it is possible to reduce the existence of optimal control to a certain linear filtering problem when the observations are incomplete. This fact is proved in Section 5.6.

Chapters 6–10 present estimation and control results for partially observable linear systems under specific dependent noises. We use three different methods to investigate the problems in these chapters.

The first method is traditional; it is based on the duality principle and the separation principle. In Chapter 6 this method is used to derive optimal estimators and optimal controls for systems under correlated white noises. We also employ this method in Chapter 9 to get the control and filtering results under shifted white noises when the state or signal noise is a delay of the observation noise. While in Chapter 6 we use the separation principle in its classic form, the control results of Chapter 9 are obtained through the extended separation principle.

The second method simplifies hard calculations that are involved in the first method; this method is based on the reduction of the originally given system to a system disturbed by correlated white noises. This method is well applicable to systems under colored noises. This is demonstrated in Chapter 7. We employ this method in Chapter 8 as well to study the control and estimation problems under wide band noises. Though the results, concerning colored noises, are familiar (see Bucy and Joseph [35]), those which concern wide band noises are recent. We derive a complete set of formulae for the respective optimal control and for the respective optimal estimators when the noise processes of the underlying system are wide band. This set of formulae includes the stochastic partial differential equations for the respective optimal filter and, hence, offers a challenge for applications of stochastic partial differential equations, which are being studied intensively nowadays. Also, the respective Riccati equation is derived as a system of equations including first order partial differential equations. We find the results of this chapter most useful in engineering and hope to see a realization of them in applications.

The reduction method successfully used in Chapters 7 and 8 is not effective for systems with shifted white noises since it leads to stochastic differential equations with a boundary noise. Therefore, we develop the third method which is based on convergence. This method is used in Chapter 10 for both the state (signal) noise delaying and anticipating the observation noise. Starting from the fact that a white noise is the limit case of wide band noises, we approximate a shifted white noise driven linear system by wide band noise driven linear systems. Taking the limit in the respective control and filtering results for wide band noise driven systems from Chapter 8, we derive the equations for the optimal control and optimal filter for shifted white noise driven systems. This set of equations includes stochastic partial differential equations as well and, moreover, the boundary conditions for them are again stochastic differential equations offering another challenge for applications of stochastic partial differential equations. The results of this chapter are most recent and they are not proved precisely. An interesting feature of this chapter is a derivation of equations for optimal filters and optimal controls for navigation of spacecraft (both Earth orbiting satellites and space probes) which can have interesting implications for engineering.

An expert in systems theory can observe that both the wide band noise and shifted white noise processes are two kinds of colored noise, when the linear equation transforming the input white noise is a differential delay equation with distributed or pointwise delays, respectively. Formally, a white noise can also be considered as a colored noise since it is an identical transformation of itself. Thus all basic noise processes are colored with specific linear transformations. This can be considered as a way to classify and specify different kinds of noise. Nevertheless, having a colored noise as a general noise model does not decrease the importance of specific wide band and shifted white noises. For a comparison recall that probability theory was also discovered as a particular case of measure theory, by Kolmogorov [65] in a monograph published in 1933.

In Chapter 11 we discuss the duality principle. This is a remarkable relation between the control and estimation problems. This relation was discovered by Kalman [60] between the linear regulator and linear filtering problems and stated as the principle of duality. We extend this duality to linear stochastic optimal control and estimation problems in which connection these interesting relations have been discovered:

- If the classical separation principle is valid for a stochastic control problem, then it is dual to a filtering problem.

- If the extended separation principle is valid for a stochastic control problem, then it is dual to a smoothing problem.

- For control problems, the role of an innovation process plays a certain random process. The major role of this process is to transform the original control problem to a new one for which the classical separation principle holds.
- In fact the reduction method mentioned above (in connection to Chapters 7 and 8) implicitly uses the dual analogue of the innovation process. This is the main reason that the optimal control results of Chapters 7 and 8 are obtained without any reference to the extended separation principle.

In this chapter we follow Bashirov [7].

Chapter 12 deals with controllability concepts. Theory of controllability originates from the famous work of Kalman [61] and was well-discussed for deterministic systems in a number of books, see Balakrishnan [4], Curtain and Pritchard [40], Bensoussan *et al.* [32], Zabczyk [95] etc. The significant achievements in controllability theory for deterministic linear systems are Kalman's rank condition, the complete controllability condition and the approximate controllability condition. Afterwards the resolvent conditions for complete and approximate controllability were discovered in Bashirov and Mahmudov [23]. Both the concepts of complete and approximate controllability lose sense for stochastic systems since now a terminal value is a random variable. The two different interconnections of controllability and randomness define the two principally different methods of extending the controllability concepts to stochastic systems. In the first method the state space in the definition of controllability concepts is replaced by a suitable space of random variables, for example, the space of square integrable random variables. Thus, attaining random variables, even those with large entropy, is necessary to be controllable in this sense. This direction is employed by Mahmudov [74, 75]. The second method is more practical: it assumes attaining only those random variables that have small entropy, excluding the needless random variables with large entropy. In this chapter we follow the second method and use the works of the author and his colleagues [20, 21, 22, 23, 12, 24]. We prove the resolvent conditions for the complete and approximate controllability of deterministic linear systems and then apply them to study the concepts of controllability for partially observable linear systems. We define two main concepts of controllability for partially observable linear systems. The concept of S-controllability is defined as a property of a system to attain an arbitrarily small neighborhood of each point in the state space with probability arbitrarily near to 1. Also, the concept of C-controllability is defined as S-controllability fortified with some uniformity. We show that a partially observable linear system is C-controllable (respectively, S-controllable) for every time moment if and only if the respective deterministic linear system is completely (approximately) controllable for every time moment.

We mentioned that this book discusses only four problems of systems theory: optimal control, estimation, duality and controllability for partially observable linear systems under dependent noises. Many other problems in conjunction with dependent noises issue are still open. Let us mention some of them:

- Stability of the optimal filters derived for wide band and shifted white noise driven linear systems.
- Infinite time horizon issue in case of dependent noises.
- Random coefficients and dependent noises.
- Nonlinear filtering under wide band and shifted white noises.
- Nonlinear control under dependent noises.
- Necessary and sufficient conditions for S- and C-controllability of wide band and shifted white noise driven systems.
- The other interesting problem concerning modelling wide band noise processes in the integral form on the basis of autocovariance functions is described in Section 8.4.

Summarizing, we define the target readers of this book to be both applied mathematicians and theoretically oriented engineers who are designing new technology, as well as students of the related branches. Especially, the complete sets of equations for the optimal controls and for the optimal filters under wide band noises and shifted white noises and their possible application to navigation of spacecraft can have interesting implications for engineering. The book may be used as a reference manual in the part of functional analysis that is needed for problems of infinite dimensional linear systems theory.

Finally, the bibliography given at the end of the book is by no means complete. It reflects mathematical sources on the subject, and mainly those which have been used by the author. The index of notation may also be found at the end of the book. One major remark about notation is that the symbol f_t is preferred for the value of the function f at t instead of $f(t)$. This simplifies line breaks in long formulae.

I am indebted to the Institute of Cybernetics of the Azerbaijan Academy of Sciences (Baku), especially, the members of Lab. no. 1 and its head J. E. Allahverdiev, the Steklov Mathematical Institute (Moscow), especially, the organizer of the seminar "Statistics and Control of Stochastic Processes" A. N. Shiryaev and active members R. S. Liptser and A. A. Novikov, the Institute of Mathematics of the Ukrainian Academy of Sciences (Kiev), especially, the organizer of the seminar "Theory of Random Processes" A. V. Skorohod and active members Yu. L. Daletskii and A. I. Ponomarenko, the Kiev State University, especially, A. A. Anisimov and A. G. Nakonechniy, the Institute of Mathematics and Mechanics of the Ukrainian Academy of Sciences (Donetsk), especially, I. I. Gihman, the Department of Mathematics of the Eastern Mediterranean University (Famagusta), especially, Zeka Mazhar who supported and encouraged me during my research. Thanks to Tuğrul Taner for numerous suggestions for improving the manuscript. I also appreciate my younger colleague N. I. Mahmudov with whom

I have been discussing our mathematical problems for more than 20 years. I am grateful to Gülşen Musayeva for helping me with proofreading of some portions of the manuscript. I thank, furthermore, Birkhäuser, especially, Thomas Hempfling who properly evaluated the manuscript and the editors, reviewers and other staff taking part in publishing this book. My special thanks go to two ladies, Güneş and Aynur, for their patience throughout my work on the book.

A. E. Bashirov
(agamirza.bashirov@emu.edu.tr)
Baku – Famagusta
May – June, 2002

Chapter 1

Basic Elements of Functional Analysis

This chapter includes the basic elements of functional analysis which are needed in discussing partially observable linear systems in separable Hilbert spaces.

1.1 Sets and Functions

The basic objects of mathematics such as sets and functions as well as some common notation are introduced in this section.

1.1.1 Sets and Quotient Sets

The concept of *set* is a primitive concept of mathematics and, therefore, taken as undefined. The words *class*, *system*, *collection* and *family* are used synonymously with set. Each set, except the *empty set* $\varnothing$, consists of its *elements*. The *membership* of the element x in the set A is indicated by $x \in A$ or $A \ni x$. In this case we say that x *belongs* to A. If x is not an element of A, then we write $x \notin A$. The expression $\{x : R(x)\}$ denotes the set of all x for which the statement R involving x is true. Given two sets A and B, A is called a *subset* (or, synonymously, *subclass*, *subsystem* etc.) of B if each element of A is an element of B; this is indicated by $A \subset B$ or $B \supset A$. In this case we also say that A is *included* in B or B *contains* A. Two sets A and B are said to be *equal* if $A \subset B$ and $B \subset A$; this is indicated by $A = B$. Two sets are said to be *disjoint* if they have no elements in common.

The *union* of a family of sets is the set of all elements belonging to at least one of these sets. The *intersection* of a family of sets is the set of all common elements of these sets. The symbols $A \cup B$ and $A \cap B$ are used for the union and for the intersection of the sets A and B, respectively. It is convenient to use the symbols $\bigcup_\alpha A_\alpha$ and $\bigcap_\alpha A_\alpha$, respectively, for the union and for the intersection of

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

the family $\{A_\alpha\}$ of sets. The *difference* of the sets A and B is the set

$$A \setminus B = \{x : x \in A,\ x \notin B\}.$$

If $B \subset A$, then $A \setminus B$ is called the *complement* of B in A.

Sometimes, we use the *universal* and *existential quantifiers* $\forall$ and $\exists$ instead of *for all* and *exists*, respectively. Similarly, the *implication* $\Rightarrow$ and the *logical equivalence* $\Leftrightarrow$ replace *follows* and *if and only if*, respectively.

A binary relation $\sim$ between the elements of a set A is called an *equivalence relation* if for all $x, y, z \in A$,

(a) $x \sim x$ (reflexivity);

(b) $x \sim y \ \Rightarrow\ y \sim x$ (symmetry);

(c) $x \sim y,\ y \sim z \ \Rightarrow\ x \sim z$ (transitivity).

An equivalence relation on a set A splits A into mutually disjoint *equivalence classes* of *equivalent elements*. The collection of all these equivalence classes is called a *quotient set* of A. Generally, the same notation is used for a set and for its quotient set, and $x \in A$ represents both the element x of the set A and the equivalence class of the quotient set A containing the element x.

1.1.2 Systems of Numbers and Cardinality

We suppose that the reader is familiar with the basic systems of numbers. Briefly recall that the *system of counting numbers* (or *positive integers*) is

$$\mathbb{N} = \{1, 2, 3, \ldots\},$$

in which the ordinary order and the ordinary algebraic operations of addition and multiplication are defined. Using the system $\mathbb{N}$, one can define the *system of rational numbers*

$$\mathbb{Q} = \{0, n/m, -n/m : n, m \in \mathbb{N}\}$$

and extend the ordinary order and the ordinary algebraic operations to all rational numbers. $\mathbb{Q}$ is an ordered field, but it does not satisfy the least upper bound property. To improve $\mathbb{Q}$, the *system of real numbers* $\mathbb{R}$ is defined as the completion of $\mathbb{Q}$. $\mathbb{R}$ is a unique ordered field having the least upper bound property and containing $\mathbb{Q}$ as its subfield. The numbers in $\mathbb{R} \setminus \mathbb{Q}$ are called *irrational numbers.* A disadvantage of $\mathbb{R}$ is the nonexistence in $\mathbb{R}$ of square roots of negative numbers. For this, $\mathbb{R}$ is extended up to the *system of complex numbers* $\mathbb{C}$ which is a field containing $\mathbb{R}$ as its subfield. While any complex number has a square root in $\mathbb{C}$, $\mathbb{C}$ is not an ordered field and, hence, the least upper bound property does not make sense in $\mathbb{C}$. Often $\mathbb{R}$ is called the *real line* as well. An extensive discussion of number systems can be found in Rudin [85].

Often, we will use the symbols $i = 1, 2, \ldots$ and $i = 1, \ldots, n$ to show that i varies in $\mathbb{N}$ and in $\{1, 2, \ldots, n\}$, respectively. The *least upper bound* and the *greatest lower bound* of a bounded set $A \subset \mathbb{R}$ will be denoted by $\sup A$ and $\inf A$, respectively. If $\sup A$ ($\inf A$) is an element of A, then $\max A = \sup A$ ($\min A = \inf A$). The symbol $|a|$ will denote the *absolute value* of $a \in \mathbb{R}$. For the *open*, *closed* and *half-closed intervals* in $\mathbb{R}$, we will use the notation

$$\begin{array}{ll} (a,b) = \{t \in \mathbb{R} : a < t < b\}, & (a,\infty) = \{t \in \mathbb{R} : a < t < \infty\}, \\ [a,b] = \{t \in \mathbb{R} : a \le t \le b\}, & [a,\infty) = \{t \in \mathbb{R} : a \le t < \infty\}, \\ [a,b) = \{t \in \mathbb{R} : a \le t < b\}, & (-\infty,b) = \{t \in \mathbb{R} : -\infty < t < b\}, \\ (a,b] = \{t \in \mathbb{R} : a < t \le b\}, & (-\infty,b] = \{t \in \mathbb{R} : -\infty < t \le b\}, \end{array}$$

where $-\infty < a < b < \infty$ and ∞ is a notation for *infinity.* Also, for $T > 0$, we denote $\mathbf{T} = [0, T]$ and

$$\Delta_T = \{(t, s) : 0 \le s \le t \le T\}.$$

If the number of elements in a set is finite, then this set is said to be *finite*. Otherwise, it is said to be *infinite*. Note that the intervals (a, b), $[a, b]$, $[a, b)$ and $(a, b]$, as defined above for $-\infty < a < b < \infty$, are also called *finite intervals* while they are infinite sets. Among infinite sets those are simplest which have a one-one onto correspondence with $\mathbb{N}$. These sets are said to be *countable*. A finite or countable set is said to be *at most countable*. If a set is not at most countable, then it is said to be *uncountable*. $\mathbb{Q}$ is a countable set, but $\mathbb{R}$ and any of intervals (a, b), $[a, b]$, $[a, b)$ and $(a, b]$ in $\mathbb{R}$ for $a < b$ are uncountable sets.

Generally, two sets are said to have the same *cardinality* if there exists a one-one onto correspondence between them. One can mark the sets of the same cardinality by a symbol and call it the *cardinal number* of these sets. The cardinal number of a finite set is the number of elements in this set. A countable set has the cardinality of $\mathbb{N}$ which is called the *countable cardinality.*

1.1.3 Systems of Sets

Infinite (especially, uncountable) sets have too many subsets. It becomes necessary to single out these subsets in systems with some useful properties.

A system Σ of subsets of a nonempty set S is called a *σ-algebra* (with the *unit* S) if

(a) $S \in \Sigma$;

(b) $A \in \Sigma \Rightarrow S \setminus A \in \Sigma$;

(c) $A_1, A_2, \ldots \in \Sigma \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \Sigma$.

The equalities

$$A \setminus B = (S \setminus B) \setminus (S \setminus A), \quad \bigcap_{n=1}^{\infty} A_n = S \setminus \bigcup_{n=1}^{\infty} (S \setminus A_n)$$

show that any σ-algebra is closed under set differences and under countable intersections.

Given a σ-algebra Σ with a unit S, each σ-algebra Σ' with the unit S and such that $\Sigma' \subset \Sigma$ is called a *sub-σ-algebra* of Σ. It is easy to prove that the intersection of any family of σ-algebras is again a σ-algebra. Therefore, the following useful concept can be introduced: given a system Σ of sets, the intersection of all σ-algebras containing Σ is called the *smallest σ-algebra generated by* Σ and it is denoted by $\sigma(\Sigma)$.

A weaker concept than σ-algebra is semialgebra. A system Σ of subsets of a nonempty set S is called a *semialgebra* if

(a) $S \in \Sigma$;

(b) $A, B \in \Sigma \;\Rightarrow\; A \cap B \in \Sigma$;

(c) $A \in \Sigma$ implies that there are sets $A_1, \dots, A_k \in \Sigma$, where $k \in \mathbb{N}$, such that $A_n \cap A_m = \varnothing$ for $n \neq m$ and $S \setminus A = \bigcup_{n=1}^{k} A_n$.

In particular, each σ-algebra is a semialgebra, but not conversely.

1.1.4 Functions and Sequences

Given two sets X and Y, a *function from X to Y* is a rule f that assigns to each element of X a unique element of Y. The words *transformation*, *mapping*, *correspondence* are used synonymously with function. A function from X to Y is denoted in the forms f, $f(\cdot)$, $f : X \to Y$ or f_x, $x \in X$. By $f(x)$ or f_x we denote the *value* of the function f at x. The sets X and $\{f_x : x \in X\}$ are called the *domain* and the *range* of the function $f : X \to Y$ and they are denoted by $D(f)$ and $R(f)$, respectively. For $A \subset X$ and $B \subset Y$, the sets

$$f(A) = \{f_x : x \in A\} \text{ and } f^{-1}(B) = \{x \in X : f_x \in B\}$$

are called the *image* of A and the *inverse image* of B under $f : X \to Y$, respectively. Obviously, $f(X) \subset Y$ and $f^{-1}(Y) = X$. A function f with $R(f) \subset \mathbb{R}$ is called a *real-valued function* or a *functional*.

Given $f : X \to Y$ and $g : Y \to Z$, the function $h : X \to Z$, defined by $h(x) = g(f(x))$, $x \in X$, is called the *composition* of g and f and it is denoted by $g \circ f$. The *characteristic function* of the set A is the function, denoted by χ_A, which satisfies $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$. The *restriction* of the function $f : X \to Y$ to the set $A \subset X$ is denoted by $f|_A$.

A function $f : X \to Y$ is said to be *one-one* if for all $x_1, x_2 \in X$, $f(x_1) = f(x_2)$ implies $x_1 = x_2$. It is said to be *onto* if $R(f) = Y$. For a one-one onto function $f : X \to Y$, it is possible to define a unique one-one onto function $f^{-1} : Y \to X$ satisfying $f^{-1}\big(f(x)\big) = x$ for all $x \in X$ and $f\big(f^{-1}(y)\big) = y$ for all $y \in Y$. This function is said to be the *inverse* of f. Obviously, $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$ if f^{-1} and g^{-1} exist and the composition $g \circ f$ is defined. Also, $(f^{-1})^{-1} = f$.

One general property of functions is the following: if X and Y are nonempty sets and Σ is a σ-algebra with the unit Y, then the system $\{f^{-1}(A) : A \in \Sigma\}$ is a σ-algebra with the unit X. Moreover, if Σ is the smallest σ-algebra generated by a system Σ_0 of sets, then $\{f^{-1}(A) : A \in \Sigma\}$ is the smallest σ-algebra generated by the system $\{f^{-1}(A) : A \in \Sigma_0\}$.

A *sequence* is a function which has a countable domain. Sequences are denoted in the form $\{x_n\}$ where it is supposed that n varies in a countable set. This notation will be used for functions with a finite domain as well, it being clear from the context what is meant.

1.2 Abstract Spaces

An *abstract space* is a nonempty set endowed with a certain structure. The useful structures are linear and metric structures which lead to linear and metric spaces, respectively. Banach, Hilbert and Euclidean spaces are endowed with both these structures. Other abstract spaces considered in this section are measurable and measure spaces. They are based on the concept of subset.

1.2.1 Linear Spaces

A nonempty set X is called a *linear space* (or *vector space*) *over* $\mathbb{R}$ if the algebraic operations of *addition* and *multiplication with real numbers* on the elements of X, denoted by $x + y$ and ax for $x, y \in X$ and for $a \in \mathbb{R}$, respectively, are defined such that the following axioms hold:

(a) $\forall x, y \in X$ and $\forall a \in \mathbb{R}$, $x + y \in X$ and $ax \in X$ (closedness);

(b) $\forall x, y \in X,\ x + y = y + x$ (commutativity);

(c) $\forall x, y, z \in X,\ (x + y) + z = x + (y + z)$ (associativity);

(d) $\exists 0 \in X$ such that $\forall x \in X,\ x + 0 = x$ (existence of zero);

(e) $\forall x \in X,\ \exists(-x) \in X$ such that $x + (-x) = 0$ (existence of negative);

(f) $\forall x, y \in X$ and $\forall a \in \mathbb{R}$, $a(x + y) = ax + ay$ (distributivity);

(g) $\forall x \in X$ and $\forall a, b \in \mathbb{R}$, $(a + b)x = ax + bx$ (distributivity);

(h) $\forall x \in X$ and $\forall a, b \in \mathbb{R}$, $a(bx) = (ab)x$ (associativity);

(i) $\forall x \in X,\ \ 1x = x$ (property of unit).

Note that in functional analysis, linear spaces over any other field, say, over the field of complex numbers $\mathbb{C}$ are being considered too. The definition of these spaces differs from the above mentioned definition only by replacing $\mathbb{R}$ by the corresponding field. Since we will consider only the linear spaces over $\mathbb{R}$, they will briefly be called *linear spaces*. An element of a linear space is often called a *vector*.

A vector $a_1x_1 + \cdots + a_nx_n$, where $n \in \mathbb{N}$ and $a_1, \ldots, a_n \in \mathbb{R}$, is called a *linear combination* of the vectors $x_1, \ldots, x_n$ in a given linear space. A subset G of a linear space is said to be *linearly independent* if the equalities $a_1 = \cdots = a_n = 0$ hold whenever $a_1x_1 + \cdots + a_nx_n = 0$ for each finite number of vectors $x_1, \ldots, x_n \in G$. Otherwise, G is said to be *linearly dependent*. A linear space is said to be *k-dimensional* (with either $k = 0$ or $k \in \mathbb{N}$) if it has k linearly independent vectors and each collection of its $k+1$ vectors is linearly dependent. Obviously, the zero-dimensional linear space contains only the zero vector. If for each $n \in \mathbb{N}$ a linear space contains n linearly independent vectors, then it is said to be *infinite dimensional*. The *dimension* of the linear space X is denoted by $\dim X$.

A subset of a linear space X is called a *linear subspace* if it is closed under the algebraic operations defined on X. If G is a subset of a linear space X, then the set of all linear combinations of vectors in G defines a linear subspace of X. This subspace is denoted by $\operatorname{span} G$ and it is called the *linear subspace spanned by* G.

Example 1.1. A simple example of a linear space is the real line $\mathbb{R}$ with the algebraic operations being the ordinary addition and the ordinary multiplication of real numbers.

Example 1.2. The set of all k-vectors

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix},$$

where $k \in \mathbb{N}$ and $x_1, \ldots, x_k \in \mathbb{R}$, is denoted by $\mathbb{R}^k$. This set is a k-dimensional linear space with the *componentwise algebraic operations*, i.e.,

$$\begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_k + y_k \end{bmatrix}, \quad a \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} ax_1 \\ \vdots \\ ax_k \end{bmatrix}.$$

Example 1.3. The set of all bounded sequences $x = \{x_n\}$ of real numbers is denoted by l_∞. This set is an infinite dimensional linear space with the *termwise algebraic operations*, i.e.,

$$\{x_n\} + \{y_n\} = \{x_n + y_n\}, \ a\{x_n\} = \{ax_n\}.$$

1.2.2 Metric Spaces

A nonempty set X is called a *metric space* if the real number $d(x, y)$, called the *metric* between x and y, is assigned to each pair of elements $x, y \in X$ such that the following axioms hold:

(a) $\forall x, y \in X,\ d(x, y) \geq 0$ (nonnegativity);

(b) $d(x, y) = 0 \ \Leftrightarrow \ x = y$ (nondegeneracy);

(c) $\forall x, y \in X,\ d(x, y) = d(y, x)$ (symmetry);

(d) $\forall x, y, z \in X,\ d(x, y) \leq d(x, z) + d(y, z)$ (triangle inequality).

An element of a metric space is also called a *point*.

A sequence $\{x_n\}$ in a metric space X is said to *converge* to a point $x \in X$ and, the point x is called the *limit* of $\{x_n\}$, if $d(x_n, x) \to 0$. This convergence is expressed by writing $\lim_{n\to\infty} x_n = x$ or $x_n \to x$. In view of the nondegeneracy axiom, the limit of a convergent sequence is unique. A sequence $\{x_n\}$ in a metric space is said to be *Cauchy* if $d(x_n, x_m) \to 0$ as $n, m \to \infty$. Each convergent sequence is Cauchy. But, the converse may not be true. A metric space is said to be *complete* if each of its Cauchy sequences converges.

Given a subset G of a metric space X, a point $x \in X$ is called a *limit point* of G if there exists a sequence $\{x_n\}$ in $G \setminus \{x\}$ such that $x_n \to x$. If G contains all its limit points, then it is said to be *closed.* A subset of a metric space X is said to be *open* if it is the complement in X of some closed set. The collection of all open subsets of a metric space X is called the *metric topology* of X and it is denoted by $\tau(X)$.

The intersection of any family (as well as the union of a finite family) of closed sets is again a closed set. Similarly, the union of any family (as well as the intersection of a finite family) of open sets is again an open set. Therefore, with a given subset G of a metric space, one can associate the closed set, denoted by $\overline{G}$, which is the intersection of all closed sets containing G, and the open set, denoted by G^0, which is the union of all open sets contained in G. The sets $\overline{G}$ and G^0 are called the *closure* and the *interior* of G, respectively. $\overline{G}$ is the smallest closed set and G^0 is the largest open set satisfying $G^0 \subset G \subset \overline{G}$.

The subsets of a metric space X of the forms $\{x \in X : d(x, x_0) < r\}$ and $\{x \in X : d(x, x_0) \leq r\}$ are called the *open* and *closed balls* with *radius* $r > 0$ centered at $x_0 \in X$ and they provide simple examples of open and closed subsets of X, respectively. A subset of a metric space is said to be *bounded* if it is contained in some ball.

A subset G of a metric space X is said to be *dense* in X if $\overline{G} = X$. A metric space is said to be *separable* if it contains a countable dense subset.

A metric space is said to be *compact* if each of its infinite subsets has a limit point. A compact metric space is necessarily complete and separable.

Each nonempty subset G of a metric space X is again a metric space with respect to the metric of X and it is called a *metric subspace* of X. If X is complete and G is closed, then G is a complete metric subspace of X.

Example 1.4. The real line $\mathbb{R}$ is a complete and separable metric space with the metric being the absolute value of the difference of real numbers. But, it is not compact. A subset of $\mathbb{R}$ is compact if and only if it is bounded and closed.

1.2.3 Banach Spaces

A linear space X is called a *normed space* if the real number $\|x\|$, called the *norm* of x, is assigned to each $x \in X$ such that the following axioms hold:

(a) $\forall x \in X,\ \|x\| \geq 0$ (nonnegativity);

(b) $\|x\| = 0 \Leftrightarrow x = 0$ (nondegeneracy);

(c) $\forall x \in X$ and $\forall a \in \mathbb{R},\ \|ax\| = |a|\ \|x\|$ (positive homogeneity);

(d) $\forall x, y \in X,\ \|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).

Each normed space is a metric space with the metric $d(x, y) = \|x-y\|$. Convergence with respect to this metric is called *convergence in norm* or *strong convergence* or simply *convergence*. A complete normed space is called a *Banach space*.

Presence of linear and metric structures in a Banach space X allows one to consider the sum of countably many vectors $x_n \in X$, $n = 1, 2, \dots$, called a *series* and denoted by $\sum_{n=1}^{\infty} x_n$. A series $\sum_{n=1}^{\infty} x_n$ is said to *converge* to a vector x and the vector x is called the *sum* of the series $\sum_{n=1}^{\infty} x_n$ if the sequence of partial sums $S_k = \sum_{n=1}^{k} x_n$ converges to x as $k \to \infty$. This convergence is expressed by writing $x = \sum_{n=1}^{\infty} x_n$. A series $\sum_{n=1}^{\infty} x_n$ is said to *converge absolutely* if the series $\sum_{n=1}^{\infty} \|x_n\|$ of real numbers is convergent. Obviously, an absolutely convergent series is convergent.

Example 1.5. The linear space l_∞ (see Example 1.3) is a nonseparable Banach space with the the norm

$$\|\{x_n\}\|_{l_\infty} = \sup_n |x_n|.$$

Example 1.6. Let $1 \leq p < \infty$. The set of all sequences $\{x_n\}$ in $\mathbb{R}$ satisfying $\sum_{n=1}^{\infty} |x_n|^p < \infty$ is denoted by l_p. This set is a separable Banach space with the termwise algebraic operations (see Example 1.3) and with the norm

$$\|\{x_n\}\|_{l_p} = \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}.$$

For $\{x_n\}, \{y_n\} \in l_p$, the triangular inequality yields

$$\left(\sum_{n=1}^{\infty} |x_n + y_n|^p\right)^{1/p} \leq \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p} + \left(\sum_{n=1}^{\infty} |y_n|^p\right)^{1/p},$$

which is called the *Minkowski inequality for sums*.

1.2.4 Hilbert and Euclidean Spaces

A linear space X is called a *scalar product space* if the number $\langle x, y\rangle$, called the *scalar product* of x and y, is assigned to each pair of vectors $x, y \in X$ such that the following axioms hold:

(a) $\forall x \in X,\ \langle x, x\rangle \geq 0$ (nonnegativity);

(b) $\langle x, x\rangle = 0 \ \Leftrightarrow\ x = 0$ (nondegeneracy);

(c) $\forall x, y \in X,\ \langle x, y\rangle = \langle y, x\rangle$ (symmetry);

(d) $\forall x, y, z \in X,\ \langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ (additivity);

(e) $\forall x \in X$ and $\forall a \in \mathbb{R},\ \langle ax, y\rangle = a\langle x, y\rangle$ (homogeneity).

Each scalar product space is a normed space with the norm defined by $\|x\| = \langle x, x\rangle^{1/2}$. If a scalar product space is complete with respect to convergence in this norm, then it is called a *Hilbert space*.

The scalar products and the norms in all Hilbert and Banach spaces will be denoted by $\langle \cdot, \cdot\rangle$ and $\|\cdot\|$, respectively. In ambiguous cases, the subscript will indicate which scalar product (norm) is meant.

Each linear subspace L of a Hilbert (Banach) space X is a scalar product (normed) space with respect to the scalar product (norm) of X. The closure of L in X is a Hilbert (Banach) space which is called a *subspace* of X. If G is a nonempty subset of X, then $\overline{\operatorname{span} G}$ is called the *subspace spanned by* G.

An important concept in Hilbert spaces is the concept of orthogonality. Two vectors x and y in a Hilbert space X are said to be *orthogonal* if $\langle x, y\rangle = 0$. If H is a subspace of X, then the set

$$H^{\perp} = \{x \in X : \langle x, h\rangle = 0 \text{ for all } h \in H\}$$

is called the *orthogonal complement* of H in X. $H^{\perp}$ is a subspace of X and each $x \in X$ can be uniquely represented in the form $x = y + z$, where $y \in H$ and $z \in H^{\perp}$. Moreover,

$$\|x\|^2 = \|y\|^2 + \|z\|^2 \ (\textit{generalized Pythagorean theorem}).$$

In particular, $X^{\perp} = \{0\}$, i.e., if $\langle x, h\rangle = 0$ for all $h \in X$, then $x = 0$.

A system $\{e_\alpha : \alpha \in A\}$ of nonzero vectors in a Hilbert space X is called an *orthogonal system* if $\langle e_\alpha, e_\beta\rangle = 0$ for all $\alpha, \beta \in A$ with $\alpha \neq \beta$. It is possible to show that any orthogonal system is linearly independent. An orthogonal system $\{e_\alpha : \alpha \in A\}$ is said to be *orthonormal* if $\|e_\alpha\| = 1$ for all $\alpha \in A$. An orthogonal system in a Hilbert space X is said to be *complete* if the subspace spanned by it coincides with X. A complete orthonormal system in a Hilbert space X is called a *basis* of X. All bases of a given Hilbert space have the same cardinality.

Theorem 1.7. *A Hilbert space is separable if and only if it has an at most countable basis. If* $\{e_n\}$ *is a basis in a separable Hilbert space* X*, then for each* $x \in X$*,*

$$x = \sum_{n=1}^{\dim X} \langle x, e_n \rangle e_n \tag{1.1}$$

and

$$\|x\|^2 = \sum_{n=1}^{\dim X} \langle x, e_n \rangle^2. \tag{1.2}$$

The right-hand side of (1.1) is called the *Fourier series* of $x \in X$ and (1.2) the *Parseval identity.*

If a Hilbert space has a finite dimension $k \in \mathbb{N}$, then it is called a *k-dimensional Euclidean space.*

Example 1.8. The real line $\mathbb{R}$ is a one-dimensional Euclidean space with the scalar product equaling the product of two real numbers.

Example 1.9. The linear space $\mathbb{R}^k$ (see Example 1.2) is a k-dimensional Euclidean space with the scalar product

$$\left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} \right\rangle = \sum_{n=1}^{k} x_n y_n.$$

Example 1.10. The space l_2 (see Example 1.6) is an infinite dimensional separable Hilbert space with the scalar product

$$\langle \{x_n\}, \{y_n\} \rangle = \sum_{n=1}^{\infty} x_n y_n.$$

The more complicated Hilbert and Banach spaces are the operator and functional spaces which will be considered later.

The separable Hilbert spaces are the main spaces we will consider. We denote the class of all separable Hilbert spaces by $\mathcal{H}$. Thus, $\mathcal{H}$ consists of all finite dimensional Euclidean spaces as well as all infinite dimensional separable Hilbert spaces.

1.2.5 Measurable and Borel Spaces

A pair (S, Σ), where S is a nonempty set and Σ is a σ-algebra with the unit S, is called a *measurable space*. It is easily seen that $\Sigma_* = \{S, \varnothing\}$ and the system Σ^* of all subsets of S are the smallest and largest σ-algebras of subsets of S, respectively. The most useful σ-algebras are the so-called Borel σ-algebras which are related with a metric (or, more generally, topological) structure in S.

Let X be a metric space with the metric topology $\tau(X)$. The smallest σ-algebra generated by $\tau(X)$ is called the *Borel σ-algebra* of subsets of X and it is denoted by $\mathcal{B}_X$. An element of $\mathcal{B}_X$ is called a *Borel measurable set*. Since each open set is the complement of some closed set, $\mathcal{B}_X$ is the smallest σ-algebra generated by the system of all closed subsets of X as well. $\mathcal{B}_X$ is also the smallest σ-algebra generated by all open balls in X if X is separable. A measurable space $(X, \mathcal{B}_X)$ is called a *Borel space*.

Given a function f from a set S to a metric space X, the σ-algebra

$$\sigma(f) = \{f^{-1}(B) : B \in \mathcal{B}_X\}$$

is called as the *σ-algebra generated by f*. If $\{f^\alpha : \alpha \in A\}$ is a family of functions from the same set S to possibly distinct metric spaces, then we set

$$\sigma(f^\alpha; \alpha \in A) = \sigma\Big(\bigcup_{\alpha\in A} \sigma(f^\alpha)\Big)$$

and call it the *σ-algebra generated by f^α*, $\alpha \in A$.

1.2.6 Measure and Probability Spaces

Given a semialgebra Σ, a function ν from Σ to $[0, \infty)$ is called a (*positive and finite*) *measure* if it is σ-additive, i.e., if

$$\nu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \nu(A_n)$$

whenever

$$A_1, A_2, \ldots \in \Sigma,\ A_n \cap A_m = \varnothing \text{ for } n \neq m,\ \bigcup_{n=1}^{\infty} A_n \in \Sigma.$$

A (positive and finite) measure, defined on a semialgebra Σ, can be uniquely extended to the σ-algebra $\sigma(\Sigma)$ as a σ-additive set function with values in $[0, \infty)$. Note that in measure theory, measures with negative as well as infinite values are allowed. Since we will consider only positive and finite measures, they will briefly be called *measures*.

A triple of objects (S, Σ, ν) is called a *measure space* if (S, Σ) is a measurable space and ν is a measure defined on Σ. A property $R(s)$ is said to hold *ν-almost everywhere on S* (briefly, *ν-a.e. on S*), if there exists $G \in \Sigma$ with $\nu(G) = 0$ such that $R(s)$ holds for all $s \in S \setminus G$. Sometimes, instead of ν-a.e. on S we write ν-a.e. $s \in S$ and read this as *ν-almost each $s \in S$*.

Let (S, Σ, ν) be a measure space and denote

$$\tilde{\Sigma} = \{A \cup B : A \in \Sigma,\ B \subset C \in \Sigma,\ \nu(C) = 0\}.$$

Define the function $\tilde{\nu} : \tilde{\Sigma} \to [0, \infty)$ by

$$\tilde{\nu}(A \cup B) = \nu(A), \ A \in \Sigma, \ B \subset C \in \Sigma, \ \nu(C) = 0.$$

One can show that $\tilde{\Sigma}$ is a σ-algebra, Σ is a sub-σ-algebra of $\tilde{\Sigma}$ and $\tilde{\nu}$ is a measure, defined on $\tilde{\Sigma}$, which is equal to ν on Σ. $\tilde{\Sigma}$ is called the *Lebesgue extension of* Σ *with respect to* ν. The measure $\tilde{\nu}$ is said to be *complete* because the Lebesgue extension of $\tilde{\Sigma}$ with respect to $\tilde{\nu}$ coincides with $\tilde{\Sigma}$. We use the same notation for the measures ν and $\tilde{\nu}$. A measure space (S, Σ, ν) is said to be *complete* if ν is a complete measure on Σ.

A measure space (S, Σ, ν) is said to be *separable* if Σ contains an at most countable subsystem Σ_0 such that for each $\varepsilon > 0$ and for each $A \in \Sigma$, there is a set $B \in \Sigma_0$ with $\nu((A \setminus B) \cup (B \setminus A)) < \varepsilon$. Note that, if Σ is the smallest σ-algebra generated by an at most countable semialgebra, then for any measure ν, defined on Σ, the measure space (S, Σ, ν) is separable.

Let (S, Σ, ν) be a measure space. If S_0 is a nonempty set in Σ, then $\Sigma_0 = \{B \in \Sigma : B \subset S_0\}$ is a σ-algebra and $\nu_0(B) = \nu(B)$, $B \in \Sigma_0$, is a measure. The measure ν_0 is called the *restriction* of the measure ν to Σ_0. Similarly, (S_0, Σ_0, ν_0) is called a *restriction* of the measure space (S, Σ, ν).

Example 1.11. Consider a finite and half-closed interval $(a, b] \subset \mathbb{R}$ where $a < b$. The collection $\mathcal{R}_{a,b}$ of all sets in the form $(\alpha, \beta]$, where α and β are rational numbers satisfying $a \le \alpha \le \beta \le b$ (the values $\alpha = a$ and $\beta = b$ are admitted even if a and b are irrational), defines a countable semialgebra of subsets of $(a, b]$. The smallest σ-algebra generated by $\mathcal{R}_{a,b}$ coincides with $\mathcal{B}_{(a,b]}$. Define the measure ℓ on $\mathcal{R}_{a,b}$ by $\ell((\alpha, \beta]) = \beta - \alpha$. The extension of ℓ to $\mathcal{B}_{(a,b]}$ is called the *one-dimensional Lebesgue measure* on $(a, b]$. The measure space $((a, b], \mathcal{B}_{(a,b]}, \ell)$ is separable because $\mathcal{R}_{a,b}$ is a countable semialgebra.

In general, a nondecreasing and right-continuous real-valued function f on $[a, b]$ generates the measure $\nu : \mathcal{B}_{(a,b]} \to [0, \infty)$ by $\nu((\alpha, \beta]) = f_\beta - f_\alpha$, where $a \le \alpha \le \beta \le b$, which is called the *Lebesgue–Stieltjes measure* on $(a, b]$ generated by f. The Lebesgue measure on $(a, b]$, as a particular case of Lebesgue–Stieltjes measures, is generated by $f_x = x$, $a \le x \le b$. The difference $h_x = f_x - g_x$, $a \le x \le b$, of two nondecreasing and right-continuous real-valued functions f and g is called a *function of bounded variation*. It has the property

$$\sup \sum_{i=0}^{n} |h_{x_{i+1}} - h_{x_i}| < \infty,$$

where it is assumed that the supremum is taken over all finite partitions

$$a = x_0 < \cdots < x_{n+1} = b$$

of $[a, b]$. This property means that the graph of a function of bounded variation has a finite length.

A measure space (S, Σ, ν) is called a *probability space* if $\nu(S) = 1$. In the sequel, we suppose that a fixed complete probability space is given and use the traditional notation $(\Omega, \mathcal{F}, \mathbf{P})$ for it. The set Ω is called a *sample space*. The elements of Ω are denoted by ω and they are called *samples* or *elementary events*. The elements of $\mathcal{F}$ are called *events*. $\mathbf{P}$ is called a *probability measure* or, briefly, a *probability*. For $A \in \mathcal{F}$, the number $\mathbf{P}(A)$ is the *probability of the event A*. If some property holds $\mathbf{P}$-a.e. on Ω, then we say that it holds *with probability 1* or, briefly, *w.p.*1.

1.2.7 Product of Spaces

Starting from some spaces one can construct other spaces. A product of spaces is one of these constructions.

The *product of the sets X and Y* is the collection of all ordered pairs (x, y) where $x \in X$ and $y \in Y$; it is denoted by $X \times Y$:

$$X \times Y = \{(x, y) : x \in X, \ y \in Y\}.$$

If both X and Y are endowed with a certain structure, then this structure can be extended to $X \times Y$. In this case $X \times Y$ is called a *product of spaces*

The *product of the linear spaces X and Y* is the linear space $X \times Y$ with the algebraic operations

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}, \quad a \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax \\ ay \end{bmatrix},$$

where $x, x_1, x_2 \in X$, $y, y_1, y_2 \in Y$ and $a \in \mathbb{R}$. Note that it is more convenient to represent the elements of a product of linear spaces as columns.

If X and Y are metric spaces with the metrics d_1 and d_2, respectively, then a metric in $X \times Y$ can be defined in various topologically equivalent forms. Among them we will prefer

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{d_1(x_1, x_2)^2 + d_2(y_1, y_2)^2},$$

where $x_1, x_2 \in X$ and $y_1, y_2 \in Y$. With this metric, $X \times Y$ is called the *product of the metric spaces X and Y*.

Similarly, if X and Y are Banach spaces, then $X \times Y$ is a linear space with the algebraic operations of the product of the linear spaces X and Y in which a norm can be defined in various topologically equivalent forms. We will prefer the norm

$$\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \sqrt{\|x\|^2 + \|y\|^2},$$

where $x \in X$ and $y \in Y$. With this norm, $X \times Y$ is called the *product of the Banach spaces X and Y*.

The *product of the Hilbert spaces X and Y* is the Hilbert space $X \times Y$ with the algebraic operations of the product of the linear spaces X and Y and with the scalar product

$$\left\langle \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \right\rangle = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle,$$

where $x_1, x_2 \in X$ and $y_1, y_2 \in Y$.

One can observe that there is a consistency between the metric, the norm and the scalar product as defined above in products of spaces. For example, the norm in a product of Hilbert spaces, that is

$$\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \left\langle \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x \\ y \end{bmatrix} \right\rangle^{1/2} = \sqrt{\langle x, x \rangle + \langle y, y \rangle} = \sqrt{\|x\|^2 + \|y\|^2},$$

coincides with the norm in the respective product of Banach spaces.

The *product of the semialgebras Σ and Γ* is the semialgebra

$$\Sigma \times \Gamma = \{A \times B : A \in \Sigma, \ B \in \Gamma\}.$$

If Σ and Γ are σ-algebras, then $\Sigma \times \Gamma$ is a semialgebra, but may not be a σ-algebra. So, the *product of the σ-algebras Σ and Γ* is defined as the smallest σ-algebra generated by $\Sigma \times \Gamma$ and it is denoted by $\Sigma \otimes \Gamma$, i.e.,

$$\Sigma \otimes \Gamma = \sigma(\{A \times B : A \in \Sigma, \ B \in \Gamma\}).$$

If ν and μ are measures on the σ-algebras Σ and Γ, respectively, then a measure $\nu \otimes \mu$ on the semialgebra $\Sigma \times \Gamma$ is defined by

$$(\nu \otimes \mu)(A \times B) = \nu(A)\mu(B), \ A \in \Sigma, \ B \in \Gamma,$$

which has a unique extension to $\Sigma \otimes \Gamma$. The measure $\nu \otimes \mu$ is called the *product of the measures ν and μ*.

The *product of the measurable spaces (S, Σ) and (R, Γ)* is the measurable space $(S \times R, \Sigma \otimes \Gamma)$. Similarly, the *product of the measure spaces (S, Σ, ν) and (R, Γ, μ)* is the measure space $(S \times R, \Sigma \otimes \Gamma, \nu \otimes \mu)$.

The product of separable metric (Banach, Hilbert, measure) spaces is again a separable space. The equality $\mathcal{B}_{X \times Y} = \mathcal{B}_X \otimes \mathcal{B}_Y$ holds for separable metric spaces X and Y.

Since the products of sets, σ-algebras and measures are associative, for the finite products of more than two sets, σ-algebras and measures, respectively, one can use the notation

$$X \times Y \times \cdots \times Z, \ \Sigma \otimes \Gamma \otimes \cdots \otimes \Pi, \ \nu \otimes \mu \otimes \cdots \otimes \lambda.$$

Among products of an infinite family of spaces we single out the linear space $F(S, X)$ of functions f from the set S to the linear space X with the *pointwise algebraic operations*, i.e., for all $f, g, \in F(S, X)$ and for all $a \in \mathbb{R}$,

$$(f + g)(s) = f(s) + g(s), \ (af)(s) = af(s), \ s \in S.$$

Example 1.12. The k-dimensional Euclidean space $\mathbb{R}^k$ (see Example 1.9) is the k-times product of the real line $\mathbb{R}$ by itself. Using this fact, a one-dimensional Lebesgue measure can be generalized in the following way. Let $(a,b]^2 = (a,b] \times (a,b]$. Since $\mathcal{B}_{(a,b]^2} = \mathcal{B}_{(a,b]} \otimes \mathcal{B}_{(a,b]}$, the measure $\ell \otimes \ell$ on $\mathcal{B}_{(a,b]^2}$ is defined. This measure is called the *two-dimensional Lebesgue measure* on $(a,b]^2$. For $(a,b]^k = (a,b] \times (a,b]^{k-1}$, the *$k$-dimensional Lebesgue measure* on $(a,b]^k$ can then be defined by induction for all $k \in \mathbb{N}$. We will denote all Lebesgue measures by ℓ. The distinction between them will follow from the context. Since $((a,b], \mathcal{B}_{(a,b]}, \ell)$ is a separable measure space, the same is true for $((a,b]^k, \mathcal{B}_{(a,b]^k}, \ell)$.

1.3 Linear Operators

A function operating from a linear space to a linear space is called an *operator*. In general, an operator may be defined on some subset of a linear space. An operator with the range contained in $\mathbb{R}$ is called a *functional*. In this section, linear operators will be discussed. The value of the linear operator A at x is denoted by Ax where the parentheses are dropped. An operator A is called a *linear operator* if $D(A)$ is a linear space and

$$\forall x, y \in D(A) \text{ and } \forall a, b \in \mathbb{R},\ A(ax+by) = aAx + bAy \text{ (linearity).}$$

The sum, the product with a real number, the composition and the inverse of linear operators (if they are defined) are again linear operators. We use the symbol AB (instead of $A \circ B$) for the composition of the linear operators A and B. A linear operator A can be uniquely decomposed in the form

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \tag{1.3}$$

if it operates from the product of linear spaces $X_1, \ldots, X_n$ to the product of linear spaces $Y_1, \ldots, Y_m$. Here A_{ij} is a linear operator from X_j to Y_i, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Note that the addition, the multiplication by a real number and the composition of decomposed linear operators obey the corresponding rules for matrices.

1.3.1 Bounded Operators

Let X and Y be Banach spaces. A linear operator A from $D(A) \subset X$ to Y is said to be *bounded* if

(a) $\overline{D(A)} = X$ (denseness of domain);

(b) $\exists c > 0$ such that $\forall x \in X$, $\|Ax\| \le c\|x\|$ (boundedness).

A bounded linear operator A, as defined above, can be uniquely extended to X preserving the linearity and boundedness properties. For a bounded linear operator A, the number

$$\|A\| = \sup_{\|x\|=1} \|Ax\|$$

is called the *norm* of A. The class of all bounded linear operators from X to Y defines a Banach space with respect to the above mentioned norm and it is denoted by $\mathcal{L}(X,Y)$. The brief notation $\mathcal{L}(X) = \mathcal{L}(X,X)$ is used as well. Note that, in general, $\mathcal{L}(X,Y)$ is neither a Hilbert space nor separable, even if $X, Y \in \mathcal{H}$.

The *identity operator* on a Banach space X is the operator I defined by $Ix = x$, $x \in X$. Irrespective of Banach space, the identity operators will be denoted by I. The distinction between them will follow from the context. The *zero operator* from a Banach space X to a Banach space Y is the operator assigning zero of Y to each vector of X. The symbol 0 will denote the number zero as well as zero vectors of all linear spaces including zero operators, it being clear from the context what is meant.

The composition BA of $B \in \mathcal{L}(Y,Z)$ and $A \in \mathcal{L}(X,Y)$, where X, Y and Z are Banach spaces, belongs to $\mathcal{L}(X,Z)$ and $\|BA\| \leq \|B\| \, \|A\|$. A linear operator A from the product of Banach spaces $X_1, \ldots, X_n$ to the product of Banach spaces $Y_1, \ldots, Y_m$, decomposed by (1.3), is bounded if and only if

$$A_{ij} \in \mathcal{L}(X_j, Y_i), \text{ for all } i = 1, \ldots, m, \text{ and for all } j = 1, \ldots, n.$$

The *uniform boundedness principle* stated in the following theorem is an important property of bounded linear operators.

Theorem 1.13. *Let X, Y be Hilbert spaces and let $\{A_\alpha\}$ be a family of bounded linear operators from X to Y satisfying* $\sup_\alpha \|A_\alpha x\| < \infty$ *for all $x \in X$. Then* $\sup_\alpha \|A_\alpha\| < \infty$.

The space of all bounded linear functionals, defined on a Banach space X, has a special name and a special notation. It is called the *dual space* of X and denoted by X^*, i.e., $X^* = \mathcal{L}(X, \mathbb{R})$.

A Banach space X is said to be *naturally embedded* into a Banach space Y if $X \subset Y$ and there exists a constant $c > 0$ such that the inequality $\|x\|_Y \leq c\|x\|_X$ holds for all $x \in X$. In this case, the operator J, defined by $Jx = x \in Y$, $x \in X$, is called an *embedding operator*. If, additionally, $\overline{X} = Y$, where $\overline{X}$ is the closure of X in Y, then X is said to be *tightly embedded* into Y.

Example 1.14. If $1 \leq p < q \leq \infty$, then $l_p \subset l_q$ is a natural embedding and

$$\forall \{x_n\} \in l_p, \ \|\{x_n\}\|_{l_q} \leq \|\{x_n\}\|_{l_p}.$$

The following proposition shows the relation between the Borel σ-algebras in a natural embedding of separable spaces.

Proposition 1.15. *If $X \subset Y$ is the natural embedding of the separable Banach spaces X and Y, then*

(a) $\forall A \in \mathcal{B}_Y,\ A \cap X \in \mathcal{B}_X$;

(b) $\mathcal{B}_X \subset \mathcal{B}_Y$.

Proof. Let J be the corresponding embedding operator. Then $A \cap X = J^{-1}(A)$ for all $A \subset Y$. So, part (a) follows from the continuity (discussed in Section 2.1.1) of J. Since, additionally, J is a one-one function, by the well-known theorem of Kuratowski (see Parthasarathy [80]), $J(A) = A \in \mathcal{B}_Y$ for all $A \in \mathcal{B}_X$. This proves part (b). □

1.3.2 Inverse Operators

Let X and Y be Banach spaces. A linear operator A from $D(A) \subset X$ to Y is said to have a *bounded inverse* if there exists $A^{-1} \in \mathcal{L}(Y, X)$. Obviously, a necessary and sufficient condition for a linear operator A from $D(A) \subset X$ to Y to have a bounded inverse is $R(A) = Y$ and $\|Ax\| \geq c\|x\|$ for some constant $c > 0$ and for all $x \in D(A)$.

The following theorem expresses a case when a bounded operator has a bounded inverse.

Theorem 1.16. *A bounded linear operator from a Banach space to a Banach space has a bounded inverse if and only if it is one-one and onto. Furthermore, if a bounded linear operator A has a bounded inverse, then $\|A^{-1}\| \geq \|A\|^{-1}$.*

Two Banach spaces X and Y are said to be *isomorphic* if there exists $J \in \mathcal{L}(X, Y)$ which has a bounded inverse. If, additionally, $\|x\| = \|Jx\|$ for all $x \in X$, then X and Y are said to be *isometric* and J an *isometry* from X onto Y. Mathematically, the isometric spaces are just different realizations of the same space. Often, the isometric spaces X and Y are identified and one writes $X = Y$.

Example 1.17. Any Banach space X and $\mathcal{L}(\mathbb{R}, X)$ are isometric under the isometry

$$X \ni x \leftrightarrow Jx \in \mathcal{L}(\mathbb{R}, X) : \ (Jx)a = ax, \ a \in \mathbb{R}.$$

This isometry is called the *natural isometry.* We identify these spaces and write $\mathcal{L}(\mathbb{R}, X) = X$.

Example 1.18. The spaces l_1^* and l_∞ are isometric under the isometry

$$l_1^* \ni f \leftrightarrow Jf = \{f_n\} \in l_\infty : \ f\{x_n\} = \sum_{n=1}^{\infty} f_n x_n, \ \{x_n\} \in l_1,$$

according to which we identify these spaces and write $l_1^* = l_\infty$.

Example 1.19. For $1 < p < \infty$ and $p^{-1}+q^{-1} = 1$, the spaces l_p^* and l_q are isometric under the isometry

$$l_p^* \ni f \leftrightarrow Jf = \{f_n\} \in l_q : \ f\{x_n\} = \sum_{n=1}^{\infty} f_n x_n, \ \{x_n\} \in l_p.$$

We identify these spaces and write $l_p^* = l_q$. In particular, this isometry yields the inequality

$$\forall \{x_n\} \in l_p, \ \forall \{f_n\} \in l_q, \ \sum_{n=1}^{\infty} |f_n x_n| \leq \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{1/p} \left(\sum_{n=1}^{\infty} |f_n|^q \right)^{1/q},$$

which is called the *Hölder inequality for sums.* When $p = q = 2$, this inequality has the form

$$\forall \{x_n\}, \{f_n\} \in l_2, \ \sum_{n=1}^{\infty} |f_n x_n| \leq \left(\sum_{n=1}^{\infty} x_n^2 \right)^{1/2} \left(\sum_{n=1}^{\infty} f_n^2 \right)^{1/2},$$

which is called the *Cauchy–Schwarz inequality for sums.*

The following theorem shows that in fact the collection of all distinct Hilbert spaces can be put into one-one correspondence with the collection of all cardinal numbers.

Theorem 1.20. *Two Hilbert spaces are isometric if and only if they have bases of the same cardinality. In particular, for fixed* $k \in \mathbb{N} \cup \{\infty\}$, *all* k*-dimensional separable Hilbert spaces are isometric.*

This theorem is the reason for the use of the common symbol $\mathbb{R}^k$ for all k-dimensional Euclidean spaces, identifying them with the k-dimensional Euclidean space from Example 1.9. But, in general, it is convenient to use distinct symbols for isometric infinite dimensional Hilbert spaces.

The next theorem presents another isometry.

Theorem 1.21 (Riesz). *Any Hilbert space* H *and its dual* H^* *are isometric under the isometry*

$$H^* \ni h^* \leftrightarrow Jh^* = h \in H : \ h^* x = \langle h, x \rangle, \ x \in H.$$

We will identify the Hilbert spaces H and H^*. Theorem 1.21 yields the inequality

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|,$$

for any two vectors x and y in a Hilbert space. This inequality is called the *Cauchy–Schwarz inequality* for abstract Hilbert spaces.

1.3.3 Closed Operators

Let X and Y be Banach spaces. A linear operator A from $D(A) \subset X$ to Y is said to be *closed* if

(a) $\overline{D(A)} = X$ (denseness of domain);

(b) for any sequence $\{x_n\}$ in $D(A)$, $x_n \to x$ and $Ax_n \to y$ imply $x \in D(A)$ and $Ax = y$ (closedness).

The class of all closed linear operators from a dense subset of X to Y will be denoted by $\tilde{\mathcal{L}}(X, Y)$. The brief notation $\tilde{\mathcal{L}}(X) = \tilde{\mathcal{L}}(X, X)$ will be used. Note that in $\tilde{\mathcal{L}}(X, Y)$ neither linear nor metric structures are defined. Obviously, $\mathcal{L}(X, Y) \subset \tilde{\mathcal{L}}(X, Y)$ and $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) = \tilde{\mathcal{L}}(\mathbb{R}^n, \mathbb{R}^m)$ for $n, m \in \mathbb{N}$. If $A \in \tilde{\mathcal{L}}(X, Y)$ and $D(A) = X$, then $A \in \mathcal{L}(X, Y)$.

The following theorem shows that any closed linear operator can be reduced to a bounded linear operator.

Theorem 1.22. *If $X, Y \in \mathcal{H}$ and $A \in \tilde{\mathcal{L}}(X, Y)$, then*

(a) *$D(A) \in \mathcal{H}$ with $\langle x, y\rangle_{D(A)} = \langle x, y\rangle_X + \langle Ax, Ay\rangle_Y$;*

(b) *$D(A) \subset X$ is a natural and tight embedding;*

(c) *$A \in \mathcal{L}(D(A), Y)$.*

1.3.4 Adjoint Operators

Let X and Y be Banach spaces. If $A \in \tilde{\mathcal{L}}(X, Y)$, then there exists a unique operator in $\tilde{\mathcal{L}}(Y^*, X^*)$, denoted by A^*, such that

$$\forall x \in D(A) \text{ and } \forall y^* \in D(A^*), \ (A^*y^*)x = y^*(Ax). \tag{1.4}$$

A^* is said to be the *adjoint* of A. If $A \in \mathcal{L}(X, Y)$, then $A^* \in \mathcal{L}(Y^*, X^*)$ and, therefore, (1.4) holds for all $x \in X$ and for all $y^* \in Y^*$.

Below, we suppose that X, Y and Z are Banach spaces and list some useful properties of adjoint operators.

(a) If $A \in \mathcal{L}(X, Y)$ and $B \in \mathcal{L}(Y, Z)$, then $BA \in \mathcal{L}(X, Z)$ and $(BA)^* = A^*B^*$.

(b) If $A \in \tilde{\mathcal{L}}(X, Y)$ and $B \in \mathcal{L}(X, Y)$, then $(A + B) \in \tilde{\mathcal{L}}(X, Y)$ and $(A + B)^* = A^* + B^*$.

(c) If $A \in \mathcal{L}(X, Y)$ has a bounded inverse, then A^* has a bounded inverse too and $(A^*)^{-1} = (A^{-1})^*$.

(d) If X and Y are Hilbert spaces and $A \in \tilde{\mathcal{L}}(X, Y)$, then $(A^*)^* = A$.

(e) If X and Y are Hilbert spaces, then $\mathcal{L}(X,Y)$ and $\mathcal{L}(Y^*,X^*)$ are isometric under the isometry

$$\mathcal{L}(X,Y) \ni A \leftrightarrow A^* \in \mathcal{L}(Y^*,X^*).$$

(f) If a bounded linear operator A from the product of Banach spaces to the product of Banach spaces is decomposed by (1.3), then

$$A^* = \begin{bmatrix} A_{11}^* & \cdots & A_{m1}^* \\ \vdots & \ddots & \vdots \\ A_{1n}^* & \cdots & A_{mn}^* \end{bmatrix}.$$

1.3.5 Projection Operators

Let H be a subspace of a Hilbert space X. It was mentioned earlier that each vector $x \in X$ can be uniquely represented in the form $x = y + z$ where $y \in H$ and $z \in H^\perp$. The operator P assigning to each $x \in X$ the vector $y \in H$ from the above mentioned representation is called the *projection operator* from X onto H. By definition, P is the projection operator from X onto H if and only if

$$\forall x \in X \text{ and } \forall h \in H,\ \langle x - Px, h\rangle = 0.$$

Obviously, $P \in \mathcal{L}(X,H)$, $P^2 = PP = P$ and $\|P\| = 1$.

An important property of Hilbert spaces is expressed in the following proposition.

Proposition 1.23. *Let X be a Hilbert space and let H be a subspace of X. Then P is the projection operator from X onto H if and only if for all $x \in X$,*

$$\|x - Px\|^2 = \min_{h\in H} \|x - h\|^2.$$

Proof. Let P be the projection operator from X onto H. For all $x \in X$ and for all $h \in H$, we have $x - h = (x - Px) + (Px - h)$, where $(x - Px) \in H^\perp$ and $(Px - h) \in H$. So, by the generalized Pythagorean theorem,

$$\|x - h\|^2 = \|x - Px\|^2 + \|Px - h\|^2 \geq \|x - Px\|^2.$$

To prove the converse, fix any $x \in X$ and define the functional f by

$$f(h) = \|x - h\|^2,\ h \in H.$$

If $x = y + z$, where $y \in H$ and $z \in H^\perp$, then we have

$$f(h) = \|x - h\|^2 = \langle h, h\rangle - 2\langle h, y\rangle + \|x\|^2.$$

So, by Proposition 2.7 (it will be proved in Section 2.2.1), $h_0 = y$ is the unique point in H satisfying $f(h_0) = \min_{h\in H} f(h)$. Therefore, $Px = y$, i.e., P is the projection operator from X onto H. ☐

1.3.6 Self-Adjoint, Nonnegative and Coercive Operators

Let X be a Hilbert space. An operator $A \in \mathcal{L}(X)$ is said to be *self-adjoint* if $A^* = A$. A self-adjoint operator $A \in \mathcal{L}(X)$ is said to be *nonnegative* (*coercive*) if $\langle Ax, x\rangle \geq 0$ (there exists $c > 0$ such that $\langle Ax, x\rangle \geq c\|x\|^2$) for all $x \in X$. The nonnegativity (coerciveness) of A is expressed by $A \geq 0$ ($A > 0$). Also, if $A - B \geq 0$ ($A - B > 0$), then we write $A \geq B$ or $B \leq A$ ($A > B$ or $B < A$). Obviously, $A > 0$ implies $A \geq 0$. The norm of a nonnegative operator A can be determined by one of the following formulae:

$$\|A\|_{\mathcal{L}} = \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\|=1} \langle Ax, x\rangle.$$

In some aspects nonnegative operators are similar to nonnegative numbers. For example, if $A \geq 0$, then there exists a unique linear operator, denoted by $A^{1/2}$, so that $A^{1/2} \geq 0$ and $(A^{1/2})^2 = A^{1/2}A^{1/2} = A$. The operator $A^{1/2}$ is called the *square root* of A. Obviously, $A^{1/2} > 0$ if $A > 0$.

It is easy to show that each $A > 0$ has a bounded inverse and $A^{-1} > 0$. The following proposition modifies this result.

Proposition 1.24. *Let X and Y be Hilbert spaces and let*

$$G = \begin{bmatrix} A & C^* \\ C & B \end{bmatrix},$$

where $A \in \mathcal{L}(X)$, $B \in \mathcal{L}(Y)$, $C \in \mathcal{L}(X, Y)$, $A > 0$ and $B < 0$. Then there exists $G^{-1} \in \mathcal{L}(X \times Y)$ and $G^{-1} = G_1 = G_2$, where

$$G_1 = \begin{bmatrix} \left(A - C^*B^{-1}C\right)^{-1} & -A^{-1}C^*\left(B - CA^{-1}C^*\right)^{-1} \\ -B^{-1}C\left(A - C^*B^{-1}C\right)^{-1} & \left(B - CA^{-1}C^*\right)^{-1} \end{bmatrix}$$

and

$$G_2 = \begin{bmatrix} \left(A - C^*B^{-1}C\right)^{-1} & -\left(A - C^*B^{-1}C\right)^{-1}C^*B^{-1} \\ -\left(B - CA^{-1}C^*\right)^{-1}CA^{-1} & \left(B - CA^{-1}C^*\right)^{-1} \end{bmatrix}.$$

Proof. First, note that $A > 0$ and $B < 0$ imply $A - C^*B^{-1}C > 0$ and, therefore, there exists $(A - C^*B^{-1}C)^{-1} \in \mathcal{L}(X)$. Similarly, there exists $(B - CA^{-1}C^*)^{-1} \in \mathcal{L}(Y)$. So, $G_1, G_2 \in \mathcal{L}(X \times Y)$. Take arbitrary $h \in X \times Y$ and let

$$h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \in X \times Y \text{ and } Gh = g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \in X \times Y.$$

We have

$$
\begin{aligned}
Gh = g &\Rightarrow \left\{ \begin{array}{l} Ah_1 + C^*h_2 = g_1 \\ Ch_1 + Bh_2 = g_2 \end{array} \right\} \\
&\Rightarrow \left\{ \begin{array}{l} h_1 = A^{-1}g_1 - A^{-1}C^*h_2 \\ h_2 = B^{-1}g_2 - B^{-1}Ch_1 \end{array} \right\} \\
&\Rightarrow \left\{ \begin{array}{l} Ah_1 + C^*B^{-1}g_2 - C^*B^{-1}Ch_1 = g_1 \\ Bh_2 + CA^{-1}g_1 - CA^{-1}C^*h_2 = g_2 \end{array} \right\} \\
&\Rightarrow \left\{ \begin{array}{l} h_1 = \left(A - C^*B^{-1}C\right)^{-1}\left(g_1 - C^*B^{-1}g_2\right) \\ h_2 = \left(B - CA^{-1}C^*\right)^{-1}\left(g_2 - CA^{-1}g_1\right) \end{array} \right\} \Rightarrow h = G_2 g.
\end{aligned}
$$

So, G_2 is the left inverse of G, i.e., $G_2G = I$. Since G is self-adjoint and $G_2^* = G_1$, we easily obtain that $GG_1 = I$, i.e., G_1 is the right inverse of G. In view of

$$
\begin{aligned}
A^{-1}C^* &= \left(A - C^*B^{-1}C\right)^{-1}\left(A - C^*B^{-1}C\right)A^{-1}C^* \\
&= \left(A - C^*B^{-1}C\right)^{-1}C^*B^{-1}\left(B - CA^{-1}C^*\right),
\end{aligned}
$$

we conclude that

$$
A^{-1}C^*\left(B - CA^{-1}C^*\right)^{-1} = \left(A - C^*B^{-1}C\right)^{-1}C^*B^{-1}.
$$

Thus, $G_1 = G_2 = G^{-1} \in \mathcal{L}(X \times Y)$. □

1.3.7 Compact, Hilbert–Schmidt and Nuclear Operators

Let $X, Y \in \mathcal{H}$. An operator $A \in \mathcal{L}(X, Y)$ is said to be *compact* if the convergence $\langle h, x_n - x\rangle \to 0$ for all $h \in X$ implies $\|Ax_n - Ax\| \to 0$. The class of all compact linear operators from X to Y will be denoted by $\mathcal{L}_\infty(X, Y)$. The brief notation $\mathcal{L}_\infty(X) = \mathcal{L}_\infty(X, X)$ will be used. $\mathcal{L}_\infty(X, Y)$ is a subspace of $\mathcal{L}(X, Y)$. Moreover, the space $\mathcal{L}_\infty(X, Y)$ is separable whereas $\mathcal{L}(X, Y)$ is not separable in general.

The following theorem expresses the structure of compact and self-adjoint operators.

Theorem 1.25. *Let $X \in \mathcal{H}$. If $A \in \mathcal{L}_\infty(X)$ and $A^* = A$, then there exist a basis $\{e_n\}$ in X and a sequence $\{\lambda_n\}$ of real numbers such that*

$$
\forall x \in X, \ Ax = \sum_{n=1}^{\dim X} \lambda_n \langle x, e_n\rangle e_n.
$$

If, in addition, $A \geq 0$, then $\lambda_n \geq 0$ for all n.

The vectors e_n and the numbers λ_n from Theorem 1.25 are called the *eigenvectors* and the corresponding *eigenvalues*, respectively, of the compact and self-adjoint operator A.

Let $X \in \mathcal{H}$. Given $A \in \mathcal{L}(X)$, one can show that the (finite or infinite) sum

$$\sum_{n=1}^{\dim X} \langle Ae_n, e_n \rangle$$

is independent of the choice of basis $\{e_n\}$ in X. This sum is called the *trace* of A and denoted by $\mathrm{tr}A$. By Theorem 1.25, $\mathrm{tr}A$ is equal to the sum of all eigenvalues of A if $A \in \mathcal{L}_\infty(X)$ and $A^* = A$.

Now let $X, Y \in \mathcal{H}$. Obviously, $A^*A \in \mathcal{L}_\infty(X)$ and $A^*A \geq 0$ for all $A \in \mathcal{L}_\infty(X,Y)$. An operator $A \in \mathcal{L}_\infty(X,Y)$ is called a *Hilbert–Schmidt operator* if

$$\mathrm{tr}(A^*A) < \infty.$$

The class of all Hilbert–Schmidt operators from X to Y is denoted by $\mathcal{L}_2(X,Y)$. An operator $A \in \mathcal{L}_\infty(X,Y)$ is called a *nuclear operator* if $\mathrm{tr}\big((A^*A)^{1/2}\big) < \infty$. The class of all nuclear operators from X to Y is denoted by $\mathcal{L}_1(X,Y)$. The brief symbols $\mathcal{L}_2(X) = \mathcal{L}_2(X,X)$ and $\mathcal{L}_1(X) = \mathcal{L}_1(X,X)$ are used as well. $\mathcal{L}_2(X,Y)$ is a separable Hilbert space with the scalar product

$$\langle A, B \rangle_{\mathcal{L}_2} = \mathrm{tr}(B^*A),$$

and $\mathcal{L}_1(X,Y)$ is a separable Banach space with the norm

$$\|A\|_{\mathcal{L}_1} = \mathrm{tr}\big((A^*A)^{1/2}\big).$$

Note that we have mentioned only three most widely used $\mathcal{L}_p$-spaces (the general definition of $\mathcal{L}_p$-spaces for $p \geq 1$ the reader can find in reference books on functional analysis): $\mathcal{L}_1$ (smallest one), $\mathcal{L}_2$ (middle one, as it is the only Hilbert space among all $\mathcal{L}_p$-spaces) and $\mathcal{L}$ (greatest one). Also, we have mentioned $\mathcal{L}_\infty$ as an important separable subspace of $\mathcal{L}$.

$\mathcal{L}_p$-spaces are similar to l_p-spaces. In particular, the natural embeddings $l_1 \subset l_2 \subset l_\infty$ have the analog $\mathcal{L}_1(X,Y) \subset \mathcal{L}_2(X,Y) \subset \mathcal{L}(X,Y)$ with

$$\forall A \in \mathcal{L}_1(X,Y), \ \|A\|_{\mathcal{L}_2} \leq \|A\|_{\mathcal{L}_1},$$

and

$$\forall A \in \mathcal{L}_2(X,Y), \ \|A\|_{\mathcal{L}} \leq \|A\|_{\mathcal{L}_2}.$$

Also, the isometric equality $l_1^* = l_\infty$ has the analog $\mathcal{L}_1(X,Y)^* = \mathcal{L}(X,Y)$ which holds under the isometry

$$\mathcal{L}_1(X,Y)^* \ni f \leftrightarrow Jf = F \in \mathcal{L}(X,Y) : fA = \mathrm{tr}(F^*A), \ A \in \mathcal{L}_1(X,Y).$$

When $X = \mathbb{R}^n$ or $Y = \mathbb{R}^m$, all the spaces $\mathcal{L}_p(X,Y)$ for $p \geq 1$ are isomorphic to $\mathcal{L}(X,Y)$. The spaces $\mathcal{L}_2(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathbb{R}^{nm}$ are isometric.

Note that for sake of similarity the space of bounded linear operators should be denoted by $\mathcal{L}_\infty$ instead of $\mathcal{L}$. However, we follow the traditional notation $\mathcal{L}$ for this space reserving $\mathcal{L}_\infty$ for the space of compact linear operators.

We will also use the following space given in Rozovskii [84]. Suppose $M \in \mathcal{L}_1(X)$ and $M \geq 0$. Denote by $\mathcal{L}_M(X,Y)$ the class of all (in general unbounded) linear operators A from $R(M^{1/2})$ to Y such that $AM^{1/2} \in \mathcal{L}_2(X,Y)$. The class $\mathcal{L}_M(X,Y)$ is a separable Hilbert space with the scalar product

$$\langle A, B\rangle_{\mathcal{L}_M} = \mathrm{tr}\big((BM^{1/2})^*(AM^{1/2})\big)$$

and $\mathcal{L}(X,Y) \subset \mathcal{L}_M(X,Y)$ is a natural embedding with

$$\|A\|_{\mathcal{L}_M} \leq \|M\|_{\mathcal{L}_1}\|A\|_{\mathcal{L}} \text{ for } A \in \mathcal{L}(X,Y).$$

Below we suppose that $X, Y, Z \in \mathcal{H}$ and list some useful facts regarding the linear operators.

(a) If $A \in \mathcal{L}_i(X,Y)$, $i = 1, 2, \infty$, $B \in \mathcal{L}(Y,Z)$ and $C \in \mathcal{L}(Z,X)$, then $BA \in \mathcal{L}_i(X,Z)$, $AC \in \mathcal{L}_i(Z,Y)$, $\|BA\|_{\mathcal{L}_i} \leq \|B\|_{\mathcal{L}}\|A\|_{\mathcal{L}_i}$ and $\|AC\|_{\mathcal{L}_i} \leq \|A\|_{\mathcal{L}_i}\|C\|_{\mathcal{L}}$.

(b) If $A \in \mathcal{L}_i(X,Y)$, $i = 1, 2, \infty$, then $A^* \in \mathcal{L}_i(Y,X)$ and $\|A\|_{\mathcal{L}_i} = \|A^*\|_{\mathcal{L}_i}$.

(c) If $A \in \mathcal{L}_2(X,Y)$ and $B \in \mathcal{L}_2(Y,Z)$, then $BA \in \mathcal{L}_1(X,Z)$ and $\|BA\|_{\mathcal{L}_1} \leq \|B\|_{\mathcal{L}_2}\|A\|_{\mathcal{L}_2}$.

(d) If $A \in \mathcal{L}_1(X)$ and $A \geq 0$, then $A^{1/2} \in \mathcal{L}_2(X)$ and $\|A\|_{\mathcal{L}_1} = \left\|A^{1/2}\right\|_{\mathcal{L}_2} = \mathrm{tr}A$.

(e) If $A \in \mathcal{L}_1(X)$ and $A > 0$, then $\dim X < \infty$.

(f) A linear operator A from the product of $X_1, \ldots, X_n \in \mathcal{H}$ to the product of $Y_1, \ldots, Y_m \in \mathcal{H}$, decomposed by (1.3), is a Hilbert–Schmidt operator if and only if $A_{ij} \in \mathcal{L}_2(X_j, Y_i)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$; moreover, $\|A\|^2_{\mathcal{L}_2} = \sum_{i,j}\|A_{ij}\|^2_{\mathcal{L}_2}$.

For fixed $u \in X$ and $v \in Y$, define the operator $u \otimes v$ by the formula

$$(u \otimes v)h = u\langle v, h\rangle,\ \ h \in Y. \tag{1.5}$$

Proposition 1.26. *Suppose that $X, Y, Z, H \in \mathcal{H}$, $u, w \in X$, $v \in Y$, $A \in \mathcal{L}(X,Z)$ and $B \in \mathcal{L}(Y,H)$. Then*

(a) $(u \otimes v) \in \mathcal{L}_1(Y,X)$;

(b) $(u \otimes v)^* = (v \otimes u)$;

(c) $\|u \otimes v\|_{\mathcal{L}_1} = \|u\|\,\|v\|$;

(d) $(u \otimes u) \geq 0$;

(e) $(u + w) \otimes v = (u \otimes v) + (w \otimes v)$;

(f) $(Au) \otimes (Bv) = A(u \otimes v)B^*$;

(g) $\operatorname{tr}(u \otimes w) = \langle u, w \rangle$;

(h) $|\operatorname{tr}(u \otimes w)| \leq \|u \otimes w\|_{\mathcal{L}_1}$;

(i) $\operatorname{tr}((Au) \otimes (Au)) = \operatorname{tr}((u \otimes u)A^*A) = \operatorname{tr}(A^*A(u \otimes u))$.

Proof. Parts (a)–(c) are obvious when $v = 0$. Let $v \neq 0$. It is clear that $(u \otimes v) \in \mathcal{L}(Y, X)$. In view of

$$\langle x, (u \otimes v)y \rangle = \langle u, x \rangle \langle v, y \rangle = \langle (v \otimes u)x, y \rangle, \ x \in X, \ y \in Y,$$

we obtain $(u \otimes v)^* = (v \otimes u)$. This proves part (b). If $\langle h, y_n - y \rangle \to 0$ for all $h \in Y$, then

$$\|(u \otimes v)y_n - (u \otimes v)y\| \leq \|u\| \, |\langle v, y_n - y \rangle| \to 0.$$

Hence, $(u \otimes v) \in \mathcal{L}_\infty(Y, X)$. Furthermore,

$$(u \otimes v)^*(u \otimes v)y = (v \otimes u)(u \otimes v)y = v\|u\|^2 \langle v, y \rangle, \ y \in Y.$$

Using this expression, it is easy to verify that

$$((u \otimes v)^*(u \otimes v))^{1/2} y = v\|u\| \, \|v\|^{-1} \langle v, y \rangle, \ y \in Y.$$

Therefore, by the Parseval identity (see Theorem 1.7), we have

$$\operatorname{tr}\big(((u \otimes v)^*(u \otimes v))^{1/2}\big) = \|u\| \, \|v\|^{-1} \sum_{n=1}^{\dim Y} \langle v, e_n \rangle^2 = \|u\| \, \|v\| < \infty,$$

where $\{e_n\}$ is any basis in Y. This proves parts (a) and (c). Parts (d) and (e) are obvious. Part (f) follows from

$$\begin{aligned} \langle ((Au) \otimes (Bv))h, z \rangle &= \langle Au, z \rangle \langle Bv, h \rangle = \langle u, A^*z \rangle \langle v, B^*h \rangle \\ &= \langle (u \otimes v)B^*h, A^*z \rangle = \langle A(u \otimes v)B^*h, z \rangle, \end{aligned}$$

where $h \in H$ and $z \in Z$ are arbitrary. For part (g),

$$\begin{aligned} \operatorname{tr}(u \otimes w) &= \sum_{n=1}^{\dim X} \langle (u \otimes w)e_n, e_n \rangle \\ &= \sum_{n=1}^{\dim X} \langle u, e_n \rangle \langle w, e_n \rangle \\ &= \left\langle u, \sum_{n=1}^{\dim X} \langle w, e_n \rangle e_n \right\rangle = \langle u, w \rangle, \end{aligned}$$

where $\{e_n\}$ is any basis of X. Part (h) follows from

$$|\mathrm{tr}(u \otimes w)| = |\langle u, w\rangle| \le \|u\| \, \|w\| = \|u \otimes v\|_{\mathcal{L}_1}.$$

Finally, for part (i), let $U = (u \otimes u)$ and let $\{e_n\}$ and $\{\lambda_n\}$ be the systems of eigenvectors and corresponding eigenvalues of U. Then by part (f),

$$\begin{aligned}
\mathrm{tr}((Au) \otimes (Au)) &= \mathrm{tr}(AUA^*) = \left\|U^{1/2}A^*\right\|_{\mathcal{L}_2}^2 \\
&= \left\|AU^{1/2}\right\|_{\mathcal{L}_2}^2 = \mathrm{tr}\big(U^{1/2}A^*AU^{1/2}\big) \\
&= \sum_{n=1}^{\dim X} \big\langle A^*AU^{1/2}e_n, U^{1/2}e_n\big\rangle \\
&= \sum_{n=1}^{\dim X} \big\langle \sqrt{\lambda_n}A^*Ae_n, \sqrt{\lambda_n}e_n\big\rangle \\
&= \sum_{n=1}^{\dim X} \langle A^*Ae_n, \lambda_n e_n\rangle \\
&= \sum_{n=1}^{\dim X} \langle A^*Ae_n, Ue_n\rangle = \mathrm{tr}(UA^*A) = \mathrm{tr}(A^*AU).
\end{aligned}$$

Thus, the proof is completed. □

The operators $u \otimes v$ for $u \in X$ and $v \in Y$ play a significant role in constructing general Hilbert–Schmidt operators. In particular, if $\{u_n\}$ and $\{v_m\}$ are bases in X and Y, respectively, then the double sequence $\{u_n \otimes v_m\}$ is a basis in $\mathcal{L}_2(Y, X)$.

1.4 Weak Convergence

Some useful properties of convergence in $\mathbb{R}^k$ are not true for strong convergence in infinite dimensional Hilbert and Banach spaces, while they do hold for weak convergence as defined below. This already demonstrates the importance of weak convergence.

1.4.1 Strong and Weak Forms of Convergence

Given a Banach space X, a sequence $\{x_n\}$ in X is said to *converge weakly* to $x \in X$ if

$$x^*(x_n - x) \to 0 \text{ for all } x^* \in X^*.$$

When X is a Hilbert space, in view of Theorem 1.21, a sequence $\{x_n\}$ in X weakly converges to $x \in X$ if

$$\langle h, x_n - x\rangle \to 0 \text{ for all } h \in X.$$

We write $x = w\text{-}\lim_{n\to\infty} x_n$ or $x_n \xrightarrow{w} x$ if the sequence $\{x_n\}$ weakly converges to the vector x.

A sequence $\{x_n\}$ in X is said to be *weakly Cauchy* if

$$x_n - x_m \xrightarrow{w} 0, \ n, m \to \infty.$$

Obviously, a weakly convergent sequence is weakly Cauchy. If each weakly Cauchy sequence in X is weakly convergent, then X is said to be *weakly complete*. Any Hilbert space is weakly complete.

Obviously, strong convergence implies weak convergence. The converse is true in $\mathbb{R}^k$, but not in infinite dimensional spaces in general. Indeed, it is easy to show that if $\{e_n : n \in \mathbb{N}\}$ is an orthonormal system in an infinite dimensional Hilbert space, then the sequence $\{e_n\}$ converges weakly to zero, but not strongly. The following theorem states a condition under which weak convergence implies strong convergence.

Theorem 1.27. *Let X be a Hilbert space. If a sequence $\{x_n\}$ in X is so that w-$\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} \|x_n\| = \|x\|$, then $\lim_{n\to\infty} x_n = x$.*

By the well-known theorem of Weierstrass, each bounded sequence in $\mathbb{R}^k$ has a convergent subsequence. This theorem is not true for strong convergence in infinite dimensional spaces. Indeed, if $\{e_n : n \in \mathbb{N}\}$ is an orthonormal system in an infinite dimensional Hilbert space, then each of its subsequences is again an orthonormal system and, therefore, has no strong limit, while any orthonormal system is bounded. But, for weak convergence, this result is true and called the *weak compactness property*.

Theorem 1.28. *Every bounded sequence in a Hilbert space has a weakly convergent subsequence.*

1.4.2 Weak Convergence and Convexity

A point x of a Hilbert space X is called a *weak limit point* of a subset G of X if there exists a sequence $\{x_n\}$ in $G \setminus \{x\}$ such that w-$\lim_{n\to\infty} x_n = x$. A set G is said to be *weakly closed* if it contains all its weak limit points. Obviously, a weakly closed set is closed. The converse can be stated under an additional convexity condition.

A nonempty subset G of a Hilbert space X is said to be *convex* if $x, y \in G$ and $0 \leq a \leq 1$ imply

$$ax + (1 - a)y \in G.$$

The vector

$$a_1 x_1 + \cdots + a_n x_n \in X,$$

where $a_i \geq 0$ and $\sum_{i=1}^{n} a_i = 1$, is called the *convex combination* of the vectors $x_1, \ldots, x_n \in X$. Obviously, a set $G \subset X$ is convex if and only if it contains all convex combinations of its elements.

Theorem 1.29 (Mazur). *Let X be a Hilbert space. The following statements hold.*

(a) *If a sequence $\{x_n\}$ in X weakly converges to $x \in X$, then there exists a sequence $\{y_n\}$ of convex combinations of $\{x_n\}$, i.e.,*

$$y_n = \sum_{i=1}^{n} a_{n,i} x_i, \ a_{n,i} \geq 0, \ \sum_{i=1}^{n} a_{n,i} = 1, \ i = 1, \ldots, n, \ n = 1, 2, \ldots,$$

such that $\lim_{n\to\infty} y_n = w\text{-}\lim_{n\to\infty} x_n = x$.

(b) *Any convex and closed subset of X is weakly closed.*

A functional f, defined on a subset G of a Hilbert space X, is said to be *convex* if G is a convex set and

$$f(ax + (1-a)y) \leq af(x) + (1-a)f(y)$$

for all $x, y \in G$ with $x \neq y$ and for all $0 < a < 1$. Similarly, a functional f, defined on a subset G of a Hilbert space X, is said to be *strictly convex* if G is a convex set and

$$f(ax + (1-a)y) < af(x) + (1-a)f(y)$$

for all $x, y \in G$ with $x \neq y$ and for all $0 < a < 1$.

Proposition 1.30. *Let X be a Hilbert space and let a functional f be defined by $f(x) = \langle Ax, x\rangle + 2\langle b, x\rangle$, $x \in X$, where $A \in \mathcal{L}(X)$ and $b \in X$. Then*

(a) $A \geq 0 \ \Rightarrow \ f$ *is convex;*

(b) $A > 0 \ \Rightarrow \ f$ *is strictly convex.*

Proof. For $0 < a < 1$, one can easily calculate that

$$f(ax + (1-a)y) = af(x) + (1-a)f(y) - a(1-a)\langle A(x-y), x-y\rangle.$$

Hence, $A \geq 0$ implies the convexity of f. Also, $A > 0$ implies the strict convexity of f. □

1.4.3 Convergence of Operators

Using the concepts of strong and weak convergence in a Banach space, one can define various concepts of convergence of operators. Let X and Y be Banach spaces. Convergence in norm of $\mathcal{L}(X,Y)$ is called *uniform operator convergence.* A sequence $\{A_n\}$ in $\mathcal{L}(X,Y)$ is said to *converge strongly* (*weakly*) to $A \in \mathcal{L}(X,Y)$ if $\{A_n x\}$ converges strongly (weakly) to Ax in Y for all $x \in X$.

A sequence $\{A_n\}$ in the space $\mathcal{L}(X,Y)$ is said to be *strongly* (*weakly*) *Cauchy* if $\lim_{n,m\to\infty} \|A_n x - A_m x\| = 0$ for all $x \in X$ ($\lim_{n,m\to\infty}(y^* A_n x - y^* A_m x) = 0$ for all $x \in X$ and for all $y^* \in Y^*$). Obviously, a strongly (weakly) convergent sequence

$\{A_n\}$ in $\mathcal{L}(X, Y)$ is strongly (weakly) Cauchy. If each strongly (weakly) Cauchy sequence in $\mathcal{L}(X, Y)$ is strongly (weakly) convergent, then $\mathcal{L}(X, Y)$ is said to be *strongly (weakly) complete.* $\mathcal{L}(X, Y)$ is strongly complete which follows from the completeness of the Banach space Y. $\mathcal{L}(X, Y)$ is weakly complete if Y is a weakly complete Banach space. In particular, this is true when Y is a Hilbert space.

Obviously, uniform operator convergence implies strong operator convergence which implies weak operator convergence. But, the converses are not true in general. In $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ all these concepts of convergence are equivalent.

The following theorem is about the strong convergence of a bounded and nonincreasing sequence of nonnegative operators.

Theorem 1.31. *Let X be a Hilbert space. If a sequence $\{A_n\}$ in $\mathcal{L}(X)$ satisfies $A_n \geq A_{n+1} \geq 0$ for all $n \in \mathbb{N}$, then $\{A_n\}$ converges strongly to some $A \in \mathcal{L}(X)$ and $A_n \geq A \geq 0$ for all $n \in \mathbb{N}$.*

Finally, we list some properties of various concepts of convergence.

Proposition 1.32. *Let X and Y be Hilbert spaces. The following statements hold.*

(a) *A weakly convergent sequence of operators $A_n \in \mathcal{L}(X, Y)$ (as well as vectors $x_n \in X$) is bounded.*

(b) *If a sequence $\{A_n\}$ in $\mathcal{L}(X, Y)$ and a sequence $\{x_n\}$ in X converge strongly to A and x, respectively, then $\{A_n x_n\}$ converges strongly to Ax.*

(c) *If a sequence $\{A_n\}$ in $\mathcal{L}(X, Y)$ converges weakly to A and a sequence $\{x_n\}$ in X converges strongly to x, then $\{A_n x_n\}$ converges weakly to Ax.*

(d) *If a sequence $\{A_n\}$ in $\mathcal{L}(X, Y)$ converges weakly to A, then $\{A_n^*\}$ converges weakly to A^*.*

(e) *If a sequence $\{A_n\}$ in $\mathcal{L}(X)$ converges strongly to A and there exists $c > 0$ such that $\langle A_n x, x\rangle \geq c\|x\|^2$ for all $n \in \mathbb{N}$ and for all $x \in X$, then $A > 0$ and $\{A_n^{-1}\}$ converges strongly to A^{-1}.*

Proof. By the weak convergence of $\{A_n\}$, we have the convergence of the sequence $\{\langle A_n x, y\rangle\}$ of real numbers for all $x \in X$ and for all $y \in Y$. Therefore, $\sup_n |\langle A_n x, y\rangle| < \infty$ for all $x \in X$ and for all $y \in Y$. Applying Theorem 1.13, we obtain $\sup_n \|A_n\| < \infty$. In a similar way, it can be proved that the weak convergence of $\{x_n\}$ in X implies $\sup_n \|x_n\| < \infty$, proving part (a). In view of the boundedness of $\{A_n\}$, the strong convergence of $\{A_n x_n\}$ to Ax in part (b) follows from

$$\|A_n x_n - Ax\| \leq \|A_n\| \, \|x_n - x\| + \|(A_n - A)x\|.$$

Similarly, the weak convergence of $\{A_n x_n\}$ to Ax in part (c) follows from

$$|\langle A_n x_n - Ax, y\rangle| \leq \|A_n\| \, \|y\| \, \|x_n - x\| + |\langle (A_n - A)x, y\rangle|, \; y \in Y.$$

Part (d) easily follows from the definition of adjoint operator. Finally,

$$\langle A_n x, x\rangle \geq c\|x\|^2, \; x \in X, \; n \in \mathbb{N},$$

implies

$$\langle Ax, x\rangle \geq c\|x\|^2, \; x \in X.$$

So, $A > 0$ and, hence, there exists $A^{-1} \in \mathcal{L}(X)$. From

$$\|A_n x\| \, \|x\| \geq \langle A_n x, x\rangle \geq c\|x\|^2, \; n \in \mathbb{N},$$

we obtain $\|A_n^{-1}\| \leq c^{-1}$ for all $n \in \mathbb{N}$, i.e., the sequence $\{A_n^{-1}\}$ is bounded. Finally, the inequality

$$\left\|\left(A_n^{-1} - A^{-1}\right) x\right\| \leq \|A_n^{-1}\| \, \left\|(A - A_n)A^{-1}x\right\|, \; x \in X,$$

completes the proof of part (e). □

Chapter 2

Basic Concepts of Analysis in Abstract Spaces

In this chapter we discuss the basic concepts of analysis such as continuity, differentiability, measurability and integrability for functions with values in abstract spaces. These concepts are based on convergence. Having the uniform, strong and weak forms of convergence in abstract spaces, we can define these concepts of analysis in different forms. The relationship between them is important when dealing with functions taking values in abstract spaces. To avoid ambiguity, we make distinction between a vector-valued function ranging in an abstract Banach space and an operator-valued function ranging in a specific Banach space of bounded linear operators.

2.1 Continuity

2.1.1 Continuity of Vector-Valued Functions

A function f from a metric space S to a metric space X is said to be *continuous on* S if the inverse image under f of each open set in X is an open set in S. Also, one can define continuity at a point. A function $f : S \to X$ is said to be *continuous at* $s_0 \in S$ if f_s converges to f_{s_0} whenever $d(s, s_0) \to 0$.

Theorem 2.1. *A function from a metric space S to a metric space X is continuous on S if and only if it is continuous at each point of S.*

A function $f : S \to X$ that is continuous on S (respectively, at $s_0 \in S$) is said to be *strongly continuous on* S (respectively, *at* $s_0 \in S$) if X is a Banach space. A function $f : S \to X$ is said to be *weakly continuous at* $s_0 \in S$ if f_s converges weakly to f_{s_0} whenever $d(s, s_0) \to 0$. If f is weakly continuous at each $s_0 \in S$, then it is said to be *weakly continuous on* S. The relationship between the concepts of strong and weak continuity is similar to the relationship between the concepts of strong and weak convergence.

One can easily verify that a bounded linear operator is continuous since it transforms each convergent sequence to a convergent sequence. One can also verify that a bounded linear operator transforms each weakly convergent sequence to a weakly convergent sequence. A compact linear operator has a stronger property: it transforms each weakly convergent sequence to a convergent sequence.

$C(S, X)$ denotes the set of all continuous functions from the metric space S to the Banach space X. It is a linear space with the pointwise algebraic operations. It is a Banach space with the norm

$$\|f\|_C = \max_{s \in S} \|f_s\|$$

when the *support* S of $C(S, X)$ is a compact metric space. The functions in the space $C(S, X)$ with compact support are *uniformly continuous*, i.e., if $f \in C(S, X)$, then

$$\forall \varepsilon > 0, \ \exists \delta > 0 \text{ such that } \|f_s - f_r\| < \varepsilon \text{ whenever } d(s, r) < \delta.$$

In the case $S = [a, b]$ the brief notation $C(a, b; X)$ is used for this space. $C(a, b; X)$ is a Banach space since $[a, b]$ is compact for all $-\infty < a < b < \infty$. If, in addition, $X \in \mathcal{H}$, then $C(a, b; X)$ is separable.

Convergence in the norm of $C(S, X)$ is called uniform convergence, which can be generalized to sequences of discontinuous functions as well. Let X be a Banach space and let S be a metric space. A sequence of functions $f^n : S \to X$ is said to *converge uniformly* to $f : S \to X$ if

$$\sup_{s \in S} \|f_s^n - f_s\| \to 0 \text{ as } n \to \infty.$$

Similarly, a series $\sum_{n=1}^{\infty} f_s^n$ of functions is said to *converge uniformly* if the sequence of partial sums $\sum_{n=1}^{k} f_s^n$ converges uniformly as $k \to \infty$. Note that if a series $\sum_{n=1}^{\infty} f_s^n$ is majorized by a convergent series of real numbers, i.e., $\|f_s^n\| \le c_n$ for all n and for all $s \in S$ with $\sum_{n=1}^{\infty} c_n < \infty$, then it converges uniformly.

A function f from $[a, b]$ to a Banach space X is said to be *right continuous* (*left continuous*) *at* s_0 if $\|f_s - f_{s_0}\| \to 0$ whenever $s \to s_0$ and $s_0 < s \le b$ ($a \le s < s_0$). A function $f : [a, b] \to X$ is said to be *right continuous* (*left continuous*) if it is right continuous (left continuous) at each $s \in [a, b)$ ($s \in (a, b]$). Note that the uniform limit of a sequence of right continuous (left continuous) functions is again right continuous (left continuous).

2.1.2 Weak Lower Semicontinuity

A functional f, defined on a subset G of a Hilbert space X, is said to be *weakly lower semicontinuous* if the weak convergence of a sequence $\{x_n\}$ in G to $x \in G$ implies

$$f(x) \le \liminf_{n \to \infty} f(x_n).$$

Theorem 2.2. *Let G be a convex subset of a Hilbert space X. If a functional $f : G \to \mathbb{R}$ is continuous and convex, then f is weakly lower semicontinuous.*

By the well-known theorem of Weierstrass, each real-valued continuous function, defined on a compact metric space, takes on its minimum and maximum values. We present the following modification of this theorem.

Proposition 2.3. *Let G be a closed and convex subset of a Hilbert space. If f is a continuous and convex functional on G satisfying $f(x) \to \infty$ whenever $\|x\| \to \infty$, then there exists a point $x_0 \in G$ satisfying $f(x_0) = \min_{x \in G} f(x)$. If, in addition, f is strictly convex, then x_0 is the unique point in G at which f takes its minimum value.*

Proof. Let $\{x_n\}$ be a minimizing sequence for the functional f, i.e.,

$$\lim_{n \to \infty} f(x_n) = \inf_{x \in G} f(x).$$

If $\{x_{n_k}\}$ is a subsequence of $\{x_n\}$ and $\|x_{n_k}\| \to \infty$, then $f(x_{n_k}) \to \infty$ and $\inf_{x \in G} f(x) = \infty$. This means $f(x) = \infty$ for all $x \in G$ which is impossible. Therefore, $\{x_n\}$ is a bounded sequence. By Theorem 1.28, there exists a weakly convergent subsequence of $\{x_n\}$ that will again be denoted by $\{x_n\}$. By Theorem 1.29(b), G is a weakly closed subset of X. Hence, $w\text{-}\lim_{n \to \infty} x_n = x_0 \in G$. Finally, by Theorem 2.2, we have

$$f(x_0) \leq \liminf_{n \to \infty} f(x_n) = \lim_{n \to \infty} f(x_n) = \inf_{x \in G} f(x) = \min_{x \in G} f(x),$$

i.e., the functional f takes its minimum value at $x_0 \in G$. Let $y_0 \in G$ be a point having the same property. If $x_0 \neq y_0$, then by the strict convexity of f we obtain

$$f\left(\frac{x_0 + y_0}{2}\right) < \frac{1}{2}\left(f(x_0) + f(y_0)\right) = \min_{x \in G} f(x).$$

Since $(x_0 + y_0)/2 \in G$, the case $x_0 \neq y_0$ is impossible, and hence $x_0 = y_0$. □

2.1.3 Continuity of Operator-Valued Functions

Let X, Y be Banach spaces and let S be a metric space. A function F from S to $\mathcal{L}(X,Y)$ is said to be *uniformly continuous at* $s_0 \in S$ if

$$\|F_s - F_{s_0}\| \to 0 \text{ whenever } d(s, s_0) \to 0.$$

A function $F : S \to \mathcal{L}(X,Y)$ is said to be *strongly* (*weakly*) *continuous at* s_0 if $F_s x$, $s \in S$, is strongly (weakly) continuous at s_0 for all $x \in X$. A function $F : S \to \mathcal{L}(X,Y)$ which is uniformly (strongly, weakly) continuous at each $s \in S$ is said to be *uniformly* (*strongly, weakly*) *continuous on* S. The relationship between these three concepts of continuity for operator-valued functions is similar to the relationship between the corresponding concepts of convergence for operators.

Remark 2.4. For given operator-valued functions

$$F : S \to \mathcal{L}(X,Y) \text{ and } G : S \to \mathcal{L}(Z,X),$$

we will often consider the functions defined as composition $F_s G_s$, as adjoint $(F_s)^*$, as inverse $(F_s)^{-1}$, as square root $(F_s)^{1/2}$, as norm $\|F_s\|$, where $s \in S$. These functions briefly will be denoted by FG, F^*, F^{-1}, $F^{1/2}$, $\|F\|$, respectively. Also, for $f : S \to X$, the symbol Ff means the function $F_s f_s$, $s \in S$.

Proposition 2.5. *Let X, Y be Hilbert spaces and let S be a metric space. The following statements hold.*

(a) *If $F : S \to \mathcal{L}(X,Y)$ is weakly continuous at $s_0 \in S$, then it is bounded on some ball centered at s_0.*

(b) *If $F : S \to \mathcal{L}(X,Y)$ and $f : S \to X$ are strongly continuous at $s_0 \in S$, then $Ff : S \to Y$ is strongly continuous at s_0.*

(c) *If $F : S \to \mathcal{L}(X,Y)$ is weakly continuous and $f : S \to X$ is strongly continuous at $s_0 \in S$, then $Ff : S \to Y$ is weakly continuous at s_0.*

(d) *If $F : S \to \mathcal{L}(X,Y)$ is weakly continuous at $s_0 \in S$, then the same holds for $F^* : S \to \mathcal{L}(Y,X)$.*

(e) *If $F : S \to \mathcal{L}(X)$ is strongly continuous at $s_0 \in S$ and there exists a constant $c > 0$ such that $\langle F_s x, x\rangle \geq c\|x\|^2$ for all $x \in X$ and for all $s \in S$, then $F^{-1} : S \to \mathcal{L}(X)$ is strongly continuous at s_0.*

Proof. Part (a) is true because otherwise there is a sequence $\{s_n\}$ in S, converging to $s_0 \in S$, such that $\|F_{s_n}\| \to \infty$. This is contrary to Proposition 1.32(a) since the sequence $\{F_{s_n}\}$ converges weakly to F_{s_0}. Also, parts (b)–(e) follow from respective parts of Proposition 1.32. □

2.2 Differentiability

2.2.1 Differentiability of Nonlinear Operators

Let X, Y be Banach spaces. An operator F from $D(F) \subset X$ to Y is said to be *strongly differentiable* or, simply, *differentiable at* $x \in D(F)$ if there exists $F'(x) \in \mathcal{L}(X,Y)$ such that

$$\frac{\|F(x+h) - F(x) - F'(x)h\|}{\|h\|} \to 0 \text{ as } \|h\| \to 0$$

with $h \neq 0$ and $x + h \in D(F)$. The operator $F'(x)$ is called the *derivative of F at* x. An operator F is said to be *differentiable on* $D(F)$ if it is differentiable at each $x \in D(F)$. The derivative of F is denoted by F' or by $(d/dx)F$. In general, the derivative of F is a nonlinear operator from $D(F)$ to $\mathcal{L}(X,Y)$. If F' is differentiable, then F is said to be *twice differentiable on* $D(F)$. The derivative of F' is called the *second derivative of* F. For the second derivative of F, the symbols F'' and $(d^2/dx^2)F$ are used. Obviously, F'' is an operator from $D(F)$ to $\mathcal{L}(X,\mathcal{L}(X,Y))$. The following equality holds:

$$\forall h, g \in X,\ (F'(x)h)'g = (F''(x)g)h. \tag{2.1}$$

The higher order derivatives can be defined by induction. Obviously, a differentiable operator is continuous. An operator F is said to be k *times continuously differentiable on* $D(F)$ if its kth derivative exists and is continuous.

For an operator F from a subset $D(F)$ of the product of Banach spaces X_1 and X_2 to a Banach space Y, we can write $F(x) = F(x_1, x_2)$, where x_1 and x_2 are the components of $x \in D(F)$ in X_1 and X_2, respectively, and define the *partial derivatives* of F with respect to x_1 (by considering x_2 as fixed) and x_2 (by considering x_1 as fixed). The symbols $(\partial/\partial x_1)F$ and $(\partial/\partial x_2)F$ are used for first order and $(\partial^2/\partial x_1^2)F$, $(\partial^2/\partial x_2^2)F$, $(\partial^2/\partial x_1\partial x_2)F$ and $(\partial^2/\partial x_2\partial x_1)F$ for second order partial derivatives.

According to the natural isometry $X = \mathcal{L}(\mathbb{R}, X)$, where X is a Banach space, the derivative of $f : [a,b] \to X$ is again a function from $[a,b]$ to X. Since $\mathcal{L}(X,\mathbb{R}) = X^* = X$, where X is a Hilbert space, the first derivative of $f : X \to \mathbb{R}$ is an operator from X to X, and the values of its second derivative lie in $\mathcal{L}(X)$. So, equality (2.1) in this case can be written as

$$\forall h, g \in X,\ \langle\langle f'(x), h\rangle', g\rangle = \langle f''(x)g, h\rangle.$$

We present the following useful application of first and second order derivatives. A point $x_0 \in X$ is called a *local minimum point* of a functional f defined on a Hilbert space X if there exists an open ball B in X centered at x_0 such that $f(x_0) \leq f(x)$ for all $x \in B$.

Theorem 2.6. *Let X be a Hilbert space and let $f : X \to \mathbb{R}$ be twice continuously differentiable on X. If $x_0 \in X$ is a local minimum point of f, then $f'(x_0) = 0$ and $f''(x_0) \geq 0$. Conversely, if $f'(x_0) = 0$ and $f''(x_0) > 0$, then x_0 is a local minimum point of f.*

Proposition 2.7. *Let X be a Hilbert space and let f be a functional given by $f(x) = \langle Ax, x\rangle + 2\langle b, x\rangle$, $x \in X$, where $A \in \mathcal{L}(X)$, $A > 0$ and $b \in X$. Then $x_0 = -A^{-1}b$ is the unique point in X satisfying $f(x_0) = \min_{x\in X} f(x)$.*

Proof. By Propositions 1.30 and 2.3, there exists a unique $x_0 \in X$ at which f takes its minimum value. A simple computation gives $f'(x) = 2(Ax + b)$ and $f''(x) = 2A > 0$. Hence, by Theorem 2.6, $x_0 = -A^{-1}b$. □

2.2.2 Differentiability of Operator-Valued Functions

Let X and Y be Banach spaces and let $[a,b]$ be a finite interval in $\mathbb{R}$. A differentiable operator-valued function is often said to be *uniformly differentiable.* If for $F : [a,b] \to \mathcal{L}(X,Y)$ there exists $F'_{t_0} \in \mathcal{L}(X,Y)$ such that the function $F_t x$, $a \le t \le b$, is differentiable at t_0 and $(Fx)'_{t_0} = F'_{t_0}x$ for all $x \in X$, then F is said to be *strongly differentiable at* t_0. An operator-valued function which is strongly differentiable at each $a \le t \le b$ is said to be *strongly differentiable on* $[a,b]$. The same notation is used for uniform and strong derivatives of an operator-valued function.

Proposition 2.8. *Let X, Y be Hilbert spaces and let $[a,b]$ be a finite interval in $\mathbb{R}$. The following statements hold.*

(a) *If $F : [a,b] \to \mathcal{L}(X,Y)$ is strongly differentiable and $f : [a,b] \to X$ is differentiable at $t_0 \in [a,b]$, then $Ff : [a,b] \to Y$ is differentiable at t_0 and $(Ff)'_{t_0} = F'_{t_0} f_{t_0} + F_{t_0} f'_{t_0}$.*

(b) *If $f : [a,b] \to X$ and $g : [a,b] \to X$ are differentiable at $t_0 \in [a,b]$, then the function $\langle f,g\rangle_t = \langle f_t, g_t\rangle$, $t \in [a,b]$, is differentiable at t_0 and $\langle f,g\rangle'_{t_0} = \langle f'_{t_0}, g_{t_0}\rangle + \langle f_{t_0}, g'_{t_0}\rangle$.*

(c) *If $f : [a,b] \to X$ and $g : [a,b] \to X$ are differentiable at $t_0 \in [a,b]$ and $A, B \in \mathcal{L}(X,Y)$, then $(Af + Bg) : [a,b] \to Y$ is differentiable at t_0 and $(Af+Bg)'_{t_0} = Af'_{t_0} + Bg'_{t_0}$.*

(d) *If $F : [a,b] \to \mathcal{L}(X,Y)$ and $F^* : [a,b] \to \mathcal{L}(Y,X)$ are strongly differentiable at $t_0 \in [a,b]$, then $(F^*)'_{t_0} = (F')^*_{t_0}$.*

(e) *If $F : [a,b] \to \mathcal{L}(X)$ is strongly differentiable at $t_0 \in [a,b]$ and there exists a constant $c > 0$ such that $\langle F_t x, x\rangle \ge c\|x\|^2$ for all $x \in X$ and for all $a \le t \le b$, then $F^{-1} : [a,b] \to \mathcal{L}(X)$ is strongly differentiable at t_0 and $\left(F^{-1}\right)'_{t_0} = -F^{-1}_{t_0} F'_{t_0} F^{-1}_{t_0}$.*

Proof. By Proposition 2.5(a), part (a) follows from

$$
\begin{aligned}
&\left\|(t-t_0)^{-1}(F_t f_t - F_{t_0} f_{t_0}) - F'_{t_0} f_{t_0} - F_{t_0} f'_{t_0}\right\| \\
&\qquad \le \|F_t\| \left\|(t-t_0)^{-1}(f_t - f_{t_0}) - f'_{t_0}\right\| \\
&\qquad\quad + \|F_t f'_{t_0} - F_{t_0} f'_{t_0}\| \\
&\qquad\quad + \left\|(t-t_0)^{-1}(F_t f_{t_0} - F_{t_0} f_{t_0}) - F'_{t_0} f_{t_0}\right\|.
\end{aligned}
$$

Part (b) is a consequence of part (a). Parts (c) and (d) are obvious. By Propositions

2.5(e) and 2.5(a), part (e) follows from

$$\begin{aligned}
&\left\|(t-t_0)^{-1}\left(F_t^{-1}-F_{t_0}^{-1}\right)x+F_{t_0}^{-1}F'_{t_0}F_{t_0}^{-1}x\right\| \\
&\qquad =\left\|(t-t_0)^{-1}F_t^{-1}(F_{t_0}-F_t)F_{t_0}^{-1}x+F_{t_0}^{-1}F'_{t_0}F_{t_0}^{-1}x\right\| \\
&\qquad \leq \left\|F_t^{-1}\right\| \left\|((t-t_0)^{-1}(F_{t_0}-F_t)+F'_{t_0})F_{t_0}^{-1}x\right\| \\
&\qquad\qquad +\left\|\left(F_{t_0}^{-1}-F_t^{-1}\right)F'_{t_0}F_{t_0}^{-1}x\right\|.
\end{aligned}$$

Thus, the proof is completed. □

2.3 Measurability

2.3.1 Measurability of Vector-Valued Functions

A function f from a measurable space (S,Σ) to a measurable space (X,Γ) is said to be (Σ,Γ)-*measurable* if $f^{-1}(A)\in\Sigma$ for all $A\in\Gamma$. When X is a metric space and $\mathcal{B}_X$ is the Borel σ-algebra of subsets of X, a $(\Sigma,\mathcal{B}_X)$-measurable function is briefly said to be Σ-*measurable*. Obviously, a continuous function $f:S\to X$ is $\mathcal{B}_S$-measurable when both S and X are metric spaces.

To define other measurability concepts, we need weaker concepts of convergence for functions rather than the uniform one. Let X be a Banach space and let (S,Σ,ν) be a measure space. A sequence $\{f^n\}$ of functions from S to X is said to *converge everywhere* or *converge pointwise on* S to a function $f:S\to X$ if $\|f^n_s-f_s\|\to 0$ as $n\to\infty$ for all $s\in S$. If $\|f^n_s-f_s\|\to 0$ as $n\to\infty$ for ν-a.e. $s\in S$, then the sequence of functions f^n is said to *converge* ν-*a.e. on* S to the function f. Uniform convergence of functions implies everywhere convergence which implies ν-a.e. convergence, but the converse statements do not hold in general.

A function f from a measurable space (S,Σ) to a Banach space X is said to be Σ-*simple* if it takes on a finite number of values $x_1,\ldots,x_k$ with $f^{-1}(\{x_n\})\in\Sigma$, $n=1,\ldots,k$. A function $f:S\to X$ is said to be *strongly* Σ-*measurable* if there exists a sequence of Σ-simple functions converging to f everywhere on S. A function $f:S\to X$ is said to be *weakly* Σ-*measurable* if x^*f_s, $s\in S$, is a Σ-measurable real-valued function for all $x^*\in X^*$. Each strongly Σ-measurable function is Σ-measurable and weakly Σ-measurable. The converses are true under the condition of the following theorem.

Theorem 2.9 (Pettis). *The concepts of measurability, strong measurability and weak measurability are equivalent for functions from a measurable space to a separable Banach space.*

The class of all Σ-measurable functions from the measurable space (S,Σ) to the Banach space X is denoted by $m(S,\Sigma,X)$. This class is a linear space with the pointwise algebraic operations and it is closed under everywhere convergence. Obviously, each everywhere convergent sequence in $m(S,\Sigma,X)$ has a unique limit.

Now let X be a Banach space, let (S, Σ, ν) be a measure space and let $\tilde{\Sigma}$ be the Lebesgue extension of Σ with respect to ν. Note that according to our convention in Section 1.2.6, by a measure we always mean a positive and finite measure. A $\tilde{\Sigma}$-measurable (respectively, strongly, weakly $\tilde{\Sigma}$-measurable) function $f : S \to X$ is said to be *ν-measurable* (respectively, *strongly*, *weakly ν-measurable*). Since $\Sigma \subset \tilde{\Sigma}$, each kind of measurability with respect to Σ implies the respective kind of measurability with respect to ν. The converse statements are true when (S, Σ, ν) is complete.

The linear space $m(S, \tilde{\Sigma}, X)$ is denoted by $m(S, \Sigma, \nu, X)$ (or, briefly, by $m(S, \nu, X)$ if there is no ambiguity about Σ). If $f \in m(S, \nu, X)$, then there exists $g \in m(S, \Sigma, X)$ so that $f_s = g_s$ for ν-a.e. $s \in S$. The function g is called a *strongly Σ-measurable modification* of f. In particular, this means that each $f \in m(S, \nu, X)$ can be defined as an everywhere limit of $\tilde{\Sigma}$-simple functions as well as a ν-a.e. limit of Σ-simple functions.

The space $m(S, \nu, X)$ is closed under ν-a.e. convergence, but if a sequence $\{f^n\}$ in $m(S, \nu, X)$ converges ν-a.e. on S to $f \in m(S, \nu, X)$, then it converges ν-a.e. on S to each function in $m(S, \nu, X)$ which is equal to f ν-a.e. on S. To have a unique ν-a.e. limit in $m(S, \nu, X)$ the following equivalence relation is introduced. Two functions $f, g \in m(S, \nu, X)$ are said to be equivalent if $f_s = g_s$ for ν-a.e. $s \in S$. The quotient set of $m(S, \nu, X)$ with respect to this equivalence relation is again denoted by $m(S, \nu, X)$. According to our agreement in Section 1.1.1 the expression $f \in m(S, \nu, X)$ now means that the function f is ν-measurable and represents the equivalence class of functions which are equal to f ν-a.e. on S. To carry out the algebraic and limit operations on the equivalence classes, it is sufficient to work with representatives of these classes and then pass to the equivalence class containing the resulting function. The quotient set $m(S, \nu, X)$ is a linear space, which is closed under ν-a.e. convergence, and each ν-a.e. convergent sequence in $m(S, \nu, X)$ has a unique limit.

A weaker concept of convergence than ν-a.e. convergence can be defined in $m(S, \nu, X)$. A sequence $\{f^n\}$ in $m(S, \nu, X)$ is said to *converge in measure* ν to $f \in m(S, \nu, X)$ if

$$\forall \varepsilon > 0, \ \nu(\{s \in S : \|f_s^n - f_s\| > \varepsilon\}) \to 0 \text{ as } n \to \infty.$$

Each ν-a.e. convergent sequence converges in measure ν. The converse is not true in general. But if a sequence $\{f^n\}$ in $m(S, \nu, X)$ converges to $f \in m(S, \nu, X)$ in measure ν, then there exists a subsequence of $\{f^n\}$ which converges ν-a.e. on S to f.

Let f be a ν-measurable function from a measure space (S, Σ, ν) to a Banach space X. A measure on $\mathcal{B}_X$, denoted by ν_f and defined by

$$\nu_f(A) = \nu\left(f^{-1}(A)\right), \ A \in \mathcal{B}_X,$$

is called the *measure generated by f*. Thus, the ν-measurable function f from S to X generates the measure space $(X, \mathcal{B}_X, \nu_f)$. With a proof similar to that of the particular case of real-valued functions (see, for example, Meyer [78], p. 10), the following theorem shows that the strong ν_f-measurability of a function is same with a function of f.

Theorem 2.10 (Doob). *Let X, Y be Banach spaces, let (S, Σ, ν) be a measure space and let $f : S \to X$ be ν-measurable. Then a function g belongs to $m(S, \sigma(f), \nu, Y)$ if and only if there exists $\varphi \in m(X, \nu_f, Y)$ such that $g = \varphi \circ f$, i.e., $g_s = \varphi(f_s)$ for ν-a.e. $s \in S$.*

2.3.2 Measurability of Operator-Valued Functions

Let X, Y be Banach spaces and let (S, Σ, ν) be a measure space. Operator-valued functions in $m(S, \Sigma, \mathcal{L}(X, Y))$ (in $m(S, \nu, \mathcal{L}(X, Y))$) are said to be *uniformly Σ-measurable* (*uniformly ν-measurable*). Obviously, the uniform measurability of operator-valued functions corresponds to the strong measurability of vector-valued functions. Most useful operator-valued functions, for example, the strongly continuous semigroups of bounded linear operators, are not uniformly measurable. Therefore, weaker concepts of measurability for operator-valued functions are needed.

An operator-valued function $F : S \to \mathcal{L}(X, Y)$ is said to be *strongly* (*weakly*) *Σ-measurable* if $F_s x$, $s \in S$, is a strongly (weakly) Σ-measurable Y-valued function for all $x \in X$. Similarly, $F : S \to \mathcal{L}(X, Y)$ is said to be *strongly* (*weakly*) *ν-measurable* if $F_s x$, $s \in S$, is a strongly (weakly) ν-measurable Y-valued function for all $x \in X$. The relationship between the concepts of uniform, strong and weak measurability for operator-valued functions is similar to the relationship between the respective concepts of convergence of operators. By Theorem 2.9, the concepts of strong and weak measurability for $\mathcal{L}(X, Y)$-valued functions are equivalent if Y is separable.

Proposition 2.11. *Let X, Y be Hilbert spaces and let (S, Σ) be a measurable space. The following statements hold.*

(a) *If $F : S \to \mathcal{L}(X, Y)$ is a strongly Σ-measurable operator-valued function and $f \in m(S, \Sigma, X)$, then $Ff \in m(S, \Sigma, Y)$.*

(b) *If $F : S \to \mathcal{L}(X, Y)$ is a weakly Σ-measurable operator valued-function, then the same holds for $F^* : S \to \mathcal{L}(Y, X)$.*

Proof. Let $\{f^n\}$ be a sequence of Σ-simple functions which converges to f on S. Obviously, $Ff^n \in m(S, \Sigma, Y)$ for all n and the sequence $\{Ff^n\}$ converges to Ff on S. Therefore, $Ff \in m(S, \Sigma, Y)$. Part (a) is proved. Part (b) is obvious. □

2.3.3 Measurability of $\mathcal{L}_1$- and $\mathcal{L}_2$-Valued Functions

$\mathcal{L}_1$- and $\mathcal{L}_2$-spaces have better properties than $\mathcal{L}$-space. In particular, they are separable Banach and Hilbert spaces, respectively. Therefore, it is convenient to study the $\mathcal{L}_1$- and $\mathcal{L}_2$-valued functions as vector-valued functions taking values in separable Banach and Hilbert spaces rather than in spaces of operators. Below we present some useful measurability properties of $\mathcal{L}_1$- and $\mathcal{L}_2$-valued functions.

Proposition 2.12. *Let $X, Y \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. The following statements hold.*

(a) *A function $F : S \to \mathcal{L}_2(X, Y)$ belongs to $m(S, \nu, \mathcal{L}_2(X, Y))$ if and only if $Fx \in m(S, \nu, Y)$ for all $x \in X$.*

(b) *A function $F : S \to \mathcal{L}_1(X, Y)$ belongs to $m(S, \nu, \mathcal{L}_1(X, Y))$ if and only if $Fx \in m(S, \nu, Y)$ for all $x \in X$.*

Proof. For part (a), let $F \in m(S, \nu, \mathcal{L}_2(X, Y))$. By Theorem 2.9, for all $A \in \mathcal{L}_2(X, Y)$ and for any basis $\{e_n\}$ of X, the real-valued function

$$\langle F_s, A\rangle_{\mathcal{L}_2} = \operatorname{tr}(A^* F_s) = \sum_{n=1}^{\dim X} \langle F_s e_n, A e_n\rangle, \quad \nu\text{-a.e. } s \in S, \tag{2.2}$$

is ν-measurable since $\mathcal{L}_2(X, Y) \in \mathcal{H}$. Let $x \in X$, $x \neq 0$, and let $y \in Y$. Consider a basis $\{e_n\}$ in X and an operator $A \in \mathcal{L}_2(X, Y)$ satisfying

$$e_1 = x\|x\|^{-1}, \quad Ae_1 = y\|x\|, \quad Ae_n = 0 \text{ for } n \geq 2.$$

From (2.2), $\langle F_s, A\rangle_{\mathcal{L}_2} = \langle F_s x, y\rangle$, ν-a.e. $s \in S$, is ν-measurable. Therefore, by the separability of Y, we have $Fx \in m(S, \nu, Y)$ for all $x \neq 0$ and as well as for $x = 0$. Let us prove the converse. Suppose $Fx \in m(S, \nu, Y)$ for all $x \in X$ and consider (2.2) for any $A \in \mathcal{L}_2(X, Y)$. Since all terms in the right-hand side of (2.2) are ν-measurable, the left-hand side of (2.2) is ν-measurable and, therefore, the function $F : S \to \mathcal{L}_2(X, Y)$ is a weakly ν-measurable vector-valued function. So, by Theorem 2.9, we obtain $F \in m(S, \nu, \mathcal{L}_2(X, Y))$ since $\mathcal{L}_2(X, Y)$ is separable. Part (a) is thus proved. Part (b) can be proved in a similar way by considering the operator A in the larger space $\mathcal{L}_1(X, Y)^* = \mathcal{L}(X, Y)$ than $\mathcal{L}_2(X, Y)$. □

Proposition 2.13. *Let $X, Y \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. The following statements hold.*

(a) *If $\Phi \in m(S, \nu, \mathcal{L}_2(X, Y))$ and if $F : S \to \mathcal{L}(Y, Z)$ is strongly ν-measurable, then $F\Phi \in m(S, \nu, \mathcal{L}_2(X, Z))$.*

(b) *If $\Phi \in m(S, \nu, \mathcal{L}_1(X, Y))$ and if $F : S \to \mathcal{L}(Y, Z)$ is strongly ν-measurable, then $F\Phi \in m(S, \nu, \mathcal{L}_1(X, Z))$.*

Proof. Consider part (a). Clearly, $F\Phi$ is an $\mathcal{L}_2$-valued function on S. By Proposition 2.12(a), we have $\Phi x \in m(S, \nu, Y)$ for all $x \in X$. Therefore, by Proposition 2.11(a), $F\Phi x \in m(S, \nu, Z)$ for all $x \in X$. Hence, again by Proposition 2.12(a), $F\Phi \in m(S, \nu, \mathcal{L}_2(X, Z))$. Part (b) can be proved in a similar way by the use of Proposition 2.12(b). □

Proposition 2.14. *Let $X, Y \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. If $f \in m(S, \nu, X)$ and $g \in m(S, \nu, Y)$, then $(f \otimes g) \in m(S, \nu, \mathcal{L}_1(Y, X))$, where the operation $\otimes$ is defined by* (1.5).

Proof. If $\{f^n\}$ and $\{g^n\}$ are sequences of Σ-simple functions which converge to the functions f and g ν-a.e. on S, respectively, then $f^n \otimes g^n$ is a Σ-simple $\mathcal{L}_1(Y, X)$-valued function for all $n \in \mathbb{N}$ and, by Proposition 1.26,

$$\begin{aligned} \|(f_s^n \otimes g_s^n) - (f_s \otimes g_s)\|_{\mathcal{L}_1} &\leq \|(f_s^n - f_s) \otimes g_s^n\|_{\mathcal{L}_1} + \|f_s \otimes (g_s^n - g_s)\|_{\mathcal{L}_1} \\ &= \|f_s^n - f_s\| \, \|g_s^n\| + \|f_s\| \, \|g_s^n - g_s\| \to 0, \; n \to \infty \end{aligned}$$

for ν-a.e. $s \in S$. Hence, $(f \otimes g) \in m(S, \nu, \mathcal{L}_1(Y, X))$. □

Proposition 2.15. *Let $X, Y \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. If $F \in m(S, \nu, \mathcal{L}_1(X))$ and $F_s \geq 0$, $s \in S$, then $F^{1/2} \in m(S, \nu, \mathcal{L}_2(X))$.*

Proof. We will use the following decomposition of F (see Da Prato and Zabczyk [42], p. 25):

$$F_s = \sum_{n=1}^{\dim X} \lambda_s^n (e_s^n \otimes e_s^n), \; \nu\text{-a.e. } s \in S,$$

where $\{\lambda^n\}$ is a sequence of real-valued nonnegative ν-measurable functions, $e^n \in m(S, \nu, X)$ for all n with

$$\|e_s^n\| = \left\{ \begin{matrix} 1, & \lambda_s^n > 0 \\ 0, & \lambda_s^n = 0 \end{matrix} \right\}, \; \langle e_s^n, e_s^m \rangle = 0, \; \nu\text{-a.e. } s \in S, \; n \neq m,$$

and the above series converges in norm of $\mathcal{L}_1(X)$ if $\dim X = \infty$. Using this decomposition, one can easily verify that

$$F_s^{1/2} = \sum_{n=1}^{\dim X} \sqrt{\lambda_s^n} (e_s^n \otimes e_s^n), \; \nu\text{-a.e. } s \in S,$$

where the series converges in norm of $\mathcal{L}_2(X)$ if $\dim X = \infty$. This implies $F^{1/2} \in m(S, \nu, \mathcal{L}_2(X))$. □

2.4 Integrability

We recall that according to our convention in Section 1.2.6, by a measure we always mean a positive and finite measure.

2.4.1 Bochner Integral

Let X be a Banach space and let (S, Σ, ν) be a measure space. The Bochner integral of the Σ-simple function

$$f_s = \sum_{n=1}^{k} x_n \chi_{A_n}(s), \quad s \in S, \ x_n \in X, \ A_n \in \Sigma, \ n = 1, \ldots, k, \ k \in \mathbb{N},$$

is defined by the sum

$$\sum_{n=1}^{k} x_n \nu(A_n).$$

Let $f \in m(S, \nu, X)$. Then f is a ν-a.e. limit of a sequence $\{f^n\}$ of Σ-simple functions. If the sequence of Bochner integrals of the Σ-simple functions f^n converges in norm of X and the limit is independent of the choice of a sequence of Σ-simple functions, then this limit is called the *Bochner integral* of the function f and it is denoted by

$$\int_S f_s \, d\nu.$$

In this case, the function f is said to be *ν-integrable in the sense of Bochner* (or, briefly, *ν-integrable*). The following theorem describes the class of all ν-integrable functions.

Theorem 2.16. *Let X be a Banach space and let (S, Σ, ν) be a measure space. A function $f : S \to X$ is ν-integrable if and only if $f \in m(S, \nu, X)$ and*

$$\int_S \|f_s\| \, d\nu < \infty.$$

Let X be a Banach space and let (S, Σ, ν) be a measure space. For given $1 \leq p < \infty$, the class of all (equivalence classes of ν-a.e. equal) functions $f \in m(S, \nu, X)$ satisfying

$$\int_S \|f_s\|^p \, d\nu < \infty$$

is denoted by $L_p(S, \Sigma, \nu, X)$ or, briefly, $L_p(S, \nu, X)$ if there is no ambiguity about Σ. The class of all (equivalence classes of ν-a.e. equal) functions $f \in m(S, \nu, X)$ satisfying

$$\operatorname*{ess\,sup}_{s \in S} \|f_s\| = \inf_{\nu(A)=0} \sup_{s \in S \setminus A} \|f_s\| < \infty$$

is denoted by $L_\infty(S, \Sigma, \nu, X)$ or $L_\infty(S, \nu, X)$. For $1 \leq p \leq \infty$, the class $L_p(S, \nu, X)$, is a Banach space with respect to the corresponding norm

$$\|f\|_{L_p} = \left(\int_S \|f_s\|^p \, d\nu \right)^{1/p}, 1 \leq p < \infty; \ \|f\|_{L_\infty} = \operatorname*{ess\,sup}_{s \in S} \|f_s\|.$$

The triangular inequality in $L_p(S,\nu,X)$, $1 \le p < \infty$, has the form

$$\left(\int_S \|f_s + g_s\|^p \, d\nu\right)^{1/p} \le \left(\int_S \|f_s\|^p \, d\nu\right)^{1/p} + \left(\int_S \|g_s\|^p \, d\nu\right)^{1/p},$$

where $f, g \in L_p(S,\nu,X)$. This is called the *Minkowski inequality for integrals.*

If the Banach space X and the measure space (S,Σ,ν) are separable, then $L_p(S,\nu,X)$ is separable for all $1 \le p < \infty$. When X is a Hilbert space, the space $L_2(S,\nu,X)$ is also a Hilbert space with the scalar product

$$\langle f, g\rangle_{L_2} = \int_S \langle f_s, g_s\rangle_X \, d\nu.$$

L_p-spaces are symmetric to l_p- and $\mathcal{L}_p$-spaces. The natural embedding, which hold for l_p- and $\mathcal{L}_p$-spaces, have symmetric analogs in case of L_p-spaces: if $1 \le p < q \le \infty$, then $L_q(S,\nu,X) \subset L_p(S,\nu,X)$ is a natural and tight embedding with

$$\begin{cases} \forall f \in L_q(S,\nu,X), \ \|f\|_{L_p} \le \nu(S)^{(q-p)/q}\|f\|_{L_q} & \text{if } q < \infty, \\ \forall f \in L_\infty(S,\nu,X), \ \|f\|_{L_p} \le \nu(S)\|f\|_{L_\infty} & \text{if } q = \infty. \end{cases}$$

The isometric equalities which take place for l_p- and $\mathcal{L}_p$-spaces have symmetric analogs for L_p-spaces: $L_1(S,\nu,X)^*$ and $L_\infty(S,\nu,X^*)$ are isometric under the isometry

$$L_1(S,\nu,X)^* \ni f^* \leftrightarrow Jf^* = f \in L_\infty(S,\nu,X^*) : f^*g = \int_S f_s g_s \, d\nu, \ g \in L_1(S,\nu,X),$$

and, if $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$, then $L_p(S,\nu,X)^*$ and $L_q(S,\nu,X^*)$ are isometric under the isometry

$$L_p(S,\nu,X)^* \ni f^* \leftrightarrow Jf^* = f \in L_q(S,\nu,X^*) : f^*g = \int_S f_s g_s \, d\nu, \ g \in L_p(S,\nu,X).$$

In particular, the last isometry yields

$$\int_S |f_s g_s| \, d\nu \le \left(\int_S \|f_s\|^q \, d\nu\right)^{1/q} \left(\int_S \|g_s\|^p \, d\nu\right)^{1/p}$$

for all $f \in L_q(S,\nu,X^*)$ and $g \in L_p(S,\nu,X)$. This inequality is called the *Hölder inequality for integrals*. For $p = q = 2$,

$$\int_S |f_s g_s| \, d\nu \le \left(\int_S \|f_s\|^2 \, d\nu\right)^{1/2} \left(\int_S \|g_s\|^2 \, d\nu\right)^{1/2}.$$

This is called the *Cauchy–Schwarz inequality for integrals*. We also mention that $C(S,X)$ is a subspace of $L_\infty(S,\nu,X)$ if S is a compact metric space and $\mathcal{B}_S \subset \Sigma$.

Convergence of functions in the norm of $L_\infty(S,\nu,X)$ is called *ν-a.e. uniform convergence*. Norm convergence in $L_2(S,\nu,X)$ (in $L_1(S,\nu,X)$) is called *mean square (mean) convergence*. Mean convergence implies convergence in measure. Therefore, if a sequence $\{f^n\}$ in $L_1(S,\nu,X)$ converges in norm to f, then there exists a subsequence of $\{f^n\}$ which converges ν-a.e. to f. But generally, mean convergence does not imply ν-a.e. convergence.

Theorem 2.17. *Let X be a Banach space X and let (S,Σ,ν) be a measure space. Then the Bochner integral is a bounded linear operator from $L_1(S,\nu,X)$ to X and*

$$\left\|\int_S f_s\,d\nu\right\| \leq \int_S \|f_s\|\,d\nu,\quad f \in L_1(S,\nu,X).$$

Theorem 2.18. *Let X, Y be Banach spaces and let (S,Σ,ν) be a measure space. If $f \in L_1(S,\nu,X)$ and $A \in \mathcal{L}(X,Y)$, then $Af \in L_1(S,\nu,Y)$ and*

$$A\int_S f_s\,d\nu = \int_S Af_s\,d\nu.$$

Theorem 2.19. *Let X be a Banach space and let (S,Σ,ν) be a measure space. If a sequence $\{f^n\}$ in $L_1(S,\nu,X)$ converges to f ν-a.e. on S and if there exists a function $g \in L_1(S,\nu,\mathbb{R})$ such that the inequality $\|f_s^n\| \leq g_s$ holds for ν-a.e. $s \in S$ and for all $n \in \mathbb{N}$, then $f \in L_1(S,\nu,X)$ and*

$$\lim_{n\to\infty}\int_S f_s^n\,d\nu = \int_S f_s\,d\nu.$$

In particular, if a series $\sum_{n=1}^{\infty}\varphi_s^n$ with $\varphi^n \in L_1(S,\nu,X)$, $n = 1,2,\ldots$, is majorized by a convergent series of real numbers, then

$$\sum_{n=1}^{\infty}\int_S \varphi_s^n\,d\nu = \int_S \sum_{n=1}^{\infty}\varphi_s^n\,d\nu.$$

We also present the following useful result.

Proposition 2.20. *Let $X \subset Y$ be a natural embedding of Banach spaces and let (S,Σ,ν) be a measure space. Then $L_p(S,\nu,X) \subset L_p(S,\nu,Y)$ is a natural embedding for all $1 \leq p \leq \infty$ as well. The Bochner integrals of a function $f \in L_1(S,\nu,X) \subset L_1(S,\nu,Y)$ in both of these spaces are equal. Furthermore, if the embedding $X \subset Y$ is tight, then the embedding $L_p(S,\nu,X) \subset L_p(S,\nu,Y)$ is tight for all $1 \leq p \leq \infty$ as well.*

Proof. This can be proved by direct verification. □

Two properties of the Bochner integral, Fubini's property and change of variable, are very important for us and they will be considered separately in Sections 2.4.2 and 2.4.3. Now let us set some useful notation.

If the measure space (S, Σ, ν) is equal to $(\Omega, \mathcal{F}, \mathbf{P})$ (fixed probability space in this book) or $(G, \mathcal{B}_G, \ell)$ or $(G \times \Omega, \mathcal{B}_G \otimes \mathcal{F}, \ell \otimes \mathbf{P})$, where $G \subset \mathbb{R}^n$, then the measure ν will be dropped in the symbols $m(S, \nu, X)$ and $L_p(S, \nu, X)$. For example, $L_2(\Omega, X) = L_2(\Omega, \mathcal{F}, \mathbf{P}, X)$. We will write *a.e.* instead of ℓ-a.e. If g is a function of bounded variation on a finite interval $[a, b]$, that is if it is the difference of two nondecreasing and right continuous functions u and v with the corresponding Lebesgue–Stieltjes measures ν and μ, respectively, then we write

$$\int_a^b f_s\, dg_s = \int_a^b f_s\, du_s - \int_a^b f_s\, dv_s = \int_{(a,b]} f_s\, d\nu - \int_{(a,b]} f_s\, d\mu.$$

In this case, the Bochner integral is also called the *Lebesgue–Stieltjes integral*. In particular, when $g_s = s$, $a \le s \le b$, the corresponding integral is called the *Lebesgue integral* which is denoted by

$$\int_a^b f_s\, ds = \int_{(a,b]} f_s\, d\ell.$$

Since the Lebesgue measure of a one-point set is equal to zero, for any fixed $1 \le p \le \infty$, the spaces $L_p([a,b], X)$, $L_p([a,b), X)$, $L_p((a,b], X)$ and $L_p((a,b), X)$ are isometric. All these spaces will be denoted by $L_p(a, b; X)$.

2.4.2 Fubini's Property

A function f of two variables can be considered as a function $f_{s,r}$ of one total variable $(s, r) \in S \times R$ as well as a function of one of the variables, say $s \in S$, with values in a space of functions with respect to another variable $r \in R$. The Bochner integral as applied to such functions is discussed in Dunford and Schwartz [45]. Below we present the important items of this discussion and complete it with Proposition 2.24.

Remark 2.21. In the sequel, to avoid possible ambiguity, we will use the following notation with brackets. Let $f_{s,r}$, $(s, r) \in S \times R$, be a function of two variables taking values in a set X. By $[f_s]$ (respectively, by $[f_r]$) we will denote the function f of the variable $r \in R$ by considering $s \in S$ as fixed (respectively, the variable $s \in S$ by considering $r \in R$ as fixed). The symbol f will denote either the function $f : S \times R \to X$ or the function $f : S \to F(R, X)$ or the function $f : R \to F(S, X)$ being easily clear what is meant from the context, where $F(S, X)$ is the set of all functions from S to X. Thus, with this notation

$$[f_s] \in F(R, X),\ \ s \in S; \quad [f_r] \in F(S, X),\ \ r \in R;$$
$$f_{s,r} = [f_s]_r = [f_r]_s,\ \ s \in S,\ \ r \in R.$$

Theorem 2.22 (Fubini). *Let X be a Banach space and let (S, Σ, ν) and (R, Γ, μ) be measure spaces. Then for all $f \in L_1(S \times R, \nu \otimes \mu, X)$,*

$$\int_{S \times R} f_{s,r}\, d(\nu \otimes \mu) = \int_S \int_R [f_s]_r\, d\mu\, d\nu = \int_R \int_S [f_r]_s\, d\nu\, d\mu.$$

Theorem 2.23. *Let X be a Banach space and let (S, Σ, ν) and (R, Γ, μ) be measure spaces. The following statements hold.*

(a) *If $f \in m(S \times R, \nu \otimes \mu, X)$, then $[f_s] \in m(R, \mu, X)$ for ν-a.e. $s \in S$.*

(b) *If $f \in m(S \times R, \nu \otimes \mu, X)$ and $[f_s] \in L_2(R, \mu, X)$ for ν-a.e. $s \in S$, then $f \in m(S, \nu, L_2(R, \mu, X))$.*

(c) *If $f \in m(S, \nu, L_2(R, \mu, X))$, then $f \in m(S \times R, \nu \otimes \mu, X)$.*

Proposition 2.24. *Let X be a Banach space and let (S, Σ, ν) and (R, Γ, μ) be measure spaces. Then $L_2(S \times R, \nu \otimes \mu, X)$ and $L_2(S, \nu, L_2(R, \mu, X))$ are isometric spaces under the isometry*

$$L_2(S \times R, \nu \otimes \mu, X) \ni f \leftrightarrow Jf = f \in L_2(S, \nu, L_2(R, \mu, X)).$$

Proof. Let $f \in L_2(S, \nu, L_2(R, \mu, X))$. Applying Theorem 2.23(c), we have $f \in m(S \times R, \nu \otimes \mu, X)$. By Tonelli's theorem (see Dunford and Schwartz [45], p. 194),

$$\int_{S \times R} \|f_{s,r}\|^2 \, d(\nu \otimes \mu) = \int_S \left(\int_R \| [f_s]_r \|^2 \, d\mu \right) d\nu < \infty. \tag{2.3}$$

Thus, $f \in L_2(S \times R, \nu \otimes \mu, X)$. Now let $f \in L_2(S \times R, \nu \otimes \mu, X)$. By Theorem 2.23(a), $[f_s] \in m(R, \mu, X)$ for ν-a.e. $s \in S$. Applying Tonelli's theorem, we obtain $[f_s] \in L_2(R, \mu, X)$ for ν-a.e. $s \in S$. Therefore, by Theorem 2.23(b), $f \in m(S, \nu, L_2(R, \mu, X))$. Applying Tonelli's theorem again, we see that $f \in L_2(S, \nu, L_2(R, \mu, X))$. Thus, the isomorphism of the spaces $L_2(S \times R, \nu \otimes \mu, X)$ and $L_2(S, \nu, L_2(R, \mu, X))$ is proved. By (2.3), these spaces are isometric. □

In the sequel, the spaces $L_2(S \times R, \nu \otimes \mu, X)$, $L_2(S, \nu, L_2(R, \mu, X))$ and $L_2(R, \mu, L_2(S, \nu, X))$ will be identified.

The following proposition completes Theorem 2.23.

Proposition 2.25. *Let $X \in \mathcal{H}$, let (S, Σ) be a measurable space, let $[a, b]$ be a finite interval in $\mathbb{R}$ and let the function $f : [a, b] \times S \to X$ be given. If $[f_t] \in m(S, \Sigma, X)$ for all $t \in [a, b]$ and $[f_s] \in C(a, b; X)$ for all $s \in S$, then $f \in m(S, \Sigma, C(a, b; X))$.*

Proof. First note that, if $\{g^n\}$ is a sequence of functions in $m(S, \Sigma, \mathbb{R})$ and $g_s = \sup_n g_s^n < \infty$, $s \in S$, then $g \in m(S, \Sigma, \mathbb{R})$. Therefore, for any $x \in C(a, b; X)$, the function

$$\| [f_s] - x \|_C = \sup \| [f_s]_t - x_t \|, \ s \in S,$$

with the supremum taken over all rational numbers t in $[a, b]$, is Σ-measurable. Hence, $\{s \in S : \| [f_s] - x \|_C < r\} \in \Sigma$ for all $r > 0$. Thus, the inverse image of each open ball in $C(a, b; X)$ under the function f belongs to Σ. Since $C(a, b; X)$ is separable, $\mathcal{B}_{C(a,b;X)}$ is the smallest σ-algebra generated by all open balls in $C(a, b; X)$. This implies the Σ-measurability of the function $f : S \to C(a, b; X)$ and, therefore, $f \in m(S, \Sigma, C(a, b; X))$ since $C(a, b; X)$ is separable. □

2.4.3 Change of Variable

Theorem 2.26. *Let X, Y be Banach spaces and let (S, Σ, ν) be a measure space. If f is a ν-measurable function from S to X and ν_f is the measure on $\mathcal{B}_X$ generated by f, then for all $\varphi \in L_1(X, \nu_f, Y)$,*

$$\int_X \varphi(x)\, d\nu_f = \int_S \varphi(f_s)\, d\nu.$$

Theorem 2.26 expresses the change of variable property of the Bochner integral and, together with Theorem 2.10, it forms a basis for the following results.

Proposition 2.27. *Let X, Y be Hilbert spaces, let (S, Σ, ν) be a measure space and let ν_f be a measure on $\mathcal{B}_X$ generated by a ν-measurable function $f : S \to X$. Then $g \in L_2(S, \sigma(f), \nu, Y)$ if and only if there exists a function $\varphi \in L_2(X, \nu_f, Y)$ such that g is the composition of φ and f. Furthermore, the spaces $L_2(X, \nu_f, Y)$ and $L_2(S, \sigma(f), \nu, Y)$ are isometric under the isometry*

$$L_2(X, \nu_f, Y) \ni \varphi \leftrightarrow J\varphi = \varphi \circ f \in L_2(S, \sigma(f), \nu, Y).$$

Proof. By Theorem 2.26, J is a bounded linear operator from $L_2(X, \nu_f, Y)$ to $L_2(S, \sigma(f), \nu, Y)$. Let us show that J has a bounded inverse. For this, take $g \in L_2(S, \sigma(f), \nu, Y)$. Then $g \in m(S, \sigma(f), \nu, Y)$ and

$$\int_S \|g_s\|^2\, d\nu < \infty.$$

By Theorems 2.10 and 2.26, there exists $\varphi \in m(X, \nu_f, Y)$ such that $g = \varphi \circ f$ and

$$\int_X \|\varphi(x)\|^2\, d\nu_f = \int_S \|g_s\|^2\, d\nu < \infty.$$

We conclude that $g = \varphi \circ f$ with $\varphi \in L_2(X, \nu_f, Y)$ and

$$\|\varphi\|_{L_2(X, \nu_f, Y)} = \|g\|_{L_2(S, \sigma(f), \nu, Y)}.$$

Hence, J has a bounded inverse and, moreover, it is the isometry mentioned in the proposition. □

Proposition 2.28. *Let X, Y, Z be Hilbert spaces and let (S, Σ, ν) be a measure space. The following statements hold.*

(a) *If Σ_0 is a sub-σ-algebra of Σ, then $L_2(S, \Sigma_0, \nu, X)$ is a subspace of $L_2(S, \Sigma, \nu, X)$.*

(b) *If $f : S \to Z$ is a ν-measurable function and $g \in m(S, \sigma(f), \nu, Y)$, then $L_2(S, \sigma(g), \nu, X)$ is a subspace of $L_2(S, \sigma(f), \nu, X)$.*

Proof. Part (a) can be proved by direct verification. Part (b) is a consequence of part (a). □

To continue, we introduce the following definition. The set

$$\int_a^b H_t\,dt = \{g \in L_2(a,b;X) : g_t \in H_t \text{ for a.e. } t \in (a,b]\}$$

is called the *Hilbertian sum* of subspaces H_t, $a < t \le b$, of a Hilbert space X.

Proposition 2.29. *Let X, Y be Hilbert spaces, let (S,Σ,ν) be a measure space, let $[a,b]$ be a finite interval in $\mathbb{R}$ and let f be a ν-measurable function on S to $L_2(a,b;Y)$. For $a < t \le b$, denote by f^t the restriction of f to $[a,t] \times S$. Let $X_t^f = L_2\big(S, \sigma(f^t), \nu, X\big)$. Then*

(a) *X_r^f is a subspace of X_t^f for all $a < r \le t \le b$;*

(b) *$\tilde{X}^f = \int_a^b X_t^f\,dt$ is a subspace of $L_2(a,b;L_2(S,\nu,X))$.*

Proof. Part (a) is a consequence of Proposition 2.28(b) since f^r is the composition of the restriction operation and f^t. For part (b), clearly, $\tilde{X}^f$ is a linear subspace of $L_2(a,b;L_2(S,\nu,X))$. Let $\{g^n\}$ be a sequence in $\tilde{X}^f$ which converges to g in norm of $L_2(a,b;L_2(S,\nu,X))$. Then some subsequence of $\{g^n\}$ converges to g a.e. on $[a,b]$ in norm of $L_2(S,\nu,X)$. Since $[g_t^n] \in X_t^f$ for a.e. $t \in (a,b]$ and for all $n \in \mathbb{N}$, we have $[g_t] \in X_t^f$ for a.e. $t \in (a,b]$, that is $g \in \tilde{X}^f$. Thus, $\tilde{X}^f$ is closed under the convergence in norm of $L_2(a,b;L_2(S,\nu,X))$ and, consequently, it is a subspace of $L_2(a,b;L_2(S,\nu,X))$. □

Proposition 2.30. *Let $Y \in \mathcal{H}$, let (S,Σ,ν) be a measure space, let $[a,b]$ be a finite interval in $\mathbb{R}$ and let $f^c \in m(S,\nu,C(a,b;Y))$. Consider f^c as a function with values in $L_2(a,b;Y)$ and denote this function by f. Then*

$$\sigma(f) = \sigma(f^c) = \sigma([f_t^c];\ a \le t \le b).$$

Proof. Since $Y \in \mathcal{H}$, $C(a,b;Y) \subset L_2(a,b;Y)$ is a natural embedding of separable spaces. Applying Proposition 1.15, we obtain $\sigma(f) = \sigma(f^c)$. The equality $\sigma(f^c) = \sigma([f_t^c];\ a \le t \le b)$ is a consequence of the fact that $\mathcal{B}_{C(a,b;Y)}$ is the smallest σ-algebra generated by the system of all cylindrical subsets of $C(a,b;Y)$, i.e., the subsets of the form $\{g \in C(a,b;Y) : g_{t_n} \in B_n,\ n = 1,\dots,k\}$, where $B_n \in \mathcal{B}_Y$, $a \le t_1 \le \cdots \le t_k \le b$ and $k \in \mathbb{N}$. □

Proposition 2.31. *Let $X, Y \in \mathcal{H}$, let (S,Σ,ν) be a measure space, let $[a,b]$ be a finite interval in $\mathbb{R}$ and let $f^c \in m(S,\nu,C(a,b;Y))$. Consider f^c as a function with values in $L_2(a,b;Y)$ and denote this function by f. For $a < t \le b$, denote* *by* $f^{c,t}$ *and* f^t *the restrictions of* f^c *and* f*, respectively, to* $[a,t] \times S$*. Let* $X_t^{c,f} = L_2(S, \sigma(f^{c,t}), \nu, X)$ *and* $X_t^f = L_2(S, \sigma(f^t), \nu, X)$*. Then*

(a) $X_t^f = X_t^{c,f}$, $a < t \leq b$;

(b) $\int_a^b X_t^f \, dt = \int_a^b X_t^{c,f} \, dt.$

Proof. By Proposition 2.30, $\sigma(f^t) = \sigma(f^{c,t})$ for all $a < t \leq b$. This implies part (a) and part (b) as well. □

2.4.4 Strong Bochner Integral

Let $X, Y \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. Operator-valued functions in $L_1(S, \nu, \mathcal{L}(X,Y))$ are said to be *uniformly ν-integrable*. The space $L_1(S, \nu, \mathcal{L}(X,Y))$ is not sufficiently large for our purposes because it is based on the concept of uniform ν-measurability and, hence, it does not contain useful examples of strongly ν-measurable operator-valued functions. Therefore, we need a generalization of the Bochner integral for the case of strongly ν-measurable operator-valued functions.

Proposition 2.32. *Let* $X, Y \in \mathcal{H}$ *and let* (S, Σ, ν) *be a measure space. If the operator-valued function* $F : S \to \mathcal{L}(X,Y)$ *is strongly ν-measurable, then the real-valued function* $\|F\|_{\mathcal{L}} : S \to \mathbb{R}$ *is ν-measurable.*

Proof. Let D be a countable dense subset of X not containing the zero vector. Since F is strongly ν-measurable, the real-valued function $\|F_s x\|$, $s \in S$, is ν-measurable for all $x \in X$. Hence, the real-valued function

$$\|F_s\|_{\mathcal{L}} = \sup_{\|x\|=1} \|F_s x\| = \sup_{x \in D} \frac{\|F_s x\|}{\|x\|}, \quad s \in S,$$

is ν-measurable as well. □

By Proposition 2.32, the following classes of strongly ν-measurable operator-valued functions can be defined. Let $X, Y \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. By $B_p(S, \Sigma, \nu, \mathcal{L}(X,Y))$ (or $B_p(S, \nu, \mathcal{L}(X,Y))$ if there is no ambiguity about Σ) we denote the class of all (equivalence classes of ν-a.e. equal) strongly ν-measurable $\mathcal{L}(X,Y)$-valued functions on S with

$$\int_S \|F_s\|^p \, d\nu < \infty \text{ if } 1 \leq p < \infty \text{ and } \operatorname*{ess\,sup}_{s \in S} \|F_s\| < \infty \text{ if } p = \infty.$$

$B_p(S, \nu, \mathcal{L}(X,Y))$ is a Banach space (see Thomas [88]) with the respective norm

$$\|F\|_{B_p} = \left(\int_S \|F_s\|^p \, d\nu \right)^{1/p}, \; 1 \leq p < \infty; \quad \|F\|_{B_\infty} = \operatorname*{ess\,sup}_{s \in S} \|F_s\|.$$

Clearly, $L_p(S,\nu,\mathcal{L}(X,Y))$ is a subspace of $B_p(S,\nu,\mathcal{L}(X,Y))$ if $1 \le p \le \infty$. Since the concepts of uniform and strong ν-measurability coincide for $\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$-valued functions,

$$L_p(S,\nu,\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)) = B_p(S,\nu,\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)),\ 1 \le p \le \infty,\ n,m \in \mathbb{N}.$$

Also,

$$B_q(S,\nu,\mathcal{L}(X,Y)) \subset B_p(S,\nu,\mathcal{L}(X,Y)),\ 1 \le p < q \le \infty,$$

is a natural embedding.

If $S \subset \mathbb{R}^n$, $\Sigma = \mathcal{B}_S$ and $\nu = \ell$ (the Lebesgue measure on $\mathcal{B}_S$), then the notation $B_p(S,\mathcal{L}(X,Y))$ for $B_p(S,\mathcal{B}_S,\ell,\mathcal{L}(X,Y))$ will be used. Also, we briefly write $B_p(a,b;\mathcal{L}(X,Y)) = B_p(S,\mathcal{L}(X,Y))$ if $S = [a,b]$.

The operator-valued functions of the space $B_1(S,\nu,\mathcal{L}(X,Y))$ are said to be *strongly ν-integrable*. If $F \in B_1(S,\nu,\mathcal{L}(X,Y))$, then $Fx \in L_1(S,\nu,Y)$ for all $x \in X$. The operator $G \in \mathcal{L}(X,Y)$, defined by

$$Gx = \int_S F_s x\, d\nu,\ x \in X,$$

is called the *strong Bochner integral* of $F \in B_1(S,\nu,\mathcal{L}(X,Y))$. The notation

$$\int_S F_s\, d\nu$$

is used for the strong Bochner integral of F.

Note that, if $F \in L_1(S,\nu,\mathcal{L}(X,Y))$, then the strong and uniform Bochner integrals of F coincide. Therefore, the above notation may denote either of these integrals without any ambiguity. Moreover, if $F \in L_1(S,\nu,\mathcal{L}_1(X,Y))$, then, by Proposition 2.20, the Bochner integrals of F as $\mathcal{L}$- as well as $\mathcal{L}_1$- and $\mathcal{L}_2$-valued functions coincide and we will use the above notation for all them.

Proposition 2.33. *Let $X,Y \in \mathcal{H}$ and let (S,Σ,ν) be a measure space. The following statements hold for all $1 \le p \le \infty$.*

(a) *If $F \in B_\infty(S,\nu,\mathcal{L}(X,Y))$ and $f \in L_p(S,\nu,X)$, then $Ff \in L_p(S,\nu,Y)$.*

(b) *If $F \in B_p(S,\nu,\mathcal{L}(X,Y))$ and $f \in L_\infty(S,\nu,X)$, then $Ff \in L_p(S,\nu,Y)$.*

(c) *If $F \in B_p(S,\nu,\mathcal{L}(X,Y))$, then $F^* \in B_p(S,\nu,\mathcal{L}(Y,X))$ and*

$$\left(\int_S F_s\, d\nu\right)^* = \int_S F_s^*\, d\nu.$$

Proof. This can be proved by direct verification using Proposition 2.11. □

Remark 2.34. From the Propositions 2.5(b), 2.8(a), 2.11(a), 2.33(a) and 2.33(b), one can observe that the strong continuity, differentiability, measurability, integrability properties of a vector-valued function f remain valid for the composition Ff when the operator-valued function F has not only the uniform, but also the strong version of the corresponding property. This shows the efficiency of strong operator convergence as a generalization of the uniform one.

Proposition 2.35. *Let $X \in \mathcal{H}$ and let (S, Σ, ν) be a measure space such that $\nu(S) > 0$. Suppose that $F \in B_\infty(S, \nu, \mathcal{L}(X))$, $F_s > 0$ for ν-a.e. $s \in S$ and $F^{-1} \in B_\infty(S, \nu, \mathcal{L}(X))$. Then $\langle x, F_s x\rangle \geq k^{-1}\|x\|^2$ for all $x \in X$ and for ν-a.e. $s \in S$, where $k = \operatorname{ess\,sup}_{s\in S}\|F_s^{-1}\|$.*

Proof. First, note that $F_s > 0$ for ν-a.e. $s \in S$ implies $F_s^{-1} > 0$ for ν-a.e. $s \in S$. Hence, $k > 0$ since $\nu(S) > 0$. Further, let us show that if $A \in \mathcal{L}(X)$ and $A > 0$, then $\langle x, Ax\rangle \geq \|A^{-1}\|^{-1}\|x\|^2$ for all $x \in X$. Indeed,

$$\|A^{-1}\| = \sup_{y\neq 0} \frac{\langle y, A^{-1}y\rangle}{\|y\|^2}$$

implies

$$\|y\|^2 \|A^{-1}\| \geq \langle y, A^{-1}y\rangle, \ y \in X.$$

Substituting $y = A^{1/2}x$, for $x \in X$, we obtain

$$\langle A^{1/2}x, A^{1/2}x\rangle\|A^{-1}\| \geq \langle A^{1/2}x, A^{-1}A^{1/2}x\rangle$$

or

$$\langle x, Ax\rangle \geq \|A^{-1}\|^{-1}\langle x, A^{1/2}A^{-1}A^{1/2}x\rangle.$$

But, $A^{1/2}A^{-1}A^{1/2} = A^{1/2}A^{-1}A^{1/2}A^{1/2}(A^{1/2})^{-1} = A^{1/2}(A^{1/2})^{-1} = I$. Therefore,

$$\langle x, Ax\rangle \geq \|A^{-1}\|^{-1} \|x\|^2, \ x \in X.$$

Using this inequality, we have

$$\langle x, F_s x\rangle \geq \|F_s^{-1}\|^{-1} \|x\|^2 \geq k^{-1}\|x\|^2$$

for all $x \in X$ and for ν-a.e. $s \in S$. □

Apart from B_p-spaces, the following class of operator-valued functions is also useful. If S is a metric space and if $X, Y \in \mathcal{H}$, then the class of all strongly $\mathcal{B}_S$-measurable functions from S to $\mathcal{L}(X, Y)$ with $\sup_{s\in S} \|F_s\| < \infty$ will be denoted by $B(S, \mathcal{L}(X, Y))$. As always, we write $B(a, b; \mathcal{L}(X, Y)) = B(S, \mathcal{L}(X, Y))$ if $S = [a, b]$.

2.4.5 Bochner Integral of $\mathcal{L}_1$- and $\mathcal{L}_2$-Valued Functions

We will now consider some results related with the integration of $\mathcal{L}_1$- and $\mathcal{L}_2$-valued functions.

Proposition 2.36. *Let $X, Y \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. If $f \in L_2(S, \nu, X)$ and $g \in L_2(S, \nu, Y)$, then $(f \otimes g) \in L_1(S, \nu, \mathcal{L}_1(Y, X))$.*

Proof. By Proposition 2.14, $(f \otimes g) \in m(S, \nu, \mathcal{L}_1(Y, X))$ and, by Proposition 1.26(c),

$$\int_S \|f_s \otimes g_s\|_{\mathcal{L}_1} \, d\nu = \int_S \|f_s\| \, \|g_s\| \, d\nu < \infty.$$

Hence, $(f \otimes g) \in L_1(S, \nu, \mathcal{L}_1(Y, X))$. □

Proposition 2.37. *Let $X \in \mathcal{H}$ and let (S, Σ, ν) be a measure space. If the function F belongs to $L_\infty(S, \nu, \mathcal{L}_1(X))$ and $F_s \geq 0$ for ν-a.e. $s \in S$, then $F^{1/2} \in L_\infty(S, \nu, \mathcal{L}_2(X))$.*

Proof. By Proposition 2.15, $F^{1/2} \in m(S, \nu, \mathcal{L}_2(X))$. Also, $\|F_s^{1/2}\|_{\mathcal{L}_2} = \|F_s\|_{\mathcal{L}_1} = \mathrm{tr} F_s$, ν-a.e. $s \in S$. Thus, $F^{1/2} \in L_\infty(S, \nu, \mathcal{L}_2(X))$. □

2.5 Integral and Differential Operators

The integral and differential operators which are based on the integration and differentiation operations, respectively, define two important classes of linear operators on L_p-spaces.

2.5.1 Integral Operators

Let $X, Y \in \mathcal{H}$, and let (S, Σ, ν) and (R, Γ, μ) be measure spaces. The operators A and B from $L_2(S, \nu, X)$ to Y and to $L_2(R, \mu, Y)$, respectively, defined by

$$Af = \int_S M_s f_s \, d\nu, \ \ f \in L_2(S, \nu, X),$$

$$[Bf]_r = \int_S N_{s,r} f_s \, d\nu, \ \ \mu\text{-a.e. } r \in R, \ \ f \in L_2(S, \nu, X),$$

where $M \in B_2(S, \nu, \mathcal{L}(X, Y))$ and $N \in B_2(S \times R, \nu \otimes \mu, \mathcal{L}(X, Y))$, are called *linear integral operators*. It can be verified that

$$A \in \mathcal{L}(L_2(S, \nu, X), Y) \text{ with } \|A\| = \left(\int_S \|M_s\|^2 \, d\nu \right)^{1/2}$$

and

$$B \in \mathcal{L}(L_2(S, \nu, X), L_2(R, \mu, Y)) \text{ with } \|B\| = \left(\int_S \int_R \|N_{s,r}\|^2 \, d\mu \, d\nu \right)^{1/2}.$$

One can compute the adjoint A^* and B^* of A and B, respectively, in the form

$$[A^*y]_s = M_s^* y, \ \nu\text{-a.e. } s \in S, \ y \in Y,$$

and

$$[B^*g]_s = \int_R N_{s,r}^* g_r \, d\mu, \ \nu\text{-a.e. } s \in S, \ g \in L_2(R, \mu, Y).$$

Related with the integral operators one can consider the equation

$$f_t = g_t + \int_0^t G_{t,s} f_s \, ds, \tag{2.4}$$

where g and G are given functions and f is unknown. The equation (2.4), called a *Volterra integral equation*, often arises in applied problems.

Theorem 2.38. *Let $X \in \mathcal{H}$, let $T > 0$ and suppose $G \in B_2(\Delta_T, \mathcal{L}(X))$. Then there exists $F \in B_2(\Delta_T, \mathcal{L}(X))$ such that for all $g \in L_2(0, T; X)$,*

$$f_t = g_t + \int_0^t F_{t,s} g_s \, ds, \ \ a.e. \ t \in [0, T],$$

is a unique solution of the equation (2.4) in $L_2(0, T; X)$. Furthermore, F is a unique solution in $B_2(\Delta_T, \mathcal{L}(X))$ of the equation

$$F_{t,s} = G_{t,s} + \int_s^t G_{t,r} F_{r,s} \, dr, \ \ a.e. \ (t, s) \in \Delta_T.$$

When $G \in B_\infty(\Delta_T, \mathcal{L}(X))$, this theorem can be proved as Theorem 3.13. In the general case the function $G \in B_2(\Delta_T, \mathcal{L}(X))$ must be approximated by the functions in $B_\infty(\Delta_T, \mathcal{L}(X))$.

2.5.2 Integral Hilbert–Schmidt Operators

The following proposition shows that Hilbert–Schmidt operators on L_2-spaces can be completely described by linear integral operators.

Proposition 2.39. *Let $X, Y \in \mathcal{H}$, let (S, Σ, ν) and (R, Γ, μ) be separable measure spaces. Then the following statements are true.*

(a) *The spaces $L_2(S, \nu, \mathcal{L}_2(X, Y))$ and $\mathcal{L}_2(X, L_2(S, \nu, Y))$ are isometric under the isometry*

$$L_2(S, \nu, \mathcal{L}_2(X, Y)) \ni \Phi \leftrightarrow J_1\Phi = \tilde{\Phi} \in \mathcal{L}_2(X, L_2(S, \nu, Y)) :$$
$$\left[\tilde{\Phi} h\right]_s = \Phi_s h, \ \nu\text{-}a.e. \ s \in S, \ h \in X.$$

(b) *The spaces* $L_2(S,\nu,\mathcal{L}_2(X,Y))$ *and* $\mathcal{L}_2(L_2(S,\nu,X),Y)$ *are isometric under the isometry*

$$L_2(S,\nu,\mathcal{L}_2(X,Y)) \ni \Phi \leftrightarrow J_2\Phi = \tilde{\Phi} \in \mathcal{L}_2(L_2(S,\nu,X),Y):$$
$$\tilde{\Phi}f = \int_S \Phi_s f_s \, d\nu, \;\; f \in L_2(S,\nu,X).$$

(c) *The spaces* $\tilde{L}_2$ *and* $\tilde{\mathcal{L}}_2$*, where* $\tilde{L}_2 = L_2(S\times R,\nu\otimes\mu,\mathcal{L}_2(X,Y))$ *and* $\tilde{\mathcal{L}}_2 = \mathcal{L}_2(L_2(S,\nu,X),L_2(R,\mu,Y))$*, are isometric under the isometry*

$$\tilde{L}_2 \ni \Phi \leftrightarrow J_3\Phi = \tilde{\Phi} \in \tilde{\mathcal{L}}_2:$$
$$\left[\tilde{\Phi}f\right]_r = \int_S \Phi_{s,r} f_s \, d\nu, \;\; \mu\text{-a.e. } r \in R, \;\; f \in L_2(S,\nu,X).$$

Proof. First, note that separability of the measure space (S,Σ,ν) and $X \in \mathcal{H}$ imply $L_2(S,\nu,X) \in \mathcal{H}$. So, the $\mathcal{L}_2$-spaces, considered above, are well-defined. For part (a), let $\Phi \in L_2(S,\nu,\mathcal{L}_2(X,Y))$. Then, by Proposition 2.12(a), $\Phi h \in m(S,\nu,Y)$ for all $h \in X$ and, by the monotone convergence theorem (see Rudin [85], p. 318),

$$\sum_{n=1}^{\dim X} \int_S \|\Phi_s e_n\|^2 \, d\nu = \int_S \sum_{n=1}^{\dim X} \|\Phi_s e_n\|^2 \, d\nu < \infty,$$

where $\{e_n\}$ is a basis in X. Hence, $\tilde{\Phi} \in \mathcal{L}_2(X, L_2(S,\nu,Y))$. Thus, J_1 is a transformation from $L_2(S,\nu,\mathcal{L}_2(X,Y))$ to $\mathcal{L}_2(X,L_2(S,\nu,Y))$. Clearly, J_1 is a bounded linear operator with $\|J_1\Phi\|_{\mathcal{L}_2} = \|\Phi\|_{L_2}$. It remains to prove that J_1 has a bounded inverse. For this, let $\tilde{\Phi} \in \mathcal{L}_2(X,L_2(S,\nu,Y))$. Then one can define the function Φ as in part (a) of the proposition. By the monotone convergence theorem, one can verify that Φ has values in $\mathcal{L}_2(X,Y)$ ν-a.e. on S. Then from Proposition 2.12(a), it is easy to see that $\Phi \in L_2(S,\nu,\mathcal{L}_2(X,Y))$. Thus, J_1 has a bounded inverse proving part (a). In view of

$$\begin{aligned}\Phi \in L_2(S,\nu,\mathcal{L}_2(X,Y)) &\Leftrightarrow \Phi^* \in L_2(S,\nu,\mathcal{L}_2(Y,X)) \\ &\Leftrightarrow J_1\Phi^* \in \mathcal{L}_2(Y,L_2(S,\nu,X)) \\ &\Leftrightarrow (J_1\Phi^*)^* \in \mathcal{L}_2(L_2(S,\nu,X),Y),\end{aligned}$$

where J_1 is defined in part (a), it follows that $\Phi \leftrightarrow (J_1\Phi^*)^*$ is an isometry between $L_2(S,\nu,\mathcal{L}_2(X,Y))$ and $\mathcal{L}_2(L_2(S,\nu,X),Y)$. One can verify that $(J_1\Phi^*)^* = J_2\Phi$. This proves part (b). Also, from

$$\begin{aligned}\Phi \in \tilde{L}_2 &\Leftrightarrow \Phi \in L_2(S,\nu,L_2(R,\mu,\mathcal{L}_2(X,Y))) \\ &\Leftrightarrow J_1\Phi \in L_2(S,\nu,\mathcal{L}_2(X,L_2(R,\mu,Y))) \\ &\Leftrightarrow J_2J_1\Phi \in \tilde{\mathcal{L}}_2,\end{aligned}$$

we obtain the isometry $\Phi \leftrightarrow J_2 J_1 \Phi$ between $\tilde{L}_2$ and $\tilde{\mathcal{L}}_2$. Note that in this isometry, J_1 is related with the variable in R and J_2 with the variable in S. One can verify that $J_2 J_1 \Phi = J_3 \Phi$. This proves part (c). □

In the sequel the spaces $L_2(S, \nu, \mathcal{L}_2(X, Y))$ and $\mathcal{L}_2(X, L_2(S, \nu, Y))$ and also their corresponding elements Φ and $J_1 \Phi$ will be identified.

2.5.3 Differential Operators

Let $X \in \mathcal{H}$, let $[a, b]$ be a finite interval in $\mathbb{R}$ and let $W^{1,p}(a, b; X)$ be the class of all functions $f : [a, b] \to X$ that can be represented in the form

$$f_t = f_a + \int_a^t g_s \, ds = f_b - \int_t^b g_s \, ds, \;\; a \le t \le b, \tag{2.5}$$

for some $g \in L_p(a, b; X)$, $1 \le p \le \infty$. The symbol $W^{n,p}(a, b; X)$, where $n = 2, 3, \ldots$ and $1 \le p \le \infty$, will denote the class of functions $f : [a, b] \to X$ which have $(n-1)$st derivative in $W^{1,p}(a, b; X)$. Under the corresponding norm, $W^{n,p}(a, b; X)$ is a Banach space. In particular, $W^{1,2}(a, b; X)$ is a Hilbert space in which a scalar product can be defined by

$$\langle f, g \rangle_{W^{1,2}} = \langle f_b, g_b \rangle + \int_a^b \langle f_t', g_t' \rangle \, dt.$$

Theorem 2.40. *Let $X \in \mathcal{H}$ and let $[a, b]$ be a finite interval in $\mathbb{R}$. If $g \in L_1(a, b; X)$, then the function f defined by (2.5) is continuous and a.e. differentiable on $[a, b]$ with $f_t' = g_t$ for a.e. $t \in [a, b]$. Furthermore, the derivative of f is defined for all t at which g is continuous. In particular, f is continuously differentiable if $g \in C(a, b; X)$.*

The following propositions generalize Theorem 2.40 in two different directions.

Proposition 2.41. *Let $X \in \mathcal{H}$, let $T > 0$ and let $f : \Delta_T \to X$ be a function satisfying*

(a) *$[f_s] \in W^{1,1}(s, T; X)$ for a.e. $s \in [a, b)$;*

(b) *$[f_t] \in L_1(0, t; X)$ for all $0 < t \le T$;*

(c) *$h \in L_1(0, T; X)$ where $h_s = f_{s,s}$ for a.e. $s \in [0.T]$;*

(d) *$(\partial/\partial t) f \in L_1(\Delta_T, X)$.*

Then the function

$$F_t = \int_0^t f_{t,s} \, ds, \;\; 0 \le t \le T,$$

belongs to $W^{1,1}(0,T;X)$ *with*

$$F_t' = f_{t,t} + \int_0^t \frac{\partial}{\partial t} f_{t,s}\, ds, \ \text{a.e. } t \in [0,T].$$

In particular, if $f, (\partial/\partial t)f \in C(\Delta_T, X)$, *then the function* F *is continuously differentiable on* $[0,T]$.

Proof. By condition (b), the function F is defined over all $[0,T]$. Using Theorem 2.22, we have

$$\begin{aligned} F_t &= \int_0^t f_{t,s}\, ds = \int_0^t \left(f_{s,s} + \int_s^t \frac{\partial}{\partial r} f_{r,s}\, dr \right) ds \\ &= \int_0^t f_{r,r} dr + \int_0^t \int_0^r \frac{\partial}{\partial r} f_{r,s}\, ds\, dr = \int_0^t \left(f_{r,r} + \int_0^r \frac{\partial}{\partial r} f_{r,s}\, ds \right) dr. \end{aligned}$$

Hence, the first part of this proposition follows from Theorem 2.40. One can verify that if $f, (\partial/\partial t)f \in C(\Delta_T, X)$, then the integrand in the integral representation of F is continuous. Hence, the second part follows from Theorem 2.40 too. □

Proposition 2.42. *Let* $X \in \mathcal{H}$ *and suppose that* τ *is an increasing or decreasing real-valued function with* $D(\tau) = [a,b]$ *and* $R(\tau) = [a_1, b_1]$. *For a given* $f \in W^{1,1}(a_1, b_1; X)$, *define a function* g *by* $g_t = f_{\tau(t)}$, $a \le t \le b$. *Then* $g \in W^{1,1}(a,b;X)$ *with* $g_t' = f_{\tau(t)}' \tau'(t)$ *for a.e.* $t \in [a,b]$. *In particular, if* f *and* τ *are continuously differentiable, then* g *is also continuously differentiable.*

Proof. Without loss of generality, assume that τ is increasing. By Theorem 2.26,

$$g_t = f_{\tau(t)} = f_{a_1} + \int_{a_1}^{\tau(t)} f_s'\, ds = f_{a_1} + \int_{\tau^{-1}(a_1)}^t f_{\tau(s)}'\, d\tau(s) = g_a + \int_a^t f_{\tau(s)}' \tau'(s)\, ds.$$

Hence, by Theorem 2.40, the first statement is obtained. If f and τ are continuously differentiable, then the integrand in the integral representation of g is continuous. Hence, the second statement follows from Theorem 2.40 too. □

Proposition 2.43. *Let* $X, Y \in \mathcal{H}$, *let* $h, g \in W^{1,2}(a,b;X)$ *and let* F *be a strongly differentiable operator-valued function from* $[a,b]$ *to* $\mathcal{L}(X,Y)$ *with the derivative* F' *in* $B_2(a,b;\mathcal{L}(X,Y))$. *Then*

(a) $$F_b h_b - F_a h_a = \int_a^b F_s' h_s\, ds + \int_a^b F_s h_s'\, ds;$$

(b) $$\langle h_b, g_b \rangle - \langle h_a, g_a \rangle = \int_a^b \langle h_s', g_s \rangle\, ds + \int_a^b \langle h_s, g_s' \rangle\, ds.$$

Proof. This follows from Propositions 2.8(a), 2.8(b) and Theorem 2.40. □

Using the spaces $W^{1,2}(a,b;X)$ and $W^{2,2}(a,b;X)$, one can define the differential operators d/dt and d^2/dt^2. In view of Proposition 2.8(c), these operators are linear. But, on $L_2(a,b;X)$ these operators are not bounded since they can not be defined over all $L_2(a,b;X)$. The next proposition shows that these operators are closed linear operators on $L_2(a,b;X)$.

Proposition 2.44. *Let $X \in \mathcal{H}$ and let $[a,b]$ be a finite interval in $\mathbb{R}$. Then the following statements hold.*

(a) *The differential operator d/dt, defined on*

$$D(d/dt) = \left\{h \in W^{1,2}(a,b;X) : h_b = 0\right\},$$

belongs to $\tilde{\mathcal{L}}(L_2(a,b;X))$ and $(d/dt)^* = -d/dt$ with

$$D(-d/dt) = \left\{h \in W^{1,2}(a,b;X) : h_a = 0\right\}.$$

(b) *The differential operator d^2/dt^2, defined on*

$$D\left(d^2/dt^2\right) = \left\{h \in W^{2,2}(a,b;X) : h_a = h_b = 0\right\},$$

belongs to $\tilde{\mathcal{L}}(L_2(a,b;X))$ and $\left(d^2/dt^2\right)^* = d^2/dt^2$.

Proof. Only part (a) will be proved. Part (b) can be proved in a similar way. Clearly, $\overline{D(d/dt)} = L_2(a,b;X)$. To show the closedness of d/dt suppose that $\{f^n\}$ is a sequence in $D(d/dt)$ such that

$$\|f^n - f\|_{L_2} \to 0 \text{ and } \|(d/dt)f^n - g\|_{L_2} \to 0 \text{ as } n \to \infty.$$

Then for all $a \le t \le b$, we have

$$\begin{aligned}\left\|f_t^n + \int_t^b g_s\,ds\right\| &= \left\|\int_t^b \left(g_s - \frac{d}{ds}f_s^n\right)ds\right\| \\ &\le (b-a)\int_a^b \left\|g_s - \frac{d}{ds}f_s^n\right\|^2 ds \to 0,\ n \to \infty.\end{aligned}$$

Since $\|f^n - f\|_{L_2} \to 0$, there is a function in the equivalence class of $f \in L_2(a,b;X)$, which will also be denoted by f, such that

$$f_t = -\int_t^b g_s\,ds,\ a \le t \le b.$$

Therefore, $f \in D(d/dt)$ and $(d/dt)f = g$, i.e., $d/dt \in \tilde{\mathcal{L}}(L_2(a,b;X))$. Now let $f, g \in W^{1,2}(a,b;X)$. By Proposition 2.43(b),

$$\langle f_b, g_b\rangle - \langle f_a, g_a\rangle = \int_a^b \left\langle \frac{d}{dt}f_t, g_t\right\rangle dt + \int_a^b \left\langle f_t, \frac{d}{dt}g_t\right\rangle dt.$$

Substituting $f_b = g_a = 0$ in the above formula we obtain $(d/dt)^* = -d/dt$. □

2.5.4 Gronwall's Inequality and Contraction Mappings

In studying integral and differential equations the following two theorems are useful in stating the existence and the uniqueness of solutions.

Theorem 2.45 (Gronwall). *Let $[a, b]$ be a finite interval in $\mathbb{R}$ and suppose that f is a nonnegative function in $L_1(a, b; \mathbb{R})$ satisfying*

$$f_t \leq c_1 + c_2 \int_a^t f_s\, ds, \quad a \leq t \leq b,$$

where $c_1 \geq 0$ and $c_2 \geq 0$ are constants. Then

$$f_t \leq c_1 e^{c_2(t-a)}, \quad a \leq t \leq b.$$

In particular, $f_t = 0$ for all $a \leq t \leq b$ when $c_1 = 0$.

Theorem 2.46 (Contraction Mapping). *Let X be a Banach space, let G be a closed and nonempty subset of X and suppose that $F : G \to G$ is a nonlinear (in general) operator. Define recursively $F^m = F \circ F^{m-1}$, $m \in \mathbb{N}$, where $F^0 = I$. If F^m is a contraction mapping for some $m \in \mathbb{N}$, i.e., if there exists $0 < c < 1$ such that for all $x, y \in X$*

$$\|F^m(x) - F^m(y)\| \leq c\|x - y\|,$$

then the equation $x = F(x)$ has a unique solution in G.

Chapter 3

Evolution Operators

This chapter deals with semigroups of bounded linear operators and mild evolution operators and transformations on them. Also, we study Riccati equations in operator form.

Convention. In this chapter it is always assumed that $X, Y, U, H \in \mathcal{H}$, and $\Delta_T = \{(t,s) : 0 \le s \le t \le T\}$ for $T > 0$.

3.1 Main Classes of Evolution Operators

Two basic classes of evolution operators, namely strongly continuous semigroups of bounded linear operators and mild evolution operators, are discussed in this section.

3.1.1 Strongly Continuous Semigroups

Definition 3.1. A function $\mathcal{U} : [0,\infty) \to \mathcal{L}(X)$ is called a *strongly continuous semigroup* of bounded linear operators if

(a) $\mathcal{U}_0 = I$;

(b) $\mathcal{U}_{t+s} = \mathcal{U}_t \mathcal{U}_s$, $t \ge 0$, $s \ge 0$ (semigroup property);

(c) $\|\mathcal{U}_t x - x\| \to 0$ as $t \to 0$ for all $x \in X$ (strong continuity at zero).

The class of all strongly continuous semigroups with values in $\mathcal{L}(X)$ will be denoted by $\mathcal{S}(X)$.

By the semigroup property, a strongly continuous semigroup can be uniquely defined if it is given on any finite interval $[0, T]$ with $T > 0$. Using Definition 3.1, one can prove that a strongly continuous semigroup is strongly continuous at each $t > 0$ as an operator-valued function.

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

Given $\mathcal{U} \in \mathcal{S}(X)$, one can consider the linear operator

$$A: \; Ax = \lim_{t\to 0} t^{-1}(\mathcal{U}_t - I)x, \tag{3.1}$$

defined for those $x \in X$ for which the limit in (3.1) exists. It can be proved that $A \in \tilde{\mathcal{L}}(X)$. The operator A is called the *infinitesimal generator* of $\mathcal{U}$. We will also say that A *generates* $\mathcal{U}$. Using (3.1), one can show that

$$x \in D(A) \;\Rightarrow\; \mathcal{U}_t x \in D(A), \; t \geq 0,$$

and

$$\frac{d}{dt}\mathcal{U}_t x = A\mathcal{U}_t x = \mathcal{U}_t Ax, \; x \in D(A), \; t \geq 0.$$

These equalities can also be written in the integral form

$$\mathcal{U}_t x = x + \int_0^t A\mathcal{U}_s x\, ds = x + \int_0^t \mathcal{U}_s Ax\, ds, \; x \in D(A), \; t \geq 0, \tag{3.2}$$

where the integrals are in the Bochner sense.

From (3.2), one can observe that a strongly continuous semigroup $\mathcal{U}$ with the infinitesimal generator A is a tool for representing the solution of the linear differential equation

$$x_t' = Ax_t, \; t > 0, \; x_0 \in D(A).$$

So, it is important to describe the class of all closed linear operators which generate a strongly continuous semigroup.

Theorem 3.2 (Hille–Yosida–Phillips). *An operator $A \in \tilde{\mathcal{L}}(X)$ generates a strongly continuous semigroup if and only if there exist the numbers M and ω such that for all $\lambda > \omega$, $(\lambda I - A)^{-1} \in \mathcal{L}(X)$ and*

$$\|(\lambda I - A)^{-n}\| \leq M(\lambda - \omega)^{-n}, \; n = 1, 2, \ldots.$$

3.1.2 Examples

The examples given below demonstrate that the solutions of different differential equations can be represented by use of strongly continuous semigroups.

Example 3.3. A bounded linear operator $A \in \mathcal{L}(X)$ generates the semigroup

$$\mathcal{U}_t = e^{At} = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!},$$

which is uniformly continuous. In particular, if $A = 0$, then $\mathcal{U}_t \equiv I$. Conversely, if a semigroup is uniformly continuous, then its infinitesimal generator is a bounded linear operator.

Example 3.4. If $A \in \tilde{\mathcal{L}}(X)$ generates $\mathcal{U} \in \mathcal{S}(X)$, then A^* generates $\mathcal{U}^* \in \mathcal{S}(X)$. The strongly continuous semigroup $\mathcal{U}^*$ is said to be the *dual* of $\mathcal{U}$.

Example 3.5. By the method of separation of variables, well-known in theory of partial differential equations, the solution of the one-dimensional heat equation

$$\frac{\partial}{\partial t} u_{t,\theta} = \frac{\partial^2}{\partial \theta^2} u_{t,\theta}, \ t > 0,$$

with the initial and boundary conditions

$$\begin{cases} u_{0,\theta} = f_\theta, \ 0 \le \theta \le 1, \ f \in W^{2,2}(0,1;\mathbb{R}), \\ u_{t,0} = u_{t,1} = 0, \ t \ge 0, \end{cases}$$

can be represented in the form

$$u_{t,\theta} = \sum_{n=1}^{\infty} 2e^{-n^2\pi^2 t} \sin(n\pi\theta) \int_0^1 f_\alpha \sin(n\pi\alpha)\, d\alpha, \ 0 \le \theta \le 1, \ t \ge 0.$$

Let $X = L_2(0,1;\mathbb{R})$ and consider the second order differential operator $d^2/d\theta^2$ with

$$D\left(d^2/d\theta^2\right) = \left\{h \in W^{2,2}(0,1;\mathbb{R}) : h_0 = h_1 = 0\right\}.$$

By Proposition 2.44(b), $d^2/d\theta^2 \in \tilde{\mathcal{L}}(X)$. Letting $x_t = [u_t]$ (see Remark 2.21) and defining $\mathcal{U} : [0,\infty) \to \mathcal{L}(X)$ by

$$[\mathcal{U}_t h]_\theta = \sum_{n=1}^{\infty} 2e^{-n^2\pi^2 t} \sin(n\pi\theta) \int_0^1 h_\alpha \sin(n\pi\alpha)\, d\alpha, \ 0 \le \theta \le 1, \ t \ge 0, \ h \in X,$$

the above mentioned problem and its solution can be written as

$$x_t' = \frac{d^2}{d\theta^2} x_t, \ t > 0, \ x_0 = f \in D\left(d^2/d\theta^2\right); \ x_t = \mathcal{U}_t f, \ t \ge 0.$$

One can show that $\mathcal{U}_t$, $t \ge 0$, is the strongly continuous semigroup generated by the second order differential operator $d^2/d\theta^2$. By Proposition 2.44(b), $\left(d^2/d\theta^2\right)^* = d^2/d\theta^2$. Hence, $\mathcal{U}_t^* = \mathcal{U}_t$, $t \ge 0$. One can verify this equality by direct computation of the dual semigroup $\mathcal{U}^*$.

Example 3.6. Consider the wave equation

$$\frac{\partial^2}{\partial t^2} u_{t,\theta} = \frac{\partial^2}{\partial \theta^2} u_{t,\theta}, \ t > 0,$$

with the initial and boundary conditions

$$\begin{cases} u_{0,\theta} = f_\theta, \ (\partial/\partial t) u_{t,\theta}\big|_{t=0} = g_\theta, \ 0 \le \theta \le 1, \ f, g \in W^{2,2}(0,1;\mathbb{R}), \\ u_{t,0} = u_{t,1} = 0, \ t \ge 0, \end{cases}$$

With f and g, we associate the respective sequences $\{\check{f}_n\}$ and $\{\check{g}_n\}$ of Fourier coefficients in the half-range Fourier sine expansions

$$f_\theta = \sum_{n=1}^{\infty} \check{f}_n \sqrt{2} \sin(n\pi\theta) \text{ and } g_\theta = \sum_{n=1}^{\infty} \check{g}_n \sqrt{2} \sin(n\pi\theta)$$

and assume that

$$\sum_{n=1}^{\infty} n^2 \check{f}_n^2 < \infty.$$

Let X be the Hilbert space of all functions

$$h = \begin{bmatrix} \xi \\ \eta \end{bmatrix} \in L_2(0,1;\mathbb{R}) \times L_2(0,1;\mathbb{R})$$

with

$$\sum_{n=1}^{\infty} n^2 \check{\xi}_n^2 < \infty,$$

endowed with the scalar product

$$\langle h, h' \rangle = \left\langle \begin{bmatrix} \xi \\ \eta \end{bmatrix}, \begin{bmatrix} \xi' \\ \eta' \end{bmatrix} \right\rangle = \sum_{n=1}^{\infty} \left(n^2 \pi^2 \check{\xi}_n \check{\xi}'_n + \check{\eta}_n \check{\eta}'_n \right),$$

where $\check{\xi}_n$, $\check{\xi}'_n$, $\check{\eta}_n$ and $\check{\eta}'_n$ are the respective Fourier coefficients of ξ, ξ', η and η'. For the operator

$$A = \begin{bmatrix} 0 & I \\ d^2/d\theta^2 & 0 \end{bmatrix},$$

where I is the identity operator on $W^{2,2}(0,1;\mathbb{R})$ and $d^2/d\theta^2$ is the second order differential operator with the domain $D\big(d^2/d\theta^2\big)$ as defined in Example 3.5, the above mentioned problem can be formulated in the abstract form

$$x'_t = Ax_t, \ t > 0,$$

where

$$[x_t]_\theta = \begin{bmatrix} u_{t,\theta} \\ (\partial/\partial t) u_{t,\theta} \end{bmatrix}, \ 0 \le \theta \le 1, \ t > 0; \ x_0 = \begin{bmatrix} f \\ g \end{bmatrix}.$$

It is known that the operator A, as defined above, belongs to $\tilde{\mathcal{L}}(X)$ and it generates the strongly continuous semigroup $\mathcal{U}$ (see Curtain and Zwart [41], p.149, and Zabczyk [95], p.180) as defined by

$$[\mathcal{U}_t h]_\theta = \sum_{n=1}^{\infty} \begin{bmatrix} \cos(n\pi t) & (n\pi)^{-1} \sin(n\pi t) \\ -n\pi \sin(n\pi t) & \cos(n\pi t) \end{bmatrix} \begin{bmatrix} \check{\xi}_n \\ \check{\eta}_n \end{bmatrix} \sqrt{2} \sin(n\pi\theta), \ 0 \le \theta \le 1, \ t \ge 0,$$

where ξ and η are components of $h \in X$, and $\check{\xi}_n$ and $\check{\eta}_n$ are Fourier coefficients of ξ and η, respectively. Moreover, $\|\mathcal{U}_t\| \le 1$ for all $t \ge 0$, i.e., $\mathcal{U}$ is a contraction semigroup, and the natural extension of $\mathcal{U}$ to $\mathbb{R}$ satisfies $\mathcal{U}_t^{-1} = \mathcal{U}_t^* = \mathcal{U}_{-t}$, i.e., $\mathcal{U}$ (as defined on $\mathbb{R}$) is a group.

Example 3.7. Let $\varepsilon > 0$. By the method of characteristics, well-known in theory of partial differential equations, the solution of the first order partial differential equation

$$\frac{\partial}{\partial t} u_{t,\theta} = \frac{\partial}{\partial \theta} u_{t,\theta}, \ t > 0,$$

with the initial and boundary conditions

$$\begin{cases} u_{0,\theta} = f_\theta, \ -\varepsilon \le \theta \le 0, \ f \in W^{1,2}(-\varepsilon, 0; X), \\ u_{t,0} = 0, \ t \ge 0, \end{cases}$$

can be represented in the form

$$u_{t,\theta} = \begin{Bmatrix} f_{\theta+t}, & \theta + t \le 0 \\ 0, & \theta + t > 0 \end{Bmatrix}, \ -\varepsilon \le \theta \le 0, \ t \ge 0.$$

Let $\tilde{X} = L_2(-\varepsilon, 0; X)$ and consider the first order differential operator $d/d\theta$ with

$$D(d/d\theta) = \{h \in W^{1,2}(-\varepsilon, 0; X) : h_0 = 0\}.$$

By Proposition 2.44(a), $d/d\theta \in \tilde{\mathcal{L}}(\tilde{X})$. Letting $x_t = [u_t]$ (see Remark 2.21) and

$$[\mathcal{T}_t^* h]_\theta = \begin{Bmatrix} h_{\theta+t}, & \theta + t \le 0 \\ 0, & \theta + t > 0 \end{Bmatrix}, \ -\varepsilon \le \theta \le 0, \ t \ge 0, \ h \in \tilde{X}, \tag{3.3}$$

the above mentioned problem and its solution can be written as

$$x_t' = \frac{d}{d\theta} x_t, \ t > 0, \ x_0 = f \in D(d/d\theta); \ x_t = \mathcal{T}_t^* f, \ t \ge 0.$$

One can show that $\mathcal{T}_t^*$, $t \ge 0$, defined by (3.3), is the strongly continuous semigroup generated by the first order differential operator $d/d\theta$. $\mathcal{T}^*$ is called a *semigroup of left translation*.

Example 3.8. Let $\varepsilon > 0$, let $\tilde{X} = L_2(-\varepsilon, 0; X)$ and let $d/d\theta \in \tilde{\mathcal{L}}(\tilde{X})$ be the differential operator from Example 3.7. By Proposition 2.44(a), $(d/d\theta)^* = -d/d\theta$ with

$$D(-d/d\theta) = \{h \in W^{1,2}(-\varepsilon, 0; X) : h_{-\varepsilon} = 0\}.$$

Computing the dual of $\mathcal{T}^*$ (see (3.3)), one can easily verify that $-d/d\theta$ generates the strongly continuous semigroup $\mathcal{T}$ defined by

$$[\mathcal{T}_t h]_\theta = \begin{Bmatrix} h_{\theta-t}, & \theta - t \ge -\varepsilon \\ 0, & \theta - t < -\varepsilon \end{Bmatrix}, \ -\varepsilon \le \theta \le 0, \ t \ge 0, \ h \in \tilde{X}, \tag{3.4}$$

which is called a *semigroup of right translation*.

Example 3.9. A significant class of strongly continuous semigroups concerns linear differential delay equations in the form

$$\frac{d}{dt}u_t = Au_t + Nu_{t-\varepsilon} + \int_{-\varepsilon}^{0} M_\theta u_{t+\theta}\, d\theta,\ t > 0,$$

with the initial condition

$$u_\theta = f_\theta,\ -\varepsilon \le \theta \le 0,\ f \in W^{1,2}(-\varepsilon, 0; X),$$

where $\varepsilon > 0$, $N \in \mathcal{L}(X)$, $M \in B_2(-\varepsilon, 0; \mathcal{L}(X))$ and $A \in \tilde{\mathcal{L}}(X)$ generates the strongly continuous semigroup $\mathcal{U} \in \mathcal{S}(X)$. We do not consider these semigroups here in detail since they could be described through bounded and unbounded perturbations from Sections 3.2 and 3.4.

3.1.3 Mild Evolution Operators

An evolution operator is a generalization of the concept of semigroup of bounded linear operators to the two-parameter case. There are several kinds of evolution operators. The reader can find a complete discussion of evolution operators in Curtain and Pritchard [39]. We will consider mild evolution operators.

Definition 3.10. Let $T > 0$. A function $\mathcal{U} : \Delta_T \to \mathcal{L}(X)$ is called a *mild evolution operator* if

(a) $\mathcal{U}_{t,t} = I$, $0 \le t \le T$;

(b) $\mathcal{U}_{t,s} = \mathcal{U}_{t,r}\mathcal{U}_{r,s}$, $0 \le s \le r \le t \le T$ (semigroup property);

(c) $[\mathcal{U}_t] : [0, t] \to \mathcal{L}(X)$ and $[\mathcal{U}_s] : [s, T] \to \mathcal{L}(X)$ are weakly continuous for all $0 < t \le T$ and for all $0 \le s < T$;

(d) $\sup_{\Delta_T} \|\mathcal{U}_{t,s}\| < \infty$.

The class of all mild evolution operators from Δ_T to $\mathcal{L}(X)$ will be denoted by $\mathcal{E}(\Delta_T, \mathcal{L}(X))$.

Each strongly continuous semigroup $\mathcal{U}$ is a mild evolution operator when it is defined in the two-parameter form $\mathcal{U}_{t,s} = \mathcal{U}_{t-s}$, $0 \le s \le t \le T$. Hence, if the one-parameter $(\mathcal{U}_t)$ and two-parameter $(\mathcal{U}_{t,s} = \mathcal{U}_{t-s})$ forms of semigroups are identified, then $\mathcal{S}(X) \subset \mathcal{E}(\Delta_T, \mathcal{L}(X))$ for all $T > 0$.

Proposition 3.11. *Let $T > 0$. The following statements hold.*

(a) $\mathcal{E}(\Delta_T, \mathcal{L}(X)) \subset B(\Delta_T, \mathcal{L}(X))$.

(b) *If $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$ and $\varphi \in L_1(0, T; X)$, then*

$$\psi_t = \int_0^t \mathcal{U}_{t,s}\varphi_s\, ds,\ 0 \le t \le T,$$

is weakly continuous.

Proof. Part (b) can be proved by direct verification. For part (a), let

$$\tilde{\mathcal{U}}_{t,s} = \begin{cases} \mathcal{U}_{t,s}, & 0 \le s \le t \le T, \\ I, & 0 \le t < s \le T, \end{cases}$$

and let $u_{t,s} = \langle \tilde{\mathcal{U}}_{t,s} x, y \rangle$ for any $x, y \in X$. The real-valued function u is defined on $[0,T] \times [0,T]$ and is continuous in each of the variables. Therefore, by Proposition 2.25, it can be considered as a $\mathcal{B}_{[0,T]}$-measurable function from $[0,T]$ to $C(0,T;\mathbb{R})$. We conclude that there is a sequence $\{u^n\}$ of $\mathcal{B}_{[0,T]}$-simple functions from $[0,T]$ to $C(0,T;\mathbb{R})$ such that

$$\max_{s \in [0,T]} |u_{t,s} - u^n_{t,s}| \to 0 \text{ as } n \to \infty \text{ for all } 0 \le t \le T.$$

It is easy to verify that each u^n is a $\mathcal{B}_{[0,T] \times [0,T]}$-measurable function. Hence, the limit function u is also $\mathcal{B}_{[0,T] \times [0,T]}$-measurable. This implies the strong $\mathcal{B}_{\Delta_T}$-measurability of $\mathcal{U}$ since X is separable. Thus, $\mathcal{U} \in B(\Delta_T, \mathcal{L}(X))$ since $\mathcal{U}$ is bounded. □

Example 3.12. Let $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, let $\tilde{X} = L_2(-\varepsilon, 0; X)$ and let $\varepsilon > 0$. Then $\tilde{\mathcal{U}} \in \mathcal{E}(\Delta_T, \mathcal{L}(X \times \tilde{X}))$ where

$$\tilde{\mathcal{U}}_{t,s} = \begin{bmatrix} \mathcal{U}_{t,s} & \mathcal{E}_{t,s} \\ 0 & \mathcal{T}_{t-s} \end{bmatrix}, \ 0 \le s \le t \le T,$$

$\mathcal{T} \in \mathcal{S}(\tilde{X})$ is the semigroup of right translation as defined by (3.4) and

$$\mathcal{E}_{t,s} g = \int_{\max(-\varepsilon, s-t)}^{0} \mathcal{U}_{t,s-r} g_r \, dr, \ 0 \le s \le t \le T, \ g \in \tilde{X}.$$

Indeed, the conditions (a), (c) and (d) of Definition 3.10 can easily be verified. For the condition (b), it is sufficient to show that

$$\mathcal{E}_{t,s} = \mathcal{U}_{t,s}\mathcal{E}_{s,r} + \mathcal{E}_{t,s}\mathcal{T}_{s-r}, \ 0 \le r \le s \le t \le T.$$

This follows from

$$\begin{aligned}
(\mathcal{U}_{t,s}\mathcal{E}_{s,r} + \mathcal{E}_{t,s}\mathcal{T}_{s-r})g &= \mathcal{U}_{t,s} \int_{\max(-\varepsilon, r-s)}^{0} \mathcal{U}_{s,r-\sigma} g_\sigma \, d\sigma \\
&\quad + \int_{\max(-\varepsilon, s-t)}^{0} \mathcal{U}_{t,s-\sigma} [\mathcal{T}_{s-r} g]_\sigma \, d\sigma \\
&= \int_{\max(-\varepsilon, r-s)}^{0} \mathcal{U}_{t,r-\sigma} g_\sigma \, d\sigma \\
&\quad + \int_{\max(-\varepsilon, r-t)}^{\max(-\varepsilon, r-s)} \mathcal{U}_{t,r-\sigma} g_\sigma \, d\sigma \\
&= \int_{\max(-\varepsilon, r-t)}^{0} \mathcal{U}_{t,r-\sigma} g_\sigma \, d\sigma = \mathcal{E}_{t,s} g,
\end{aligned}$$

where $g \in \tilde{X}$. In fact, $\tilde{\mathcal{U}}$ is the perturbation of $\mathcal{U} \odot \mathcal{T}$ by the bounded operator

$$\begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}).$$

This will be studied in the next section.

Actually, in applied problems we meet mild evolution operators which have some differentiability property. Such an evolution operator is called a *strong evolution operator*. The class of strong evolution operators is not closed even under simple transformations and that makes this class inconvenient in studying control and estimation problems. But, the class of mild evolution operators is closed with respect to the basic transformations which are considered in the next section.

3.2 Transformations of Evolution Operators

3.2.1 Bounded Perturbations

An important transformation on evolution operators is their bounded perturbation.

Theorem 3.13. *The equations*

$$\mathcal{Y}_{t,s} = \mathcal{U}_{t,s} + \int_s^t \mathcal{Y}_{t,r} N_r \mathcal{U}_{r,s}\, dr, \;\; 0 \le s \le t \le T, \tag{3.5}$$

$$\mathcal{Y}_{t,s} = \mathcal{U}_{t,s} + \int_s^t \mathcal{U}_{t,r} N_r \mathcal{Y}_{r,s}\, dr, \;\; 0 \le s \le t \le T, \tag{3.6}$$

where $T > 0$, $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $N \in B_\infty(0,T;\mathcal{L}(X))$ and the integrals are in the strong Bochner sense, are equivalent and have a unique solution in $\mathcal{E}(\Delta_T, \mathcal{L}(X))$.

Proof. Denote $c_1 = \operatorname{ess\,sup}_{[0,T]} \|N_s\|$, $c_2 = \sup_{\Delta_T} \|\mathcal{U}_{t,s}\|$, and consider the sequence of operator-valued functions

$$\mathcal{Y}^n_{t,s} = \int_s^t \mathcal{U}_{t,r} N_r \mathcal{Y}^{n-1}_{r,s}\, dr, \;\; n = 1, 2, \ldots, \;\; \mathcal{Y}^0_{t,s} = \mathcal{U}_{t,s}, \;\; 0 \le s \le t \le T.$$

By induction, it can be proved that

$$\|\mathcal{Y}^n_{t,s}\| \le c_2 (c_1 c_2)^n \frac{(t-s)^n}{n!}, \;\; n = 1, 2, \ldots.$$

Therefore, the series $\sum_{n=0}^\infty \mathcal{Y}^n_{t,s}$ is majorized by a convergent series of real numbers and, consequently, it converges uniformly in $(t,s) \in \Delta_T$ with

$$\sum_{n=0}^\infty \|\mathcal{Y}^n_{t,s}\| \le c_2 e^{c_1 c_2 T}.$$

By Theorem 2.19, we have

$$\sum_{n=0}^{\infty} \mathcal{Y}_{t,s}^{n} = \mathcal{U}_{t,s} + \sum_{n=1}^{\infty} \mathcal{Y}_{t,s}^{n} = \mathcal{U}_{t,s} + \int_s^t \mathcal{U}_{t,r} N_r \sum_{n=0}^{\infty} \mathcal{Y}_{r,s}^{n}\, dr.$$

Thus, $\mathcal{Y}_{t,s} = \sum_{n=0}^{\infty} \mathcal{Y}_{t,s}^{n}$, $0 \le s \le t \le T$, is a solution of the equation (3.6). One can observe that $\mathcal{Y}^n$ is weakly continuous in each of the variables for all $n = 0, 1, \ldots$. So, by uniform convergence, we obtain the weak continuity of $\mathcal{Y}$ in each of the variables. Let us prove that this solution is unique. Let $\tilde{\mathcal{Y}}$ be another solution and $\mathcal{R} = \mathcal{Y} - \tilde{\mathcal{Y}}$. Then

$$\mathcal{R}_{t,s} = \int_s^t \mathcal{U}_{t,r} N_r \mathcal{R}_{r,s}\, dr.$$

Hence,

$$\|\mathcal{R}_{t,s}\| \le c_1 c_2 \int_s^t \|\mathcal{R}_{r,s}\|\, dr.$$

So, by Theorem 2.45, $\mathcal{R} = \mathcal{Y} - \tilde{\mathcal{Y}} = 0$, i.e., the equation (3.6) has a unique solution. Now consider the equation (3.5). Let

$$\tilde{\mathcal{Y}}_{t,s}^{n} = \int_s^t \tilde{\mathcal{Y}}_{t,r}^{n-1} N_r \mathcal{U}_{r,s}\, dr,\ n = 1, 2, \ldots,\ \tilde{\mathcal{Y}}_{t,s}^{0} = \mathcal{U}_{t,s},\ 0 \le s \le t \le T.$$

Similarly, one can show that the function $\tilde{\mathcal{Y}}_{t,s} = \sum_{n=0}^{\infty} \tilde{\mathcal{Y}}_{t,s}^{n}$, $0 \le s \le t \le T$, is a unique solution of the equation (3.5). For $\mathcal{Y} = \tilde{\mathcal{Y}}$, it is sufficient to show that $\mathcal{Y}^n = \tilde{\mathcal{Y}}^n$ for all $n = 0, 1, \ldots$. When $n = 0$ or $n = 1$, this is obvious. Suppose that $\mathcal{Y}^n = \tilde{\mathcal{Y}}^n$ for $n = k - 1$ and $n = k - 2$. Then we have

$$\begin{aligned}
\tilde{\mathcal{Y}}_{t,s}^{k} &= \int_s^t \tilde{\mathcal{Y}}_{t,r}^{k-1} N_r \mathcal{U}_{r,s}\, dr \\
&= \int_s^t \mathcal{Y}_{t,r}^{k-1} N_r \mathcal{U}_{r,s}\, dr \\
&= \int_s^t \int_r^t \mathcal{U}_{t,\sigma} N_\sigma \mathcal{Y}_{\sigma,r}^{k-2} N_r \mathcal{U}_{r,s}\, d\sigma\, dr \\
&= \int_s^t \int_s^\sigma \mathcal{U}_{t,\sigma} N_\sigma \tilde{\mathcal{Y}}_{\sigma,r}^{k-2} N_r \mathcal{U}_{r,s}\, dr\, d\sigma \\
&= \int_s^t \mathcal{U}_{t,\sigma} N_\sigma \tilde{\mathcal{Y}}_{\sigma,s}^{k-1}\, d\sigma \\
&= \int_s^t \mathcal{U}_{t,\sigma} N_\sigma \mathcal{Y}_{\sigma,s}^{k-1}\, d\sigma = \mathcal{Y}_{t,s}^{k}.
\end{aligned}$$

By induction, $\mathcal{Y}^n = \tilde{\mathcal{Y}}^n$ for all $n = 0, 1, \ldots$ and, hence, $\mathcal{Y} = \tilde{\mathcal{Y}}$. Thus, the equations (3.5) and (3.6) are equivalent. Finally, let us show that $\mathcal{Y} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$.

Obviously, $\mathcal{Y}_{t,t} = \mathcal{U}_{t,t} = I$, $0 \le t \le T$. To show the semigroup property, let $0 \le s \le r \le t \le T$. Then from (3.6) and from

$$\mathcal{Y}_{t,r}\mathcal{Y}_{r,s} = \mathcal{U}_{t,r}\mathcal{U}_{r,s} + \int_s^r \mathcal{U}_{t,r}\mathcal{U}_{r,\sigma}N_\sigma\mathcal{Y}_{\sigma,s}\,d\sigma + \int_r^t \mathcal{U}_{t,\sigma}N_\sigma\mathcal{Y}_{\sigma,r}\mathcal{Y}_{r,s}\,d\sigma,$$

we obtain

$$\mathcal{Y}_{t,r}\mathcal{Y}_{r,s} - \mathcal{Y}_{t,s} = \int_r^t \mathcal{U}_{t,\sigma}N_\sigma(\mathcal{Y}_{\sigma,r}\mathcal{Y}_{r,s} - \mathcal{Y}_{\sigma,s})\,d\sigma.$$

Hence,

$$\|\mathcal{Y}_{t,r}\mathcal{Y}_{r,s} - \mathcal{Y}_{t,s}\| \le c_1c_2\int_r^t \|\mathcal{Y}_{\sigma,r}\mathcal{Y}_{r,s} - \mathcal{Y}_{\sigma,s}\|\,d\sigma.$$

By Theorem 2.45, we conclude $\mathcal{Y}_{t,r}\mathcal{Y}_{r,s} = \mathcal{Y}_{t,s}$, $0 \le s \le r \le t \le T$. Thus, we have $\mathcal{Y} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$. □

Definition 3.14. The solution $\mathcal{Y}$ of the equivalent equations (3.5) and (3.6) is called the *bounded perturbation* of $\mathcal{U}$ by N. The notation $\mathcal{P}_N(\mathcal{U})$ will be used to denote this bounded perturbation.

Theorem 3.15. *Suppose that $\mathcal{U} \in \mathcal{S}(X)$ and $N_t \equiv N \in \mathcal{L}(X)$. Then $\mathcal{Y} = \mathcal{P}_N(\mathcal{U}) \in \mathcal{S}(X)$ and $\mathcal{Y}$ in the one-parameter form is a unique solution of the equivalent equations*

$$\mathcal{Y}_t = \mathcal{U}_t + \int_0^t \mathcal{Y}_r N\mathcal{U}_{t-r}\,dr, \ t \ge 0, \tag{3.7}$$

$$\mathcal{Y}_t = \mathcal{U}_t + \int_0^t \mathcal{U}_r N\mathcal{Y}_{t-r}\,dr, \ t \ge 0. \tag{3.8}$$

Furthermore, if A is the infinitesimal generator of $\mathcal{U}$, then $A+N$ is the infinitesimal generator of $\mathcal{Y}$.

Proof. Let $T > 0$ and consider the solution $\mathcal{Y}$ of the equivalent equations (3.5) and (3.6) when $\mathcal{U}_{t,s} = \mathcal{U}_{t-s}$, $0 \le s \le t \le T$, and $N_t \equiv N$, $0 \le t \le T$. By Theorem 3.13, $\mathcal{Y} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$. One can verify that $\mathcal{Y}_{t,s}$, $0 \le s \le t \le T$, is a function of the difference $t - s$. Denote $\mathcal{Y}_{t-s} = \mathcal{Y}_{t,s}$, $0 \le s \le t \le T$. On the interval $[0, T]$, the equations (3.5) and (3.6) can be easily transformed to the equations (3.7) and (3.8), respectively, and vice versa. The semigroup property of $\mathcal{Y}$ in the one-parameter form can be obtained from the same property in the two-parameter form. The strong continuity of $\mathcal{Y}$ follows from the strong continuity of $\mathcal{U}$ and from (3.7) or (3.8). Furthermore, letting T tend to ∞, we conclude that $\mathcal{Y}_t$, $t \ge 0$, is a unique solution of the equivalent equations (3.7) and (3.8) and $\mathcal{Y} \in \mathcal{S}(X)$. Also, using Proposition 2.41, from (3.7), for $x \in D(A)$, one can compute

$$\begin{aligned}\frac{d}{dt}\mathcal{Y}_t x &= \frac{d}{dt}\mathcal{U}_t x + \int_0^t \mathcal{Y}_r N\frac{d}{dt}\mathcal{U}_{t-r}x\,dr + \mathcal{Y}_t Nx \\ &= \mathcal{U}_t Ax + \int_0^t \mathcal{Y}_r N\mathcal{U}_{t-r}Ax\,dr + \mathcal{Y}_t Nx = \mathcal{Y}_t(A+N)x.\end{aligned}$$

Hence, $A+N$ is the infinitesimal generator of $\mathcal{Y}$. □

The next result explains the meaning of bounded perturbations.

Proposition 3.16. *Let $T>0$, let $x_0\in X$ and let $\mathcal{Y}=\mathcal{P}_N(\mathcal{U})$ where $\mathcal{U}\in\mathcal{E}(\Delta_T,\mathcal{L}(X))$ and $N\in B_\infty(0,T;\mathcal{L}(X))$. Then the function $x_t=\mathcal{Y}_{t,0}x_0$, $0\le t\le T$, is a unique weakly continuous solution of the equation*

$$x_t=\mathcal{U}_{t,0}x_0+\int_0^t \mathcal{U}_{t,s}N_s x_s\,ds,\ \ 0\le t\le T.$$

Proof. This is a direct consequence of Theorem 3.13. □

We also present the following properties of bounded perturbations.

Proposition 3.17. *Let $T>0$, let $\mathcal{U}\in\mathcal{E}(\Delta_T,\mathcal{L}(X))$ and let $N,M\in B_\infty(0,T;\mathcal{L}(X))$. Then the following equalities hold:*

(a) $\mathcal{P}_{N+M}(\mathcal{U})=\mathcal{P}_N(\mathcal{P}_M(\mathcal{U}))$;

(b) $\mathcal{Y}=\mathcal{P}_N(\mathcal{U})\ \Rightarrow\ \mathcal{U}=\mathcal{P}_{-N}(\mathcal{Y})$.

Proof. Part (b) is obvious. To prove part (a) let $\mathcal{R}=\mathcal{P}_M(\mathcal{U})$ and $\mathcal{K}=\mathcal{P}_N(\mathcal{R})$. Then

$$\mathcal{R}_{t,s}=\mathcal{U}_{t,s}+\int_s^t \mathcal{R}_{t,r}M_r\mathcal{U}_{r,s}\,dr,\ \ 0\le s\le t\le T,$$
$$\mathcal{K}_{t,s}=\mathcal{R}_{t,s}+\int_s^t \mathcal{K}_{t,r}N_r\mathcal{R}_{r,s}\,dr,\ \ 0\le s\le t\le T.$$

Using these equalities, we have

$$\begin{aligned}
\mathcal{K}_{t,s}&=\mathcal{U}_{t,s}+\int_s^t \mathcal{R}_{t,r}M_r\mathcal{U}_{r,s}\,dr\\
&\quad+\int_s^t \mathcal{K}_{t,r}N_r\Big(\mathcal{U}_{r,s}+\int_s^r \mathcal{R}_{r,\sigma}M_\sigma\mathcal{U}_{\sigma,s}\,d\sigma\Big)dr\\
&=\mathcal{U}_{t,s}+\int_s^t \mathcal{K}_{t,r}N_r\mathcal{U}_{r,s}\,dr\\
&\quad+\int_s^t\Big(\mathcal{R}_{t,r}+\int_r^t \mathcal{K}_{t,\sigma}N_\sigma\mathcal{R}_{\sigma,r}\,d\sigma\Big)M_r\mathcal{U}_{r,s}\,dr\\
&=\mathcal{U}_{t,s}+\int_s^t \mathcal{K}_{t,r}(N_r+M_r)\mathcal{U}_{r,s}\,dr.
\end{aligned}$$

Hence, $\mathcal{K}=\mathcal{P}_{N+M}(\mathcal{U})$. □

3.2.2 Some Other Transformations

Suppose $0 < t \le T$ and $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$. Then $\mathcal{U}|_{\Delta_t} \in \mathcal{E}(\Delta_t, \mathcal{L}(X))$. The restriction of $\mathcal{U}$ to Δ_t will be considered as a transformation of $\mathcal{U}$ and it will be denoted by $\mathcal{U}|_{\Delta_t}$. The restriction of $\mathcal{U} \in \mathcal{S}(X)$ to Δ_t in the one-parameter form is $\mathcal{U}|_{[0,t]}$. Since $\mathcal{U} \in \mathcal{S}(X)$ is uniquely defined by its values on $[0,t]$ for $t > 0$, it is convenient to identify $\mathcal{U} \in \mathcal{S}(X)$ and its restrictions.

Suppose $T > 0$, $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$ and $\mathcal{R} \in \mathcal{E}(\Delta_T, \mathcal{L}(Y))$. Then the function $\mathcal{Y}$, defined by

$$\mathcal{Y}_{t,s} = \begin{bmatrix} \mathcal{U}_{t,s} & 0 \\ 0 & \mathcal{R}_{t,s} \end{bmatrix},$$

belongs to $\mathcal{E}(\Delta_T, \mathcal{L}(X \times Y))$. The mild evolution operator $\mathcal{Y}$ will be denoted by $\mathcal{U} \odot \mathcal{R}$. For $\mathcal{U} \in \mathcal{S}(X)$ and $\mathcal{R} \in \mathcal{S}(Y)$, we readily obtain that $\mathcal{U} \odot \mathcal{R} \in \mathcal{S}(X \times Y)$. If A_1 and A_2 are the respective infinitesimal generators of $\mathcal{U}$ and $\mathcal{R}$, then

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

is the infinitesimal generator of $\mathcal{U} \odot \mathcal{R}$.

Suppose $0 < t \le T$ and $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$. Then one can verify that the function $\mathcal{Y}$, defined by

$$\mathcal{Y}_{s,r} = \mathcal{U}^*_{t-r,t-s}, \quad 0 \le r \le s \le t,$$

belongs to $\mathcal{E}(\Delta_t, \mathcal{L}(X))$. The mild evolution operator $\mathcal{Y}$ is called the *dual* of $\mathcal{U}$ and it is denoted by $\mathcal{D}_t(\mathcal{U})$. Obviously, if $\mathcal{U} \in \mathcal{S}(X)$, then $\mathcal{D}_t(\mathcal{U}) = \mathcal{U}^*$ (see Example 3.4).

If $0 < t \le T$ and if $N \in B_\infty(0,T; \mathcal{L}(X,Y))$, then by $D_t(N)$ we denote the function in $B_\infty(0,t; \mathcal{L}(Y,X))$ that is defined by

$$[D_t(N)]_s = N^*_{t-s}, \quad 0 \le s \le t.$$

Also, for a function $\mu : [0,T] \to \mathbb{R}$, we will denote

$$[D_t(\mu)]_s = t - \mu_{t-s}, \quad 0 \le s \le t.$$

In the sequel, the difference between these two meanings of the transformation D_t will follow from the context.

Proposition 3.18. *Let* $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X,Y))$, *let* $\mathcal{Y} \in \mathcal{E}(\Delta_T, \mathcal{L}(X,Y))$, *let* $N \in B_\infty(0,T; \mathcal{L}(X,Y))$, *let* $\mu : [0,T] \to \mathbb{R}$ *and let* $0 < t \le T$. *Then*

(a) $\mathcal{P}_N(\mathcal{U})|_{\Delta_t} = \mathcal{D}_t(\mathcal{P}_{D_t(N)}(\mathcal{D}_t(\mathcal{U})))$;

(b) $\mathcal{D}_t(\mathcal{U} \odot \mathcal{Y}) = \mathcal{D}_t(\mathcal{U}) \odot \mathcal{D}_t(\mathcal{Y})$;

(c) $\mathcal{D}_t(\mathcal{D}_t(\mathcal{U})) = \mathcal{U}|_{\Delta_t}$;

(d) $D_t(D_t(N)) = N|_{[0,t]}$;

(e) $D_t(D_t(\mu)) = \mu|_{[0,t]}$;

(f) $\nu = D_t(\mu)$ *is continuous and increasing with*

$$\nu_s^{-1} = t - \mu_{t-s}^{-1},\ t - \mu_t \le s \le t - \mu_0,$$

if μ *is continuous and increasing.*

Proof. Part (a) is a consequence of the equivalence of the equations (3.5) and (3.6). Parts (b)–(e) can be proved by direct verification. Also,

$$\nu_s = t - \mu_{t-s} \Rightarrow \nu_{t-\mu_{t-s}^{-1}} = t - \mu_{t-t+\mu_{t-s}^{-1}} = s \Rightarrow \nu_s^{-1} = t - \mu_{t-s}^{-1}$$

proves part (f). □

3.3 Operator Riccati Equations

The equation

$$Q_t = \mathcal{U}_{T,t}^* Q_T \mathcal{U}_{T,t} + \int_t^T \mathcal{U}_{s,t}^*(F_s - (Q_s B_s + L_s^*) G_s^{-1} (B_s^* Q_s + L_s)) \mathcal{U}_{s,t}\, ds,\ 0 \le t \le T, \tag{3.9}$$

where Q is an unknown function, is called an *operator Riccati equation* and it often arises in applied problems. In this section, bounded perturbations of mild evolution operators are used to study this equation.

3.3.1 Existence and Uniqueness of Solution

The equation (3.9) will be studied under the conditions

$$\begin{cases} T > 0,\ \mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X)),\ B \in B_\infty(0,T;\mathcal{L}(U,X)), \\ F \in B_\infty(0,T;\mathcal{L}(X)),\ L \in B_\infty(0,T;\mathcal{L}(X,U)), \\ G, G^{-1} \in B_\infty(0,T;\mathcal{L}(U)),\ Q_T \in \mathcal{L}(X),\ Q_T \ge 0, \\ G_t > 0 \text{ and } F_t - L_t^* G_t^{-1} L_t \ge 0 \text{ for a.e. } t \in [0,T]. \end{cases} \tag{3.10}$$

We will also use the notation $M = F - L^* G^{-1} L$, $K = BG^{-1}B^*$ and $\mathcal{R} = \mathcal{P}_{-BG^{-1}L}(\mathcal{U})$.

Consider the sequence of operator-valued functions defined by

$$Q_t^n = \mathcal{R}_{T,t}^{n*} Q_T \mathcal{R}_{T,t}^n + \int_t^T \mathcal{R}_{s,t}^{n*}(M_s + Q_s^{n-1} K_s Q_s^{n-1}) \mathcal{R}_{s,t}^n\, ds,\ Q_t^0 = 0,\ 0 \le t \le T,\ n = 1, 2, \dots, \tag{3.11}$$

where $\mathcal{R}^n = \mathcal{P}_{-KQ^{n-1}}(\mathcal{R})$. By Proposition 3.11(a), we have $Q^n \in B(0,T;\mathcal{L}(X))$ for all $n = 1,2,\ldots$. Also, $Q_t^n \geq 0$ for all $0 \leq t \leq T$ and for all $n = 1,2,\ldots$. To find an equation for the difference $\Delta Q^n = Q^{n+1} - Q^n$, we need a representation of Q^n in terms of $\mathcal{R}^{n+1}$.

Lemma 3.19. *Under the above conditions and notation,*

$$\begin{aligned} Q_t^n &= \mathcal{R}_{T,t}^{n+1*} Q_T \mathcal{R}_{T,t}^n + \int_t^T \mathcal{R}_{s,t}^{n+1*} \big(M_s \\ &\quad + Q_s^{n-1} K_s Q_s^{n-1} + \Delta Q_s^{n-1} K_s Q_s^n\big) \mathcal{R}_{s,t}^n \, ds, \ 0 \leq t \leq T, \ n = 1,2,\ldots. \end{aligned}$$

Proof. Since

$$\mathcal{R}^n = \mathcal{P}_{-KQ^{n-1}}(\mathcal{R}),$$

by Propositions 3.17(b) and 3.17(c), we have

$$\mathcal{R}^n = \mathcal{P}_{K\Delta Q^{n-1}}\big(\mathcal{R}^{n+1}\big).$$

Hence, the following equality holds:

$$\mathcal{R}_{t,s}^{n*} = \mathcal{R}_{t,s}^{n+1*} + \int_s^t \mathcal{R}_{r,s}^{n+1*} \Delta Q_r^{n-1} K_r \mathcal{R}_{t,r}^{n*} \, dr, \ 0 \leq s \leq t \leq T.$$

Substituting $\mathcal{R}^{n*}$ from this equality in (3.11), we have

$$\begin{aligned} Q_t^n &= \left(\mathcal{R}_{T,t}^{n+1*} + \int_t^T \mathcal{R}_{r,t}^{n+1*} \Delta Q_r^{n-1} K_r \mathcal{R}_{T,r}^{n*} \, dr \right) Q_T \mathcal{R}_{T,t}^n \\ &\quad + \int_t^T \mathcal{R}_{s,t}^{n+1*} \big(M_s + Q_s^{n-1} K_s Q_s^{n-1}\big) \mathcal{R}_{s,t}^n \, ds \\ &\quad + \int_t^T \int_t^s \mathcal{R}_{r,t}^{n+1*} \Delta Q_r^{n-1} K_r \mathcal{R}_{s,r}^{n*} \big(M_s + Q_s^{n-1} K_s Q_s^{n-1}\big) \mathcal{R}_{s,t}^n \, dr \, ds \\ &= \mathcal{R}_{T,t}^{n+1*} Q_T \mathcal{R}_{T,t}^n + \int_t^T \mathcal{R}_{s,t}^{n+1*} \big(M_s + Q_s^{n-1} K_s Q_s^{n-1}\big) \mathcal{R}_{s,t}^n \, ds \\ &\quad + \int_t^T \mathcal{R}_{r,t}^{n+1*} \Delta Q_r^{n-1} K_r \\ &\qquad \times \left(\mathcal{R}_{T,r}^{n*} Q_T \mathcal{R}_{T,r}^n + \int_r^T \mathcal{R}_{s,r}^{n*} \left(M_s + Q_s^{n-1} K_s Q_s^{n-1}\right) \mathcal{R}_{s,r}^n \, ds \right) \mathcal{R}_{r,t}^n \, dr. \end{aligned}$$

Thus by (3.11), we obtain the statement. □

Lemma 3.20. *Under the above conditions and notation,*

$$\begin{aligned} Q_t^n &= \mathcal{R}_{T,t}^{n+1*} Q_T \mathcal{R}_{T,t}^{n+1} + \int_t^T \mathcal{R}_{s,t}^{n+1*} \big(M_s \\ &\quad + Q_s^n K_s Q_s^n + \Delta Q_s^{n-1} K_s \Delta Q_s^{n-1}\big) \mathcal{R}_{s,t}^{n+1} \, ds, \ 0 \leq t \leq T, \ n = 1,2,\ldots. \end{aligned}$$

Proof. We start from the expression for Q_t^n in Lemma 3.19 and substitute

$$\mathcal{R}_{t,s}^{n} = \mathcal{R}_{t,s}^{n+1} + \int_s^t \mathcal{R}_{t,r}^{n} K_r \Delta Q_r^{n-1} \mathcal{R}_{r,s}^{n+1}\, dr,\ 0 \le s \le t \le T.$$

Then, we have

$$\begin{aligned} Q_t^n &= \mathcal{R}_{T,t}^{n+1*} Q_T \left(\mathcal{R}_{T,t}^{n+1} + \int_t^T \mathcal{R}_{T,r}^{n} K_r \Delta Q_r^{n-1} \mathcal{R}_{r,t}^{n+1}\, dr \right) \\ &\quad + \int_t^T \mathcal{R}_{s,t}^{n+1*} (M_s + Q_s^{n-1} K_s Q_s^{n-1} + \Delta Q_s^{n-1} K_s Q_s^n) \\ &\qquad \times \left(\mathcal{R}_{s,t}^{n+1} + \int_t^s \mathcal{R}_{s,r}^{n} K_r \Delta Q_r^{n-1} \mathcal{R}_{r,t}^{n+1}\, dr \right) ds \\ &= \mathcal{R}_{T,t}^{n+1*} Q_T \mathcal{R}_{T,t}^{n+1} \\ &\quad + \int_t^T \mathcal{R}_{s,t}^{n+1*} (M_s + Q_s^{n-1} K_s Q_s^{n-1} + \Delta Q_s^{n-1} K_s Q_s^n) \mathcal{R}_{s,t}^{n+1}\, ds \\ &\quad + \int_t^T \mathcal{R}_{r,t}^{n+1*} \left(\mathcal{R}_{T,r}^{n+1*} Q_T \mathcal{R}_{T,r}^{n} + \int_r^T \mathcal{R}_{s,r}^{n+1*} (M_s \right. \\ &\qquad \left. + Q_s^{n-1} K_s Q_s^{n-1} + \Delta Q_s^{n-1} K_s Q_s^n) \mathcal{R}_{s,r}^{n}\, ds \right) K_r \Delta Q_r^{n-1} \mathcal{R}_{r,t}^{n+1}\, dr. \end{aligned}$$

Thus by Lemma 3.19, we obtain the statement. □

Lemma 3.21. *Under the above conditions and notation, the sequence of operators Q_t^n converges strongly to some $Q_t \ge 0$ for all $0 \le t \le T$. Furthermore, the operator-valued function Q belongs to $B(0,T;\mathcal{L}(X))$ and it is a unique solution of the equation*

$$Q_t = \mathcal{Y}_{T,t}^{*} Q_T \mathcal{Y}_{T,t} + \int_t^T \mathcal{Y}_{s,t}^{*} (M_s + Q_s K_s Q_s) \mathcal{Y}_{s,t}\, ds,\ 0 \le t \le T, \tag{3.12}$$

where $\mathcal{Y} = \mathcal{P}_{-KQ}(\mathcal{R})$.

Proof. From (3.11) and Lemma 3.20, we have

$$Q_t^n - Q_t^{n+1} = \int_t^T \mathcal{R}_{s,t}^{n+1*} \Delta Q_s^{n-1} K_s \Delta Q_s^{n-1} \mathcal{R}_{s,t}^{n+1}\, ds \ge 0.$$

Therefore, $\{Q_t^n\}$ is a nonincreasing sequence of nonnegative operators for all $0 \le t \le T$. Hence, by Theorem 1.31, there exists $Q_t \ge 0$ such that Q_t^n converges strongly to Q_t and $Q_t^n \ge Q_t$. Since $Q^n \in B(0,T;\mathcal{L}(X))$ and

$$\|Q_t\| \le \|Q_t^n\| \le \left\|Q_t^1\right\| \le c_0 = \sup_{[0,T]} \left\|Q_t^1\right\|, \tag{3.13}$$

we conclude that $Q \in B(0,T;\mathcal{L}(X))$. Hence, $\mathcal{Y} = \mathcal{P}_{-KQ}(\mathcal{R})$ is well-defined and

$$\mathcal{R}^n_{t,s} - \mathcal{Y}_{t,s} = \int_s^t \mathcal{R}_{t,r} K_r \left(Q_r \mathcal{Y}_{r,s} - Q_r^{n-1} \mathcal{R}^n_{r,s} \right) dr.$$

So, for arbitrary $h \in X$, we have

$$\begin{aligned} \|\mathcal{R}^n_{t,s}h - \mathcal{Y}_{t,s}h\| \ &\le \ c_1 c_2 \int_s^t \left\| \left(Q_r - Q_r^{n-1}\right) \mathcal{Y}_{r,s} h \right\| dr \\ &\quad + c_0 c_1 c_2 \int_s^t \|\mathcal{Y}_{r,s}h - \mathcal{R}^n_{r,s}h\| \, dr, \end{aligned}$$

where $c_1 = \sup_{\Delta_T} \|\mathcal{R}_{t,r}\|$ and $c_2 = \text{ess sup}_{[0,T]} \|K_r\|$. By Theorem 2.45,

$$\|\mathcal{R}^n_{t,s}h - \mathcal{Y}_{t,s}h\| \le c_1 c_2 e^{c_0 c_1 c_2 (t-s)} \int_s^t \left\| \left(Q_r - Q_r^{n-1}\right) \mathcal{Y}_{r,s} h \right\| dr. \tag{3.14}$$

Hence, by the strong convergence of Q_t^n to Q_t and by (3.13), we can apply Theorem 2.19 in (3.14) and obtain the strong convergence of $\mathcal{R}^n_{t,s}$ to $\mathcal{Y}_{t,s}$ for all $0 \le s \le t \le T$. Similarly, the strong convergence of $\mathcal{R}^{n*}_{t,s}$ to $\mathcal{Y}^*_{t,s}$ for all $0 \le s \le t \le T$ can be shown. From $\mathcal{R}^n = \mathcal{P}_{-KQ^{n-1}}(\mathcal{R})$, we have

$$\left\|\mathcal{R}^n_{t,s}\right\| \le c_1 + c_0 c_1 c_2 \int_s^t \left\|\mathcal{R}^n_{r,s}\right\| dr.$$

By Theorem 2.45, this implies

$$\left\|\mathcal{R}^{n*}_{t,s}\right\| = \left\|\mathcal{R}^n_{t,s}\right\| \le c_1 e^{c_0 c_1 c_2 T}, \ 0 \le s \le t \le T. \tag{3.15}$$

Thus, using the estimations (3.13) and (3.15) and applying Proposition 1.32(b) and Theorem 2.19 in (3.11), we conclude that Q is a solution of the equation (3.12). To prove that this solution is unique, let P be also a solution of the equation (3.12). Obviously, $P_t \ge 0$, $0 \le t \le T$. Since $\mathcal{Y}$ in (3.12) is related with Q, we see that P satisfies

$$P_t = \mathcal{K}^*_{T,t} Q_T \mathcal{K}_{T,t} + \int_t^T \mathcal{K}^*_{s,t} (M_s + P_s K_s P_s) \mathcal{K}_{s,t} \, ds, \ 0 \le t \le T,$$

where $\mathcal{K} = \mathcal{P}_{-KP}(\mathcal{R})$. Using the relation $\mathcal{K} = \mathcal{P}_{K(Q-P)}(\mathcal{Y})$, in a similar way as in the proof of Lemmas 3.19 and 3.20, one can show that

$$P_t - Q_t = \int_t^T \mathcal{Y}^*_{s,t} (Q_s - P_s) K_s (Q_s - P_s) \mathcal{Y}_{s,t} \, ds \ge 0, \ 0 \le t \le T.$$

Also, by symmetry,

$$Q_t - P_t = \int_t^T \mathcal{K}^*_{s,t} (P_s - Q_s) K_s (P_s - Q_s) \mathcal{K}_{s,t} \, ds \ge 0, \ 0 \le t \le T.$$

Combining the last two inequalities, we conclude that $Q_t = P_t$, $0 \le t \le T$. □

Theorem 3.22. *Under the conditions* (3.10), *the Riccati equation* (3.9) *is equivalent to the equation*

$$Q_t = \mathcal{Y}_{T,t}^* Q_T \mathcal{Y}_{T,t} + \int_t^T \mathcal{Y}_{s,t}^* (F_s - L_s^* G_s^{-1} L_s + Q_s B_s G_s^{-1} B_s^* Q_s) \mathcal{Y}_{s,t}\, ds,\ 0 \le t \le T, \tag{3.16}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, *and there exists a unique solution* Q *of these equations in* $B(0,T;\mathcal{L}(X))$ *satisfying* $Q_t \ge 0,\ 0 \le t \le T$.

Proof. One can observe that the equation (3.16) is same as equation (3.12) in view of $M = F - L^*G^{-1}L$ and $K = BG^{-1}B^*$. So, by Lemma 3.21, there exists a unique solution Q of the equation (3.16) in $B(0,T;\mathcal{L}(X))$ which satisfies $Q_t \ge 0$, $0 \le t \le T$. For $\mathcal{U} = \mathcal{P}_{BG^{-1}(B^*Q+L)}(\mathcal{Y})$, in a similar way as in the proof of Lemmas 3.19 and 3.20, the equation (3.16) can be transformed to the equation (3.9). So, Q is a solution of the equation (3.9) too. Taking Q as a solution of the equation (3.9), by inverse transformations, the equation (3.9) can be transformed to the equation (3.16). So, each solution of the equation (3.9) is a solution of the equation (3.16). Thus, the equation (3.9) and the equation (3.16) are equivalent. □

The Riccati equation (3.9) has different equivalent forms. One of them is the equation (3.16). We present also the following form of the equation (3.9) which will be used in the sequel.

Proposition 3.23. *The Riccati equation* (3.9) *is equivalent to the equation*

$$Q_t = \mathcal{Y}_{T,t}^* Q_T \mathcal{U}_{T,t} + \int_t^T \mathcal{Y}_{s,t}^* (F_s - (Q_s B_s + L_s^*) G_s^{-1} L_s) \mathcal{U}_{s,t}\, ds,\ 0 \le t \le T,$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$.

Proof. This can be proved in a similar way as the equivalence of the equations (3.9) and (3.16) in Theorem 3.22. □

3.3.2 Dual Riccati Equation

The equation (3.9) includes the final value Q_T of its solution. There is a modification of the operator Riccati equation (3.9) which uses the initial value of its solution. This equation has the form

$$P_t = \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s} (W_s - (P_s C_s^* + R_s) V_s^{-1} (C_s P_s + R_s^*)) \mathcal{U}_{t,s}^*\, ds,\ 0 \le t \le T, \tag{3.17}$$

and it is called a *dual operator Riccati equation*. This equation will be considered under the conditions

$$\begin{cases} T > 0,\ \mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X)),\ C \in B_\infty(0,T;\mathcal{L}(X,U)), \\ W \in B_\infty(0,T;\mathcal{L}(X)),\ R \in B_\infty(0,T;\mathcal{L}(U,X)), \\ V, V^{-1} \in B_\infty(0,T;\mathcal{L}(U)),\ P_0 \in \mathcal{L}(X),\ P_0 \geq 0, \\ V_t > 0 \text{ and } W_t - R_t V_t^{-1} R_t^* \geq 0 \text{ for a.e. } t \in [0,T]. \end{cases} \tag{3.18}$$

Theorem 3.24. *Under the conditions* (3.18), *there exists a unique solution P of the dual operator Riccati equation* (3.17) *in $B(0,T;\mathcal{L}(X))$ satisfying $P_t \geq 0$ for all $0 \leq t \leq T$.*

Proof. In the equation (3.9), replace $\mathcal{U}$ by $\mathcal{D}_T(\mathcal{U})$ and substitute $Q_T = P_0$, $B = D_T(C)$, $F = D_T(W)$, $L = D_T(R)$ and $G = D_T(V)$ (see Section 3.2.2 for the transformations $\mathcal{D}_T$ and D_T) to obtain

$$\begin{aligned} Q_t = \mathcal{U}_{T-t,0} P_0 \mathcal{U}_{T-t,0}^* + \int_t^T \mathcal{U}_{T-t,T-s}\big(W_{T-s} \\ - (Q_s C_{T-s}^* + R_{T-s}) V_{T-s}^{-1} (C_{T-s} Q_s + R_{T-s}^*)\big) \mathcal{U}_{T-t,T-s}^* \, ds. \end{aligned}$$

The conditions of Theorem 3.22 hold and, therefore, this final equation has a unique solution $Q \in B(0,T;\mathcal{L}(X))$ with $Q_t \geq 0$, $0 \leq t \leq T$. Replacing t in it by $T - t$ and simplifying, we obtain

$$\begin{aligned} Q_{T-t} &= \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_{T-t}^T \mathcal{U}_{t,T-s}\big(W_{T-s} \\ &\quad - (Q_s C_{T-s}^* + R_{T-s}) V_{T-s}^{-1} (C_{T-s} Q_s + R_{T-s}^*)\big) \mathcal{U}_{t,T-s}^* \, ds \\ &= \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s}\big(W_s \\ &\quad - (Q_{T-s} C_s^* + R_s) V_s^{-1} (C_s Q_{T-s} + R_s^*)\big) \mathcal{U}_{t,s}^* \, ds. \end{aligned}$$

Hence, the function $P_t = Q_{T-t}$, $0 \leq t \leq T$, is a solution of the equation (3.17). Clearly, $P \in B(0,T;\mathcal{L}(X))$ and $P_t \geq 0$, $0 \leq t \leq T$, because the same properties hold for Q. Finally, the uniqueness of P as a solution of the equation (3.17) follows from the uniqueness of the solution of the equation (3.9). □

Remark 3.25. One can easily verify that the equation (3.17) is equivalent to the equation

$$\begin{aligned} P_t = \mathcal{R}_{t,0} P_0 \mathcal{R}_{t,0}^* + \int_0^t \mathcal{R}_{t,s}\big(W_s \\ - R_s V_s^{-1} R_s^* + P_s C_s^* V_s^{-1} C_s P_s\big) \mathcal{R}_{t,s}^* \, ds,\ 0 \leq t \leq T, \end{aligned}$$

where $\mathcal{R} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$.

The equation (3.17) arises in studying estimation problems under more restrictive conditions than (3.18). For this case, we present the following result.

Proposition 3.26. *Suppose*

$$\begin{cases} T > 0,\ \mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X)),\ C \in B_\infty(0,T;\mathcal{L}(X,\mathbb{R}^n)), \\ W \in L_\infty(0,T;\mathcal{L}_1(X)),\ R \in B_\infty(0,T;\mathcal{L}(\mathbb{R}^n, X)), \\ V, V^{-1} \in L_\infty(0,T;\mathcal{L}(\mathbb{R}^n)),\ P_0 \in \mathcal{L}_1(X),\ P_0 \geq 0, \\ V_t > 0 \text{ and } W_t - R_t V_t^{-1} R_t^* \geq 0 \text{ for a.e. } t \in [0,T]. \end{cases}$$

Then there exists a unique solution P of the equation (3.17) *in* $L_\infty(0,T;\mathcal{L}_1(X))$ *satisfying* $P_t \geq 0$, $0 \leq t \leq T$.

Proof. One can verify that the conditions of Theorem 3.24 hold. Therefore, there exists a unique solution P of the equation (3.17) in $B(0,T;\mathcal{L}(X))$ which satisfies $P_t \geq 0$, $0 \leq t \leq T$. Let us show that $P_t \in \mathcal{L}_1(X)$ for all $0 \leq t \leq T$. Obviously, the first term in the right-hand side of (3.17) belongs to $\mathcal{L}_1(X)$ for all $0 \leq t \leq T$. To show the same property for the second term consider the integrand in (3.17). It is an $\mathcal{L}_1(X)$-valued function. Since this integrand is also a strongly measurable $\mathcal{L}(X)$-valued function, by Proposition 2.13(b), it is a measurable function with values in $\mathcal{L}_1(X)$. So, using the relation between the norms of the spaces $\mathcal{L}(X)$ and $\mathcal{L}_1(X)$ (see Section 1.3.7), one can verify that this integrand belongs to $L_\infty(0,t;\mathcal{L}_1(X))$ for fixed $0 < t \leq T$. Thus, the integral in (3.17) can be interpreted as the Bochner integral of an $\mathcal{L}_1(X)$-valued function rather than a strong Bochner integral and, consequently, the integral in (3.17) is an operator in $\mathcal{L}_1(X)$ for fixed $0 \leq t \leq T$. Thus, the solution P of the equation (3.17) is an $\mathcal{L}_1(X)$-valued function and, consequently, $P \in B(0,T;\mathcal{L}(X))$ implies $P \in L_\infty(0,T;\mathcal{L}_1(X))$ in view of Proposition 2.13(b). □

3.3.3 Riccati Equations in Differential Form

Let $\mathcal{U} \in \mathcal{S}(X)$ and let A be the infinitesimal generator of $\mathcal{U}$. With the Riccati equations (3.9) and (3.17) one can associate the differential equations

$$\frac{d}{dt}Q_t + Q_t A + A^* Q_t + F_t - (Q_t B_t + L_t^*) G_t^{-1} (B_t^* Q_t + L_t) = 0,\ 0 \leq t < T, \quad (3.19)$$

$$\frac{d}{dt}P_t - P_t A^* - A P_t - W_t + (P_t C_t^* + R_t) V_t^{-1} (C_t P_t + R_t^*) = 0,\ 0 < t \leq T. \quad (3.20)$$

The equations (3.19) and (3.20) are called differential Riccati equations. Under a solution of the equation (3.19) we mean a function $Q : [0,T] \to \mathcal{L}(X)$ that is strongly continuous and satisfies

$$\frac{d}{dt}\langle Q_t x, y\rangle + \langle Ax, Q_t y\rangle + \langle Q_t x, Ay\rangle$$
$$+ \left\langle \left(F_t - (Q_t B_t + L_t^*) G_t^{-1} (B_t^* Q_t + L_t)\right) x, y \right\rangle = 0,$$

for all $x, y \in D(A)$ and for a.e. $t \in [0,T)$ provided that Q_T is given. Similarly, a solution of the equation (3.20) is a function $P : [0,T] \to \mathcal{L}(X)$ that is strongly continuous and satisfies

$$\frac{d}{dt}\langle P_t x^*, y^*\rangle - \langle A^* x^*, P_t y^*\rangle - \langle P_t x^*, A^* y^*\rangle \\ - \left\langle \left(W_t - (P_t C_t^* + R_t) V_t^{-1}(C_t P_t + R_t^*)\right) x^*, y^* \right\rangle = 0,$$

for all $x^*, y^* \in D(A^*)$ and for a.e. $t \in (0,T]$ provided that P_0 is given. The solutions of the equations (3.19) and (3.20) in the above mentioned sense are called *scalar product solutions.*

Theorem 3.27. *Assume that the conditions in* (3.10) *hold with* $\mathcal{U} \in \mathcal{S}(X)$ *and* A *is the infinitesimal generator of* $\mathcal{U}$. *Then the solution* Q *of the operator Riccati equation* (3.9) *is a unique scalar product solution of the equation* (3.19) *with the final value* Q_T.

Proof. Under the conditions of the theorem the functions Q^n, $n = 1, 2, \dots$, defined by (3.11), are strongly continuous. Hence, by Lemma 3.21, Q is strongly continuous as it is the strong limit of Q^n. Let $x, y \in D(A)$. From (3.9), we have

$$\langle Q_t x, y\rangle = \langle Q_T \mathcal{U}_{T-t} x, \mathcal{U}_{T-t} y\rangle \\ + \int_t^T \left\langle (F_s - (Q_s B_s + L_s^*) G_s^{-1}(B_s^* Q_s + L_s))\mathcal{U}_{s-t} x, \mathcal{U}_{s-t} y \right\rangle ds. \quad (3.21)$$

By Propositions 2.8(b), 2.41 and 2.42, the right-hand side of (3.21) is a.e. differentiable in t with

$$\begin{aligned}\frac{d}{dt}\langle Q_t x, y\rangle = &- \langle Q_T \mathcal{U}_{T-t} A x, \mathcal{U}_{T-t} y\rangle - \langle Q_T \mathcal{U}_{T-t} x, \mathcal{U}_{T-t} A y\rangle \\ &- \left\langle \left(F_t - (Q_t B_t + L_t^*) G_t^{-1}(B_t^* Q_t + L_t)\right) x, y\right\rangle \\ &- \int_t^T \left\langle \left(F_s - (Q_s B_s + L_s^*) G_s^{-1}(B_s^* Q_s + L_s)\right) \mathcal{U}_{s-t} A x, \mathcal{U}_{s-t} y \right\rangle ds \\ &- \int_t^T \left\langle \left(F_s - (Q_s B_s + L_s^*) G_s^{-1}(B_s^* Q_s + L_s)\right) \mathcal{U}_{s-t} x, \mathcal{U}_{s-t} A y \right\rangle ds \\ = &- \langle A x, Q_t y\rangle - \langle Q_t x, A y\rangle \\ &- \left\langle \left(F_t - (Q_t B_t + L_t^*) G_t^{-1}(B_t^* Q_t + L_t)\right) x, y\right\rangle.\end{aligned}$$

So, Q is a scalar product solution of the equation (3.19). For the uniqueness of this solution, we refer to Curtain and Pritchard [40]. □

Theorem 3.28. *Assume that the conditions in* (3.18) *hold with* $\mathcal{U} \in \mathcal{S}(X)$ *and* A *is the infinitesimal generator of* $\mathcal{U}$. *Then the solution* P *of the dual operator Riccati equation* (3.17) *is a unique scalar product solution of the equation* (3.20) *with the initial value* P_0.

Proof. This theorem can be proved in a similar way as Theorem 3.27. □

Finally, we present the following theorems showing the explicit form of the solutions of the equations (3.19) and (3.20).

Theorem 3.29. *If A is the infinitesimal generator of $\mathcal{U} \in \mathcal{S}(X)$, $T > 0$, $B_t \equiv B \in \mathcal{L}(U,X)$, $F_t \equiv F \in \mathcal{L}(X)$, $L_t \equiv L \in \mathcal{L}(X,U)$, $G_t \equiv G \in \mathcal{L}(U)$, $G > 0$, $F = L^*G^{-1}L$ and $Q_T > 0$, then the scalar product solution of the equation* (3.19) *with the final value Q_T has the explicit form*

$$Q_t = \mathcal{R}^*_{T-t}\left(Q_T^{-1} + \int_t^T \mathcal{R}_{T-s}BG^{-1}B^*\mathcal{R}^*_{T-s}\,ds\right)^{-1}\mathcal{R}_{T-t},\ 0 \le t \le T, \tag{3.22}$$

where $\mathcal{R} = \mathcal{P}_{-BG^{-1}L}(\mathcal{U})$.

Proof. By Theorem 3.15, $\mathcal{R} \in \mathcal{S}(X)$ and $A - BG^{-1}L$ is the infinitesimal generator of $\mathcal{R}$. Therefore,

$$\frac{d}{dt}\mathcal{R}_t x = (A - BG^{-1}L)\mathcal{R}_t x = \mathcal{R}_t(A - BG^{-1}L)x,\ x \in D(A).$$

Let

$$\mathcal{W}_t = \left(Q_T^{-1} + \int_t^T \mathcal{R}_{T-s}BG^{-1}B^*\mathcal{R}^*_{T-s}\,ds\right)^{-1}.$$

Since $Q_T > 0$, we have $Q_T^{-1} > 0$. So, by Proposition 2.8(e) and Theorem 2.40,

$$\mathcal{W}_t'h = \mathcal{W}_t\mathcal{R}_{T-t}BG^{-1}B^*\mathcal{R}^*_{T-t}\mathcal{W}_t h,\ h \in X.$$

Hence, by Propositions 2.8(a) and 2.8(b), for all $x, y \in D(A)$,

$$\begin{aligned}
\langle Q_t x, y\rangle' &= \langle \mathcal{W}_t\mathcal{R}_{T-t}x, \mathcal{R}_{T-t}y\rangle' \\
&= \langle \mathcal{W}_t'\mathcal{R}_{T-t}x, \mathcal{R}_{T-t}y\rangle + \langle \mathcal{W}_t(\mathcal{R}_{T-t}x)', \mathcal{R}_{T-t}y\rangle \\
&\quad + \langle \mathcal{W}_t\mathcal{R}_{T-t}x, (\mathcal{R}_{T-t}y)'\rangle \\
&= \left\langle \mathcal{W}_t\mathcal{R}_{T-t}BG^{-1}B^*\mathcal{R}^*_{T-t}\mathcal{W}_t\mathcal{R}_{T-t}x, \mathcal{R}_{T-t}y\right\rangle \\
&\quad - \left\langle \mathcal{W}_t\mathcal{R}_{T-t}\left(A - BG^{-1}L\right)x, \mathcal{R}_{T-t}y\right\rangle \\
&\quad - \left\langle \mathcal{W}_t\mathcal{R}_{T-t}x, \mathcal{R}_{T-t}\left(A - BG^{-1}L\right)y\right\rangle \\
&= \left\langle Q_tBG^{-1}B^*Q_t x, y\right\rangle - \left\langle Q_t\left(A - BG^{-1}L\right)x, y\right\rangle \\
&\quad - \left\langle Q_t x, \left(A - BG^{-1}L\right)y\right\rangle \\
&= -\langle Ax, Q_t y\rangle - \langle Q_t x, Ay\rangle \\
&\quad - \left\langle \left(L^*G^{-1}L - (Q_tB + L^*)G^{-1}(B^*Q_t + L)\right)x, y\right\rangle.
\end{aligned}$$

So, the operator-valued function Q, defined by (3.22), is a scalar product solution of the equation (3.19). One can also verify that this function at $t = T$ is equal to the given operator Q_T. □

Theorem 3.30. *If A is the infinitesimal generator of $\mathcal{U} \in \mathcal{S}(X)$, $T > 0$, $C_t \equiv C \in \mathcal{L}(X,U)$, $W_t \equiv W \in \mathcal{L}(X)$, $R_t \equiv R \in \mathcal{L}(U,X)$, $V_t \equiv V \in \mathcal{L}(U)$, $V > 0$, $W = RV^{-1}R^*$ and $P_0 > 0$, then the scalar product solution of the equation* (3.20) *with the initial value P_0 has the following explicit form:*

$$P_t = \mathcal{R}_t \left(P_0^{-1} + \int_0^t \mathcal{R}_s^* C^* V^{-1} C \mathcal{R}_s \, ds \right)^{-1} \mathcal{R}_t^*, \ 0 \le t \le T,$$

where $\mathcal{R} = \mathcal{P}_{-RV^{-1}C}(\mathcal{U})$.

Proof. Replacing $\mathcal{U}$ by $\mathcal{D}_T(\mathcal{U})$ and letting $Q_T = P_0$, $B = C^*$, $F = W$, $L = R^*$, $G = V$, this theorem can be obtained from Theorem 3.29. □

3.4 Unbounded Perturbation

In this section a perturbation of evolution operators by an unbounded operator, directed to a representation of solutions of differential delay equations, is considered.

In this section, $0 < \varepsilon < T$ and we use the notation $\tilde{X} = L_2(-\varepsilon, 0; X)$ and $\check{X} = W^{1,2}(-\varepsilon, 0; X)$. Also, it is supposed that $\mathcal{T}$ is the strongly continuous semigroup of right translation, defined by (3.4), and

$$\Gamma \in \mathcal{L}(\check{X}, X) : \ \Gamma h = h_0, \ h \in \check{X}. \tag{3.23}$$

Recall that $\check{X}$ is a Hilbert space and the scalar product in $\check{X}$ will be defined by

$$\langle h, g \rangle_{\check{X}} = \langle h_0, g_0 \rangle_X + \int_{-\varepsilon}^0 \langle h'_\theta, g'_\theta \rangle_X \, d\theta.$$

3.4.1 Preliminaries

Consider the equation

$$x_t^s = \mathcal{R}_{t,s} l + \int_s^t \mathcal{R}_{t,r} N_r \begin{Bmatrix} x_{\nu_r}^s, & \nu_r > s \\ f_{\nu_r - s}, & \nu_r \le s \end{Bmatrix} dr, \ 0 \le s \le t \le T, \tag{3.24}$$

where $l \in X$ and $f \in \tilde{X}$, and suppose that the following conditions hold:

$$\begin{cases} \mathcal{R} \in \mathcal{E}(\Delta_T, \mathcal{L}(X)), \ N \in B_\infty(0,T;\mathcal{L}(X)), \\ \nu \in W^{1,\infty}(0,T;\mathbb{R}), \ t - \varepsilon \le \nu_t \le t \text{ for } 0 \le t \le T, \\ \nu_s < \nu_t \text{ for } 0 \le s < t \le T. \end{cases} \tag{3.25}$$

We would like to prove an analogue of Proposition 3.16 for the equation (3.24). Since the equation (3.24) contains the translation of an unknown function, it is impossible to express its solution by use of bounded perturbations of $\mathcal{R}$. For this, we will consider a modification of bounded perturbations to an unbounded case.

Below we will use integrands of the operator-valued functions containing $\Gamma\mathcal{T}$ and $\mathcal{T}^*\Gamma^*$. Since Γ and Γ^* are not bounded operators from $\tilde{X}$ to X and from X to $\tilde{X}$, respectively, we have to give a sense to these integrals. For this, let $h \in \tilde{X}$. Though Γ is defined on $\check{X}$, we can still consider

$$\Gamma\mathcal{T}_t h = \begin{Bmatrix} h_{-t}, & t \le \varepsilon \\ 0, & t > \varepsilon \end{Bmatrix}, \text{ a.e. } t \in [0,T], \tag{3.26}$$

as a function in $L_2(0,T;X)$. If $G \in B_\infty(0,T;\mathcal{L}(X,Y))$ and $0 \le s \le t \le T$, then

$$\left\| \int_s^t G_r \Gamma \mathcal{T}_{t-r} h \, dt \right\|^2 \le T \operatorname*{ess\,sup}_{r\in[0,T]} \|G_r\|^2 \|h\|^2_{\tilde{X}}.$$

Thus

$$J = \int_s^t G_r \Gamma \mathcal{T}_{t-r} \, dr$$

is well-defined as an operator in $\mathcal{L}(\tilde{X}, Y)$ by

$$Jh = \left(\int_s^t G_r \Gamma \mathcal{T}_{t-r} \, dr \right) h = \int_s^t G_r \Gamma \mathcal{T}_{t-r} h \, dr, \;\; h \in \tilde{X}.$$

The adjoint $J^* \in \mathcal{L}(Y, \tilde{X})$ of J will be denoted by

$$J^* = \int_s^t \mathcal{T}^*_{t-r} \Gamma^* G^*_r \, dr.$$

Remark 3.31. The usage of the above notation for J^* is suggested by Proposition 2.33(c) for strong Bochner integrals. But neither J nor J^* should be interpreted as a strong Bochner integral of $\mathcal{L}(\tilde{X}, Y)$- and $\mathcal{L}(Y, \tilde{X})$-valued functions, respectively.

Lemma 3.32. *With the notation introduced above, the equality*

$$\left[\int_s^t \mathcal{T}^*_{t-r} \Gamma^* G^*_r g \, dr \right]_\theta = \begin{Bmatrix} G^*_{t+\theta} g, & t+\theta > s \\ 0, & t+\theta \le s \end{Bmatrix}, \;\; a.e. \; \theta \in [-\varepsilon, 0], \tag{3.27}$$

holds for all $G \in B_\infty(0,T;\mathcal{L}(X,Y))$, *for all* $g \in Y$ *and for all* $0 \le s \le t \le T$.

Proof. Let $h \in \tilde{X}$. By (3.26), we have

$$\begin{aligned} \langle Jh, g \rangle &= \left\langle \int_s^t G_r \Gamma \mathcal{T}_{t-r} h \, dr, g \right\rangle \\ &= \int_s^t \left\langle \begin{Bmatrix} h_{r-t}, & t-r \le \varepsilon \\ 0, & t-r > \varepsilon \end{Bmatrix}, G^*_r g \right\rangle dr \\ &= \int_{-\varepsilon}^0 \left\langle h_\theta \begin{Bmatrix} G^*_{t+\theta} g, & t+\theta > s \\ 0, & t+\theta \le s \end{Bmatrix} \right\rangle d\theta, \end{aligned}$$

proving the lemma. □

Remark 3.33. Though J^* is not a strong Bochner integral of $\mathcal{L}(Y,\check{X})$-valued function, it can be interpreted as a strong Bochner integral of $\mathcal{L}(Y,\check{X}^*)$-valued function. Indeed, let $h \in \check{X}$. Then

$$\begin{aligned}\|\Gamma\mathcal{T}_t h\|_X^2 &= \|h_{-t}\|_X^2 \\ &= \left\| h_0 - \int_{-t}^{0} h'_\theta \, d\theta \right\|_X^2 \\ &\le 2\left(\|h_0\|_X^2 + \varepsilon \int_{-\varepsilon}^{0} \|h'_\theta\|_X^2 \, d\theta \right) \\ &\le 2\max(1,\varepsilon)\|h\|_{\check{X}}^2.\end{aligned}$$

We conclude that $\Gamma\mathcal{T}$ belongs to $B_\infty(0,T;\mathcal{L}(\check{X},X))$. By Proposition 2.33(c), this implies that $(\Gamma\mathcal{T})^* = \mathcal{T}^*\Gamma^* \in B_\infty(0,T;\mathcal{L}(X,\check{X}^*))$. Thus, the strong Bochner integral of the $\mathcal{L}(Y,\check{X}^*)$-valued function $\mathcal{T}^*\Gamma^*G^*$ exists.

3.4.2 Λ^*-Perturbation

Given $g \in L_2(a,b;X)$ and $\mu \in \tilde{X}$, we call the functions $\bar{g} : [a,b] \to \tilde{X}$ and $\tilde{g} : [a,b] \to X \times \tilde{X}$, defined by

$$[\bar{g}_t]_\theta = \begin{Bmatrix} g_{t+\theta}, & t+\theta > a \\ \lambda_{t+\theta-a}, & t+\theta \le a \end{Bmatrix}, \text{ a.e. } \theta \in [-\varepsilon, 0], \; a \le t \le b,$$

and

$$\tilde{g}_t = \begin{bmatrix} g_t \\ \bar{g}_t \end{bmatrix}, \; a \le t \le b,$$

the *bar* and *tilde functions*, respectively, over g with the initial distribution λ. One can observe that $\bar{g}$ expresses the past of g on $[t-\varepsilon, t]$ at instant t, and $\tilde{g}$ jointly expresses g and $\bar{g}$. Obviously, one has $\bar{g} \in L_2(a,b;\tilde{X})$ and $\tilde{g} \in L_2(a,b;X\times\tilde{X})$.

Now consider the function x^s defined by (3.24). Under the conditions (3.25), we have $x^s \in L_2(s,T;X)$ for all $0 \le s < T$. Let $\bar{x}^s$ and $\tilde{x}^s$ be the bar and tilde functions, respectively, over x^s with the initial distribution f. By (3.3) and Lemma 3.32, $\bar{x}^s$ can be represented in the form

$$\bar{x}_t^s = \mathcal{T}_{t-s}^* f + \int_s^t \mathcal{T}_{t-r}^* \Gamma^* x_r^s \, dr, \; 0 \le s \le t \le T. \tag{3.28}$$

Let

$$\tilde{\mathcal{R}} = \mathcal{R} \odot \mathcal{T}^*, \; \Lambda_r^*(N,\nu) = \begin{bmatrix} 0 & N_r\Gamma\mathcal{T}_{r-\nu_r} \\ \Gamma^* & 0 \end{bmatrix}. \tag{3.29}$$

Then, by (3.24) and (3.28), we have

$$\tilde{x}_t^s = \tilde{\mathcal{R}}_{t,s}\tilde{l} + \int_s^t \tilde{\mathcal{R}}_{t,r}\Lambda_r^*(N,\nu)\tilde{x}_r^s \, dr, \; 0 \le s \le t \le T, \tag{3.30}$$

where

$$\tilde{l} = \begin{bmatrix} l \\ f \end{bmatrix} \in X \times \tilde{X}$$

and the integral in (3.30) is the sense defined in Section 3.4.1.

In the following proposition we will use the operator-valued function

$$\mathcal{K}_{t,s} = \begin{bmatrix} \mathcal{K}^{00}_{t,s} & \mathcal{K}^{01}_{t,s} \\ \mathcal{K}^{10}_{t,s} & \mathcal{K}^{11}_{t,s} \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}),\ 0 \le s \le t \le T. \tag{3.31}$$

Here $\mathcal{K}^{00}$ is a unique, weakly continuous in each of the variables, solution of the following equivalent equations:

$$\mathcal{K}^{00}_{t,s} = \mathcal{R}_{t,s} + \int_{\min(\nu_s^{-1},t)}^{t} \mathcal{K}^{00}_{t,r} N_r \mathcal{R}_{\nu_r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.32}$$

$$\mathcal{K}^{00}_{t,s} = \mathcal{R}_{t,s} + \int_{\min(\nu_s^{-1},t)}^{t} \mathcal{R}_{t,r} N_r \mathcal{K}^{00}_{\nu_r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.33}$$

$\mathcal{K}^{01}$ and $\mathcal{K}^{10}$ are defined by

$$\mathcal{K}^{01}_{t,s} = \int_{s}^{\min(\nu_s^{-1},t)} \mathcal{K}^{00}_{t,r} N_r \Gamma \mathcal{T}_{s-\nu_r}\, dr,\ 0 \le s \le t \le T, \tag{3.34}$$

and

$$\mathcal{K}^{10}_{t,s} = \int_{s}^{t} \mathcal{T}^{*}_{t-r} \Gamma^{*} \mathcal{K}^{00}_{r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.35}$$

$\mathcal{K}^{11}$ is defined by either

$$\mathcal{K}^{11}_{t,s} = \mathcal{T}^{*}_{t-s} + \int_{s}^{\min(\nu_s^{-1},t)} \mathcal{K}^{10}_{t,r} N_r \Gamma \mathcal{T}_{s-\nu_r}\, dr,\ 0 \le s \le t \le T, \tag{3.36}$$

or

$$\mathcal{K}^{11}_{t,s} = \mathcal{T}^{*}_{t-s} + \int_{s}^{t} \mathcal{T}^{*}_{t-r} \Gamma^{*} \mathcal{K}^{01}_{r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.37}$$

where the integrals in (3.34)–(3.37) are in the respective senses defined in Section 3.4.1.

Proposition 3.34. *Under the above conditions and notation, the function* $\tilde{x}_t^s = \mathcal{K}_{t,s}\tilde{l}$, $0 \le s \le t \le T$, *is a unique weakly continuous solution of the equation* (3.30).

Proof. The existence of a unique weakly continuous solution of the equation (3.24) can be proved by use of the contraction mapping principle (see Theorem 2.46). This implies that the equation (3.30) has a unique weakly continuous solution as well. Let us show that the function $\tilde{x}_t^s = \mathcal{K}_{t,s}\tilde{l}$, $0 \le s \le t \le T$, is this solution.

First, note that the equivalence of the equations (3.32) and (3.33), the existence and the uniqueness of their solution and the weak continuity of this solution in each of the variables can be proved in a similar way as the proof of Theorem 3.13. Also, the equivalence of the formulae (3.36) and (3.37) can be proved easily. Hence, the formulae (3.31)–(3.37) uniquely define the operator-valued function $\mathcal{K}$ that is weakly continuous in each of the variables. Let

$$\begin{bmatrix} \xi_t^s \\ \eta_t^s \end{bmatrix} = \begin{bmatrix} \mathcal{K}_{t,s}^{00} l + \mathcal{K}_{t,s}^{01} f \\ \mathcal{K}_{t,s}^{10} l + \mathcal{K}_{t,s}^{11} f \end{bmatrix}.$$

It must be shown that $\xi_t^s = x_t^s$ and $\eta_t^s = \bar{x}_t^s$ for all $0 \le s \le t \le T$. From (3.33) and (3.34), we have

$$\begin{aligned} \xi_t^s &= \mathcal{K}_{t,s}^{00} l + \mathcal{K}_{t,s}^{01} f \\ &= \mathcal{R}_{t,s} l + \int_{\min(\nu_s^{-1},t)}^t \mathcal{R}_{t,r} N_r \mathcal{K}_{\nu_r,s}^{00} l \, dr \\ &\quad + \int_s^{\min(\nu_s^{-1},t)} \mathcal{R}_{t,r} N_r \Gamma \mathcal{T}_{s-\nu_r} f \, dr \\ &\quad + \int_s^{\min(\nu_s^{-1},t)} \int_{\min(\nu_r^{-1},t)}^t \mathcal{R}_{t,\alpha} N_\alpha \mathcal{K}_{\nu_\alpha,r}^{00} N_r \Gamma \mathcal{T}_{s-\nu_r} f \, d\alpha \, dr. \end{aligned}$$

Changing the order of integration yields

$$\begin{aligned} \xi_t^s &= \mathcal{R}_{t,s} l + \int_s^t \mathcal{R}_{t,r} N_r \begin{Bmatrix} \mathcal{K}_{\nu_r,s}^{00} l + \mathcal{K}_{\nu_r,s}^{01} f & \nu_r > s \\ \Gamma \mathcal{T}_{s-\nu_r} f, & \nu_r \le s \end{Bmatrix} dr \\ &= \mathcal{R}_{t,s} l + \int_s^t \mathcal{R}_{t,r} N_r \begin{Bmatrix} \xi_{\nu_r}^s & \nu_r > s \\ f_{\nu_r - s}, & \nu_r \le s \end{Bmatrix} dr. \end{aligned}$$

Since the equation (3.24) has a unique solution, we conclude that $\xi_t^s = x_t^s$ for all $0 \le s \le t \le T$. Now let us show that $\eta_t^s = \bar{x}_t^s$, $0 \le s \le t \le T$. From (3.35) and (3.37), we have

$$\begin{aligned} \eta_t^s &= \mathcal{K}_{t,s}^{10} l + \mathcal{K}_{t,s}^{11} f \\ &= \mathcal{T}_{t-s}^* f + \int_s^t \mathcal{T}_{t-r}^* \Gamma^* \left(\mathcal{K}_{r,s}^{00} l + \mathcal{K}_{r,s}^{01} f \right) dr \\ &= \mathcal{T}_{t-s}^* f + \int_s^t \mathcal{T}_{t-r}^* \Gamma^* x_r^s \, dr, \end{aligned}$$

implying

$$[\eta_t^s]_\theta = \begin{Bmatrix} x_{t+\theta}^s, & t + \theta > s \\ f_{t-s+\theta}, & t + \theta \le s \end{Bmatrix} = [\bar{x}_t^s]_\theta.$$

The proposition is proved. □

Theorem 3.35. *Let $\mathcal{R}$, N, ν satisfy* (3.25) *and let $\tilde{\mathcal{R}}$, Λ^* be defined by* (3.29). *Then the equation*

$$\mathcal{K}_{t,s} = \tilde{\mathcal{R}}_{t,s} + \int_s^t \tilde{\mathcal{R}}_{t,r}\Lambda_r^*(N,\nu)\mathcal{K}_{r,s}dr,\ 0 \le s \le t \le T, \tag{3.38}$$

has a unique solution $\mathcal{K}$ in $\mathcal{E}\big(\Delta_T, \mathcal{L}(X \times \tilde{X})\big)$ as defined by (3.31)–(3.37).

Proof. Using Proposition 3.34, one can show that the equation (3.38) has a unique solution $\mathcal{K}$ as defined by (3.31)–(3.37) and this solution is weakly continuous in each of the variables. Obviously, $\mathcal{K}_{t,t} = I$ for $0 \le t \le T$. Let us show the semigroup property for $\mathcal{K}$. By Proposition 3.34, for $0 \le s \le r \le t \le T$, we have

$$\mathcal{K}_{t,s}\tilde{l} = \tilde{x}_t^s = \mathcal{K}_{t,r}\tilde{x}_r^s = \mathcal{K}_{t,r}\mathcal{K}_{r,s}\tilde{l}.$$

Since $\tilde{l}$ is arbitrary in $X \times \tilde{X}$, we obtain that $\mathcal{K}$ satisfies the semigroup property. Thus, $\mathcal{K} \in \mathcal{E}\big(\Delta_T, \mathcal{L}(X \times \tilde{X})\big)$. □

Definition 3.36. The mild evolution operator $\mathcal{K}$, defined by (3.31)–(3.37), will be called the *dual unbounded perturbation* of $\mathcal{R} \odot \mathcal{T}^*$ by $\Lambda^*(N,\nu)$ or, briefly, Λ^*-*perturbation* of $\mathcal{R} \odot \mathcal{T}^*$. The notation $\mathcal{P}^*_{\Lambda^*(N,\nu)}(\mathcal{R} \odot \mathcal{T}^*)$ will be used to denote this perturbation.

Remark 3.37. By (3.26)–(3.27), the formulae (3.34)–(3.37) can be rewritten in the form

$$\begin{aligned}
\mathcal{K}^{01}_{t,s}h^1 &= \int_s^{\min(\nu_s^{-1},t)} \mathcal{K}^{00}_{t,r}N_r h^1_{\nu_r - s}\,dr,\ 0 \le s \le t \le T,\ h^1 \in \tilde{X},\\
\left[\mathcal{K}^{10}_{t,s}h^0\right]_\theta &= \begin{Bmatrix} \mathcal{K}^{00}_{t+\theta,s}h^0, & t+\theta > s \\ 0, & t+\theta \le s \end{Bmatrix},\ -\varepsilon \le \theta \le 0,\ 0 \le s \le t \le T,\ h^0 \in X,\\
\mathcal{K}^{11}_{t,s}h^1 &= \mathcal{T}^*_{t-s}h^1 + \int_s^{\min(\nu_s^{-1},t)} \mathcal{K}^{10}_{t,r}N_r h^1_{\nu_r - s}\,dr,\ 0 \le s \le t \le T,\ h^1 \in \tilde{X},\\
\left[\mathcal{K}^{11}_{t,s}h^1\right]_\theta &= \begin{Bmatrix} \mathcal{K}^{01}_{t+\theta,s}h^1, & t+\theta > s \\ h^1_{t-s+\theta}, & t+\theta \le s \end{Bmatrix},\ -\varepsilon \le \theta \le 0,\ 0 \le s \le t \le T,\ h^1 \in \tilde{X}.
\end{aligned}$$

3.4.3 Λ-Perturbation

Now suppose that $\mathcal{U}$, M and μ satisfy

$$\begin{cases} \mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X)),\ M \in B_\infty(0,T;\mathcal{L}(X)), \\ \mu \in W^{1,\infty}(0,T;\mathbb{R}),\ t \le \mu_t \le t+\varepsilon \text{ for } 0 \le t \le T, \\ \mu_s < \mu_t \text{ for } 0 \le s < t \le T. \end{cases} \tag{3.39}$$

Consider the operator-valued function

$$\mathcal{Y}_{t,s} = \begin{bmatrix} \mathcal{Y}^{00}_{t,s} & \mathcal{Y}^{01}_{t,s} \\ \mathcal{Y}^{10}_{t,s} & \mathcal{Y}^{11}_{t,s} \end{bmatrix} \in \mathcal{L}\big(X \times \tilde{X}\big),\ 0 \le s \le t \le T, \tag{3.40}$$

where $\mathcal{Y}^{00}$ is a solution of the equivalent equations

$$\mathcal{Y}^{00}_{t,s} = \mathcal{U}_{t,s} + \int_s^{\max\left(\mu_t^{-1},s\right)} \mathcal{Y}^{00}_{t,\mu_r} M_r \mathcal{U}_{r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.41}$$

$$\mathcal{Y}^{00}_{t,s} = \mathcal{U}_{t,s} + \int_s^{\max\left(\mu_t^{-1},s\right)} \mathcal{U}_{t,\mu_r} M_r \mathcal{Y}^{00}_{r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.42}$$

$\mathcal{Y}^{01}$ and $\mathcal{Y}^{10}$ are defined by

$$\mathcal{Y}^{01}_{t,s} = \int_s^t \mathcal{Y}^{00}_{t,r} \Gamma \mathcal{T}_{r-s}\, dr,\ 0 \le s \le t \le T, \tag{3.43}$$

and

$$\mathcal{Y}^{10}_{t,s} = \int_{\max\left(\mu_t^{-1},s\right)}^t \mathcal{T}^*_{\mu_r - t} \Gamma^* M_r \mathcal{Y}^{00}_{r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.44}$$

and $\mathcal{Y}^{11}$ is defined by either

$$\mathcal{Y}^{11}_{t,s} = \mathcal{T}_{t-s} + \int_s^t \mathcal{Y}^{10}_{t,r} \Gamma \mathcal{T}_{r-s}\, dr,\ 0 \le s \le t \le T, \tag{3.45}$$

or

$$\mathcal{Y}^{11}_{t,s} = \mathcal{T}_{t-s} + \int_{\max\left(\mu_t^{-1},s\right)}^t \mathcal{T}^*_{\mu_r - t} \Gamma^* M_r \mathcal{Y}^{01}_{r,s}\, dr,\ 0 \le s \le t \le T. \tag{3.46}$$

Note that in (3.43)–(3.46) the integrals are in the respective senses defined in Section 3.4.1.

Theorem 3.38. *Let* $\mathcal{U}$, M, μ *satisfy* (3.39). *Then the function* $\mathcal{Y}$, *defined by* (3.40)–(3.46), *belongs to* $\mathcal{E}\big(\Delta_T, \mathcal{L}(X \times \tilde{X})\big)$.

Proof. Let $\mathcal{R} = \mathcal{D}_T(\mathcal{U})$, $N = D_T(M)$ and $\nu = D_T(\mu)$ (see Section 3.2.2 for these symbols). Consider $\mathcal{K} \in \mathcal{E}\big(\Delta_T, \mathcal{L}(X \times \tilde{X})\big)$, defined by (3.31)–(3.37), where in turn $\mathcal{R}$, N and ν are defined through $\mathcal{U}$, M and μ as above. Using Proposition 3.18, one can verify that $\mathcal{Y} = \mathcal{D}_T(\mathcal{K})$. Hence, $\mathcal{Y} \in \mathcal{E}\big(\Delta_T, \mathcal{L}(X \times \tilde{X})\big)$. □

To obtain an equation for $\mathcal{Y}$, substitute $\mathcal{K} = \mathcal{D}_T(\mathcal{Y})$, $N = D_T(M)$ and $\nu = D_T(\mu)$ in (3.38). Then one can get the following equation for $\mathcal{Y}$:

$$\mathcal{Y}_{t,s} = (\mathcal{U} \odot \mathcal{T})_{t,s} + \int_s^t \mathcal{Y}_{t,r} \Lambda_r(M,\mu) (\mathcal{U} \odot \mathcal{T})_{r,s}\, dr,\ 0 \le s \le t \le T, \tag{3.47}$$

where

$$\Lambda_r(M,\mu) = \begin{bmatrix} 0 & \Gamma \\ \mathcal{T}^*_{\mu_r - r} \Gamma^* M_r & 0 \end{bmatrix},\ 0 \le r \le T. \tag{3.48}$$

Based on (3.47) and (3.48), we can introduce the following.

Definition 3.39. The mild evolution operator $\mathcal{Y}$, defined by (3.40)–(3.46), will be called the *unbounded perturbation* of $\mathcal{U}\odot\mathcal{T}$ by $\Lambda(M,\mu)$ or Λ*-perturbation* of $\mathcal{U}\odot\mathcal{T}$. The notation $\mathcal{P}_{\Lambda(M,\mu)}(\mathcal{U}\odot\mathcal{T})$ will be used to denote this perturbation.

Remark 3.40. Similar to Remark 3.37, we can rewrite the formulae (3.43)–(3.46) in the form

$$\mathcal{Y}^{01}_{t,s}g^1 = \int_{\max(-\varepsilon,s-t)}^{0} \mathcal{Y}^{00}_{t,s-r}g^1_r\,dr,\ 0\le s\le t\le T,\ g^1\in\tilde{X},$$

$$\big[\mathcal{Y}^{10}_{t,s}g^0\big]_\theta = \chi_{[t-\mu_t,\min(0,t-\mu_s)]}(\theta)\big(\mu^{-1}\big)'_{t-\theta}M_{\mu^{-1}_{t-\theta}}\mathcal{Y}^{00}_{\mu^{-1}_{t-\theta},s}g^0,$$

$$-\varepsilon\le\theta\le 0,\ 0\le s\le t\le T,\ g^0\in X,$$

$$\mathcal{Y}^{11}_{t,s}g^1 = \mathcal{T}_{t-s}g^1 + \int_{\max(-\varepsilon,s-t)}^{0} \mathcal{Y}^{10}_{t,s-r}g^1_r\,dr,\ 0\le s\le t\le T,\ g^1\in\tilde{X},$$

$$\big[\mathcal{Y}^{11}_{t,s}g^1\big]_\theta = \chi_{[t-\mu_t,\min(0,t-\mu_s)]}(\theta)\big(\mu^{-1}\big)'_{t-\theta}M_{\mu^{-1}_{t-\theta}}\mathcal{Y}^{01}_{\mu^{-1}_{t-\theta},s}g^1$$

$$+\chi_{[\min(0,t-s-\varepsilon),0)}(\theta)g^1_{\theta-t+s},\ -\varepsilon\le\theta\le 0,\ 0\le s\le t\le T,\ g^1\in\tilde{X}.$$

Here χ_G denotes the characteristic function of the set G.

The following proposition is a modification of Proposition 3.18(a) to the cases of Λ- and Λ^*-perturbations.

Proposition 3.41. *Suppose* $\mathcal{U}$, M *and* μ *satisfy* (3.39), $\mathcal{R}$, N *and* ν *satisfy* (3.25) *and* $0<t\le T$. *Then*

(a) $\mathcal{P}_{\Lambda(M,\mu)}(\mathcal{U}\odot\mathcal{T})|_{\Delta_t} = \mathcal{D}_t\big(\mathcal{P}^*_{\Lambda^*(D_t(M),D_t(\mu))}(\mathcal{D}_t(\mathcal{U})\odot\mathcal{T}^*)\big)$;

(b) $\mathcal{P}^*_{\Lambda^*(N,\nu)}(\mathcal{R}\odot\mathcal{T}^*)|_{\Delta_t} = \mathcal{D}_t\big(\mathcal{P}_{\Lambda(D_t(N),D_t(\nu))}(\mathcal{D}_t(\mathcal{R})\odot\mathcal{T})\big)$.

Proof. This can be seen in the proof of Theorem 3.38. □

3.4.4 Examples

In this section we present the examples of Λ- and Λ^*-perturbations which are strongly continuous semigroups.

Example 3.42. Suppose that

$$\mathcal{R}\in\mathcal{S}(X),\ N_t\equiv N\in\mathcal{L}(X)\text{ and }\nu_t = t-\varepsilon,\ t\ge 0.$$

An inspection of the formulae (3.31)–(3.37) shows that $\mathcal{K} = \mathcal{P}^*_{\Lambda^*(N,\nu)}(\mathcal{R}\odot\mathcal{T}^*)$ belongs to $\mathcal{S}\big(X\times\tilde{X}\big)$ and in the one-parameter form $\mathcal{K}$ can be decomposed as

$$\mathcal{K}_t = \begin{bmatrix}\mathcal{K}^{00}_t & \mathcal{K}^{01}_t\\ \mathcal{K}^{10}_t & \mathcal{K}^{11}_t\end{bmatrix}\in\mathcal{L}\big(X\times\tilde{X}\big),\ t\ge 0, \tag{3.49}$$

where $\mathcal{K}^{00}$ is a unique strongly continuous solution of the following equivalent equations:

$$\mathcal{K}_t^{00} = \mathcal{R}_t + \int_0^{\max(t-\varepsilon,0)} \mathcal{K}_r^{00} N \mathcal{R}_{t-r-\varepsilon}\, dr,\ t \geq 0, \tag{3.50}$$

$$\mathcal{K}_t^{00} = \mathcal{R}_t + \int_0^{\max(t-\varepsilon,0)} \mathcal{R}_{t-r-\varepsilon} N \mathcal{K}_r^{00}\, dr,\ t \geq 0, \tag{3.51}$$

$\mathcal{K}^{01}$, $\mathcal{K}^{01}$ and $\mathcal{K}^{11}$ are defined by

$$\mathcal{K}_t^{01} = \int_{\max(t-\varepsilon,0)}^{t} \mathcal{K}_r^{00} N \Gamma \mathcal{T}_{r-t+\varepsilon}\, dr,\ t \geq 0, \tag{3.52}$$

$$\mathcal{K}_t^{10} = \int_0^t \mathcal{T}_r^* \Gamma^* \mathcal{K}_{t-r}^{00}\, dr,\ t \geq 0, \tag{3.53}$$

$$\mathcal{K}_t^{11} = \mathcal{T}_t^* + \int_{\max(t-\varepsilon,0)}^{t} \mathcal{K}_r^{10} N \Gamma \mathcal{T}_{r-t+\varepsilon}\, dr,\ t \geq 0, \tag{3.54}$$

$$\mathcal{K}_t^{11} = \mathcal{T}_t^* + \int_0^t \mathcal{T}_r^* \Gamma^* \mathcal{K}_{t-r}^{01}\, dr,\ t \geq 0. \tag{3.55}$$

Furthermore, the infinitesimal generator $\tilde{A}^*$ of $\mathcal{K}$ has the form

$$\begin{cases} \tilde{A}^* = \begin{bmatrix} A^* & N\Gamma\mathcal{T}_\varepsilon \\ \Gamma^* & d/d\theta \end{bmatrix} \in \tilde{\mathcal{L}}(X \times \check{X}), \\ D(\tilde{A}^*) = \left\{ h = \begin{bmatrix} h^0 \\ h^1 \end{bmatrix} : h^0 \in D(A^*),\ h^1 \in \check{X},\ h_0^1 = h^0 \right\}, \end{cases} \tag{3.56}$$

where $A^* \in \tilde{\mathcal{L}}(X)$ is the infinitesimal generator of $\mathcal{R}$, $d/d\theta$ is the infinitesimal generator of $\mathcal{T}^*$ with $D(d/d\theta) = \{h^1 \in \check{X} : h_0^1 = 0\}$ and the operator Γ^* should be understood in the following sense:

$$\begin{cases} \Gamma^* h^0 + (d/d\theta) h^1 = (d/d\theta)\left(h^1 - \Pi h^0\right), \\ h^0 \in D(A^*),\ h^1 \in \check{X},\ h_0^1 = h^0, \end{cases} \tag{3.57}$$

with

$$\Pi \in \mathcal{L}(X, \check{X}) : [\Pi x]_\theta = x,\ -\varepsilon \leq \theta \leq 0,\ x \in X. \tag{3.58}$$

Indeed, consider

$$\mathcal{K}_t h = \begin{bmatrix} \mathcal{K}_t^{00} h^0 + \mathcal{K}_t^{01} h^1 \\ \mathcal{K}_t^{10} h^0 + \mathcal{K}_t^{11} h^1 \end{bmatrix},$$

where h^0 and h^1 are the respective components of $h \in D(\tilde{A}^*)$. Denote

$$x_t = \mathcal{R}_t h^0 + \int_0^t \mathcal{R}_r N h_{t-r-\varepsilon}^1\, dt,\ 0 \leq t \leq \varepsilon.$$

By Proposition 2.41, x_t is differentiable and its derivative at $t = 0$ is equal to $A^*h^0 + Nh^1_{-\varepsilon} = A^*h^0 + N\Gamma\mathcal{T}_\varepsilon h^1$. Since for small t, $\mathcal{K}_t^{00} = \mathcal{R}_t$ and

$$\mathcal{K}_t^{01}h^1 = \int_0^t \mathcal{R}_r N h^1_{t-r-\varepsilon}\, dr,$$

we obtain

$$\mathcal{K}_t^{00}h^0 + \mathcal{K}_t^{01}h^1 = \mathcal{R}_t h^0 + \int_0^t \mathcal{R}_r N h^1_{t-r-\varepsilon}\, dr = x_t,$$

proving that $\mathcal{K}_t^{00}h^0 + \mathcal{K}_t^{01}h^1$ is differentiable and its derivative at $t = 0$ is equal to $A^*h^0 + N\Gamma\mathcal{T}_\varepsilon h^1$. Also, from (3.27), we have

$$\begin{aligned}\left[\mathcal{K}_t^{10}h^0 + \mathcal{K}_t^{11}h^1\right]_\theta &= \left\{\begin{matrix} \mathcal{K}_{t+\theta}^{00}h^0 + \mathcal{K}_{t+\theta}^{01}h^1, & t+\theta > 0 \\ h^1_{t+\theta}, & t+\theta \le 0 \end{matrix}\right\} \\ &= \left\{\begin{matrix} x_{t+\theta}, & t+\theta > 0 \\ h^1_{t+\theta}, & t+\theta \le 0 \end{matrix}\right\},\end{aligned}$$

which implies that for a.e. $\theta \in [-\varepsilon, 0]$, $\left[\mathcal{K}_t^{10}h^0 + \mathcal{K}_t^{11}h^1\right]_\theta$ is differentiable with respect to t and its derivative at $t = 0$ is equal to $(d/d\theta)\big(h^1_\theta - h^0\big)$. Finally, since $(d/d\theta)\big(h^1 - \Pi h^0\big) \in \tilde{X}$, we have

$$\begin{aligned}&\lim_{t\to 0}\int_{-\varepsilon}^0 \left\| t^{-1}\left[\mathcal{K}_t^{10}h^0 + \mathcal{K}_t^{11}h^1\right]_\theta - \left[\Gamma^* h^0 + \frac{d}{d\theta}h^1\right]_\theta \right\|^2 d\theta \\ &\quad = \int_{-\varepsilon}^0 \lim_{t\to 0}\left\| t^{-1}\left[\mathcal{K}_t^{10}h^0 + \mathcal{K}_t^{11}h^1\right]_\theta - \frac{d}{d\theta}\big(h^1_\theta - h^0\big)\right\|^2 d\theta = 0,\end{aligned}$$

proving that $\mathcal{K}_t^{10}h^0 + \mathcal{K}_t^{11}h^1$ is differentiable as an $\tilde{X}$-valued function and its derivative at $t = 0$ is equal to $\Gamma^*h^0 + (d/d\theta)h^1$. Thus, $\tilde{A}^*$ is the infinitesimal generator of $\mathcal{K}$.

Example 3.43. Suppose that

$$\mathcal{R} \in \mathcal{S}(X),\ N_t \equiv 0 \text{ and } \nu_t = t - \varepsilon,\ t \ge 0.$$

Then from Example 3.42 it follows that $\mathcal{K} = \mathcal{P}^*_{\Lambda^*(0,\nu)}(\mathcal{R} \odot \mathcal{T}^*) \in \mathcal{S}\big(X \times \tilde{X}\big)$ where $\mathcal{K}$ can be decomposed as

$$\mathcal{K}_t = \begin{bmatrix} \mathcal{R}_t & 0 \\ \mathcal{K}_t^{10} & \mathcal{T}_t^* \end{bmatrix} \in \mathcal{L}\big(X \times \tilde{X}\big),\ t \ge 0,$$

with

$$\mathcal{K}_t^{10} = \int_0^t \mathcal{T}_r^* \Gamma^* \mathcal{R}_{t-r}\, dr,\ t \ge 0,$$

and its infinitesimal generator $\tilde{A}^*$ is defined by

$$\begin{cases} \tilde{A}^* = \begin{bmatrix} A^* & 0 \\ \Gamma^* & d/d\theta \end{bmatrix} \in \tilde{\mathcal{L}}(X \times \tilde{X}), \\ D(\tilde{A}^*) = \left\{ h = \begin{bmatrix} h^0 \\ h^1 \end{bmatrix} : h^0 \in D(A^*),\ h^1 \in \check{X},\ h^1_0 = h^0 \right\}, \end{cases}$$

where A^*, $d/d\theta$ and Γ^* are as in Example 3.42.

Example 3.44. Suppose that

$$\mathcal{U} \in \mathcal{S}(X),\ M_t \equiv M \in \mathcal{L}(X) \text{ and } \mu_t = t + \varepsilon,\ t \geq 0.$$

Then $\mathcal{Y} = \mathcal{P}_{\Lambda(M,\mu)}(\mathcal{U} \odot \mathcal{T})$ belongs to $\mathcal{S}(X \times \tilde{X})$ and in the one-parameter form $\mathcal{Y}$ can be decomposed as

$$\mathcal{Y}_t = \begin{bmatrix} \mathcal{Y}_t^{00} & \mathcal{Y}_t^{01} \\ \mathcal{Y}_t^{10} & \mathcal{Y}_t^{11} \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}),\ t \geq 0, \tag{3.59}$$

where $\mathcal{Y}^{00}$ is a unique strongly continuous solution of the following equivalent equations:

$$\mathcal{Y}_t^{00} = \mathcal{U}_t + \int_0^{\max(t-\varepsilon,0)} \mathcal{Y}_r^{00} M \mathcal{U}_{t-r-\varepsilon}\, dr,\ t \geq 0, \tag{3.60}$$

$$\mathcal{Y}_t^{00} = \mathcal{U}_t + \int_0^{\max(t-\varepsilon,0)} \mathcal{U}_{t-r-\varepsilon} M \mathcal{Y}_r^{00}\, dr,\ t \geq 0, \tag{3.61}$$

$\mathcal{Y}^{01}$, $\mathcal{Y}^{10}$ and $\mathcal{Y}^{11}$ are defined by

$$\mathcal{Y}_t^{01} = \int_0^t \mathcal{Y}_r^{00} \Gamma \mathcal{T}_{t-r}\, dr,\ t \geq 0, \tag{3.62}$$

$$\mathcal{Y}_t^{10} = \int_{\max(0,\varepsilon-t)}^{\varepsilon} \mathcal{T}_r^* \Gamma^* M \mathcal{Y}_{t+r-\varepsilon}^{00}\, dr,\ t \geq 0, \tag{3.63}$$

$$\mathcal{Y}_t^{11} = \mathcal{T}_t + \int_0^t \mathcal{Y}_r^{10} \Gamma \mathcal{T}_{t-r}\, dr,\ t \geq 0, \tag{3.64}$$

$$\mathcal{Y}_t^{11} = \mathcal{T}_t + \int_{\max(0,\varepsilon-t)}^{\varepsilon} \mathcal{T}_r^* \Gamma^* M \mathcal{Y}_{t+r-\varepsilon}^{01}\, dr,\ t \geq 0. \tag{3.65}$$

The infinitesimal generator $\tilde{A}$ of $\mathcal{Y}$ has the form

$$\begin{cases} \tilde{A} = \begin{bmatrix} A & \Gamma \\ \mathcal{T}_\varepsilon^* \Gamma^* M & -d/d\theta \end{bmatrix} \in \tilde{\mathcal{L}}(X \times \tilde{X}), \\ D(\tilde{A}) = \left\{ g = \begin{bmatrix} g^0 \\ g^1 \end{bmatrix} : g^0 \in D(A),\ g^1 \in \check{X},\ g^1_{-\varepsilon} = M g^0 \right\}, \end{cases} \tag{3.66}$$

where $A \in \tilde{\mathcal{L}}(X)$ is the infinitesimal generator of $\mathcal{U}$, $-d/d\theta$ is the infinitesimal generator of $\mathcal{T}$ with $D(-d/d\theta) = \{g^1 \in \check{X} : g^1_{-\varepsilon} = 0\}$ and the operator $\mathcal{T}_\varepsilon^{*}\Gamma^{*}M$ should be understood in the following sense:

$$\begin{cases} \mathcal{T}_\varepsilon^{*}\Gamma^{*}Mg^0 + (-d/d\theta)g^1 = (-d/d\theta)\left(g^1 - \Pi M g^0\right), \\ g^0 \in D(A),\ g^1 \in \check{X},\ g^1_{-\varepsilon} = Mg^0, \end{cases} \tag{3.67}$$

with Π defined by (3.58). Indeed, let $\mathcal{R} = \mathcal{D}_{\mathcal{T}}(\mathcal{U})$ and $N = M^*$. Then, for $\mathcal{Y}$ and $\mathcal{K}$ defined by (3.59)–(3.65) and (3.49)–(3.55), respectively, we have $\mathcal{Y} = \mathcal{K}^*$. Therefore, the infinitesimal generator of $\mathcal{Y}$ must be adjoint to the infinitesimal generator of $\mathcal{K}$. So, it is sufficient to show that $\tilde{A}$, defined by (3.66), is adjoint to $\tilde{A}^*$, defined by (3.56). By Proposition 2.43(b), for $h \in D(\tilde{A}^*)$ with the components h^0 and h^1 and for $g \in D(\tilde{A})$ with the components g^0 and g^1, we have

$$\begin{aligned} 0 &= \int_{-\varepsilon}^{0} \frac{d}{d\theta} \langle h^1_\theta - h^0, g^1_\theta - Mg^0\rangle \, d\theta \\ &= \int_{-\varepsilon}^{0} \left\langle \frac{d}{d\theta}\left(h^1_\theta - h^0\right), g^1_\theta - Mg^0\right\rangle d\theta \\ &\quad + \int_{-\varepsilon}^{0} \left\langle h^1_\theta - h^0, \frac{d}{d\theta}\left(g^1_\theta - Mg^0\right)\right\rangle d\theta. \end{aligned}$$

Hence,

$$\begin{aligned} &\langle h^1_{-\varepsilon} - h^0, Mg^0\rangle + \int_{-\varepsilon}^{0} \left\langle \frac{d}{d\theta}\left(h^1_\theta - h^0\right), g^1_\theta\right\rangle d\theta \\ &\quad = \langle h^0, g^1_0 - Mg^0\rangle - \int_{-\varepsilon}^{0} \left\langle h^1_\theta, \frac{d}{d\theta}\left(g^1_\theta - Mg^0\right)\right\rangle d\theta, \end{aligned}$$

and, therefore,

$$\begin{aligned} &\langle A^*h^0 + M^*h^1_{-\varepsilon}, g^0\rangle + \int_{-\varepsilon}^{0} \left\langle \frac{d}{d\theta}\left(h^1_\theta - h^0\right), g^1_\theta\right\rangle d\theta \\ &\quad = \langle h^0, Ag^0 + g^1_0\rangle - \int_{-\varepsilon}^{0} \left\langle h^1_\theta, \frac{d}{d\theta}\left(g^1_\theta - Mg^0\right)\right\rangle d\theta. \end{aligned}$$

Thus, $\langle \tilde{A}^*h, g\rangle_{X\times\tilde{X}} = \langle h, \tilde{A}g\rangle_{X\times\tilde{X}}$.

Example 3.45. Suppose that

$$\mathcal{U} \in \mathcal{S}(X),\ M_t \equiv 0 \text{ and } \mu_t = t + \varepsilon,\ t \ge 0.$$

Then from Example 3.44 it follows that $\mathcal{Y} = \mathcal{P}_{\Lambda(0,\mu)}(\mathcal{U}\odot\mathcal{T}) \in \mathcal{S}(X\times\tilde{X})$. Moreover, $\mathcal{Y}$ is defined by

$$\mathcal{Y}_t = \begin{bmatrix} \mathcal{U}_t & \mathcal{Y}^{01}_t \\ 0 & \mathcal{T}_t \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}),\ t \ge 0,$$

with

$$\mathcal{Y}_t^{01} = \int_0^t \mathcal{U}_r \Gamma \mathcal{T}_{t-r}\, dr,\ t \geq 0,$$

and has the infinitesimal generator

$$\begin{cases} \tilde{A} = \begin{bmatrix} A & \Gamma \\ 0 & -d/d\theta \end{bmatrix} \in \tilde{\mathcal{L}}(X \times \check{X}), \\ D(\tilde{A}) = \left\{ g = \begin{bmatrix} g^0 \\ g^1 \end{bmatrix} : g^0 \in D(A),\ g^1 \in \check{X},\ g^1_{-\varepsilon} = 0 \right\}, \end{cases}$$

where $A \in \tilde{\mathcal{L}}(X)$ is the infinitesimal generator of $\mathcal{U}$ and $-d/d\theta$ is the infinitesimal generator of $\mathcal{T}$ with $D(-d/d\theta) = \{g^1 \in \check{X} : g^1_{-\varepsilon} = 0\}$.

Chapter 4

Partially Observable Linear Systems

This chapter deals with stochastic differential equations and partially observable linear systems. We also present a basic estimation in Hilbert spaces and discuss white, colored and wide band noise processes.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $X, Y, Z, U, H \in \mathcal{H}$, $T > 0$, $\mathbf{T} = [0, T]$ is a finite time interval and $\Delta_t = \{(s, r) : 0 \le r \le s \le t\}$ for $t > 0$.

4.1 Random Variables and Processes

A random variable and a random process are the analogs of a variable and a function from analysis in studying random phenomena.

4.1.1 Random Variables

An X-valued *random variable* (or, briefly, a random variable) is a function in $m(\Omega, X)$. Two random variables ξ and η are said to be *equal* if $\xi_\omega = \eta_\omega$ w.p.1. The equality of the random variables ξ and η is written as $\xi = \eta$.

Given a random variable $\xi \in m(\Omega, X)$, the measure $\mathbf{P}_\xi$ on $\mathcal{B}_X$ generated by ξ, i.e., $\mathbf{P}_\xi(A) = \mathbf{P}(\xi^{-1}(A))$, $A \in \mathcal{B}_X$, is called the *distribution* of ξ. If $\xi \in L_1(\Omega, X)$, then the integral

$$\mathbf{E}\xi = \int_\Omega \xi \, d\mathbf{P} = \int_X x \, d\mathbf{P}_\xi$$

is called the *expectation* or the *mean value* of ξ.

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

With a random variable $\xi \in m(\Omega, X)$, one can associate the function

$$\varphi_\xi(x) = \mathbf{E}e^{i\langle x,\xi\rangle} = \int_\Omega e^{i\langle x,\xi\rangle}\, d\mathbf{P} = \int_X e^{i\langle x,y\rangle}\, d\mathbf{P}_\xi(y),\ \ x \in X,$$

where i is the imaginary unit ($i^2 = -1$). This function is called the *characteristic function* of ξ. Note that the values of φ_ξ are the complex numbers. By

$$\mathbb{C} \ni (ai + b) \leftrightarrow (a, b) \in \mathbb{R}^2,$$

the space $\mathbb{C}$ of all complex numbers and $\mathbb{R}^2$ are isometric and, hence, the integral in the definition of φ_ξ can be interpreted as an integral of an $\mathbb{R}^2$-valued function. Most properties of random variables are easily obtained if they are formulated in terms of their characteristic functions. For instance, convergence of random variables in measure $\mathbf{P}$, which in probability theory is called *convergence in probability*, could be formulated as follows.

Theorem 4.1. *A sequence $\{\xi_n\}$ of random variables in $m(\Omega, X)$ converges in probability to a random variable ξ if and only if*

$$\varphi_{\xi_n}(x) \to \varphi_\xi(x) \text{ as } n \to \infty \text{ for all } x \in X.$$

By Proposition 2.36, $((\xi - \mathbf{E}\xi) \otimes (\eta - \mathbf{E}\eta)) \in L_1(\Omega, \mathcal{L}_1(Y, X))$ for $\xi \in L_2(\Omega, X)$ and for $\eta \in L_2(\Omega, Y)$. Hence, one can define $\mathbf{E}((\xi - \mathbf{E}\xi) \otimes (\eta - \mathbf{E}\eta))$ as an operator in $\mathcal{L}_1(Y, X)$. This operator is called the *covariance* of ξ and η and is denoted by $\mathrm{cov}(\xi, \eta)$. In case $\xi = \eta$, one briefly writes $\mathrm{cov}\xi = \mathrm{cov}(\xi, \xi)$. If $X = \mathbb{R}$, then $\mathrm{cov}\xi$ is also called the *variance* of ξ. Using Proposition 1.26, one can verify the following properties of covariance.

Proposition 4.2. *Let $\xi, \zeta \in L_2(\Omega, X)$, let $\eta \in L_2(\Omega, Y)$, let $\Phi \in \mathcal{L}(X, Z)$ and let $\Psi \in \mathcal{L}(Y, H)$. Then*

(a) $\mathrm{cov}(\xi, \eta)^* = \mathrm{cov}(\eta, \xi)$;

(b) $\mathrm{cov}\xi \geq 0$;

(c) $\mathrm{cov}(\xi + \zeta, \eta) = \mathrm{cov}(\xi, \eta) + \mathrm{cov}(\zeta, \eta)$;

(d) $\mathrm{cov}(\Phi\xi, \Psi\eta) = \Phi\, \mathrm{cov}(\xi, \eta)\, \Psi^*$;

(e) $\mathrm{tr}(\mathrm{cov}(\xi, \zeta)) = \mathbf{E}\langle \xi, \zeta\rangle - \langle \mathbf{E}\xi, \mathbf{E}\zeta\rangle$;

(f) $\mathrm{tr}(\mathrm{cov}\xi) = \mathbf{E}(\|\xi\|^2) - \|\mathbf{E}\xi\|^2$;

(g) $\mathrm{tr}(\mathrm{cov}(\Phi\xi)) = \mathrm{tr}((\mathrm{cov}\xi)\Phi^*\Phi) = \mathrm{tr}(\Phi^*\Phi(\mathrm{cov}\xi))$.

Proof. Parts (a)–(d) can be easily obtained from Proposition 1.26. For part (e), we use Proposition 1.26(g):

$$\begin{aligned}\operatorname{tr}(\operatorname{cov}(\xi,\zeta)) &= \operatorname{tr}(\mathbf{E}((\xi-\mathbf{E}\xi)\otimes(\zeta-\mathbf{E}\zeta))) \\ &= \mathbf{E}(\operatorname{tr}(\xi\otimes\zeta)-\operatorname{tr}(\xi\otimes\mathbf{E}\zeta)-\operatorname{tr}(\mathbf{E}\xi\otimes\zeta)+\operatorname{tr}(\mathbf{E}\xi\otimes\mathbf{E}\zeta)) \\ &= \mathbf{E}\langle\xi,\zeta\rangle-\langle\mathbf{E}\xi,\mathbf{E}\zeta\rangle.\end{aligned}$$

Part (f) is a particular case of part (e). Finally, part (g) follows from Proposition 1.26(i). □

4.1.2 Conditional Expectation and Independence

Lemma 4.3. *Let $\xi \in L_1(\Omega, X)$ and let $\mathcal{F}'$ be a sub-σ-algebra of $\mathcal{F}$. Then there exists a unique random variable ζ in $L_1(\Omega, \mathcal{F}', \mathbf{P}, X)$ such that*

$$\forall G \in \mathcal{F}', \ \int_G \xi \, d\mathbf{P} = \int_G \zeta \, d\mathbf{P}.$$

Proof. First, suppose that $\xi \in L_2(\Omega, X)$. Define the functional J on the space $L_2(\Omega, \mathcal{F}', \mathbf{P}, X)$ by

$$J\eta = \int_\Omega \langle \xi, \eta\rangle \, d\mathbf{P}, \ \ \eta \in L_2(\Omega, \mathcal{F}', \mathbf{P}, X).$$

Obviously, J is a bounded linear functional. By Theorem 1.21, there exists a unique $\zeta \in L_2(\Omega, \mathcal{F}', \mathbf{P}, X)$ such that

$$J\eta = \int_\Omega \langle \xi, \eta\rangle \, d\mathbf{P} = \int_\Omega \langle \zeta, \eta\rangle \, d\mathbf{P}.$$

Selecting $\eta = h\chi_G$, where $h \in X$ and $G \in \mathcal{F}'$, we obtain

$$\left\langle h, \int_G (\xi - \zeta) \, d\mathbf{P} \right\rangle = 0.$$

By arbitrariness of h in X, the statement is true for all $\xi \in L_2(\Omega, X)$. If $\xi \in L_1(\Omega, X)$, then the statement can be proved by approximating ξ by the random variables from $L_2(\Omega, X)$. □

The random variable ζ in Lemma 4.3 is called the *conditional expectation* of ξ with respect to $\mathcal{F}'$ and is denoted by $\zeta = \mathbf{E}(\xi|\mathcal{F}')$. For the family $\{\eta_\alpha : \alpha \in A\}$ of random variables, we let

$$\mathbf{E}(\xi|\eta_\alpha; \alpha \in A) = \mathbf{E}(\xi|\sigma(\eta_\alpha; \alpha \in A)).$$

Note that expectation is a particular case of conditional expectation when $\mathcal{F}' = \{\Omega, \varnothing\}$, i.e., $\mathbf{E}\xi = \mathbf{E}(\xi|\{\Omega, \varnothing\})$.

Proposition 4.4. *Let $\mathcal{F}'$ be a sub-σ-algebra of $\mathcal{F}$. Then $\mathbf{E}(\cdot|\mathcal{F}')$ is the projection operator from $L_2(\Omega, X)$ onto its subspace $L_2(\Omega, \mathcal{F}', \mathbf{P}, X)$.*

Proof. By Proposition 2.28(a), $L_2(\Omega, \mathcal{F}', \mathbf{P}, X)$ is a subspace of $L_2(\Omega, X)$. By definition of a projection operator, the equality $\mathbf{E}\langle \xi - \mathbf{E}(\xi|\mathcal{F}'), \eta\rangle = 0$ must be proved for all $\xi \in L_2(\Omega, X)$ and for all $\eta \in L_2(\Omega, \mathcal{F}', \mathbf{P}, X)$. For an $\mathcal{F}'$-simple function η, this equality follows from the definition of conditional expectation. For arbitrary $\eta \in L_2(\Omega, \mathcal{F}', \mathbf{P}, X)$, this equality can be proved by approximating η by $\mathcal{F}'$-simple functions. □

The events $F_\alpha \in \mathcal{F}$, $\alpha \in A$, are said to be *independent* if the equality

$$\mathbf{P}\left(\bigcap_{i=1}^{n} F_{\alpha_i}\right) = \mathbf{P}(F_{\alpha_1}) \cdots \mathbf{P}(F_{\alpha_n})$$

holds for each finite collection $\{F_{\alpha_1}, \dots, F_{\alpha_n}\}$ of them. The sub-σ-algebras $\mathcal{F}_\alpha$, $\alpha \in A$, of $\mathcal{F}$ are said to be *independent* if the events F_α, $\alpha \in A$, are independent for all $F_\alpha \in \mathcal{F}_\alpha$. The *independence* of the families of random variables is defined as independence of the σ-algebras generated by these families.

Note that the sure event Ω and the impossible event $\varnothing$ are independent of each event $A \in \mathcal{F}$ since $\mathbf{P}(\Omega \cap A) = \mathbf{P}(A) = \mathbf{P}(\Omega)\mathbf{P}(A)$ and $\mathbf{P}(A \cap \varnothing) = \mathbf{P}(\varnothing) = 0 = \mathbf{P}(A)\mathbf{P}(\varnothing)$. Hence, a constant random variable is independent of each random variable since the σ-algebra generated by a constant random variable is $\{\Omega, \varnothing\}$.

The concepts of conditional expectation and independence are very important in probability theory. We list some of their properties.

(a) If $\xi, \eta \in L_1(\Omega, X)$, then $\mathbf{E}(\xi + \eta|\mathcal{F}') = \mathbf{E}(\xi|\mathcal{F}') + \mathbf{E}(\eta|\mathcal{F}')$.

(b) If $\xi \in L_1(\Omega, X)$ and $A \in \mathcal{L}(X, Y)$, then $\mathbf{E}(A\xi|\mathcal{F}') = A\mathbf{E}(\xi|\mathcal{F}')$.

(c) If $\xi \in L_1(\Omega, X)$ and $\mathcal{F}''$ is a sub-σ-algebra of $\mathcal{F}'$, then

$$\mathbf{E}(\xi|\mathcal{F}'') = \mathbf{E}(\mathbf{E}(\xi|\mathcal{F}')|\mathcal{F}'') = \mathbf{E}(\mathbf{E}(\xi|\mathcal{F}'')|\mathcal{F}').$$

(d) If $\xi \in L_1(\Omega, \mathcal{F}', \mathbf{P}, X)$, then $\mathbf{E}(\xi|\mathcal{F}') = \xi$.

(e) If $\xi \in L_1(\Omega, X)$ is independent of $\mathcal{F}'$, then $\mathbf{E}(\xi|\mathcal{F}') = \mathbf{E}\xi$.

(f) If $\xi, \eta \in L_2(\Omega, X)$ are independent random variables, then the equality $\mathbf{E}\langle \xi, \eta\rangle = \langle \mathbf{E}\xi, \mathbf{E}\eta\rangle$ holds.

(g) If $\xi \in L_2(\Omega, X)$ and $\eta \in L_2(\Omega, Y)$ are independent random variables, then $\operatorname{cov}(\xi, \eta) = 0$.

(h) If $\xi \in L_2(\Omega, X)$ and $\eta \in L_2(\Omega, Y)$ are independent random variables, $\varphi \in m(X, \mathcal{B}_X, Z)$ and $\psi \in m(Y, \mathcal{B}_Y, H)$, then $\varphi \circ \xi$ and $\psi \circ \eta$ are independent too.

4.1.3 Gaussian Systems

A real-valued random variable ξ is said to be *Gaussian* (or *normal*) if its distribution has the form

$$\mathbf{P}_\xi((a,b]) = \frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}}dx, \quad -\infty < a < b < \infty,$$

where $m \in \mathbb{R}$ and $\sigma^2 > 0$ are parameters. For this random variable, we write $\xi \sim \mathcal{N}(m,\sigma^2)$. A constant random variable, i.e., a random variable $\xi = m = \text{const.}$, is also Gaussian and it is said to be *degenerate*. For this random variable, we write $\xi \sim \mathcal{N}(m,0)$. A Gaussian random variable $\xi \sim \mathcal{N}(m,\sigma^2)$ belongs to $L_2(\Omega,\mathbb{R})$ where the parameters m and σ^2 stand for the expectation and the variance of ξ, respectively, i.e., $m = \mathbf{E}\xi$ and $\sigma^2 = \text{cov}\xi$. If $\xi \sim \mathcal{N}(m_1,\sigma_1^2)$ and $\eta \sim \mathcal{N}(m_2,\sigma_2^2)$ are independent, then we have $a\xi + b\eta \sim \mathcal{N}(am_1 + bm_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ where $a,b \in \mathbb{R}$.

The following theorem explains a wide use of Gaussian random variables.

Theorem 4.5 (Central Limit Theorem). *Suppose $\{\xi_n\}$ is a sequence of independent and equidistributed real-valued random variables with $m = \mathbf{E}\xi_n$ and $\sigma^2 = \text{cov}\,\xi_n$. Let*

$$S_n = \frac{\sum_{i=1}^n \xi_i - nm}{\sigma\sqrt{n}}, \quad n = 1,2,\ldots.$$

Then S_n converges in distribution to $\mathcal{N}(0,1)$-distributed random variable, i.e., for all $-\infty < a < b < \infty$,

$$\mathbf{P}_{S_n}((a,b]) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx, \quad n \to \infty.$$

In practice the real objects are subjected to a large number of small random actions generated by the independent sources. By Theorem 4.5, the sum of these random actions forms approximately a Gaussian random variable. This is a reason for a wide use of Gaussian random variables in applications. On the other hand, the Gaussian random variables have a series of properties, which lead to some convenient mathematical methods to deal with them.

A family $\mathcal{N} = \{\xi_\alpha : \alpha \in A\}$ of real-valued random variables is called a *Gaussian system* if all linear combinations of random variables from $\mathcal{N}$ are Gaussian. We list the basic properties of Gaussian systems.

(a) Any subsystem of a Gaussian system is a Gaussian system.

(b) The union of independent Gaussian systems is a Gaussian system.

(c) The union of a Gaussian system and a family of constant random variables is a Gaussian system.

(d) If $\mathcal{N}$ is a Gaussian system, then the subspace $\overline{\text{span}\,\mathcal{N}}$ of $L_2(\Omega,\mathbb{R})$ is also a Gaussian system.

(e) The independence of the Gaussian systems $\mathcal{N}_1$ and $\mathcal{N}_2$ implies the independence of $\overline{\operatorname{span}\mathcal{N}_1}$ and $\overline{\operatorname{span}\mathcal{N}_2}$.

Note that the union of arbitrary Gaussian systems may not be a Gaussian system (a corresponding counterexample is given in Shiryaev [86]).

With a given X-valued random variable ξ, one can associate the system $\mathcal{N}_\xi = \{\langle \xi, h\rangle : h \in X\}$ of real-valued random variables. If $\mathcal{N}_\xi$ is a Gaussian system, then ξ is called an *X-valued Gaussian* (or *normal*) *random variable.* A family $\{\xi_\alpha : \alpha \in A\}$ of random variables is said to *define a Gaussian system* if $\bigcup_{\alpha\in A}\mathcal{N}_{\xi_\alpha}$ is a Gaussian system.

Theorem 4.6. *Suppose $\xi \in L_2(\Omega, X)$ and $\eta \in L_2(\Omega, Y)$ define a Gaussian system. Then ξ and η are independent if and only if* $\operatorname{cov}(\xi, \eta) = 0$.

Theorem 4.7. *The characteristic function of an X-valued Gaussian random variable ξ has the form*

$$\varphi_\xi(x) = e^{i\langle \mathbf{E}\,\xi, x\rangle - \frac{1}{2}\langle(\operatorname{cov}\xi)x, x\rangle}, \quad x \in X.$$

4.1.4 Random Processes

An X-valued *random process* (or, briefly, a *random process*) is a family of random variables $\xi_t \in m(\Omega, X)$, $t \geq 0$. Here the parameter t is interpreted as time. We will consider the random processes on a finite time interval, say, on $\mathbf{T} = [0, T]$, and suppose that they are $\ell \otimes \mathbf{P}$-measurable, i.e., belong to $m(\mathbf{T} \times \Omega, X)$. For a random process ξ, a function $[\xi_\omega]$ (see Remark 2.21 for this symbol), where $\omega \in \Omega$ is considered as fixed, is called its *path*. The dependence of random processes (and random variables as well) on samples ω will be indicated only in exceptional cases.

Random processes ξ and η with values in the same space are said to be *indistinguishable* if

$$\mathbf{P}\Bigl(\bigcup_{t\in\mathbf{T}}\{\omega : \xi_t \neq \eta_t\}\Bigr) = 0.$$

Obviously, the paths of indistinguishable random processes coincide w.p.1. Actually, indistinguishable random processes are considered to be equal. According to this, we say that a random process has a given property if it is indistinguishable from a random process, all the paths of which have this property.

Often, in theory of random processes the following weaker criterion of equality of random processes is useful. Random processes ξ and η are said to be a *modification* of each other if they take values in the same space and if $\xi_t = \eta_t$ for all $t \in \mathbf{T}$. For comparison, note that the equality in the space $L_1(\Omega, C(\mathbf{T}, X))$ corresponds to indistinguishability of random processes, and in the space $C(\mathbf{T}, L_1(\Omega, X))$ to modification. If two random processes that are modifications of each other have right continuous (left continuous) paths w.p.1, then they are indistinguishable.

A family $\{\mathcal{F}_t\} = \{\mathcal{F}_t : 0 \le t \le T\}$ of sub-σ-algebras of $\mathcal{F}$ is called a *filtration* if $\mathcal{F}_s \subset \mathcal{F}_t$ for all $0 \le s \le t \le T$. A filtration $\{\mathcal{F}_t\}$ is said to be *complete* if for $G \in \mathcal{F}$, $\mathbf{P}(G) = 0$ implies $G \in \mathcal{F}_0$, and is said to be *right continuous* if

$$\mathcal{F}_t = \mathcal{F}_t^+, \ \mathcal{F}_t^+ = \bigcap_{t<s\le T} \mathcal{F}_s, \ 0 \le t \le T.$$

Each filtration can be extended up to the smallest complete and right continuous filtration. Therefore, it is convenient beforehand to suppose that a given filtration is complete and right continuous. We say that a random process $\eta \in m(\mathbf{T} \times \Omega, X)$ is *adapted* with respect to a filtration $\{\mathcal{F}_t\}$ (or $\mathcal{F}_t$*-adapted*) if $\eta_t \in m(\Omega, \mathcal{F}_t, X)$ for all $0 \le t \le T$. Given a random process $\eta \in m(\mathbf{T} \times \Omega, X)$, we denote by $\{\mathcal{F}_t^\eta\}$ the smallest complete and right continuous filtration generated by $\{\sigma(\eta_s; 0 \le s \le t)\}$ and call it the *natural filtration* of η. Obviously, each random process η is adapted with respect to its natural filtration $\{\mathcal{F}_t^\eta\}$.

A random process $m \in m(\mathbf{T} \times \Omega, X)$ together with a filtration $\{\mathcal{F}_t\}$ is called an X-valued *square integrable martingale* on $\mathbf{T}$ (or, briefly, a square integrable martingale) if

(a) $m_t \in L_2(\Omega, \mathcal{F}_t, \mathbf{P}, X)$, $0 \le t \le T$;

(b) $\mathbf{E}(m_t | \mathcal{F}_s) = m_s$, $0 \le s < t \le T$.

For a square integrable martingale m with respect to a filtration $\{\mathcal{F}_t\}$, the notation $\{m_t, \mathcal{F}_t\}$ is used as well. Obviously, in the above definition a square integrable martingale is defined with accuracy up to modification. By $M_2(\mathbf{T}, X)$ or by $M_2(0, T; X)$, we denote the class of all X-valued square integrable martingales $\{m_t, \mathcal{F}_t\}$ on $\mathbf{T}$ for which $m_0 = 0$, $\{\mathcal{F}_t\}$ is a complete and right continuous filtration and m has right continuous paths. Obviously, a martingale from $M_2(\mathbf{T}, X)$ is defined up to indistinguishability. One can observe that if $\{m_t, \mathcal{F}_t\} \in M_2(\mathbf{T}, X)$ for some filtration $\{\mathcal{F}_t\}$, then $\{m_t, \mathcal{F}_t^m\} \in M_2(\mathbf{T}, X)$ too, where $\{\mathcal{F}_t^m\}$ is the natural filtration of m. The subclass of $M_2(\mathbf{T}, X)$ consisting of the martingales with continuous paths is denoted by $M_2^c(\mathbf{T}, X)$.

For $m \in M_2(\mathbf{T}, X)$, we present a particular case of the *Doob inequality* (see, for example, Rozovskii [84], p. 44):

$$\mathbf{E}\left(\sup_{t\in\mathbf{T}} \|m_t\|\right)^2 \le 4\mathbf{E}\|m_T\|^2, \tag{4.1}$$

which will be used in proving the following technical detail.

Proposition 4.8. $M_2^c(\mathbf{T}, X) \subset L_2(\Omega, C(\mathbf{T}, X))$.

Proof. Let $m \in M_2^c(\mathbf{T}, X)$. By Proposition 2.25, the function $m : \Omega \to C(\mathbf{T}, X)$ is $\mathbf{P}$-measurable. Using also (4.1), we have $m \in L_2(\Omega, C(\mathbf{T}, X))$. □

Let $\{m_t, \mathcal{F}_t\} \in M_2(\mathbf{T}, X)$. The sets of the form

$$(s,t] \times F,\ F \in \mathcal{F}_s,\ 0 \le s \le t \le T;\ [0,t] \times F_0,\ F_0 \in \mathcal{F}_0, 0 \le t \le T,$$

are called *predictable rectangles*. The system of all predictable rectangles defines a semialgebra of subsets of $\mathbf{T} \times \Omega$. The smallest σ-algebra over this semialgebra is called the *σ-algebra of predictable subsets* of $\mathbf{T} \times \Omega$ with respect to the filtration $\{\mathcal{F}_t\}$. This σ-algebra will be denoted by $\mathcal{P}$. Define the function λ on predictable rectangles by

$$\lambda((s,t] \times F) = \mathbf{E}\left(\chi_F \|m_t - m_s\|^2\right);\ \lambda([0,t] \times F_0) = \mathbf{E}\left(\chi_{F_0} \|m_t\|^2\right).$$

For $\{m_t, \mathcal{F}_t\} \in M_2(\mathbf{T}, X)$, the function λ is a measure. The extension of λ to $\mathcal{P}$ is called the *Dolean measure* of m.

The following theorem, proved in Metivier [77], p. 141, expresses a significant property of square integrable martingales.

Theorem 4.9. *Let $\{m_t, \mathcal{F}_t\} \in M_2(\mathbf{T}, X)$, let $\mathcal{P}$ be the σ-algebra of predictable subsets of $\mathbf{T} \times \Omega$ with respect to the filtration $\{\mathcal{F}_t\}$ and let λ be the Dolean measure of m on $\mathcal{P}$. Then there exists a unique function $M \in L_\infty(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_1(X))$ such that $M_{t,\omega} \ge 0$, $\mathrm{tr} M_{t,\omega} = 1$ for λ-a.e. $(t,\omega) \in \mathbf{T} \times \Omega$ and*

$$\mathbf{E}\left(\chi_F((m_t - m_s) \otimes (m_t - m_s))\right) = \int_{(s,t] \times F} M \, d\lambda$$

for each predictable rectangle $(s,t] \times F$, $F \in \mathcal{F}_s$, $0 \le s \le t \le T$.

The function M in Theorem 4.9 is called the *covariance function* of m.

A Wiener process which is defined below is a very useful particular case of square integrable martingales.

Definition 4.10. A random process $\{w_t, \mathcal{F}_t\} \in M_2^c(\mathbf{T}, X)$ is called an X-valued *Wiener process* on $\mathbf{T}$ (or, briefly, a Wiener process) if the covariance function M and the Dolean measure λ of w have the form

$$M_{t,\omega} = (\mathrm{tr} W)^{-1} W,\ (t,\omega) \in \mathbf{T} \times \Omega;\ \lambda = (\mathrm{tr} W)(\ell \otimes \mathbf{P}),$$

where $W \in \mathcal{L}_1(X)$, $W \ge 0$ and ℓ is the Lebesgue measure on $\mathbf{T}$. The operator W is called the *covariance operator* of the Wiener process w.

A random process $\xi : \mathbf{T} \times \Omega \to X$ is said to be *Gaussian* if the family $\{\xi_t : 0 \le t \le T\}$ of random variables defines a Gaussian system. A Wiener process is an example of a Gaussian process. This will be considered in greater detail in the next section.

4.2 Stochastic Modelling of Real Processes

An adequate modelling of real processes requires a consideration of all influences on them. These influences can be collected into two separate groups. The first group contains the deterministic influences that are characterized by the property that their magnitude at each instant and at each space point can be determined beforehand. The influences of the second group are called *fluctuations*. Their effects on real processes are random and, therefore, cannot be determined exactly. Often, only the first group of influences is considered and fluctuations are ignored. This leads to deterministic models of real processes such as ordinary and partial differential equations. In this section we discuss how to use fluctuations in modelling of real processes.

4.2.1 Brownian Motion

In 1827 the botanist R. Brown observed that a microscopic particle suspended in liquid makes very strange and highly irregular movements, and he reported the results of this observation in 1928. The motion of this particle was called *Brownian motion*.

Molecular physics explains Brownian motion as a motion of a microscopic particle determined by its collisions with liquid molecules. Since the mass of a microscopic particle is small enough, each of these collisions has an effect on its path.

The state of a particle exhibiting Brownian motion on the surface of a liquid can be expressed by a vector in $\mathbb{R}^2$. Below we consider one-dimensional Brownian motion. It can be imagined by projection of two-dimensional Brownian motion to one of the coordinate axes.

A deterministic analysis of Brownian motion cannot be carried out as this requires full information about molecules of a given liquid at some fixed instant including their position, velocity etc. N. Wiener suggested a stochastic approach to Brownian motion. To explain this approach, consider a microscopic particle suspended in liquid during the time interval $\mathbf{T} = [0, T]$ and let Ω be the collection of all sample cases which lead to the distinct paths of this particle. Then the position of this particle can be modelled as a function $w_{t,\omega}$, $t \in \mathbf{T}$, $\omega \in \Omega$, of two variables. Since the paths of the particle are continuous, one can select $\Omega = C(\mathbf{T}, \mathbb{R})$ and suppose that a sample $\omega \in \Omega = C(\mathbf{T}, \mathbb{R})$ leads to the path $w_{t,\omega} = \omega_t$, $t \in \mathbf{T}$. Obviously, distinct groups of paths may occur with distinct probabilities. So, a posteriori one can suppose that a probability distribution $\mathbf{P}_w$ on Ω is defined so that $\mathbf{P}_w(A)$ shows the occurrence probability of a path of the particle in A, where $A \subset \Omega$. Thus, we conclude that there is a sample space Ω and a probability distribution on Ω for which the motion of the particle may be considered as a random process $w : \mathbf{T} \times \Omega \to \mathbb{R}$.

To study the properties of this random process, at first, we will consider a microscopic particle suspended in liquid satisfying the following conditions.

(a) The liquid is homogeneous and has some fixed viscosity that will be determined below.

(b) The liquid is not disturbed by any outside influence.

The meaning of the second condition is that the motion of the particle on this liquid is a result of the collisions of liquid molecules and the particle only. The random process w has the following properties.

(A) $\mathbf{E}(w_t - w_s) = 0$. This follows from the equality of the occurrence probabilities of the particle at instant t in any two regions on the surface of the liquid, which are symmetric with respect to the position of the particle at instant s, since the conditions (a) and (b) hold.

(B) $\mathbf{E}(w_t - w_s)^2$ depends only on $|t - s|$. This is again a consequence of the conditions (a) and (b). The variance of $w_t - w_s$ is independent of the location of the particle at instant s and, hence, is a function of the time run from s.

Note that in the theory of random processes, the random process w which has the properties (A) and (B) is said to be *stationary in a wide sense.*

(C) $w_t - w_s$ and $w_r - w_\sigma$ are independent for all $\sigma < r \le s < t$. This follows from the fact that the liquid contains a very large number of molecules and, in fact, in disjoint time intervals $(s, t]$ and $(\sigma, r]$ the particle has collisions with distinct molecules. So, the increments $w_t - w_s$ and $w_r - w_\sigma$ are formed by independent collisions.

The property (C) will be revised in Section 4.6.1. In the theory of random processes, the random process w which has the property (C) is called a *process with independent increments*. At this point, we interrupt listing the properties of the process w to prove the following result.

Lemma 4.11. *Let ξ be a stationary in a wide sense random process with independent increments and let $\psi_t = \mathbf{E}(\xi_t - \xi_0)^2$, $0 \le t \le T$. Then $\psi_t = \lambda t$, $0 \le t \le T$, for some $\lambda =$ const. ≥ 0.*

Proof. For $0 \le s < t \le T$, we have

$$\psi_t - \psi_s = \mathbf{E}(\xi_t - \xi_s + \xi_s - \xi_0)^2 - \mathbf{E}(\xi_s - \xi_0)^2 = \mathbf{E}(\xi_t - \xi_s)^2 \ge 0. \tag{4.2}$$

Hence, ψ is a nondecreasing function. Substituting $t = s + r$ in (4.2), we have

$$\psi_{s+r} = \psi_s + \mathbf{E}(\xi_{s+r} - \xi_s)^2 = \psi_s + \mathbf{E}(\xi_r - \xi_0)^2 = \psi_s + \psi_r, \ \psi_0 = 0. \tag{4.3}$$

Any nondecreasing function on $\mathbf{T}$ satisfying (4.3) has the representation $\psi_t = \lambda t$, $0 \le t \le T$, with $\lambda = T^{-1}\psi_T \ge 0$. Indeed, this holds for $t = 0$. Let $0 < t \le T$. Then for a positive integer n with $0 < nt \le T$, we have $\psi_{nt} = n\psi_t$. This implies $\psi_{s/n} = \psi_s/n$ if we substitute $s = nt$. Thus, for any positive rational number a

represented as a ratio n/m of two positive integers such that $0 < at \le T$, we have $\psi_{at} = \psi_{(n/m)t} = (n/m)\psi_t = a\psi_t$. If a is an irrational number satisfying $0 < at \le T$, then we can take increasing and decreasing, respectively, sequences $\{b_n\}$ and $\{c_n\}$ of rational numbers such that $\lim_{n\to\infty} b_n = \lim_{n\to\infty} c_n = a/2$. Since ψ is nondecreasing, we have

$$b_n\psi_t = \psi_{b_n t} \le \psi_{at/2} \le \psi_{c_n t} = c_n\psi_t$$

for large n. Tending n to ∞, we obtain $\psi_{at} = a\psi_t$ for every real number a satisfying $0 < at \le T$. Now $\psi_t = tT^{-1}\psi_T = \lambda t$ implies the statement for $0 < t \le T$. □

Thus, the properties (A)–(C) imply that

$$\mathbf{E}(w_t - w_s)^2 = \lambda(t-s), \ \ \lambda > 0, \ \ 0 \le s \le t \le T.$$

Note that the limit case $\lambda = 0$ implies $w_t \equiv$ const. This holds when the liquid is so thick that it can be considered as solid. Hence, we can suppose that $\lambda > 0$ for the liquid under consideration.

The next property has a connection with the central limit theorem.

(D) $w_t - w_s$ is a Gaussian random variable. This is a consequence of Theorem 4.5, since $w_t - w_s$ can be considered as the sum of a large number of independent and equidistributed small increments. Therefore, $(w_t - w_s) \sim \mathcal{N}(m, \sigma^2)$ where $m = \mathbf{E}(w_t - w_s) = 0$ and $\sigma^2 = \mathbf{E}(w_t - w_s)^2 = \lambda|t-s|$.

(E) w has continuous paths. This needs no discussion.

Finally, to the properties (A)–(E), we add the following normalizing condition.

(F) $w_0 = 0$ and $\lambda = 1$. The first of these equalities means that the origin is selected so that it coincides with the position of the particle at $t = 0$, and the second equality shows that the liquid has the corresponding viscosity.

The random process $w : \mathbf{T} \times \Omega \to \mathbb{R}$ that has the above mentioned properties (A)–(F) is called a *standard process of Brownian motion* on $\mathbf{T}$. The first question that arises is whether a standard process of Brownian motion exists. The following theorem answers this question positively.

Theorem 4.12 (Wiener). *There exists a unique probability measure $\mathbf{P}_w$ on the measurable space $(C(\mathbf{T},\mathbb{R}), \mathcal{B}_C)$ such that the coordinate process $w_{t,\omega} = \omega_t$, $t \in \mathbf{T}$, $\omega \in C(\mathbf{T},\mathbb{R})$, is a standard process of Brownian motion on $\mathbf{T}$.*

In larger probability spaces than $(C(\mathbf{T},\mathbb{R}), \mathcal{B}_C, \mathbf{P}_w)$ it is possible to consider infinitely many independent standard processes of Brownian motion. Below we suppose that $(\Omega, \mathcal{F}, \mathbf{P})$ is sufficiently large to introduce the following definition.

Definition 4.13. A random process

$$w_t = \sum_{n=1}^{\dim X} \sqrt{\mu_n}\, w_t^n e_n,\ 0 \le t \le T, \tag{4.4}$$

where $\{e_n\}$ is a basis in X, $\{w^n\}$ is a sequence (finite or infinite depending on $\dim X$) of independent standard processes of Brownian motion, $\sum_{n=1}^{\dim X} \mu_n < \infty$, $\mu_n \ge 0$ and the convergence in (4.4) is in the norm of $L_2(\Omega, X)$ if $\dim X = \infty$, is called an X-valued *process of Brownian motion.*

4.2.2 Wiener Process Model of Brownian Motion

The following theorem shows that in fact a process of Brownian motion is a simplest martingale.

Theorem 4.14 (Levi). *Each process of Brownian motion is a Wiener process with respect to its natural filtration, and vice versa.*

In addition note that, if the X-valued Wiener process and the X-valued process of Brownian motion from Definitions 4.10 and 4.13, respectively, are identified, then $\{\mu_n\}$ and $\{e_n\}$ stand for the systems of eigenvalues and corresponding eigenvectors of the covariance operator W.

Using Theorem 4.14, the following important property of Wiener processes can be proved.

Proposition 4.15. *Any Wiener process is Gaussian.*

Proof. Let w be an X-valued Wiener process on $\mathbf{T}$ and let $\{e_n\}$ be the system of eigenvectors of the respective covariance operator. For each n, let

$$\mathcal{N}^n = \{\langle w_t, e_n\rangle : 0 \le t \le T\}.$$

Since w has independent increments and

$$\langle \alpha_1 w_{t_1} + \cdots + \alpha_m w_{t_m}, e_n\rangle = \sum_{k=1}^{m}\sum_{j=k}^{m} \alpha_j \langle w_{t_k} - w_{t_{k-1}}, e_n\rangle$$

for $0 = t_0 < t_1 < \cdots < t_m \le T$ and for $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$, we obtain that $\mathcal{N}^n$ is a Gaussian system for all n. By Definition 4.13, $\{\mathcal{N}^n\}$ is a sequence of independent Gaussian systems. Hence,

$$\mathcal{N} = \overline{\operatorname{span} \bigcup_{n=1}^{\dim X} \mathcal{N}^n}$$

is also a Gaussian system. Obviously, $\langle w_t, h\rangle \in \mathcal{N}$ for all $h \in X$ and for all $0 \le t \le T$. This means that w is a Gaussian random process. $\square$

Setting Wiener processes which define a Gaussian system is important in estimation theory. A simple example of such Wiener processes is a pair of independent Wiener processes. Practically, such Wiener processes are generated by distinct sources. More generally, a pair of Wiener processes w^1 and w^2 which are linear transformations of some third Wiener process w, i.e., $w_t^1 = \Phi w_t$ and $w_t^2 = \Psi w_t$, $0 \le t \le T$, where $\Phi \in \mathcal{L}(X,Y)$ and $\Psi \in \mathcal{L}(X,Z)$ and w has its values in X, form a Gaussian system. In this case, $\operatorname{cov}(w_t^1, w_s^2) = \Phi W \Psi^* \min(t,s)$, where W is the covariance operator of w. If $\Phi W \Psi^* = 0$, then w^1 and w^2 are independent. This means that w^1 and w^2 are transformations of independent components of w. Consequently, the source generating w can be separated into two independent sources generating w^1 and w^2, respectively. If $\Phi W \Psi^* \neq 0$, then such separation is impossible. In this case, w^1 and w^2 are said to be *correlated Wiener processes.*

More complicated examples of Wiener processes which define a Gaussian system can be obtained by special transformations on them. Below we present two such transformations which express the invariance of Wiener processes under translations and rotations.

Proposition 4.16. *Let w be an X-valued Wiener process on $\mathbf{T}$ and let*

(a) $w_t^1 = w_{t+\varepsilon} - w_\varepsilon$, $0 \le t \le T - \varepsilon$, *where* $0 < \varepsilon < T$;

(b) $w_t^2 = \frac{1}{\sqrt{c}} w_{ct}$, $0 \le t \le c^{-1}T$, *where* $c > 0$.

Then the random processes $\{w_t^1, \mathcal{F}_t^{w^1}\}$ *and* $\{w_t^2, \mathcal{F}_t^{w^2}\}$ *are Wiener processes on* $[0, T-\varepsilon]$ *and* $[0, c^{-1}T]$, *respectively. Furthermore,* w, w^1 *and* w^2 *have the same covariance operator and define a Gaussian system.*

Proof. This can be proved by direct verification. □

4.2.3 Diffusion Processes

Now we will consider the motion of a suspended microscopic particle in the general case. Let $F(t,x)$ and $g(t,x)$ be the velocity and the viscosity of the liquid under consideration at instant t and at the location $x = (x^1, x^2)$ on its surface. Note that $F(t,x)$ is a two-dimensional vector, but $g(t,x)$ is a scalar and we denote

$$G(t,x) = \begin{bmatrix} g(t,x) & 0 \\ 0 & g(t,x) \end{bmatrix}.$$

Then the displacement

$$\Delta x_t = \begin{bmatrix} \Delta x_t^1 \\ \Delta x_t^2 \end{bmatrix} = \begin{bmatrix} x_{t+\Delta t}^1 - x_t^1 \\ x_{t+\Delta t}^2 - x_t^2 \end{bmatrix}$$

of the suspended particle can be approximately written as

$$\Delta x_t = F(t,x)\Delta t + G(t,x)\Delta w_t,$$

where

$$w_t = \begin{bmatrix} w_t^1 \\ w_t^2 \end{bmatrix}, \quad \Delta w_t = \begin{bmatrix} w_{t+\Delta t}^1 - w_t^1 \\ w_{t+\Delta t}^2 - w_t^2 \end{bmatrix}$$

and w^1 and w^2 are independent standard processes of Brownian motion related with the x^1- and x^2-components of the suspended particle, respectively. Replacing increments by differentials, we obtain the equation

$$dx_t = F(t, x_t)dt + G(t, x_t)dw_t, \tag{4.5}$$

which can be written in the integral form

$$x_t = x_0 + \int_0^t F(s, x_s)\, ds + \int_0^t G(s, x_s)\, dw_s,$$

where the first integral is a Lebesgue integral by considering $\omega \in \Omega$ as fixed, but the second is a stochastic integral which will be studied in the next section. If we suppose that F and G are functions from $\mathbf{T} \times X$ to X and $\mathcal{L}(Z, X)$, respectively, w is a Z-valued Wiener process on $\mathbf{T}$ and $x_0 \in L_2(\Omega, X)$, then we obtain an abstract version of the equation (4.5). The functions F and G in (4.5) are called *drift* and *diffusion coefficients*, respectively. The equation (4.5) is called a *stochastic differential equation*, or an *Ito equation*, or a *diffusion equation*. A solution of the equation (4.5) is called a *diffusion process*. Recall that the diffusion phenomenon is a motion of a suspended microscopic particle in liquid or in gas under collisions with molecules of the medium.

It is known that a solution of the equation (4.5) is an adequate model for many real processes arising in branches which are far from diffusion, for example, in electronics, communication, geophysics, economics, finance etc. Nevertheless, more rigorous criteria imposed by mathematical models of real processes make it necessary to improve the process of modelling Brownian motion. We will return to this problem in Section 4.6.

4.3 Stochastic Integration in Hilbert Spaces

In this section we relate a stochastic integral to a Hilbert space-valued square integrable martingale. We assume always that $\{m_t, \mathcal{F}_t\} \in M_2(\mathbf{T}, X)$, $\mathcal{P}$ is the σ-algebra of the predictable subsets of $\mathbf{T} \times \Omega$ with respect to the filtration $\{\mathcal{F}_t\}$, λ is the Dolean measure of m on $\mathcal{P}$ and M is the covariance function of m (see Section 4.1.4). Also, we suppose that (S, Σ, ν) is a separable measure space. Recall that according to our convention in Section 1.2.6, all measures considered in this book are positive and finite.

4.3.1 Stochastic Integral

Consider the random function $M_t^{1/2}$, $0 \le t \le T$. According to Theorem 4.9 and Proposition 2.37, we have $M^{1/2} \in L_\infty(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_2(X))$. Let $\tilde{\Lambda}_m^2(\mathbf{T}; X, Z)$ be the set of all (in general unbounded) operator-valued functions Φ defined on $\mathbf{T} \times \Omega$ and satisfying $\Phi M^{1/2} \in L_2(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_2(X, Z))$. Consider the equivalence relation in $\tilde{\Lambda}^2_m(\mathbf{T}; X, Z)$ defined by calling $\Phi, \Psi \in \tilde{\Lambda}^2_m(\mathbf{T}; X, Z)$ equivalent if

$$\left\| \Phi_t M_t^{1/2} - \Psi_t M_t^{1/2} \right\|_{\mathcal{L}_2} = 0 \text{ for } \lambda\text{-a.e. } (t, \omega) \in \mathbf{T} \times \Omega.$$

Denote the quotient set of $\tilde{\Lambda}^2_m(\mathbf{T}; X, Z)$ with respect to this equivalence relation by $\Lambda^2_m(\mathbf{T}; X, Z)$.

Theorem 4.17. *$\Lambda^2_m(\mathbf{T}; X, Z)$ is a Hilbert space with the scalar product*

$$\langle \Phi, \Psi \rangle_{\Lambda^2_m} = \int_{\mathbf{T} \times \Omega} \left\langle \Phi_t M_t^{1/2}, \Psi_t M_t^{1/2} \right\rangle_{\mathcal{L}_2} d\lambda. \tag{4.6}$$

Proof. In Metivier [77], pp. 142–143, it is shown that the bilinear form (4.6) on $\tilde{\Lambda}^2_m(\mathbf{T}; X, Z)$ satisfies all the axioms of scalar product except the nondegeneracy axiom. Moreover, it is shown that each Cauchy sequence in $\tilde{\Lambda}^2_m(\mathbf{T}; X, Z)$ is convergent, but in general the limit is not unique because (4.6) does not satisfy the nondegeneracy axiom. Replacing $\tilde{\Lambda}^2_m(\mathbf{T}; X, Z)$ by its quotient set $\Lambda^2_m(\mathbf{T}; X, Z)$, we see that (4.6) on $\Lambda^2_m(\mathbf{T}; X, Z)$ satisfies all the axioms of scalar product (including the nondegeneracy axiom), and, as a consequence, each Cauchy sequence in $\Lambda^2_m(\mathbf{T}; X, Z)$ has a unique limit. So, $\Lambda^2_m(\mathbf{T}; X, Z)$ is a Hilbert space. □

Let $\mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$ be the class of functions of the form

$$\Psi = \sum_{i=1}^{k} \chi_{(s_i, t_i] \times F_i} \Psi^i : \mathbf{T} \times \Omega \to \mathcal{L}(X, Z), \tag{4.7}$$

where $k \in \mathbb{N}$, $\Psi^i \in \mathcal{L}(X, Z)$, $(s_i, t_i] \times F_i$ are disjoint predictable rectangles for $i = 1, \ldots, k$.

Theorem 4.18. *$\mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$ and $L_2(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_2(X, Z))$ are dense subsets of $\Lambda^2_m(\mathbf{T}; X, Z)$.*

Proof. Suppose that $\Phi \in \Lambda^2_m(\mathbf{T}; X, Z)$. Let

$$\Phi_t^n = \Phi_t M_t^{1/2} \left(I n^{-1} + M_t^{1/2} \right)^{-1}, \ t \in \mathbf{T}, \ n = 1, 2, \ldots.$$

In Rozovskii [84], pp. 56–57, it is shown that $\Phi^n \in L_2(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_2(X, Z))$ for all n and $\|\Phi^n - \Phi\|_{\Lambda^2_m} \to 0$ as $n \to \infty$. On the other hand in Metivier [77], pp. 143–145, it is proved that each function in $L_2(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_2(X, Z))$ can be approximated in the norm of $\Lambda^2_m(\mathbf{T}; X, Z)$ by functions from $\mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$. □

For the function Ψ defined by (4.7), the stochastic integral is defined as

$$\int_{\mathbf{T}} \Psi_t \, dm_t = \sum_{i=1}^{k} \chi_{F_i} \Psi^i (m_{t_i} - m_{s_i}). \tag{4.8}$$

Theorem 4.19. *The formula* (4.8) *defines a bounded linear operator from the subset* $\mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$ *of the space* $\Lambda_m^2(\mathbf{T}; X, Z)$ *to* $L_2(\Omega, Z)$ *with the properties*

$$\mathbf{E}\int_{\mathbf{T}} \Psi_t \, dm_t = 0 \text{ and } \mathbf{E}\left\| \int_{\mathbf{T}} \Psi_t \, dm_t \right\|_Z^2 = \|\Psi\|_{\Lambda_m^2}^2, \tag{4.9}$$

which has a unique extension to $\Lambda_m^2(\mathbf{T}; X, Z)$ *as a bounded linear operator preserving the properties in* (4.9).

Proof. For $\Psi \in \mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$, the properties in (4.9) can be easily proved. The second of these properties implies the boundedness of the stochastic integral (4.8) on $\mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$. Since, by Theorem 4.18, $\mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$ is dense in $\Lambda_m^2(\mathbf{T}; X, Z)$, the stochastic integral (4.8) has a unique extension to the space $\Lambda_m^2(\mathbf{T}; X, Z)$ as a bounded linear operator preserving the properties in (4.9). □

Theorem 4.20. *Suppose that* $\Phi \in \Lambda_m^2(\mathbf{T}; X, Z)$ *and* $A \in \mathcal{L}(Z, Y)$. *Then* $A\Phi \in \Lambda_m^2(\mathbf{T}; X, Y)$ *and*

$$A \int_{\mathbf{T}} \Phi_t \, dm_t = \int_{\mathbf{T}} A\Phi_t \, dm_t.$$

Proof. This is a consequence of the linearity and the boundedness of stochastic integrals. □

4.3.2 Martingale Property

The stochastic integral of $\Phi \in \Lambda_m^2(\mathbf{T}; X, Z)$ on $(a, b] \subset \mathbf{T}$ is defined by

$$\int_a^b \Phi_t \, dm_t = \int_{\mathbf{T}} \chi_{(a,b]}(t) \Phi_t \, dm_t, \quad 0 \le a \le b \le T.$$

The next theorem states a basic property of stochastic integrals.

Theorem 4.21. *Let* $\Phi \in \Lambda_m^2(\mathbf{T}; X, Z)$ *and let*

$$n_t = \int_0^t \Phi_s \, dm_s, \quad 0 \le t \le T. \tag{4.10}$$

Then

(a) $\{m_t, \mathcal{F}_t\} \in M_2(\mathbf{T}, X) \Rightarrow \{n_t, \mathcal{F}_t\} \in M_2(\mathbf{T}, Z)$;

(b) $\{m_t, \mathcal{F}_t\} \in M_2^c(\mathbf{T}, X) \Rightarrow \{n_t, \mathcal{F}_t\} \in M_2^c(\mathbf{T}, Z)$.

Proof. First, note that (4.10) defines the random process n up to modification. At the same time the random processes in $M_2(\mathbf{T}, Z)$ and $M_2^c(\mathbf{T}, Z)$ are defined up to indistinguishability. Therefore, the statement of this theorem should be interpreted as the existence of a right continuous (continuous) modification of n in $M_2(\mathbf{T}, Z)$ (in $M_2^c(\mathbf{T}, Z)$). This theorem is obvious for $\Phi \in \mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$. Let

$\Phi \in \Lambda_m^2(\mathbf{T}; X, Z)$. By definition of stochastic integral, there exists a sequence $\{\Phi^k\}$ in $\mathcal{A}(\mathbf{T}, \mathcal{L}(X, Z))$ such that for all $0 \le t \le T$,

$$\left\|\Phi^k - \Phi\right\|_{\Lambda_m^2} \to 0 \text{ and } \mathbf{E}\left\|n_t^k - n_t\right\|^2 \to 0 \text{ as } k \to \infty, \tag{4.11}$$

where

$$n_t^k = \int_0^t \Phi_s^k \, dm_s, \ 0 \le t \le T, \ k = 1, 2, \ldots.$$

Using (4.11), it is easy to show that n is a square integrable martingale. Let us prove that it has a right continuous (continuous if $\{m_t, \mathcal{F}_t\} \in M_2^c(\mathbf{T}, X)$) modification. For this, we will show that a stronger convergence than (4.11) takes place. Indeed, using (4.1) and (4.9), we have

$$\mathbf{E}\left(\sup_{t \in \mathbf{T}}\left\|n_t^k - n_t^l\right\|\right)^2 \le 4\mathbf{E}\left\|n_T^k - n_T^l\right\|^2 = 4\left\|\Phi^k - \Phi^l\right\|_{\Lambda_m^2}^2 \to 0, \ k, l \to \infty.$$

This means that for some subsequence (which for simplicity will be identified with the original one) of $\{n^k\}$, there exists $\tilde{\Omega} \subset \Omega$ with $\mathbf{P}(\tilde{\Omega}) = 0$ such that

$$\sup_{t \in \mathbf{T}}\left\|n_{t,\omega}^k - n_{t,\omega}^l\right\| \to 0 \text{ as } k, l \to \infty \text{ for all } \omega \in \Omega \setminus \tilde{\Omega}.$$

So, there is a random variable $\tilde{n}$ such that

$$\sup_{t \in \mathbf{T}}\left\|n_{t,\omega}^k - \tilde{n}_{t,\omega}\right\| \to 0 \text{ as } k \to \infty \text{ for all } \omega \in \Omega \setminus \tilde{\Omega}. \tag{4.12}$$

The uniform convergence in t in (4.12) implies that $\tilde{n}$ has right continuous (continuous if $\{m_t, \mathcal{F}_t\} \in M_2^c(\mathbf{T}, X)$) paths. Finally, comparing (4.11) and (4.12), we conclude that $\tilde{n}$ is a needed modification of n. □

In the sequel, under the process n defined by (4.10) we will always mean its right continuous (continuous if $\{m_t, \mathcal{F}_t\} \in M_2^c(\mathbf{T}, X)$) modification.

4.3.3 Fubini's Property

Let (S, Σ, ν) be a (positive, finite and) separable measure space. The Hilbert spaces $\Lambda_m^2(\mathbf{T}; X, L_2(S, \nu, Z))$ and $L_2(S, \nu, \Lambda_m^2(\mathbf{T}; X, Z))$ will be considered in this section. Note that since the measure space (S, Σ, ν) is separable and $Z \in \mathcal{H}$, we have $L_2(S, \nu, Z) \in \mathcal{H}$. Hence, $\Lambda_m^2(\mathbf{T}; X, L_2(S, \nu, Z))$ is well defined.

Lemma 4.22. *The spaces $\Lambda_m^2(\mathbf{T}; X, L_2(S, \nu, Z))$ and $L_2(S, \nu, \Lambda_m^2(\mathbf{T}; X, Z))$ are isometric under the isometry*

$$\Lambda_m^2(\mathbf{T}; X, L_2(S, \nu, Z)) \ni \Phi \leftrightarrow J\Phi = \Phi \in L_2\big(S, \nu, \Lambda_m^2(\mathbf{T}; X, Z)\big).$$

Proof. By Theorem 4.18, $L_2(\mathbf{T}\times\Omega,\lambda,\mathcal{L}_2(X,Z))$ is dense in $\Lambda^2_m(\mathbf{T};X,Z)$. Hence, by Proposition 2.20, the space $L_2(S,\nu,L_2(\mathbf{T}\times\Omega,\lambda,\mathcal{L}_2(X,Z)))$ is dense in the space $L_2(S,\nu,\Lambda^2_m(\mathbf{T};X,Z))$. On the other hand, by Propositions 2.24 and 2.39(a),

$$L_2(S,\nu,L_2(\mathbf{T}\times\Omega,\lambda,\mathcal{L}_2(X,Z)))=L_2(\mathbf{T}\times\Omega,\lambda,\mathcal{L}_2(X,L_2(S,\nu,Z))).$$

So, $L_2(\mathbf{T}\times\Omega,\lambda,\mathcal{L}_2(X,L_2(S,\nu,Z)))$ is dense in $L_2(S,\nu,\Lambda^2_m(\mathbf{T};X,Z))$. Also, the space $L_2(\mathbf{T}\times\Omega,\lambda,\mathcal{L}_2(X,L_2(S,\nu,Z)))$ is dense in $\Lambda^2_m(\mathbf{T};X,L_2(S,\nu,Z))$ by Theorem 4.18. Thus, to prove this lemma, it is sufficient to show that for any function $\Phi\in L_2(\mathbf{T}\times\Omega,\lambda,\mathcal{L}_2(X,L_2(S,\nu,Z)))$, its norms in the spaces $\Lambda^2_m(\mathbf{T};X,L_2(S,\nu,Z))$ and $L_2(S,\nu,\Lambda^2_m(\mathbf{T};X,Z))$ are same. This follows from

$$\begin{aligned}
\|\Phi\|^2_{\Lambda^2_m(\mathbf{T};X,L_2(S,\nu,Z))} &= \int_{\mathbf{T}\times\Omega}\left\|[\Phi_{t,\omega}]M_t^{1/2}\right\|^2_{\mathcal{L}_2(X,L_2(S,\nu,Z))}d\lambda\\
&= \int_{\mathbf{T}\times\Omega}\left\|[\Phi_{t,\omega}]M_t^{1/2}\right\|^2_{L_2(S,\nu,\mathcal{L}_2(X,Z))}d\lambda\\
&= \int_{\mathbf{T}\times\Omega}\int_S\left\|\Phi_{s,t,\omega}M_t^{1/2}\right\|^2_{\mathcal{L}_2(X,Z)}d\nu\,d\lambda\\
&= \int_S\int_{\mathbf{T}\times\Omega}\left\|\Phi_{s,t,\omega}M_t^{1/2}\right\|^2_{\mathcal{L}_2(X,Z)}d\lambda\,d\nu\\
&= \|\Phi\|^2_{L_2(S,\nu,\Lambda^2_m(\mathbf{T};X,Z))}.
\end{aligned}$$

Thus, the proof is completed. □

The spaces

$$\Lambda^2_m(\mathbf{T};X,L_2(S,\nu,Z))\quad\text{and}\quad L_2(S,\nu,\Lambda^2_m(\mathbf{T};X,Z))$$

will be identified and will be denoted by $\Lambda^2_{\nu,m}(S,\mathbf{T};X,Z)$. Thus, for any function Φ in $\Lambda^2_{\nu,m}(S,\mathbf{T};X,Z)$, the following repeated stochastic integrals can be defined:

$$\int_{\mathbf{T}}\int_S\Phi_{s,t}\,d\nu\,dm_t\ \text{and}\ \int_S\int_{\mathbf{T}}\Phi_{s,t}\,dm_t\,d\nu. \tag{4.13}$$

Theorem 4.23. *For $\Phi\in\Lambda^2_{\nu,m}(S,\mathbf{T};X,Z)$, the repeated stochastic integrals in* (4.13) *are equal w.p.1.*

Proof. It is easy to prove that, for $\Phi\in\mathcal{A}(\mathbf{T},\mathcal{L}(X,L_2(S,\nu,Z)))$, the repeated stochastic integrals in (4.13) are equal and

$$\mathbf{E}\left\|\int_{\mathbf{T}}\int_S\Phi_{s,t}\,d\nu\,dm_t\right\|^2=\mathbf{E}\left\|\int_S\int_{\mathbf{T}}\Phi_{s,t}\,dm_t\,d\nu\right\|^2\le\nu(S)\|\Phi\|^2_{\Lambda^2_{\nu,m}}.$$

Hence, since $\mathcal{A}(\mathbf{T},\mathcal{L}(X,L_2(S,\nu,Z)))$ is dense in $\Lambda^2_{\nu,m}(S,\mathbf{T};X,Z)$, the repeated stochastic integrals in (4.13) define the same bounded linear operator from the set $\mathcal{A}(\mathbf{T},\mathcal{L}(X,L_2(S,\nu,Z)))$ to $L_2(\Omega,Z)$ which can be uniquely extended to the space $\Lambda^2_{\nu,m}(S,\mathbf{T};X,Z)$ as a bounded linear operator. We obtain that the repeated stochastic integrals in (4.13) are equal w.p.1 for all $\Phi\in\Lambda^2_{\nu,m}(S,\mathbf{T};X,Z)$. □

Proposition 4.24. *The following statements hold.*

(a) $B(\mathbf{T}, \mathcal{L}(X,Z)) \subset \Lambda^2_m(\mathbf{T}; X, Z)$.

(b) $B(S \times \mathbf{T}, \mathcal{L}(X,Z)) \subset \Lambda^2_{\ell,m}(S, \mathbf{T}; X, Z)$ *where* $S = [a,b]$.

(c) *If* $\Phi \in B(\Delta_T, \mathcal{L}(X,Z))$ *and*

$$\varphi_t = \int_0^t \Phi_{t,s}\, dm_s,\ \ 0 \le t \le T,$$

then $\varphi \in L_\infty(\mathbf{T}, L_2(\Omega, Z))$.

Proof. Consider predictable rectangles of the form

$$(s,t] \times \Omega,\ \ (s,t] \times \varnothing,\ \ 0 \le s \le t \le T;\ \ [0,t] \times \Omega,\ \ [0,t] \times \varnothing,\ \ 0 \le t \le T.$$

The smallest σ-algebra generated by these predictable rectangles coincides with $\mathcal{B}_\mathbf{T} \otimes \{\Omega, \varnothing\}$. Therefore, $\mathcal{B}_\mathbf{T} \otimes \{\Omega, \varnothing\} \subset \mathcal{P}$ and a $\mathcal{B}_\mathbf{T}$-measurable nonrandom function can be considered as a $\mathcal{P}$-measurable function. We see that each $\Phi \in B(\mathbf{T}, \mathcal{L}(X,Z))$ is strongly $\mathcal{P}$-measurable and bounded. By Proposition 2.13(a), $\Phi M^{1/2} \in L_\infty(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_2(X,Z))$ since $M^{1/2} \in L_\infty(\mathbf{T} \times \Omega, \lambda, \mathcal{L}_2(X))$. Therefore, $B(\mathbf{T}, \mathcal{L}(X,Z)) \subset \Lambda^2_m(\mathbf{T}; X, Z)$. Part (a) is proved. Part (b) can be proved in a similar way. To prove part (c), denote

$$\tilde{\Phi}_{t,s} = \begin{cases} \Phi_{t,s}, & 0 \le s \le t \le T, \\ 0, & 0 \le t < s \le T. \end{cases}$$

Obviously, $\tilde{\Phi} \in B(\mathbf{T} \times \mathbf{T}, \mathcal{L}(X,Z))$. By part (b), $\tilde{\Phi} \in \Lambda^2_{\ell,m}(\mathbf{T}, \mathbf{T}; X, Z)$, and, hence, $\tilde{\Phi} \in \Lambda^2_m(\mathbf{T}; X, L_2(\mathbf{T}, Z))$. Therefore, the function

$$\varphi_t = \int_0^t \Phi_{t,s}\, dm_s = \int_0^T \tilde{\Phi}_{t,s}\, dm_s,\ \ 0 \le t \le T,$$

belongs to $L_2(\Omega, L_2(\mathbf{T}, Z)) = L_2(\mathbf{T}, L_2(\Omega, Z))$. Finally, from the boundedness of Φ, we obtain $\varphi \in L_\infty(\mathbf{T}, L_2(\Omega, Z))$. □

4.3.4 Stochastic Integration with Respect to Wiener Processes

Now let $\{w_t, \mathcal{F}_t\}$ be an X-valued Wiener process on $\mathbf{T}$ and let W be the covariance operator of w. By use of the Hilbert space $\mathcal{L}_W(X,Z)$ (see Section 1.3.7), we have $\Lambda^2_w(\mathbf{T}; X, Z) = L_2(\mathbf{T} \times \Omega, \mathcal{P}, \ell \otimes \mathbf{P}, \mathcal{L}_W(X,Z))$ which is a subspace of $L_2(\mathbf{T} \times \Omega, \mathcal{L}_W(X,Z))$ consisting of all $\mathcal{F}_t$-adapted processes. Using the Hilbertian summation (see Section 2.4.3), one can write

$$\Lambda^2_w(\mathbf{T}; X, Z) = \int_0^T L_2(\Omega, \mathcal{F}_t, \mathbf{P}, \mathcal{L}_W(X,Z))\, dt.$$

Proposition 4.25. *Let $\{w_t, \mathcal{F}_t\}$ be an X-valued Wiener process on $\mathbf{T}$. Then*

(a) $B_2(\mathbf{T}, \mathcal{L}(X,Z)) \subset \Lambda^2_w(\mathbf{T}; X, Z)$;

(b) $B_2(S \times \mathbf{T}, \mathcal{L}(X,Z)) \subset \Lambda^2_{\ell,w}(S, \mathbf{T}; X, Z)$ *where* $S = [a,b]$.

Proof. In proving Proposition 4.24(a), it was shown that $\mathcal{B}_\mathbf{T} \otimes \{\Omega, \varnothing\} \subset \mathcal{P}$. Since, for a Wiener process, the Dolean measure is $\lambda = c(\ell \otimes \mathbf{P})$ where $c = \text{const.} \geq 0$, then the Lebesgue extension of $\mathcal{B}_\mathbf{T} \otimes \{\Omega, \varnothing\}$ with respect to $\lambda = c(\ell \otimes \mathbf{P})$ is equal to $\tilde{\mathcal{B}}_\mathbf{T} \otimes \{\Omega, \varnothing\}$ where $\tilde{\mathcal{B}}_\mathbf{T}$ is the Lebesgue extension of $\mathcal{B}_\mathbf{T}$ with respect to ℓ. Hence, a Lebesgue measurable nonrandom function on $\mathbf{T}$ can be considered as a λ-measurable function on $\mathbf{T} \times \Omega$. So, in a similar way as in the proof of Proposition 4.24(a), we have

$$B_2(\mathbf{T}, \mathcal{L}(X,Z)) \subset \Lambda^2_w(\mathbf{T}; X, Z).$$

Part (a) is proved. Part (b) can be proved in a similar way. □

In the sequel we will use stochastic integrals of nonrandom functions with respect to Wiener processes. Below we list some of their useful properties.

Proposition 4.26. *Let w and v be X- and Y-valued Wiener processes on $\mathbf{T}$ with* $\operatorname{cov} w_t = Wt$ *and with* $\operatorname{cov}(w_t, v_t) = Rt$, *let* $\Phi \in B_2(\mathbf{T}, \mathcal{L}(X,Z))$ *and let* $\Psi \in B_2(\mathbf{T}, \mathcal{L}(Y,U))$. *Then*

(a) $\displaystyle\int_0^t \Phi_s\, dw_s$ *is* $\mathcal{F}^w_t$ *– measurable for all* $0 \leq t \leq T$;

(b) $\displaystyle\int_t^T \Phi_s\, dw_s$ *is independent of* $\mathcal{F}^w_t$ *for all* $0 \leq t \leq T$;

(c) *if* $R = 0$, *then* $\displaystyle\int_0^T \Phi_s\, dw_s$ *and* $\displaystyle\int_0^T \Psi_s\, dv_s$ *are independent*;

(d) $\left\{ \displaystyle\int_0^T \Phi_s\, dw_s : \Phi \in B_2(\mathbf{T}, \mathcal{L}(X,Z)),\ Z \in \mathcal{H} \right\}$ *is a Gaussian system*;

(e) $\operatorname{cov}\left(\displaystyle\int_0^t \Phi_r\, dw_r, \int_s^T \Psi_r\, dv_r \right) = \int_s^t \Phi_r R \Psi_r^*\, dr,\ 0 \leq s \leq t \leq T$;

(f) $\operatorname{tr}\left(\operatorname{cov} \displaystyle\int_0^T \Phi_r\, dw_r \right) = \operatorname{tr} \int_0^T W \Phi_r^* \Phi_r\, dr = \operatorname{tr} \int_0^T \Phi_r^* \Phi_r W\, dr.$

Proof. Part (a) follows from Theorem 4.21. Part (b) is a consequence of independence of the increments of w. Part (c) is obvious. To prove part (d), consider the Gaussian system $\mathcal{N}$ defined in proving Proposition 4.15. Obviously, for $\Phi \in B_2(\mathbf{T}, \mathcal{L}(X,Z))$ and for $h \in Z$,

$$\left\langle \int_0^T \Phi_s dw_s, h \right\rangle \in \mathcal{N}.$$

This proves part (d). Parts (e) and (f) follow from Propositions 4.2(d) and 4.2(g), respectively. □

4.4 Partially Observable Linear Systems

4.4.1 Solution Concepts

Consider the linear stochastic differential equation

$$dx_t = (Ax_t + \varphi_t)dt + dm_t,\ 0 < t \le T, \tag{4.14}$$

where we assume that $A \in \tilde{\mathcal{L}}(X)$, $\varphi : \mathbf{T} \times \Omega \to X$, $m \in M_2(\mathbf{T}, X)$. A random process $x : \mathbf{T} \to L_2(\Omega, X)$ will be called a *solution* (*in ordinary sense*) of the equation (4.14) for given $x_0 \in L_2(\Omega, X)$ if $x_t \in D(A)$ w.p.1 for all $0 \le t \le T$ and if x satisfies the integral equation

$$x_t = x_0 + \int_0^t (Ax_s + \varphi_s)\, ds + m_t,\ 0 \le t \le T. \tag{4.15}$$

Two solutions of the equation (4.14) will be said to be *equal* if they are modifications of each other.

We recall that by Theorem 1.22, with a given closed linear operator A on X, one can associate the separable Hilbert space $D(A)$ with the scalar product

$$\langle x, y\rangle_{D(A)} = \langle x, y\rangle_X + \langle Ax, Ay\rangle_X$$

such that $A \in \mathcal{L}(D(A), X)$ and $D(A) \subset X$ is a natural and tight embedding.

Theorem 4.27. *Suppose that $A \in \tilde{\mathcal{L}}(X)$ generates a strongly continuous semigroup $\mathcal{U} \in \mathcal{S}(X)$, $x_0 \in L_2(\Omega, D(A))$, $\varphi \in L_1(\mathbf{T}, L_2(\Omega, D(A)))$ and $m \in M_2(\mathbf{T}, D(A))$. Then the random process*

$$x_t = \mathcal{U}_t x_0 + \int_0^t \mathcal{U}_{t-s}\varphi_s\, ds + \int_0^t \mathcal{U}_{t-s}\, dm_s,\ 0 \le t \le T, \tag{4.16}$$

is a unique solution of the equation (4.14).

Proof. First note that by Proposition 2.20, the conditions of this theorem on x_0, φ and m imply $x_0 \in L_2(\Omega, X)$, $\varphi \in L_1(\mathbf{T}, L_2(\Omega, X))$ and $m \in M_2(\mathbf{T}, X)$. So, the integrals in (4.16) are Bochner and stochastic integrals of X- and $\mathcal{L}(X)$-valued functions, respectively. We will show that these integrals can be interpreted in a stronger sense. Since $\mathcal{U}$ is a strongly continuous semigroup, $\mathcal{U} \in B(\mathbf{T}, \mathcal{L}(X))$. Let us show that $\mathcal{U} \in B(\mathbf{T}, \mathcal{L}(D(A)))$. Since A is the infinitesimal generator of $\mathcal{U}$, we have $\mathcal{U}_t(D(A)) \subset D(A)$ for all $0 \le t \le T$. Let $0 \le t \le T$ and let $h \in D(A)$. We have

$$\begin{aligned}\|\mathcal{U}_t h\|^2_{D(A)} &= \|A\mathcal{U}_t h\|^2_X + \|\mathcal{U}_t h\|^2_X = \|\mathcal{U}_t Ah\|^2_X + \|\mathcal{U}_t h\|^2_X \\ &\le \|\mathcal{U}_t\|^2_{\mathcal{L}(X)}\left(\|Ah\|^2_X + \|h\|^2_X\right) = \|\mathcal{U}_t\|^2_{\mathcal{L}(X)}\|h\|^2_{D(A)}.\end{aligned}$$

Therefore, $\mathcal{U}_t \in \mathcal{L}(D(A))$, $0 \le t \le T$, and

$$\sup_{t\in\mathbf{T}} \|\mathcal{U}_t\|_{\mathcal{L}(D(A))} \le \sup_{t\in\mathbf{T}} \|\mathcal{U}_t\|_{\mathcal{L}(X)} < \infty.$$

Since $D(A) \subset X$ is a natural embedding of separable Hilbert spaces, by Proposition 1.15(b), we have $\mathcal{B}_{D(A)} \subset \mathcal{B}_X$. Hence, the strong $\mathcal{B}_{\mathbf{T}}$-measurability of the function $\mathcal{U} : \mathbf{T} \to \mathcal{L}(X)$ implies the strong $\mathcal{B}_{\mathbf{T}}$-measurability of the function $\mathcal{U} : \mathbf{T} \to \mathcal{L}(D(A))$. Thus, $\mathcal{U} \in B(\mathbf{T}, \mathcal{L}(D(A)))$. So, the Bochner and stochastic integrals in (4.16) can be interpreted as integrals of $D(A)$- and $\mathcal{L}(D(A))$-valued functions, respectively. By Proposition 4.24(c), we obtain that the formula (4.16) defines the random process x from $\mathbf{T}$ to $L_2(\Omega, D(A))$. Let us show that this random process satisfies the equation (4.15). Fix $0 < t \le T$ and substitute (4.16) in the right-hand side of (4.15). Applying Theorems 2.18, 2.22, 4.20, 4.23 and Proposition 4.24(b) and using (3.2), we have

$$\begin{aligned}
x_0 &+ \int_0^t \left(A \left(\mathcal{U}_s x_0 + \int_0^s \mathcal{U}_{s-r}\varphi_r\, dr + \int_0^s \mathcal{U}_{s-r}\, dm_r \right) + \varphi_s \right) ds + m_t \\
&= \mathcal{U}_t x_0 + \int_0^t \varphi_r\, dr + \int_0^t \int_0^s A\mathcal{U}_{s-r}\varphi_r\, dr\, ds \\
&\quad + \int_0^t dm_r + \int_0^t \int_0^s A\mathcal{U}_{s-r}\, dm_r\, ds \\
&= \mathcal{U}_t x_0 + \int_0^t \left(\varphi_r + \int_r^t A\mathcal{U}_{s-r}\varphi_r\, ds \right) dr + \int_0^t \left(I + \int_r^t A\mathcal{U}_{s-r}\, ds \right) dm_r \\
&= \mathcal{U}_t x_0 + \int_0^t \mathcal{U}_{t-r}\varphi_r\, dr + \int_0^t \mathcal{U}_{t-r}\, dm_r = x_t.
\end{aligned}$$

Therefore, x in (4.16) is a solution of the equation (4.15). If y is also a solution of the equation (4.15), then for $z_t = x_t - y_t$, $0 \le t \le T$, by (3.2), we have

$$\begin{aligned}
z_t &= \int_0^t Az_s\, ds = \int_0^t A \left(\mathcal{U}_{t-s} z_s - \int_0^{t-s} \mathcal{U}_r A z_s\, dr \right) ds \\
&= \int_0^t A\mathcal{U}_{t-s} z_s\, ds - \int_0^t A\mathcal{U}_r \int_0^{t-r} Az_s\, ds\, dr \\
&= \int_0^t A\mathcal{U}_{t-r} z_r\, dr - \int_0^t A\mathcal{U}_r z_{t-r}\, dr = 0.
\end{aligned}$$

Therefore, (4.16) is a unique solution of the equation (4.14). □

Now consider the random process (4.16) for $\mathcal{U} \in \mathcal{S}(X)$ and for x_0, φ and m satisfying the following conditions which are weaker than in Theorem 4.27:

$$x_0 \in L_2(\Omega, X), \;\; \varphi \in L_1(\mathbf{T}, L_2(\Omega, X)), \;\; m \in M_2(\mathbf{T}, X). \tag{4.17}$$

Obviously, x is still well defined by (4.16), while it may not satisfy the equation (4.15) since in general $x_t \notin D(A)$. The process x, defined by (4.16) under the above mentioned weaker conditions, is called a *mild solution* of the equation (4.14). In particular, when the conditions of Theorem 4.27 hold, a mild solution becomes a solution in the ordinary sense.

4.4.2 Linear Stochastic Evolution Systems

Developing the concept of mild solution, consider the random process

$$x_t = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\varphi_s\, ds + \int_0^t \mathcal{U}_{t,s}\, dm_s,\ 0 \le t \le T, \tag{4.18}$$

where $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$ and (4.17) holds. Generally, (4.18) has no relation with a differential equation since a mild evolution operator does not have any generator. But, (4.18) could be considered as a generalization of (4.16) to the case of mild evolution operators. The system (4.18) under the weaker condition $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$ will be called a *linear stochastic evolution system.*

Some useful properties of linear stochastic evolution systems are given below.

Proposition 4.28. *Let $x_0 \in L_2(\Omega, X)$, let $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, let $N \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, let $\varphi \in L_1(\mathbf{T}, L_2(\Omega, X))$ and let $m \in M_2(\mathbf{T}, X)$. Then the equation*

$$x_t = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,r}(N_r x_r + \varphi_r)\, dr + \int_0^t \mathcal{U}_{t,r}\, dm_r,\ 0 \le t \le T, \tag{4.19}$$

has

$$x_t = \mathcal{Y}_{t,0}x_0 + \int_0^t \mathcal{Y}_{t,r}\varphi_r\, dr + \int_0^t \mathcal{Y}_{t,r}\, dm_r,\ 0 \le t \le T, \tag{4.20}$$

as its solution, where $\mathcal{Y} = \mathcal{P}_N(\mathcal{U})$. This solution is unique up to modification and belongs to $L_\infty(\mathbf{T}, L_2(\Omega, X))$.

Proof. By Propositions 3.17(b) and 4.24(c), the formula (4.20) defines a function x in $L_\infty(\mathbf{T}, L_2(\Omega, X))$. Substituting (4.20) in the right-hand side of (4.19), we obtain

$$\begin{aligned}\mathcal{U}_{t,0}x_0 &+ \int_0^t \mathcal{U}_{t,r}N_r\mathcal{Y}_{r,0}x_0\, dr \\ &+ \int_0^t \mathcal{U}_{t,r}\varphi_r\, dr + \int_0^t \mathcal{U}_{t,s}N_s\int_0^s \mathcal{Y}_{s,r}\varphi_r\, dr\, ds \\ &+ \int_0^t \mathcal{U}_{t,r}\, dm_r + \int_0^t \mathcal{U}_{t,s}N_s\int_0^s \mathcal{Y}_{s,r}\, dm_r\, ds.\end{aligned}$$

By Theorems 2.22 and 4.23, the last expression is equal to

$$\mathcal{Y}_{t,0}x_0 + \int_0^t \left(\mathcal{U}_{t,r} + \int_r^t \mathcal{U}_{t,s}N_s\mathcal{Y}_{s,r}\,ds\right)\varphi_r\,dr + \int_0^t \left(\mathcal{U}_{t,r} + \int_r^t \mathcal{U}_{t,s}N_s\mathcal{Y}_{s,r}\,ds\right)dm_r.$$

Finally, by $\mathcal{Y} = \mathcal{P}_N(\mathcal{U})$ and by (4.20), we obtain that the substitution of x_t from (4.20) into the right-hand side of (4.19) gives the left-hand side of (4.19). Thus, the function x_t defined by (4.20) is a solution of the equation (4.19). The uniqueness of this solution can be proved via Theorem 2.45. □

Proposition 4.29. *Let* $x_0 \in L_2(\Omega, X)$, *let* $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, *let* $\varphi \in L_1(\mathbf{T}, L_2(\Omega, X))$, *let* $m \in M_2(\mathbf{T}, X)$ *and let* x *be defined by* (4.18). *Then*

$$x_t = \mathcal{U}_{t,s}x_s + \int_s^t \mathcal{U}_{t,r}\varphi_r\,dr + \int_s^t \mathcal{U}_{t,r}\,dm_r,\ \ 0 \le s \le t \le T.$$

Proof. By (4.18),

$$\begin{aligned} x_t &= \mathcal{U}_{t,s}\left(\mathcal{U}_{s,0}x_0 + \int_0^s \mathcal{U}_{s,r}\varphi_r\,dr + \int_0^s \mathcal{U}_{s,r}\,dm_r\right) \\ &\quad + \int_s^t \mathcal{U}_{t,r}\varphi_r\,dr + \int_s^t \mathcal{U}_{t,r}\,dm_r \\ &= \mathcal{U}_{t,s}x_s + \int_s^t \mathcal{U}_{t,r}\varphi_r\,dr + \int_s^t \mathcal{U}_{t,r}\,dm_r. \end{aligned}$$

This proves the proposition. □

Let us introduce the following useful concept. If a random process ξ has the form

$$\xi_t = \xi_0 + \int_0^t f_s\,ds + \int_0^t \Phi_s\,dm_s,\ \ 0 \le t \le T,$$

where we suppose $\xi_0 \in L_2(\Omega, X)$, $f \in L_1(\mathbf{T}, L_2(\Omega, X))$, $\Phi \in \Lambda_m^2(\mathbf{T}; Z, X)$ and $m \in M_2(\mathbf{T}, Z)$, then ξ is said to have

$$d\xi_t = f_t dt + \Phi_t dm_t$$

as its *stochastic differential*. In this case we use the notation

$$\int_0^T K_t\,d\xi_t = \int_0^T K_t f_t\,dt + \int_0^T K_t\Phi_t\,dm_t, \tag{4.21}$$

where K is an operator valued-function so that the integrals in the right-hand side of (4.21) are defined.

4.4.3 Partially Observable Linear Systems

A stochastic equation, when it is used as a mathematical model of some real process, is called a *state* or *signal system*. Respectively, its solution is called a *state* or *signal process*. Since a solution of a stochastic equation (denote it by x) is a random process (in general), x presents probabilistic distributions of states of a corresponding real process, but it does not exactly determine them. To improve information contained in x, another random process z is considered so that z is related with x and contains partial information about the states of the real process under consideration, which is not available from x. An equation that relates z and x is called an *observation system* and, respectively, z is called an *observation process* or, briefly, *observations*. State (or signal) and observation systems taken together define a *partially observable system*. Often, a partially observable system is called a *state-observation* or *signal-observation system* too.

When a state system is deterministic and, consequently, x is nonrandom, the states of the corresponding real process can be determined exactly by the values of the function x. In this case there is no need for an observation system and this reduces the corresponding partially observable system to only a state system.

The main object of our study is a *partially observable linear system* in which the state or signal system is a linear stochastic evolution system and the observation process is a linear transformation of the state process disturbed by a random process (called a *noise process*).

4.5 Basic Estimation in Hilbert Spaces

Often, we are faced with a problem to say something about one object while another object, related with the first one, is observed. The mathematical theory of studying this problem based on probabilistic methods is called estimation theory.

4.5.1 Estimation of Random Variables

Let $\eta \in L_2(\Omega, X)$ and let $\xi \in L_2(\Omega, Z)$. Denote by $\mathbf{P}_\xi$ the measure on $\mathcal{B}_Z$ generated by ξ. In estimation theory any function in $m(Z, \mathbf{P}_\xi, X)$ is called an *estimator*, and for $\varphi \in m(Z, \mathbf{P}_\xi, X)$, the random variable $\varphi \circ \xi$ is called an *estimate of* η *based on* ξ. The function $\varphi_0 \in m(Z, \mathbf{P}_\xi, X)$ is called an *optimal estimator* and, respectively, $\varphi_0 \circ \xi$ is called a *best estimate of* η *based on* ξ if

$$\mathbf{E}\|\eta - (\varphi_0 \circ \xi)\|^2 = \inf_\varphi \mathbf{E}\|\eta - (\varphi \circ \xi)\|^2, \tag{4.22}$$

where the infimum is taken over all estimators $\varphi \in m(Z, \mathbf{P}_\xi, X)$ for which $(\varphi \circ \xi) \in L_2(\Omega, X)$. In fact, by Proposition 2.27, the infimum in (4.22) is taken over the space $L_2(Z, \mathbf{P}_\xi, X)$.

A function in the form $\varphi(x) = \Lambda x + b$, where $\Lambda \in \mathcal{L}(Z, X)$ and $b \in X$, is called a *linear estimator* and, respectively, a random variable $\varphi \circ \xi = \Lambda\xi + b$ is called a *linear estimate of* η *based on* ξ. The function $\varphi_0(x) = \Lambda_0 x + b_0$, where $\Lambda_0 \in \mathcal{L}(Z, X)$ and $b_0 \in X$, is called an *optimal linear estimator* and, respectively, the random variable $\varphi_0 \circ \xi = \Lambda_0 \xi + b_0$ is called a *best linear estimate of* η *based on* ξ if

$$\mathbf{E}\|\eta - \Lambda_0\xi - b_0\|^2 = \inf_{\Lambda, b} \mathbf{E}\|\eta - \Lambda\xi - b\|^2,$$

where the infimum is taken over all $\Lambda \in \mathcal{L}(Z, X)$ and all $b \in X$.

Proposition 4.30. *Given $\eta \in L_2(\Omega, X)$ and $\xi \in L_2(\Omega, Z)$, there exists a best estimate of η based on ξ which is unique in $L_2(\Omega, \sigma(\xi), \mathbf{P}, X)$ and it is equal to the conditional expectation $\mathbf{E}(\eta|\xi)$. Furthermore, the corresponding optimal estimator is unique in $L_2(Z, \mathbf{P}_\xi, X)$.*

Proof. By Proposition 2.27, the space $\{\varphi \circ \xi : \varphi \in L_2(Z, \mathbf{P}_\xi, X)\}$ of all estimates is equal to $L_2(\Omega, \sigma(\xi), \mathbf{P}, X)$. Hence, by (4.22), a best estimate of η based on ξ is a random variable at which the functional $J(\zeta) = \|\eta - \zeta\|_{L_2}^2$ takes its minimum value on $L_2(\Omega, \sigma(\xi), \mathbf{P}, X)$. Since J is a strictly convex functional (see Proposition 1.30(b)), by Proposition 2.3, J takes on minimum value at some unique random variable $\zeta_0 \in L_2(\Omega, \sigma(\xi), \mathbf{P}, X)$ and by Theorem 2.6,

$$\forall \zeta \in L_2\big(\Omega, \sigma(\xi), \mathbf{P}, X\big), \ \langle J'(\zeta_0), \zeta\rangle = 2\mathbf{E}\langle \eta - \zeta_0, \zeta\rangle = 0.$$

Hence, ζ_0 is the projection of $\eta \in L_2(\Omega, X)$ to $L_2(\Omega, \sigma(\xi), \mathbf{P}, X)$. So, we conclude that ζ_0 is a unique best estimate of η based on ξ and by Proposition 4.4, $\zeta_0 = \mathbf{E}(\eta|\xi)$. Finally, by Proposition 2.27, $\mathbf{E}(\eta|\xi)$ can be represented as $\mathbf{E}(\eta|\xi) = \varphi_0 \circ \xi$ for some unique optimal estimator $\varphi_0 \in L_2(Z, \mathbf{P}_\xi, X)$. □

Thus, the problem on existence and uniqueness of optimal estimators is positively solved. But, designing optimal estimators as a function φ_0 of ξ is a central problem in estimation theory.

Another situation takes place with optimal linear estimators. The existence and uniqueness of optimal linear estimators can not always be solved positively. But, if they exist, then they can be easily designed in the form $\Lambda_0\xi + b_0$.

Proposition 4.31. *If a best linear estimate $\hat{\eta}$ of $\eta \in L_2(\Omega, X)$ based on $\xi \in L_2(\Omega, Z)$ exists, then it is unbiased, i.e., $\mathbf{E}(\eta - \hat{\eta}) = 0$.*

Proof. If $\hat{\eta}$ is the best linear estimate of η based on ξ, then

$$\forall \Lambda \in \mathcal{L}(Z, X) \text{ and } \forall b \in X, \ \mathbf{E}\langle \eta - \hat{\eta}, \Lambda\xi + b\rangle = 0.$$

Taking $\Lambda = 0$ and varying b in X, we obtain $\mathbf{E}(\eta - \hat{\eta}) = 0$. □

By Proposition 4.31, a best linear estimate $\hat{\eta}$ (if it exists) of $\eta \in L_2(\Omega, X)$ based on $\xi \in L_2(\Omega, Z)$ has the form $\hat{\eta} = \mathbf{E}\eta + \Lambda_0(\xi - \mathbf{E}\xi)$ for some $\Lambda_0 \in \mathcal{L}(Z, X)$. The next proposition determines a condition on Λ_0.

Proposition 4.32. *Suppose that* $\eta \in L_2(\Omega, X)$, $\xi \in L_2(\Omega, Z)$, $P = \operatorname{cov}\xi$ *and* $Q = \operatorname{cov}(\xi, \eta)$. *Then a random variable* $\hat{\eta} = \mathbf{E}\eta + \Lambda_0(\xi - \mathbf{E}\xi)$ *is a best linear estimate of* η *based on* ξ *if and only if* $\Lambda = \Lambda_0$ *is a solution of the operator equation*

$$\Lambda P = Q^*. \tag{4.23}$$

Proof. Let $\Lambda_0 \in \mathcal{L}(Z, X)$ be a solution of the equation (4.23). For arbitrary $\Lambda \in \mathcal{L}(Z, X)$ and $b \in X$, we have

$$\begin{aligned}
&\mathbf{E}\|\eta - \Lambda\xi - b\|^2 - \mathbf{E}\|\eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi)\|^2 \\
&\quad = \operatorname{tr}(\operatorname{cov}(\eta - \Lambda\xi - b)) - \operatorname{tr}(\operatorname{cov}(\eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi))) + \|\mathbf{E}(\eta - \Lambda\xi - b)\|^2 \\
&\quad = \operatorname{tr}(\Lambda P \Lambda^* - Q^*\Lambda^* - \Lambda Q - \Lambda_0 P \Lambda_0^* + Q^*\Lambda_0^* + \Lambda_0 Q) + \|\mathbf{E}(\eta - \Lambda\xi - b)\|^2 \\
&\quad = \operatorname{tr}((\Lambda - \Lambda_0)P(\Lambda - \Lambda_0)^*) + \|\mathbf{E}(\eta - \Lambda\xi - b)\|^2 \geq 0,
\end{aligned}$$

where we used the fact that Λ_0 is a solution of the equation (4.23). Hence, $\hat{\eta} = \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi)$ is a best linear estimate of η based on ξ. Conversely, let $\hat{\eta} = \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi)$ be a best linear estimate of η based on ξ. We have to show that Λ_0 satisfies (4.23). For $\lambda \in \mathbb{R}$, let

$$\eta_\lambda = \mathbf{E}\eta + (\Lambda_0 + \lambda(\Lambda_0 P - Q^*))(\xi - \mathbf{E}\xi).$$

We have

$$\begin{aligned}
0 &\leq \mathbf{E}\|\eta - \mathbf{E}\eta - (\Lambda_0 + \lambda(\Lambda_0 P - Q^*))(\xi - \mathbf{E}\xi)\|^2 \\
&\quad - \mathbf{E}\|\eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi)\|^2 \\
&= \operatorname{tr}(\operatorname{cov}(\eta - \mathbf{E}\eta - (\Lambda_0 + \lambda(\Lambda_0 P - Q^*))(\xi - \mathbf{E}\xi))) \\
&\quad - \operatorname{tr}(\operatorname{cov}(\eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi))) \\
&= \lambda^2 \operatorname{tr}((\Lambda_0 P - Q^*)P(\Lambda_0 P - Q^*)^*) + 2\lambda \operatorname{tr}((\Lambda_0 P - Q^*)(\Lambda_0 P - Q^*)^*).
\end{aligned}$$

Since $\lambda \in \mathbb{R}$ is arbitrary, we obtain $\operatorname{tr}((\Lambda_0 P - Q^*)(\Lambda_0 P - Q^*)^*) = 0$. Therefore,

$$\|\Lambda_0 P - Q^*\|_{\mathcal{L}} \leq \|\Lambda_0 P - Q^*\|_{\mathcal{L}_2} = \operatorname{tr}((\Lambda_0 P - Q^*)(\Lambda_0 P - Q^*)^*) = 0,$$

i.e., Λ_0 is a solution of the equation (4.23). □

The next proposition shows that in the Gaussian case a best linear estimate coincides with the corresponding best estimate.

Proposition 4.33. *Suppose that* $\eta \in L_2(\Omega, X)$ *and* $\xi \in L_2(\Omega, Z)$ *define a Gaussian system. If there exists a best linear estimate* $\hat{\eta}$ *of* η *based on* ξ, *then it is equal to the best estimate of* η *based on* ξ.

Proof. In view of Propositions 4.4, 4.30 and 4.32, it is sufficient to show that

$$\forall \varphi \in L_2(Z, \mathbf{P}_\xi, X), \ \mathbf{E}\langle \eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi), \varphi \circ \xi \rangle = 0, \tag{4.24}$$

where Λ_0 is a solution of the equation (4.23) with $P = \text{cov}\xi$ and $Q = \text{cov}(\xi, \eta)$. One can compute

$$\text{cov}(\eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi), \xi) = Q^* - \Lambda_0 P = 0.$$

Since ξ and η define a Gaussian system, by Theorem 4.6, the random variables $\eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi)$ and ξ are independent and, as a consequence, the random variables $\eta - \mathbf{E}\eta - \Lambda_0(\xi - \mathbf{E}\xi)$ and $\varphi \circ \xi$ are independent for all $\varphi \in L_2(Z, \mathbf{P}_\xi, X)$. Thus, (4.24) holds. □

4.5.2 Estimation of Random Processes

Often, a random variable is estimated based on a random process. For this case, the results of Section 4.5.1 can be modified in a respective way.

Let η be a random variable in $L_2(\Omega, X)$ and let ξ_t, $0 \le t \le T$, be a random process in $L_2(\Omega, C(\mathbf{T}, Z))$. Consider ξ as a random variable with values in $L_2(\mathbf{T}, Z)$ and denote by $\sigma(\xi)$ the σ-algebra generated by this random variable. By Proposition 2.30, $\sigma(\xi) = \sigma(\xi_t; 0 \le t \le T)$. Therefore,

$$\mathbf{E}(\eta|\xi) = \mathbf{E}(\eta|\xi_t; 0 \le t \le T).$$

So, by Proposition 4.30, the best estimate of η based on ξ, provided that ξ is an $L_2(\mathbf{T}, Z)$-valued random variable, is equal to $\mathbf{E}(\eta|\xi_t; 0 \le t \le T)$. Thus, the *best estimate of* η *based on the random process* ξ_t, $0 \le t \le T$, can be defined as the best estimate of η based on the random variable ξ and it is equal to $\mathbf{E}(\eta|\xi_t; 0 \le t \le T)$.

Now suppose that a random variable η in $L_2(\Omega, X)$ and a random process ξ_t, $0 \le t \le T$, in $L_2(\Omega, C(\mathbf{T}, Z))$ define a Gaussian system, i.e.,

$$\mathcal{N} = \{\langle \eta, h\rangle, \langle \xi_t, g\rangle : h \in X,\ g \in Z,\ 0 \le t \le T\}$$

is a Gaussian system. Obviously,

$$\forall \varphi \in L_2(\mathbf{T}, Z), \quad \int_0^T \langle \xi_t, \varphi_t\rangle\, dt \in \overline{\text{span}\,\mathcal{N}}.$$

Hence, the random variable η and the $L_2(\mathbf{T}, Z)$-valued random variable ξ define a Gaussian system. So, by Proposition 4.33, if there exists a best linear estimate of η based on the random variable ξ, then it is equal to

$$\mathbf{E}(\eta|\xi) = \mathbf{E}(\eta|\xi_t, 0 \le t \le T).$$

Thus, in this case,

$$\mathbf{E}(\eta|\xi_t, 0 \le t \le T) = \mathbf{E}\eta + \Lambda_0(\xi_t - \mathbf{E}\xi_t, 0 \le t \le T)$$

for some $\Lambda_0 \in \mathcal{L}(L_2(\mathbf{T}, Z), X)$.

In estimating a random variable based on a random process, we do not prefer linear estimates defined as above. Instead, we use estimates in a linear feedback form. A random variable

$$\zeta = b + \int_0^T K_t \, d\xi_t, \tag{4.25}$$

where $b \in X$ and $K \in B_2(\mathbf{T}, \mathcal{L}(Z, X))$, is called an *estimate of η based on ξ_t, $0 \le t \le T$, in a linear feedback form* provided that ξ has a stochastic differential and the stochastic integral in (4.25) is defined. Obviously, a vector $b \in X$ and a function $K \in B_2(\mathbf{T}, \mathcal{L}(Z, X))$ determine a *linear feedback estimator*. If the paths of the process ξ are a.e. differentiable w.p.1 and $\xi' \in L_2(\Omega, L_2(\mathbf{T}, Z))$, then (4.25) can be written as

$$\zeta = b + \int_0^T K_t \, d\xi_t = b + \int_0^T K_t \xi_t' \, dt.$$

In this case, since

$$Jf = \int_0^T K_t f_t \, dt, \ \ f \in L_2(\mathbf{T}, Z),$$

defines a bounded linear operator from $L_2(\mathbf{T}, Z)$ to X, the estimate (4.25) in a linear feedback form can be interpreted as a linear estimate of η based on the derivative of ξ. But, as it will follow from Theorem 4.35, in most useful cases the random process ξ has nowhere differentiable paths w.p.1, although the estimate in the linear feedback form (4.25) is well defined when the random process ξ has a stochastic differential.

Proposition 4.34. *Let $\eta \in L_2(\Omega, X)$ be a random variable with $\mathbf{E}\eta = 0$ and let ξ be a random process which has the stochastic differential*

$$d\xi_t = f_t dt + \Phi_t dv_t, \ \ 0 < t \le T, \ \ \xi_0 = 0,$$

where $f \in L_2(\mathbf{T} \times \Omega, Z)$, $\mathbf{E}f_t = 0$ for a.e. $t \in \mathbf{T}$, $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(Y, Z))$ and v is a Y-valued Wiener process on $\mathbf{T}$. Suppose that η and ξ define a Gaussian system. Then there exists a function $K \in B_2(\mathbf{T}, \mathcal{L}(Z, X))$ such that

$$\mathbf{E}(\eta | \xi_t; 0 \le t \le T) = \int_0^T K_t \, d\xi_t \tag{4.26}$$

if and only if

$$\forall G \in B_2(\mathbf{T}, \mathcal{L}(Z)), \ \operatorname{cov}\left(\eta - \int_0^T K_t \, d\xi_t, \int_0^T G_t \, d\xi_t\right) = 0. \tag{4.27}$$

Proof. Since η and ξ_t, $0 \le t \le T$, define a Gaussian system, then

$$\eta - \mathbf{E}(\eta | \xi_t, 0 \le t \le T) \text{ and } \sigma(\xi_t; 0 \le t \le T)$$

are independent. So, (4.26) implies (4.27). Let us prove the converse. Selecting the function G in (4.27) as $G_s = I$ if $0 \le s \le t$ and $G_s = 0$ if $t < s \le T$, we see that (4.27) implies the independence of

$$\eta - \int_0^T K_t \, d\xi_t$$

and $\sigma(\xi_t;\ 0 \le t \le T)$ (see Theorem 4.6). So,

$$\begin{aligned}\mathbf{E}(\eta|\xi_t, 0 \le t \le T) - \int_0^T K_t \, d\xi_t &= \mathbf{E}\left(\eta - \int_0^T K_t \, d\xi_t \,\middle|\, \xi_t;\ 0 \le t \le T\right) \\ &= \mathbf{E}\left(\eta - \int_0^T K_t \, d\xi_t\right) = 0.\end{aligned}$$

The proof is completed. □

Now consider a partially observable system with a signal process ξ and an observation process η. The estimation problem of finding the best estimate of η_t based on ξ_s, $0 \le s \le \tau$, is called a *filtering problem* in case $t = \tau$. When $t < \tau$ ($t > \tau$), this problem is called a *smoothing* or *interpolation* (*prediction* or *extrapolation*) *problem*. Estimators in these problems are called a *filter*, a *smoother* and a *predictor*, respectively. Among these estimation problems, the important one is the filtering problem. Smoothing and prediction problems can be reduced to filtering problems.

4.6 Improving the Brownian Motion Model

In Section 4.2 Wiener processes were presented as a suitable model for Brownian motion and the equation (4.5) was derived for a diffusion process. In this section we will continue this discussion.

4.6.1 White, Colored and Wide Band Noise Processes

Let us turn back to Section 4.2 where a Brownian motion was discussed. A microscopic particle suspended in liquid has a finite velocity at each time moment and covers a finite distance on each finite time interval. Therefore, a Wiener process that was taken as a mathematical model of a Brownian motion would be expected to have differentiable paths of finite variation. However, the following takes place instead.

Theorem 4.35. *The paths of a (one-dimensional) Wiener process are nowhere differentiable w.p.1 and have infinite variation over each finite time interval.*

Accordingly, the derivative w' of a Wiener process w could be understood only in a generalized sense. In fact, w' is a generalized Gaussian random process (i.e., the paths of w' are generalized functions) with zero expectation and the covariance $\operatorname{cov}(w'_t, w'_s) = \delta(t-s)$ where δ is the Dirac delta-function (see Hida [53] and Krylov [67]). To imagine the Dirac delta function, the reader can suppose that it is the generalized density function of the probability distribution $\mathbf{P}_\delta$ on $\mathbb{R}$ concentrated at the origin, i.e.,

$$\mathbf{P}_\delta(\{0\}) = 1 \text{ and } \mathbf{P}_\delta(\mathbb{R} \setminus \{0\}) = 0.$$

The generalized derivative of a Wiener process is called a *Gaussian white noise process* or a *white noise process* or, simply, *white noise*.

Often, in engineering for a Wiener process w, the informal representation

$$w_t = \int_0^t w'_s \, ds$$

is used and, consequently, the equation (4.5) is replaced by

$$x'_t = F(t, x_t) + G(t, x_t) w'_t.$$

Below sometimes we use this informal representation as well, keeping in mind that w' is not a random process in the ordinary sense.

Disagreement between Theorem 4.35 and the real nature of Brownian motion makes it necessary to revise the properties (A)–(F) of a Brownian motion considered in Section 4.2. A vulnerable one among them is the property (C). Indeed, when the disjoint intervals $(s,t]$ and $(\sigma, r]$ are close to each other (for example, when $s = r$) we could expect (probably weak) dependence of the increments $w_t - w_s$ and $w_r - w_\sigma$. Taking this dependence into consideration, we can correct the property (C) in the following form.

(C′) There is a number $\varepsilon > 0$ such that the increments $w_t - w_s$ and $w_r - w_\sigma$ are independent for all $0 \le \sigma < r < r + \varepsilon \le s < t \le T$.

Closely related with the modified property (C′), let us introduce the following definition from Fleming and Rishel [48], p. 126.

Definition 4.36. A random process $\varphi : \mathbf{T} \to X$ is called an X-valued *wide band noise process* or, simply, a *wide band noise*, if there exists a number $\varepsilon > 0$ such that

$$\operatorname{cov}(\varphi_t, \varphi_s) = \begin{cases} \Lambda_{t,s}, & 0 \le t - s < \varepsilon, \\ 0, & t - s \ge \varepsilon, \end{cases} \tag{4.28}$$

where Λ is a nonzero function. If in addition $\mathbf{E}\varphi_t = 0$ and $\Lambda_{t,s} = \Lambda_{t-s}$, then φ is said to be stationary in a wide sense. The function Λ is called the *autocovariance function* of φ.

For $0 \le \sigma < r \le s < t \le T$ and for the random process

$$\xi_t = \int_0^t \varphi_s \, ds, \ 0 \le t \le T,$$

where φ is the wide band noise process from Definition 4.36, one can easily calculate

$$\begin{aligned}\operatorname{cov}(\xi_t - \xi_s, \xi_r - \xi_\sigma) &= \operatorname{cov}\left(\int_s^t \varphi_\alpha \, d\alpha, \int_\sigma^r \varphi_\beta \, d\beta\right) \\ &= \int_s^{\min(t,r+\varepsilon)} \operatorname{cov}\left(\varphi_\alpha, \int_\sigma^r \varphi_\beta \, d\beta\right) d\alpha \\ &= \int_s^{\min(t,r+\varepsilon)} \int_{\max(\sigma,\alpha-\varepsilon)}^r \operatorname{cov}(\varphi_\alpha, \varphi_\beta) \, d\beta \, d\alpha \\ &= \int_s^{\min(t,r+\varepsilon)} \int_{\max(\sigma,\alpha-\varepsilon)}^r \Lambda_{\alpha,\beta} \, d\beta \, d\alpha\end{aligned}$$

if $s < r+\varepsilon$ and $\operatorname{cov}(\xi_t - \xi_s, \xi_r - \xi_\sigma) = 0$ if $s \geq r+\varepsilon$. So, if in addition φ is Gaussian, then the random process ξ satisfies the above mentioned condition (C$'$).

To compare white and wide band noise processes, let w be a standard Wiener process and define

$$\varphi_t^\varepsilon = \frac{w_{t+\varepsilon} - w_t}{\varepsilon}, \ 0 \leq t \leq T, \ \varepsilon > 0.$$

One can easily compute that $\mathbf{E}\varphi_t^\varepsilon = 0$ and

$$\Lambda_{t,s} = \operatorname{cov}(\varphi_t^\varepsilon, \varphi_s^\varepsilon) = \begin{cases} \varepsilon^{-2}(\varepsilon - (t-s)), & 0 < t-s < \varepsilon, \\ 0, & t-s \geq \varepsilon. \end{cases}$$

Hence, the process φ^ε, that is an approximation to the physically impossible white noise process w', is a stationary in a wide sense, wide band, noise process. Thus, the white noise process w' is an ideal case of wide band noise processes φ^ε when ε is infinitely small making $\Lambda_{t,t}$ infinitely large.

Note that it is convenient to work with Gaussian white noise (or Wiener) processes and, therefore, basic mathematical methods of control and estimation of stochastic systems has been developed for white noise driven systems. At the same time, the practical problems involve wide band noise.

Note that there are noise processes which are close to wide band noise, while the estimation results for them are similar to the respective results for white noise. They are called colored noise processes. A colored noise is an output of a linear system under an additive white noise disturbance, i.e., it is a solution of the linear stochastic differential equation

$$d\xi_t = A\xi_t dt + \Phi_t dw_t, \ \xi_0 = 0, \ 0 < t \leq T.$$

A mild solution of this equation has the form

$$\xi_t = \int_0^t \mathcal{U}_{t,s} \Phi_s \, dw_s, \ 0 \leq t \leq T, \tag{4.29}$$

where $\mathcal{U}$ is the strongly continuous semigroup generated by A. Therefore, in general by X-valued *colored noise* we will mean a random process ξ in the form (4.29) where it is assumed that $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(Y, X))$ and w is a Y-valued Wiener process on $\mathbf{T}$.

4.6.2 Integral Representation of Wide Band Noises

There are different approaches to a wide band noise phenomenon. For example, Kushner in [69] uses an approximative approach. Here we discuss an approach based on a certain integral representation from Bashirov [9].

Consider the random process

$$\varphi_t = \int_{\max(0,t-\varepsilon)}^{t} \Phi_{t,\theta-t}\, dw_\theta, \ 0 \le t \le T, \tag{4.30}$$

where $0 < \varepsilon < T$, $\Phi \in B_2(\mathbf{T} \times [-\varepsilon, 0], \mathcal{L}(Y, X))$ and w is a Y-valued Wiener process on $\mathbf{T}$. Obviously,

$$\operatorname{cov}(\varphi_t, \varphi_s) = \int_{\max(0,t-\varepsilon)}^{s} \Phi_{t,\theta-t} W \Phi^*_{s,\theta-s}\, d\theta \ \text{ if } \ 0 \le t - s < \varepsilon,$$

and $\operatorname{cov}(\varphi_t, \varphi_s) = 0$ if $t - s \ge \varepsilon$, where W is the covariance operator of w. Hence, the random process φ defined by (4.30) is a wide band noise process.

In real systems, the presence of the wide band noise process (4.30), which can also informally be represented as

$$\varphi_t = \int_{\max(0,t-\varepsilon)}^{t} \Phi_{t,\theta-t}\, dw_\theta = \int_{\max(0,t-\varepsilon)}^{t} \Phi_{t,\theta-t} w'_\theta\, d\theta = \int_{-\min(\varepsilon,t)}^{0} \Phi_{t,\theta} w'_{t+\theta}\, d\theta,$$

can be interpreted by vibration. At moment t a vibration that is formed by the action of white noise during the time between $t-\varepsilon$ and t affects the system. Values of white noise until the moment $t - \varepsilon$ do not take part in the formation of the vibration at the moment t, because their weight is sufficiently small and we can neglect them in the mathematical model (4.30). Consequently, the parameters Φ and ε of the wide band noise process (4.30) have the following meaning: Φ stands for the coefficient of relaxing the initial effect of white noise at different time moments (and, therefore, if $X = Y = \mathbb{R}$, it is natural to assume that under fixed t the function Φ is increasing in θ and, consequently, it is a.e. differentiable in θ with $\Phi_{t,-\varepsilon} = 0$) and ε represents the interval outside of which the consequences of the disturbing noises are not valid. The function Φ will be called a *relaxing function*. By this interpretation, the wide band noise process (4.30) corresponds to real cases when a vibration generated by white noise stands to affect the systems starting at the initial time $t = 0$ (a reason for this may be switching the systems on from resting state to dynamic state that changes their sensitivity to random disturbances) and, hence, when $0 < t < \varepsilon$ the wide band noise φ is formed by w'_θ, $0 \le \theta \le t$.

Proposition 4.37. *Let* $0 < \varepsilon < T$, *let* $W \in \mathcal{L}_1(Y)$ *with* $W \geq 0$ *and suppose that* $\Phi \in B_2(-\varepsilon, 0; \mathcal{L}(Y, X))$ *is a solution of the equation*

$$\int_{-\varepsilon}^{-s} \Phi_\theta W \Phi_{\theta+s}^* \, d\theta = \Lambda_s, \ 0 \leq s \leq \varepsilon, \tag{4.31}$$

where $\Lambda : [0, \varepsilon] \to \mathcal{L}_1(X)$ *is a given function. Then the random process*

$$\varphi_t = \int_{\max(0,t-\varepsilon)}^{t} \Phi_{\theta-t} \, dw_\theta, \ 0 \leq t \leq T, \tag{4.32}$$

where w *is a* Y*-valued Wiener process on* $\mathbf{T}$ *and* W *is its covariance operator, is a Gaussian wide band noise process. Furthermore, the restriction of* φ *to* $[\varepsilon, T]$ *is stationary in a wide sense and has the autocovariance function* Λ.

Proof. Comparing (4.30) and (4.32), we see that (4.32) is a Gaussian wide band noise process. For $\varepsilon \leq t \leq T$ and for $0 \leq s \leq \varepsilon$, $\mathbf{E}\varphi_t = 0$ and the autocovariance function of (4.32) is equal to

$$\operatorname{cov}(\varphi_{t+s}, \varphi_t) = \int_{t+s-\varepsilon}^{t} \Phi_{\theta-t-s} W \Phi_{\theta-t}^* \, d\theta = \int_{-\varepsilon}^{-s} \Phi_\theta W \Phi_{\theta+s}^* \, d\theta,$$

which is independent of t. So, since the time moment ε the process φ defined by (4.32) is stationary in a wide sense and has Λ as its autocovariance function. □

Practically, engineers meet the Gaussian stationary in wide sense wide band noise processes which are observed by their autocovariance functions. Therefore, given an autocovariance function Λ_s, $0 \leq s \leq \varepsilon$, to construct a wide band noise process in the form (4.32), one must solve the equation (4.31) in Φ. In the one-dimensional case, when $X = Y = \mathbb{R}$ and $W = 1$, the equation (4.31) reduces to

$$\int_{-\varepsilon}^{-s} \phi_\theta \phi_{\theta+s} \, d\theta = \lambda_s, \ 0 \leq s \leq \varepsilon, \tag{4.33}$$

where the operator-valued functions Φ and Λ are replaced by the real-valued functions ϕ and λ. The following theorem gives a method of construction of infinitely many relaxing functions ϕ corresponding to a given autocovariance function λ.

Theorem 4.38. *Let* $\varepsilon > 0$ *and let* $\lambda \in L_2(0, \varepsilon; \mathbb{R})$. *Define the function* λ^* *as the even extension of* λ *to the real line vanishing outside of* $[-\varepsilon, \varepsilon]$. *Assume that* λ^* *is positive definite, i.e.,*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{t_i - t_j}^* z_i \bar{z}_j \geq 0,$$

for all finite collections $\{t_1, \ldots, t_n\}$ *of real numbers and* $\{z_1, \ldots, z_n\}$ *of complex numbers, where* $\bar{z}_j$ *is the conjugate of the complex number* z_j, *and*

$$\int_0^\infty [\mathcal{F}(\lambda^*)]_\omega \, d\omega < \infty, \tag{4.34}$$

where $\mathcal{F}(\lambda^*)$ is the Fourier transformation of λ^*. Then there exists a solution ϕ of the equation (4.33) *in the space $L_2(-\varepsilon, 0; \mathbb{R})$. If λ is a nonzero function, then the number of distinct solutions of the equation* (4.33) *is infinite.*

Proof. With each $\phi \in L_2(-\varepsilon, 0; \mathbb{R})$ associate two functions ϕ^* and ϕ^{**}. Let ϕ^* be the extension of ϕ to the real line vanishing outside of $[-\varepsilon, 0]$ and define ϕ^{**} by $\phi^{**}_r = \phi^*_{-r}$ for $-\infty < r < \infty$. Then the equation (4.33) can be written in the form $\lambda^* = \phi^* \star \phi^{**}$, where $\phi^* \star \phi^{**}$ is the convolution of ϕ^* and ϕ^{**}. From the properties of the Fourier integral (see Papoulis [79]), it follows that $\mathcal{F}(\lambda^*) = \mathcal{F}(\phi^*)\mathcal{F}(\phi^{**})$ or

$$[\mathcal{F}(\lambda^*)]_\omega = [\mathcal{F}(\phi^*)]_\omega \overline{[\mathcal{F}(\phi^*)]_\omega}, \; -\infty < \omega < \infty.$$

Therefore, if

$$[\mathcal{F}(\phi^*)]_\omega = x_\omega + iy_\omega, \; -\infty < \omega < \infty, \tag{4.35}$$

where x and y are unknown real-valued functions and i is the imaginary unit, then

$$[\mathcal{F}(\lambda^*)]_\omega = x_\omega^2 + y_\omega^2, \; -\infty < \omega < \infty. \tag{4.36}$$

Note that $\mathcal{F}(\lambda^*)$ is a nonnegative even function of a real variable, since λ^* is even and positive definite. In order to obtain ϕ^* to be real-valued, we must find an even function x and an odd function y satisfying (4.36). This can be done in the following way. Let $0 \le \alpha \le 1$. Construct a measurable even function x and a measurable odd function y satisfying

$$x_\omega^2 = \alpha[\mathcal{F}(\lambda^*)]_\omega \text{ and } y_\omega^2 = (1-\alpha)[\mathcal{F}(\lambda^*)]_\omega, \; -\infty < \omega < \infty.$$

This can be done easily by considering different branches of the square root. By the condition (4.34), for each such pair (x, y), the inverse Fourier transformation

$$\phi^* = \mathcal{F}^{-1}(\mathcal{F}(\phi^*)) = \mathcal{F}^{-1}(x + iy)$$

exists and is a real-valued square integrable function vanishing outside of $[-\varepsilon, 0]$ (otherwise, λ^* will take nonzero values for $|s| > \varepsilon$). The restriction of each ϕ^*, constructed in the above mentioned way, to $[-\varepsilon, 0]$ is a solution of the equation (4.33). From the above construction of ϕ^*, it is clear that the number of solutions of the equation (4.33) is infinite if λ is a nonzero function. □

Note that the condition on positive definiteness of λ^* in Theorem 4.38 is ordinary since λ is an autocovariance function. The condition (4.34) guarantees the existence of $\mathcal{F}^{-1}(x + iy)$ as a square integrable function. The non-uniqueness of solution of the equation (4.33) demonstrates that the covariance function does not provide complete information about the respective wide band noise process. Theorem 4.38 implies that among all the wide band noise processes that are stationary in a wide sense and have the given autocovariance function λ (start with a small time moment ε), there is a sufficiently wide class of such processes which have the integral representation in which w is a standard Wiener process and the relaxing function ϕ belongs to the space $L_2(\varepsilon, 0; \mathbb{R})$. In the sequel, we will consider the Hilbert space-valued wide band noise processes represented in the integral form for which the respective relaxing functions are differentiable and vanish at $-\varepsilon$. By the above mentioned interpretation, these conditions are reasonable.

Chapter 5

Separation Principle

The main result of this chapter is an extension of the separation principle for linear quadratic optimal control problems under partial observations to a case when the noise processes of the state and the observations are dependent.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $X, Z, Z_1, Z_2, U, U_1, U_2, H \in \mathcal{H}$, $T > 0$, $\mathbf{T} = [0, T]$ is a finite time interval and $\Delta_t = \{(s, r) : 0 \le r \le s \le t\}$ for $t > 0$.

5.1 Setting of Control Problem

Setting of each optimal control problem requires a determination of three objects: a state-observation system, a set of admissible controls and a cost functional. In this section these three objects are defined for a linear quadratic optimal control problem under partial observations, which is the main control problem in this chapter.

5.1.1 State-Observation System

Define a linear state-observation system by

$$x_t^u = \mathcal{U}_{t,0} l + \int_0^t \mathcal{U}_{t,s}(B_s u_s + b_s)\, ds + \int_0^t \mathcal{U}_{t,s}\, dm_s, \ \ 0 \le t \le T, \tag{5.1}$$

$$z_t^u = \int_0^t (C_s x_s^u + c_s)\, ds + n_t, \ \ 0 \le t \le T, \tag{5.2}$$

where u is a control taken from the set of admissible controls U_{ad} which will be defined below, and x^u and z^u are the state and observation processes corresponding to the control u. The following conditions are supposed to hold:

$(\mathbf{C_1})$ $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $B \in B_\infty(\mathbf{T}, \mathcal{L}(U, X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, Z))$;

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

($\mathbf{C_2}$) $l \in L_2(\Omega, X)$, $b \in L_2(\mathbf{T} \times \Omega, X)$, $c \in L_2(\mathbf{T} \times \Omega, Z)$, $m \in M_2(\mathbf{T}, X)$, $n \in M_2(\mathbf{T}, Z)$.

Below we will use the following notation. For $h \in L_2(\mathbf{T} \times \Omega; Z)$ and for $0 < t < T$, we denote by h^t and h^{t+} the $L_2(0,t;Z)$- and $L_2(t,T;Z)$-valued random variables obtained by restriction of h to $[0,t] \times \Omega$ and to $(t,T] \times \Omega$, respectively.

Remark 5.1. By Proposition 4.24(c), for given $u \in L_2(\mathbf{T} \times \Omega, U)$, the state system (5.1) defines the random process x^u in $L_\infty(\mathbf{T}, L_2(\Omega, X))$. Hence, the random process z^u, defined by (5.2), belongs to $m(\Omega, L_2(\mathbf{T}, Z))$. Moreover, if the stronger condition $n \in M_2^c(\mathbf{T}, Z)$ holds, then by Proposition 4.8, $z^u \in m(\Omega, C(\mathbf{T}, Z))$.

Remark 5.2. The conditions ($\mathbf{C_1}$)–($\mathbf{C_2}$) do not contain anything about the relation between l, b, c, m and n. They may be independent as well as dependent. Under the conditions ($\mathbf{C_1}$)–($\mathbf{C_2}$), the observations make sense even if $C = 0$ because observing only the noise processes c and n, useful information about the state process can be obtained. Moreover, the conditions ($\mathbf{C_1}$)–($\mathbf{C_2}$) accept the case when the observations until the current time moment t depend on the future of the signal noise, i.e., when the σ-algebras $\sigma\big(z^{u,t}\big)$ and $\sigma\big(b^{t+}, m_s - m_t;\ t < s \le T\big)$ are dependent.

5.1.2 Set of Admissible Controls

Let $z^u \in m(\Omega, L_2(\mathbf{T}, Z))$ be the random process corresponding to the control $u \in L_2(\mathbf{T} \times \Omega, U)$ by (5.1)–(5.2) (see Remark 5.1). As above, for $0 < t \le T$, the restriction of z^u to $[0,t] \times \Omega$, considered as an $L_2(0,t;Z)$-valued random variable, is denoted by $z^{u,t}$. Consider

$$U_t^u = L_2\big(\Omega, \sigma\big(z^{u,t}\big), \mathbf{P}, U\big).$$

We denote $z^{0,t} = z^{u,t}$ and $U_t^0 = U_t^u$ if $u = 0$. Let

$$\tilde{U}^0 = \int_0^T U_t^0 \, dt = \big\{u \in L_2(\mathbf{T} \times \Omega, U) : u_t \in U_t^0 \text{ for a.e. } t \in \mathbf{T}\big\},$$

$$\tilde{U} = \{u \in L_2(\mathbf{T} \times \Omega, U) : u_t \in U_t^u \text{ for a.e. } t \in \mathbf{T}\},$$

where the integral means a Hilbertian sum of subspaces (see Section 2.4.3). By Proposition 2.29(a), for fixed $u \in L_2(\mathbf{T} \times \Omega, U)$, $\{U_t^u : 0 < t \le T\}$ is a nondecreasing family of subspaces of $L_2(\Omega, U)$. Additionally, by Proposition 2.29(b), $\tilde{U}^0$ is a subspace of $L_2(\mathbf{T} \times \Omega, U)$. The same can not be said about $\tilde{U}$.

Since $\tilde{U}$ consists of functions $u \in L_2(\mathbf{T} \times \Omega, U)$ adapted with respect to the partial observations, it could be chosen as a set of admissible controls. But, the complicated structure of $\tilde{U}$ makes it inconvenient in studying optimal control problems. On the other hand, $\tilde{U}^0$ has an excellent structure, but the choice of $\tilde{U}^0$ as a set of admissible controls means disregarding the partial observations. Hence, following Bensoussan and Viot [33], the set of admissible controls will be defined by

$$U_{\text{ad}} = \tilde{U}^0 \cap \tilde{U}. \tag{5.3}$$

Note that $U_{\text{ad}} \neq \varnothing$ since at least it contains the nonrandom controls. The properties of U_{ad}, studied below, show that U_{ad} is a good trade-off $\tilde{U}$ against $\tilde{U}^0$.

5.1.3 Quadratic Cost Functional

To complete setting of the main optimal control problem, we also consider the cost functional

$$\begin{aligned} J(u) = \mathbf{E}\bigg(\langle x_T^u, Q_T x_T^u \rangle &+ \int_0^T \left\langle \begin{bmatrix} x_t^u \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix} \begin{bmatrix} x_t^u \\ u_t \end{bmatrix} \right\rangle dt \\ &+ 2\langle q, x_T^u \rangle + 2\int_0^T \left\langle \begin{bmatrix} f_t \\ g_t \end{bmatrix}, \begin{bmatrix} x_t^u \\ u_t \end{bmatrix} \right\rangle dt \bigg) \end{aligned} \tag{5.4}$$

and suppose that the following conditions hold:

($\mathbf{C_3'}$) $Q_T \in \mathcal{L}(X)$, $Q_T^* = Q_T$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$, $L \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $F_t^* = F_t$ and $G_t^* = G_t$ for a.e. $t \in \mathbf{T}$;

($\mathbf{C_4}$) $q \in L_2(\Omega, X)$, $f \in L_2(\mathbf{T} \times \Omega, X)$, $g \in L_2(\mathbf{T} \times \Omega, U)$.

Below we will use the Riccati equation (3.9). The above conditions do not guarantee the existence of a solution of the equation (3.9) (see Theorem 3.22). Therefore, we will also need the following condition:

(RE) The equation (3.9) has a solution in $B(\mathbf{T}, \mathcal{L}(X))$.

Note that if $Q \in B(\mathbf{T}, \mathcal{L}(X))$ is a solution of the equation (3.9), then in view of ($\mathbf{C_3'}$), $Q_t^* = Q_t$, $0 \le t \le T$.

Minimizing the cost functional (5.4) over the state-observation system (5.1)–(5.2) on the set of admissible controls U_{ad} defined by (5.3) is the main optimal control problem in this chapter. Briefly, this problem will be called the problem (5.1)–(5.4). A control in U_{ad}, at which the functional (5.4) takes on its minimum value, is called an *optimal control*.

5.2 Separation Principle

In this section the separation principle, which is essential in studying optimal control problems for partially observable systems, is extended to dependent noise disturbances in the problem (5.1)–(5.4). In this section it is always assumed that the conditions ($\mathbf{C_1}$), ($\mathbf{C_2}$), ($\mathbf{C_3'}$), ($\mathbf{C_4}$) and (**RE**), defined in Section 5.1, hold.

5.2.1 Properties of Admissible Controls

Lemma 5.3. *If $u \in \tilde{U}$, then $U_t^0 \subset U_t^u$, $0 < t \leq T$.*

Proof. Let $u \in \tilde{U}$. Define

$$\xi_t^u = \int_0^t \mathcal{U}_{t,s} B_s u_s \, ds, \ \eta_t^u = \int_0^t C_s \xi_s^u \, ds, \ 0 \leq t \leq T.$$

We have

$$x_t^u = x_t^0 + \xi_t^u, \ z_t^u = z_t^0 + \eta_t^u, \ 0 \leq t \leq T,$$

where x^0 and z^0 are the random processes (5.1) and (5.2), respectively, corresponding to $u = 0$. Fix arbitrary $0 < t \leq T$ and define the operator

$$\Psi_t : L_2(0, t; U) \to L_2(0, t; Z)$$

by

$$[\Psi_t v]_s = \int_0^s \int_0^r C_r \mathcal{U}_{r,\sigma} B_\sigma v_\sigma \, d\sigma \, dr, \ 0 \leq s \leq t, \ v \in L_2(0, t; U).$$

Obviously, $\Psi_t \in \mathcal{L}(L_2(0, t; U), L_2(0, t; Z))$. Using the operator Ψ_t, one can write $\eta^{u,t} = \Psi_t u^t$. Therefore,

$$z^{u,t} = z^{0,t} + \Psi_t u^t, \ 0 < t \leq T. \tag{5.5}$$

On the other hand $u \in \tilde{U}$ implies

$$u_s \in U_s^u \subset U_t^u = L_2\big(\Omega, \sigma\big(z^{u,t}\big), \mathbf{P}, U\big)$$

for a.e. $s \in (0, t]$. Therefore,

$$u^t \in L_2\big(\Omega, \sigma\big(z^{u,t}\big), \mathbf{P}, L_2(0, t; U)\big).$$

By Proposition 2.27, there exists

$$\Phi_t \in L_2(L_2(0, t; Z), \mathbf{P}_{z^{u,t}}, L_2(0, t; U)),$$

where $\mathbf{P}_{z^{u,t}}$ is the distribution of $z^{u,t}$, such that $u^t = \Phi_t\big(z^{u,t}\big)$. So, using (5.5), we have

$$z^{0,t} = (I - \Psi_t \Phi_t)\big(z^{u,t}\big), \ 0 < t \leq T.$$

Thus, by Proposition 2.28(b),

$$L_2\big(\Omega, \sigma\big(z^{0,t}\big), \mathbf{P}, U\big) \subset L_2\big(\Omega, \sigma\big(z^{u,t}\big), \mathbf{P}, U\big),$$

i.e., $U_t^0 \subset U_t^u$, $0 < t \leq T$. □

Lemma 5.4. *If $u \in \tilde{U}^0$, then $U_t^u \subset U_t^0$, $0 < t \leq T$.*

Proof. Under notation introduced in proving Lemma 5.3, consider the equality (5.5). In a similar way as in proving Lemma 5.3 we have $u^t = \tilde{\Phi}_t(z^{0,t})$ for some $\tilde{\Phi}_t \in L_2(L_2(0,t;Z), \mathbf{P}_{z^{0,t}}, L_2(0,t;U))$, where $\mathbf{P}_{z^{0,t}}$ is the distribution of $z^{0,t}$. Hence, by (5.5),

$$z^{u,t} = \big(I + \Psi_t\tilde{\Phi}_t\big)\big(z^{0,t}\big), \ 0 < t \le T.$$

This implies $U_t^u \subset U_t^0$, $0 < t \le T$. □

Proposition 5.5. *If* $u \in U_{\rm ad}$, *then* $U_t^u = U_t^0$, $0 < t \le T$.

Proof. Since $U_{\rm ad} = \tilde{U}^0 \cap \tilde{U}$, the statement follows from Lemmas 5.3 and 5.4. □

Remark 5.6. In Bensoussan and Viot [33] it is shown that the condition $u \in \tilde{U}$ is not sufficient for $U_t^u = U_t^0$, $0 < t \le T$.

For the further study of admissible controls, we define the Hilbertian sum

$$\tilde{U}_\varepsilon^u = \int_0^T U_{t-\varepsilon}^u \, dt,$$

where $\varepsilon > 0$, $u \in L_2(\mathbf{T} \times \Omega, U)$ and $U_{t-\varepsilon}^u = U$ if $0 < t \le \varepsilon$. If $u = 0$, then we write $\tilde{U}_\varepsilon^0 = \tilde{U}_\varepsilon^u$.

Lemma 5.7. *If* $\varepsilon > 0$ *and if* $u \in \tilde{U}_\varepsilon^0$, *then* $U_t^u = U_t^0$ *for all* $0 < t \le T$.

Proof. Suppose $u \in \tilde{U}_\varepsilon^0$. First, we will show that $U_t^0 \subset U_t^u$ for $0 < t \le T$. Since $U_t^u = U_t^0 = U$ for $-\varepsilon < t \le 0$, it is sufficient to show that $U_t^0 \subset U_t^u$ for $-\varepsilon < t \le a$ implies $U_t^0 \subset U_t^u$ for $a < t \le a + \varepsilon$, where $-\varepsilon < a \le T - \varepsilon$. Fix t satisfying $a < t \le a + \varepsilon$. By this assumption, we have

$$u_s \in U_{s-\varepsilon}^0 \subset U_{s-\varepsilon}^u \subset U_s^u \subset U_t^u, \text{ a.e. } s \in [0,t].$$

Therefore, $u^t = \Phi_t(z^{u,t})$ for some

$$\Phi_t \in L_2(L_2(0,t;Z), \mathbf{P}_{z^{u,t}}, L_2(0,t;U))$$

(see proof of Lemma 5.3). Substituting this in (5.5), we have $z^{0,t} = (I - \Psi_t\Phi_t)\big(z^{u,t}\big)$ which implies $U_t^0 \subset U_t^u$ for all $a < t \le a + \varepsilon$. So, $U_t^0 \subset U_t^u$ for all $0 < t \le T$. On the other hand by Lemma 5.4, $u \in \tilde{U}_\varepsilon^0 \subset \tilde{U}^0$ implies $U_t^u \subset U_t^0$, $0 < t \le T$. Thus, $U_t^u = U_t^0$ for all $0 < t \le T$. □

Lemma 5.8. *If* $\varepsilon > 0$ *and if* $u \in \tilde{U}_\varepsilon^u$, *then* $U_t^u = U_t^0$ *for all* $0 < t \le T$.

Proof. This can be proved in a similar way as Lemma 5.7. □

Remark 5.9. By Lemma 5.3 (Lemma 5.4), the condition $u \in \tilde{U}$ ($u \in \tilde{U}^0$) implies only one side inclusion $U_t^0 \subset U_t^u$ ($U_t^u \subset U_t^0$) for $0 < t \le T$. But, under the condition of Lemma 5.7 or 5.8, we have the equality $U_t^0 = U_t^u$ for $0 < t \le T$.

Proposition 5.10. *U_{ad} is dense in $\tilde{U}^0$.*

Proof. Let $u \in \tilde{U}^0$. For $0 < \varepsilon \le T$, denote

$$u_t^\varepsilon = \begin{cases} u_{t-\varepsilon}, & \varepsilon < t \le T, \\ 0, & 0 \le t \le \varepsilon. \end{cases}$$

Obviously, u^ε converges to u in $L_2(\mathbf{T} \times \Omega, U)$ and, hence, in $\tilde{U}^0$ as $\varepsilon \to 0$. On the other hand $u \in \tilde{U}^0$ implies $u^\varepsilon \in \tilde{U}_\varepsilon^0$. Hence, $u^\varepsilon \in \tilde{U}_\varepsilon^0 \subset \tilde{U}^0$ and, by Lemma 5.7, $u^\varepsilon \in \tilde{U}_\varepsilon^{u^\varepsilon} \subset \tilde{U}^{u^\varepsilon}$. This yields $u^\varepsilon \in \tilde{U}^0 \cap \tilde{U}^{u^\varepsilon}$, i.e., $u^\varepsilon \in U_{\mathrm{ad}}$. Thus, U_{ad} is dense in $\tilde{U}^0$. □

Proposition 5.11. *If $\varepsilon > 0$ and if $u \in \tilde{U}_\varepsilon^u$, then $u \in U_{\mathrm{ad}}$.*

Proof. Let $u \in \tilde{U}_\varepsilon^u$. Then, by Lemma 5.8, we have $u \in \tilde{U}_\varepsilon^0 \subset \tilde{U}^0$. Also, $u \in \tilde{U}_\varepsilon^u \subset \tilde{U}^u$. Hence, $u \in \tilde{U}^0 \cap \tilde{U}^u = U_{\mathrm{ad}}$. □

Thus, by Proposition 5.11, the set of admissible controls U_{ad} contains all functions in $L_2(\mathbf{T} \times \Omega, U)$ which are adapted with respect to the partial observations having a small delay. Since in practice, for physical reasons, designing controls for current time takes away some time for processing observation data, the use of $\tilde{U}^0 \cap \tilde{U}$ as a set of admissible controls rather than $\tilde{U}$ is quite reasonable. The properties of $U_{\mathrm{ad}} = \tilde{U}^0 \cap \tilde{U}$ presented in Propositions 5.5 and 5.10 make U_{ad} convenient in studying optimal control problems.

Remark 5.12. If in the condition $(\mathbf{C_2})$ we take $n \in M_2^c(\mathbf{T}, Z)$ instead of $n \in M_2(\mathbf{T}, Z)$, then $z^u \in m(\Omega, C(\mathbf{T}, Z))$ (see Remark 5.1). At the same time $z^u \in m(\Omega, L_2(\mathbf{T}, Z))$. Let $z^{c,u,t}$ and $z^{u,t}$ be random variables with values in the spaces $C(0, t; Z)$ and $L_2(0, t; Z)$, respectively, obtained by restriction of z^u from $\mathbf{T} \times \Omega$ to $[0, t] \times \Omega$. Similar to the definitions of U_t^u, $\tilde{U}^0$ and $\tilde{U}$, one can define

$$U_t^{c,u} = L_2\left(\Omega, \sigma\left(z^{c,u,t}\right), \mathbf{P}, U\right), \quad \tilde{U}^{c,0} = \int_0^T U_t^{c,0}\, dt,$$

$$\tilde{U}^c = \left\{u \in L_2(\mathbf{T} \times \Omega, U) : u_t \in U_t^{c,u} \text{ for a.e. } t \in \mathbf{T}\right\}.$$

In Bensoussan and Viot [33] (as well as in Curtain and Ichikawa [38] and in Bashirov [7]) the set $U_{\mathrm{ad}}^c = \tilde{U}^{c,0} \cap \tilde{U}^c$ is taken as a set of admissible controls. By Proposition 2.31(b), we have $U_{\mathrm{ad}}^c = U_{\mathrm{ad}}$, i.e., the sets U_{ad}^c and U_{ad} are equal. But, U_{ad}^c can be defined only for observations with continuous paths, whereas the definition of U_{ad} does not use the continuity of observation paths. Therefore, we conclude that U_{ad} is an extension of U_{ad}^c so that this extension allows observations with continuous as well as discontinuous paths.

Proposition 5.13. $\inf_{U_{\mathrm{ad}}} J(u) = \inf_{\tilde{U}^0} J(u)$.

Proof. One can observe that (5.1) defines a continuous operator $u \to x^u$ from $\tilde{U}^0$ to $L_\infty(\mathbf{T}, L_2(\Omega, X))$. So, J is continuous on $\tilde{U}^0$. Hence, the statement follows from Proposition 5.10. □

5.2.2 Extended Separation Principle

Lemma 5.14. *If the functional J takes its minimum value on $\tilde{U}^0$ at $u^* \in \tilde{U}^0$, then*

$$u_t^* = G_t^{-1}\mathbf{E}_t^0(B_t^* y_t - L_t x_t^* - g_t), \ \text{a.e. } t \in \mathbf{T}, \tag{5.6}$$

where $x^* = x^{u^*}$, $\mathbf{E}_t^0 = \mathbf{E}(\cdot\,|z^{0,t})$ and

$$y_t = \ -\mathcal{U}_{T,t}^*(Q_T x_T^* + q) - \int_t^T \mathcal{U}_{s,t}^*(F_s x_s^* + L_s^* u_s^* + f_s)\, ds, \ 0 \le t \le T. \tag{5.7}$$

Proof. Let $u \in \tilde{U}^0$. Since $\tilde{U}^0$ is a Hilbert space, we have $u^* + \lambda u \in \tilde{U}^0$ for all $\lambda \in \mathbb{R}$. Denote

$$h_t = \int_0^t \mathcal{U}_{t,s} B_s u_s\, ds, \ 0 \le t \le T.$$

One can verify that

$$\begin{aligned} 0 &\le J(u^* + \lambda u) - J(u^*) \\ &= 2\lambda \mathbf{E} \int_0^T \langle G_t u_t^* - B_t^* y_t + L_t x_t^* + g_t, u_t\rangle\, dt \\ &\quad + \lambda^2 \mathbf{E} \int_0^T \left\langle h_t, \left(F_t - L_t^* G_t^{-1} L_t\right) h_t\right\rangle\, dt \\ &\quad + \lambda^2 \mathbf{E} \int_0^T \left\langle u_t + G_t^{-1} L_t h_t, G_t\left(u_t + G_t^{-1} L_t h_t\right)\right\rangle\, dt + \lambda^2 \mathbf{E}\langle h_T, Q_T h_T\rangle. \end{aligned}$$

Dividing both sides of the above inequality consequently by $\lambda > 0$ and by $\lambda < 0$ and then tending λ to 0, we obtain

$$\mathbf{E} \int_0^T \langle G_t u_t^* - B_t^* y_t + L_t x_t^* + g_t, u_t\rangle\, dt = 0.$$

Since u is arbitrary in $\tilde{U}^0$, the equality (5.6) holds. □

Now let $u^* \in U_{\text{ad}}$ be an optimal control in the problem (5.1)–(5.4). By Proposition 5.13, the functional J takes its minimum value on $\tilde{U}^0$ at the control $u^* \in U_{\text{ad}} \subset \tilde{U}^0$. Hence, by Lemma 5.14, the equalities (5.6) and (5.7) hold. Let $0 < \tau \le t \le T$. Substituting (5.6) in (5.1) and (5.7) and using Proposition 4.29, we have

$$\begin{aligned} x_t^* &= \mathcal{U}_{t,\tau} x_\tau^* + \int_\tau^t \mathcal{U}_{t,s}\, dm_s + \int_\tau^t \mathcal{U}_{t,s}\left(B_s G_s^{-1}\mathbf{E}_s^0(B_s^* y_s - L_s x_s^* - g_s) + b_s\right) ds, \\ y_t &= -\mathcal{U}_{T,t}^*(Q_T x_T^* + q) - \int_t^T \mathcal{U}_{s,t}^*\left(F_s x_s^* + L_s^* G_s^{-1}\mathbf{E}_s^0(B_s^* y_s - L_s x_s^* - g_s) + f_s\right) ds. \end{aligned}$$

Since $\mathbf{E}^0_\tau \mathbf{E}^0_s \xi = \mathbf{E}^0_\tau \xi$ for $0 < \tau \le s \le T$, we obtain

$$\mathbf{E}^0_\tau x^*_t = \mathbf{E}^0_\tau \Big(\mathcal{U}_{t,\tau} x^*_\tau + \int_\tau^t \mathcal{U}_{t,s}\, dm_s + \int_\tau^t \mathcal{U}_{t,s} \left(B_s G_s^{-1} (B^*_s y_s - L_s x^*_s - g_s) + b_s \right) ds \Big), \tag{5.8}$$

$$\mathbf{E}^0_\tau y_t = -\,\mathbf{E}^0_\tau \Big(\mathcal{U}^*_{T,t} (Q_T x^*_T + q) + \int_t^T \mathcal{U}^*_{s,t} \big(F_s x^*_s + L^*_s G_s^{-1} (B^*_s y_s - L_s x^*_s - g_s) + f_s \big) ds \Big). \tag{5.9}$$

Lemma 5.15. *The following equality holds:*

$$\mathbf{E}^0_\tau (y_t + Q_t x^*_t + \alpha_t) = 0, \;\; 0 < \tau \le t \le T, \tag{5.10}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q *is a solution of the equation* (3.9) *and*

$$\alpha_t = \mathcal{Y}^*_{T,t} q + \int_t^T \mathcal{Y}^*_{s,t} \left(Q_s b_s - (Q_s B_s + L^*_s) G_s^{-1} g_s + f_s \right) ds + \int_t^T \mathcal{Y}^*_{s,t} Q_s\, dm_s, \;\; 0 \le t \le T. \tag{5.11}$$

Proof. See Section 5.2.4. □

Theorem 5.16 (Extended Separation Principle). *Under the conditions* $(\mathbf{C_1})$, $(\mathbf{C_2})$, $(\mathbf{C'_3})$, $(\mathbf{C_4})$ *and* $(\mathbf{RE})$, *let* $u^* \in U_{\rm ad}$ *be an optimal control in the problem* (5.1)–(5.4). *Denote*

$$x^* = x^{u^*}, \;\; \mathbf{E}^*_t = \mathbf{E}\big(\cdot \,|z^{u^*,t} \big), \;\; 0 < t \le T.$$

Then u^* *has the form*

$$u^*_t = -\,G_t^{-1} (B^*_t Q_t + L_t)\, \mathbf{E}^*_t x^*_t - G_t^{-1} B^*_t\, \mathbf{E}^*_t \alpha_t - G_t^{-1}\, \mathbf{E}^*_t g_t, \;\; \text{a.e. } t \in \mathbf{T}, \tag{5.12}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q *is a solution of the equation* (3.9) *and* α *is defined by* (5.11). *This optimal control is unique if the equation* (3.9) *has a unique solution.*

Proof. Substituting $\tau = t$ in (5.10), we have

$$\mathbf{E}^0_t y_t = -\,\mathbf{E}^0_t (Q_t x^*_t + \alpha_t), \;\; 0 < t \le T.$$

Since $u^* \in U_{\rm ad}$, by Proposition 5.5, $U^{u^*}_t = U^0_t$ for all $0 < t \le T$. This implies $\mathbf{E}^*_t = \mathbf{E}^0_t$ for all $0 < t \le T$. So, from (5.6), we obtain (5.12). Obviously, (5.12) determines a unique optimal control (if it exists) when Q is a unique solution of the equation (3.9). □

Theorem 5.16 expresses the separation principle when the relation between the noise processes acting on state and observations is arbitrary. By (5.12), designing an optimal control is separated into two steps, the first of them being computing the conditional expectations $\mathbf{E}_t^* x_t^*$, $\mathbf{E}_t^* \alpha_t$, $\mathbf{E}_t^* g_t$ and the second finding an optimal control by a deterministic method using the results of the first step.

By (5.12) and (5.11), an optimal control at current time t depends on random processes b and m on $(t, T]$. This is natural as we consider arbitrarily related noise processes and, therefore, observations available up to time t may contain information about the future of signal noise, i.e., the σ-algebras $\sigma\big(z^{u^*,t}\big)$ and $\sigma\big(b^{t+}, m_s - m_t;\, t < s \le T\big)$ may be dependent (see Remark 5.2). The formulae (5.12) and (5.11) show how this information must be used in optimal control.

The statement of Theorem 5.16 will be called the *extended separation principle*.

5.2.3 Classical Separation Principle

The following particular case of Theorem 5.16 will be called a *classical separation principle*.

Theorem 5.17 (Classical Separation Principle). *Under the conditions* $(\mathbf{C_1})$, $(\mathbf{C_2})$, $(\mathbf{C_3'})$, $(\mathbf{RE})$ *and*

$(\mathbf{C_4'})$ $q \in X$, $f \in L_2(\mathbf{T}, X)$, $g \in L_2(\mathbf{T}, U)$;

$(\mathbf{C_5})$ $\sigma\big(l, b^t, c^t, n_s, m_s;\, 0 \le s \le t\big)$ *and* $\sigma\big(b^{t+}, m_s - m_t;\, t < s \le T\big)$ *are independent for all* $0 < t \le T$;

let $u^* \in U_{\mathrm{ad}}$ *be an optimal control in the problem* (5.1)–(5.4). *Denote*

$$x^* = x^{u^*},\ \mathbf{E}_t^* = \mathbf{E}\big(\cdot\,|z^{u^*,t}\big),\ 0 < t \le T.$$

Then u^* *has the form*

$$u_t^* = u_t^0 - G_t^{-1}(B_t^* Q_t + L_t)\,\mathbf{E}_t^* x_t^*,\ \text{a.e. } t \in \mathbf{T},$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q *is a solution of the equation* (3.9) *and* u^0 *is a function in* $L_2(\mathbf{T}, U)$ *defined by*

$$\begin{aligned} u_t^0 = &- G_t^{-1} g_t - G_t^{-1} B_t^* \mathcal{Y}_{T,t}^* q \\ &- G_t^{-1} B_t^* \int_t^T \mathcal{Y}_{s,t}^* \big(Q_s \mathbf{E} b_s - (Q_s B_s + L_s^*) G_s^{-1} g_s + f_s\big) ds,\ \text{a.e. } t \in \mathbf{T}. \end{aligned}$$

This optimal control is unique if the equation (3.9) *has a unique solution.*

Proof. By $(\mathbf{C_5})$, we have

$$\mathbf{E}_t^* \left(\int_t^T \mathcal{Y}_{s,t}^* Q_s b_s\, ds + \int_t^T \mathcal{Y}_{s,t}^* Q_s\, dm_s \right) = \int_t^T \mathcal{Y}_{s,t}^* Q_s\, \mathbf{E} b_s ds.$$

So, the statement follows from Theorem 5.16. □

Comparing Theorems 5.16 and 5.17, one can observe that Theorem 5.16 removes the only essential restriction (that is the condition $(\mathbf{C_5})$) under which the classical separation principle is true. This weakening completes the form of an optimal control with additional terms.

5.2.4 Proof of Lemma 5.15

First, we will derive an expression for $\mathbf{E}^0_\tau Q_t x^*_t$. For convenience, let

$$\tilde{F}_t = F_t - (Q_t B_t + L^*_t) G^{-1}_t (B^*_t Q_t + L_t), \ \text{a.e. } t \in \mathbf{T}$$

and let

$$h_t = B_t G^{-1}_t (B^*_t y_t - L_t x^*_t - g_t) + b_t, \ \text{a.e. } t \in \mathbf{T}.$$

Then, by (3.9) and (5.8), we have

$$\begin{aligned}
\mathbf{E}^0_\tau Q_t x^*_t &= \mathbf{E}^0_\tau \Big(\mathcal{U}^*_{T,t} Q_T \mathcal{U}_{T,t} x^*_t + \int_t^T \mathcal{U}^*_{s,t} \tilde{F}_s \mathcal{U}_{s,t} x^*_t \, ds \Big) \\
&= \mathbf{E}^0_\tau \Big(\mathcal{U}^*_{T,t} Q_T \Big(x^*_T - \int_t^T \mathcal{U}_{T,r} \, dm_r - \int_t^T \mathcal{U}_{T,r} h_r \, dr \Big) \\
&\quad + \int_t^T \mathcal{U}^*_{s,t} \tilde{F}_s \Big(x^*_s - \int_t^s \mathcal{U}_{s,r} \, dm_r - \int_t^s \mathcal{U}_{s,r} h_r \, dr \Big) ds \Big) \\
&= \mathbf{E}^0_\tau \Big(\mathcal{U}^*_{T,t} Q_T x^*_T + \int_t^T \mathcal{U}^*_{s,t} \tilde{F}_s x^*_s \, ds \\
&\quad - \int_t^T \mathcal{U}^*_{r,t} \Big(\mathcal{U}^*_{T,r} Q_T \mathcal{U}_{T,r} + \int_r^T \mathcal{U}^*_{s,r} \tilde{F}_s \mathcal{U}_{s,r} \, ds \Big) dm_r \\
&\quad - \int_t^T \mathcal{U}^*_{r,t} \Big(\mathcal{U}^*_{T,r} Q_T \mathcal{U}_{T,r} + \int_r^T \mathcal{U}^*_{s,r} \tilde{F}_s \mathcal{U}_{s,r} \, ds \Big) h_r \, dr \Big) \\
&= \mathbf{E}^0_\tau \Big(\mathcal{U}^*_{T,t} Q_T x^*_T + \int_t^T \mathcal{U}^*_{s,t} \tilde{F}_s x^*_s \, ds \\
&\quad - \int_t^T \mathcal{U}^*_{s,t} Q_s \, dm_s - \int_t^T \mathcal{U}^*_{s,t} Q_s h_s \, ds \Big) \\
&= \mathbf{E}^0_\tau \Big(\mathcal{U}^*_{T,t} Q_T x^*_T - \int_t^T \mathcal{U}^*_{s,t} Q_s \, dm_s \\
&\quad + \int_t^T \mathcal{U}^*_{s,t} \left(F_s - (Q_s B_s + L^*_s) G^{-1}_s B^*_s Q_s - L^*_s G^{-1}_s L_s \right) x^*_s \, ds \\
&\quad + \int_t^T \mathcal{U}^*_{s,t} \big(Q_s B_s G^{-1}_s g_s - Q_s B_s G^{-1}_s B^*_s y_s - Q_s b_s \big) ds \Big). \qquad (5.13)
\end{aligned}$$

Also, since $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, by Proposition 4.28, for the random variable α defined by (5.11), we have

$$\alpha_t = \mathcal{U}_{T,t}^* q + \int_t^T \mathcal{U}_{s,t}^* \big(Q_s b_s - (Q_s B_s + L_s^*) G_s^{-1} g_s + f_s \\ - (Q_s B_s + L_s^*) G_s^{-1} B_s^* \alpha_s\big) ds + \int_t^T \mathcal{U}_{s,t}^* Q_s \, dm_s. \tag{5.14}$$

Therefore, from (5.9), (5.13) and (5.14), we obtain

$$\mathbf{E}_\tau^0 (y_t + Q_t x_t^* + \alpha_t) = \int_t^T \mathcal{U}_{s,t}^* (Q_s B_s + L_s^*) G_s^{-1} B_s^* \, \mathbf{E}_\tau^0 (y_s + Q_s x_s^* + \alpha_s) \, ds.$$

If

$$\lambda_{t,\tau} = \left\| \mathbf{E}_\tau^0 (y_t + Q_t x_t^* + \alpha_t) \right\|,$$

then

$$\lambda_{t,\tau} \le c \int_t^T \lambda_{s,\tau} \, ds, \;\; c = \text{const.} \ge 0,$$

which by Theorem 2.45, implies $\lambda_{t,\tau} = 0$, $0 < \tau \le t \le T$. The proof is completed.

5.3 Generalization to a Game Problem

In fact, an optimal control problem is a simple game of one player which tries to minimize (or maximize) a cost functional. Generally, a game is played by two or more players each of which has his (her) own profit. In this section a game of two players will be considered. One of them will try to minimize and the other to maximize a given functional. Under a linear state-observation system and a quadratic functional, the results of Section 5.2 will be generalized to this game.

5.3.1 Setting of Game Problem

Consider the state system (5.1), in which the Hilbert space U is decomposed into the product of two Hilbert spaces U_1 and U_2. In this case, each function $u \in L_2(\mathbf{T} \times \Omega, U)$ can be represented as a pair of $u_1 \in L_2(\mathbf{T} \times \Omega, U_1)$ and $u_2 \in L_2(\mathbf{T} \times \Omega, U_2)$. The functions u_1 and u_2 will be considered as control actions of the first and second players, respectively. In view of decomposition $U = U_1 \times U_2$, the functions B, L, G and g from (5.1) and (5.4) could be decomposed in a respective way. Also, it will be supposed that each player has his (her) own observation system in the form (5.2). Realizing the above, we obtain the state-observation system

$$
\begin{aligned}
x_t^{u_1,u_2} &= \mathcal{U}_{t,0} l + \int_0^t \mathcal{U}_{t,s} \left([B_{1,s}\ B_{2,s}] \begin{bmatrix} u_{1,s} \\ u_{2,s} \end{bmatrix} + b_s \right) ds \\
&\quad + \int_0^t \mathcal{U}_{t,s}\, dm_s,\ 0 \le t \le T, \qquad (5.15)\\
z_{i,t}^{u_1,u_2} &= \int_0^t (C_{i,s} x_s^{u_1,u_2} + c_{i,s})\, ds + n_{i,t},\ 0 \le t \le T,\ i = 1,2, \qquad (5.16)
\end{aligned}
$$

and the quadratic functional

$$
\begin{aligned}
J(u_1,u_2) = \mathbf{E}\Bigg(& \langle x_T^{u_1,u_2}, Q_T x_T^{u_1,u_2} \rangle \\
& + \int_0^T \left\langle \begin{bmatrix} x_t^{u_1,u_2} \\ u_{1,t} \\ u_{2,t} \end{bmatrix}, \begin{bmatrix} F_t & L_{1,t}^* & L_{2,t}^* \\ L_{1,t} & G_{1,t} & R_t^* \\ L_{2,t} & R_t & G_{2,t} \end{bmatrix} \begin{bmatrix} x_t^{u_1,u_2} \\ u_{1,t} \\ u_{2,t} \end{bmatrix} \right\rangle dt \\
& + 2\,\langle q, x_T^{u_1,u_2} \rangle + 2\int_0^T \left\langle \begin{bmatrix} f_t \\ g_{1,t} \\ g_{2,t} \end{bmatrix}, \begin{bmatrix} x_t^{u_1,u_2} \\ u_{1,t} \\ u_{2,t} \end{bmatrix} \right\rangle dt \Bigg). \qquad (5.17)
\end{aligned}
$$

In this section we suppose that the following conditions hold:

($\mathbf{G_1}$) $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $B_i \in B_\infty(\mathbf{T}, \mathcal{L}(U_i, X))$, $C_i \in B_\infty(\mathbf{T}, \mathcal{L}(X, Z_i))$, $i = 1,2$;

($\mathbf{G_2}$) $l \in L_2(\Omega, X)$, $b \in L_2(\mathbf{T} \times \Omega, X)$, $c_i \in L_2(\mathbf{T} \times \Omega, Z_i)$, $m \in M_2(\mathbf{T}, X)$, $n_i \in M_2(\mathbf{T}, Z_i)$, $i = 1,2$;

($\mathbf{G_3'}$) $Q_T \in \mathcal{L}(X)$, $Q_T^* = Q_T$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G_i, G_i^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U_i))$, $R \in B_\infty(\mathbf{T}, \mathcal{L}(U_1, U_2))$, $F_t^* = F_t$, $G_{1,t} > 0$ and $G_{2,t} < 0$ for a.e. $t \in \mathbf{T}$, $L_i \in B_\infty(\mathbf{T}, \mathcal{L}(X, U_i))$, $(G_1 - R^* G_2^{-1} R)^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U_1))$, $(G_2 - R G_1^{-1} R^*)^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U_2))$, $i = 1,2$;

($\mathbf{G_4}$) $q \in L_2(\Omega, X)$, $f \in L_2(\mathbf{T} \times \Omega, X)$, $g_i \in L_2(\mathbf{T} \times \Omega, U_i)$, $i = 1,2$.

As it was mentioned above, we denote $U = U_1 \times U_2$ and

$$
u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},\ B = [B_1\ B_2],\ L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix},\ G = \begin{bmatrix} G_1 & R^* \\ R & G_2 \end{bmatrix},\ g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}.
$$

By Proposition 1.24, the condition ($\mathbf{G_3'}$) implies the existence of G_t^{-1} for a.e. $t \in \mathbf{T}$ and $G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$. Therefore, we can consider the operator Riccati equation (3.9). We suppose that

(**RE**) The equation (3.9) has a solution in $B(\mathbf{T}, \mathcal{L}(X))$.

The set of admissible controls of each player we define as in the problem (5.1)–(5.4). For $u_1 \in L_2(\mathbf{T}\times\Omega, U_1)$ and $u_2 \in L_2(\mathbf{T}\times\Omega, U_2)$, consider

$$U_{i,t}^{u_1,u_2} = L_2\big(\Omega, \sigma\big(z_i^{u_1,u_2,t}\big), \mathbf{P}, U_i\big), \ i=1,2, \ 0<t\le T,$$

where $z_i^{u_1,u_2,t}$ denotes an $L_2(0,t;Z_i)$-valued random variable, obtained by restriction of $z_i^{u_1,u_2}$ from $\mathbf{T}\times\Omega$ to $[0,t]\times\Omega$. Let

$$\tilde{U}_1^{0,u_2} = \int_0^T U_{1,t}^{0,u_2}\,dt = \big\{u_1 \in L_2(\mathbf{T}\times\Omega, U_1) : u_{1,t} \in U_{1,t}^{0,u_2} \text{ for a.e. } t\in\mathbf{T}\big\},$$

$$\tilde{U}_2^{u_1,0} = \int_0^T U_{2,t}^{u_1,0}\,dt = \big\{u_2 \in L_2(\mathbf{T}\times\Omega, U_2) : u_{2,t} \in U_{2,t}^{u_1,0} \text{ for a.e. } t\in\mathbf{T}\big\}.$$

Also, define

$$\tilde{U}_1^{u_2} = \big\{u_1 \in L_2(\mathbf{T}\times\Omega, U_1) : u_{1,t} \in U_{1,t}^{u_1,u_2} \text{ for a.e. } t\in\mathbf{T}\big\},$$
$$\tilde{U}_2^{u_1} = \big\{u_2 \in L_2(\mathbf{T}\times\Omega, U_2) : u_{2,t} \in U_{2,t}^{u_1,u_2} \text{ for a.e. } t\in\mathbf{T}\big\}.$$

The set of admissible controls in the considered game will be defined as

$$U_{\text{ad}} = \big\{(u_1,u_2) : u_1 \in \big(\tilde{U}_1^{0,u_2}\cap\tilde{U}_1^{u_2}\big), u_2 \in \big(\tilde{U}_2^{u_1,0}\cap\tilde{U}_2^{u_1}\big)\big\}, \tag{5.18}$$

where u_1 and u_2 are admissible controls of the first and second players, respectively. Note that for fixed u_1, the cross section $\tilde{U}_2^{u_1,0}\cap\tilde{U}_2^{u_1}$ of U_{ad} agrees with the set of admissible controls defined for the control problem (5.1)–(5.4). The same holds for the cross sections at fixed u_2.

Our aim in this section is to study a saddle point of the functional (5.17) on U_{ad}. A *saddle point* is a pair $(u_1^*, u_2^*) \in U_{\text{ad}}$ such that

$$J(u_1^*, u_2) \le J(u_1^*, u_2^*) \le J(u_1, u_2^*) \tag{5.19}$$

for all $u_1 \in \big(\tilde{U}_1^{0,u_2^*}\cap\tilde{U}_1^{u_2^*}\big)$ and for all $u_2 \in \big(\tilde{U}_2^{u_1^*,0}\cap\tilde{U}_2^{u_1^*}\big)$. Briefly, this problem will be called the game (5.15)–(5.18). The technique used in studying the control problem (5.1)–(5.4) will be applied to study the game (5.15)–(5.18). Two cases will be considered. In the first case the observations of the first player will be observed by the second player as well. In the second case the players will have the same observations.

5.3.2 Case 1: The First Player Has Worse Observations

The mathematical condition corresponding to the considered case is formulated as

($\mathbf{G}_5'$) $\sigma\left(z_1^{u_1,u_2,t}\right) \subset \sigma\left(z_2^{u_1,u_2,t}\right)$ for all $(u_1,u_2)\in U_{\text{ad}}$ and for all $0<t\le T$.

The condition $(\mathbf{G_5'})$ means that the second player is completely informed about the observations of the first player, i.e., the first player has worse observations than the second one.

Consider the game (5.15)–(5.18) under the conditions $(\mathbf{G_1})$, $(\mathbf{G_2})$, $(\mathbf{G_3'})$, $(\mathbf{RE})$, $(\mathbf{G_4})$ and $(\mathbf{G_5'})$. Suppose $(u_1^*, u_2^*) \in U_{\text{ad}}$ is a saddle point in the game (5.15)–(5.18). Denote

$$u^* = \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix}, \ x^* = x^{u_1^*, u_2^*}, \ \mathbf{E}_{i,t}^{u_1,u_2} = \mathbf{E}\big(\cdot \,|z_i^{u_1,u_2,t}\big), \ i = 1, 2, \ 0 < t \le T,$$

and consider the random process y defined by (5.7). Since the functional (5.17) takes its minimum value at

$$u_1^* \in \big(\tilde{U}_1^{0,u_2^*} \cap \tilde{U}_1^{u_2^*}\big) \subset \tilde{U}_1^{0,u_2^*}$$

under fixed $u_2 = u_2^*$, then by Lemma 5.14,

$$u_{1,t}^* = G_{1,t}^{-1} \mathbf{E}_{1,t}^{0,u_2^*} (B_{1,t}^* y_t - L_{1,t} x_t^* - g_{1,t} - R_t^* u_{2,t}^*), \text{ a.e. } t \in \mathbf{T}. \tag{5.20}$$

In a similar way, we have

$$u_{2,t}^* = G_{2,t}^{-1} \mathbf{E}_{2,t}^{u_1^*,0} (B_{2,t}^* y_t - L_{2,t} x_t^* - g_{2,t} - R_t u_{1,t}^*), \text{ a.e. } t \in \mathbf{T}. \tag{5.21}$$

According to the condition $(\mathbf{G_5'})$ and Proposition 5.5,

$$U_{1,t}^{0,u_2^*} = U_{1,t}^{u_1^*,u_2^*} \subset U_{2,t}^{u_1^*,u_2^*} = U_{2,t}^{u_1^*,0}, \ 0 < t \le T. \tag{5.22}$$

Hence,

$$\mathbf{E}_{2,t}^{u_1^*,0} u_{1,t}^* = \mathbf{E}_{2,t}^{u_1^*,0} \left(\mathbf{E}_{1,t}^{0,u_2^*} u_{1,t}^*\right) = \mathbf{E}_{1,t}^{0,u_2^*} u_{1,t}^* = u_{1,t}^*, \text{ a.e. } t \in \mathbf{T}.$$

So, by (5.21),

$$u_{2,t}^* = G_{2,t}^{-1} \mathbf{E}_{2,t}^{u_1^*,0} (B_{2,t}^* y_t - L_{2,t} x_t^* - g_{2,t}) - G_{2,t}^{-1} R_t u_{1,t}^*, \text{ a.e. } t \in \mathbf{T}.$$

Using this equality in (5.20), we obtain

$$\begin{aligned} u_{1,t}^* = {} & \big(G_{1,t} - R_t^* G_{2,t}^{-1} R_t\big)^{-1} \mathbf{E}_{1,t}^{0,u_2} \big(\big(B_{1,t}^* - R_t^* G_{2,t}^{-1} B_{2,t}^*\big) y_t \\ & - \big(L_{1,t} - R_t^* G_{2,t}^{-1} L_{2,t}\big) x_t^* - g_{1,t} + R_t^* G_{2,t}^{-1} g_{2,t}\big), \text{ a.e. } t \in \mathbf{T}. \end{aligned} \tag{5.23}$$

Now let $0 < \tau \le t \le T$. Using (5.22) in the formulae (5.20) and (5.21), we have

$$\mathbf{E}_{1,\tau}^{0,u_2^*} u_{1,t}^* = G_{1,t}^{-1} \mathbf{E}_{1,\tau}^{0,u_2^*} (B_{1,t}^* y_t - L_{1,t} x_t^* - g_{1,t} - R_t^* u_{2,t}^*), \text{ a.e. } t \in [\tau, T],$$
$$\mathbf{E}_{1,\tau}^{0,u_2^*} u_{2,t}^* = G_{2,t}^{-1} \mathbf{E}_{1,\tau}^{0,u_2^*} (B_{2,t}^* y_t - L_{2,t} x_t^* - g_{2,t} - R_t u_{1,t}^*), \text{ a.e. } t \in [\tau, T],$$

which imply

$$\mathbf{E}_{1,\tau}^{0,u_2^*}(G_{1,t}u_{1,t}^* + R_t^* u_{2,t}^*) = \mathbf{E}_{1,\tau}^{0,u_2^*}(B_{1,t}^* y_t - L_{1,t}x_t^* - g_{1,t}), \text{ a.e. } t \in [\tau, T],$$
$$\mathbf{E}_{1,\tau}^{0,u_2^*}(R_1 u_{1,t}^* + G_{2,t}u_{2,t}^*) = \mathbf{E}_{1,\tau}^{0,u_2^*}(B_{2,t}^* y_t - L_{2,t}x_t^* - g_{2,t}), \text{ a.e. } t \in [\tau, T].$$

Using the notation introduced above, we can write

$$G_t \mathbf{E}_{1,\tau}^{0,u_2^*} u_t^* = \mathbf{E}_{1,\tau}^{0,u_2^*} \left(B_t^* y_t - L_t x_t^* - g_t\right), \text{ a.e. } t \in [\tau, T],$$

which implies

$$\mathbf{E}_{1,\tau}^{0,u_2^*} u_t^* = G_t^{-1} \mathbf{E}_{1,\tau}^{0,u_2^*} \left(B_t^* y_t - L_t x_t^* - g_t\right), \text{ a.e. } t \in [\tau, T]. \tag{5.24}$$

Similar to the derivation of the formulae (5.8) and (5.9), substituting (5.24) in (5.15) and (5.7), for all $0 < \tau \le t \le T$, we have

$$\begin{aligned}
\mathbf{E}_{1,\tau}^{0,u_2^*} x_t^* &= \mathbf{E}_{1,\tau}^{0,u_2^*} \bigg(\mathcal{U}_{t,\tau} x_\tau^* + \int_\tau^t \mathcal{U}_{t,s}\, dm_s \\
&\quad + \int_\tau^t \mathcal{U}_{t,s} \big(B_s G_s^{-1}(B_s^* y_s - L_s x_s^* - g_s) + b_s\big) ds \bigg),
\end{aligned} \tag{5.25}$$

$$\begin{aligned}
\mathbf{E}_{1,\tau}^{0,u_2^*} y_t &= -\mathbf{E}_{1,\tau}^{0,u_2^*} \bigg(\mathcal{U}_{T,t}^* (Q_T x_T^* + q) \\
&\quad + \int_t^T \mathcal{U}_{s,t}^* \big(F_s x_s^* + L_s^* G_s^{-1}(B_s^* y_s - L_s x_s^* - g_s) + f_s\big) ds \bigg).
\end{aligned} \tag{5.26}$$

Applying Lemma 5.15 to (5.25)–(5.26), we obtain

$$\mathbf{E}_{1,\tau}^{0,u_2^*} y_t = -\mathbf{E}_{1,\tau}^{0,u_2^*}(Q_t x_t^* + \alpha_t), \ 0 < \tau \le t \le T, \tag{5.27}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q is a solution of the equation (3.9) and α is defined by (5.11). Using (5.27) in (5.23), we have

$$\begin{aligned}
u_{1,t}^* = &-\big(G_{1,t} - R_t^* G_{2,t}^{-1} R_t\big)^{-1} \mathbf{E}_{1,t}^{0,u_2^*} \big(\big((B_{1,t}^* - R_t^* G_{2,t}^{-1} B_{2,t}^*)Q_t \\
&+ L_{1,t} - R_t^* G_{2,t}^{-1} L_{2,t}\big)x_t^* + \big(B_{1,t}^* - R_t^* G_{2,t}^{-1} B_{2,t}^*\big)\alpha_t \\
&+ g_{1,t} - R_t^* G_{2,t}^{-1} g_{2,t}\big), \text{ a.e. } t \in \mathbf{T}.
\end{aligned} \tag{5.28}$$

Thus, we can state the following theorem.

Theorem 5.18. *Under the conditions* $(\mathbf{G_1}), (\mathbf{G_2}), (\mathbf{G_3'}), (\mathbf{RE}), (\mathbf{G_4})$ *and* $(\mathbf{G_5'})$, *let* $(u_1^*, u_2^*) \in U_{\rm ad}$ *be a saddle point in the game* (5.15)–(5.18). *Denote*

$$x^* = x^{u_1^*, u_2^*}, \ \mathbf{E}_{1,t}^{u_1^*, u_2^*} = \mathbf{E}\big(\cdot \,\big|z_1^{u_1^*, u_2^*, t}\big).$$

Then u_1^* has the form

$$u_{1,t}^* = -\big(G_{1,t} - R_t^* G_{2,t}^{-1} R_t\big)^{-1} \mathbf{E}_{1,t}^{u_1^*,u_2^*} \big(\big((B_{1,t}^* - R_t^* G_{2,t}^{-1} B_{2,t}^*)Q_t + L_{1,t} - R_t^* G_{2,t}^{-1} L_{2,t}\big)x_t^* + \big(B_{1,t}^* - R_t^* G_{2,t}^{-1} B_{2,t}^*\big)\alpha_t + g_{1,t} - R_t^* G_{2,t}^{-1} g_{2,t}\big), \quad a.e.\ t \in \mathbf{T}, \tag{5.29}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q is a solution of the equation (3.9) *and α is defined by* (5.11).

Proof. By (5.22), $\mathbf{E}_{1,t}^{0,u_2^*} = \mathbf{E}_{1,t}^{u_1^*,u_2^*}$, $0 < t \le T$. Hence, (5.29) follows from (5.28). □

Theorem 5.18 can be interpreted in the following way: since by $(\mathbf{G}_5')$, the second player is completely informed about the observations of the first player, he (she) can compute the u_1^* component of a saddle point. But, generally, the first player can not do the same about u_2^* as he (she) has worse observations.

5.3.3 Case 2: The Players Have the Same Observations

The mathematical condition corresponding to this case is formulated as

$(\mathbf{G}_5)$ $\sigma\left(z_1^{u_1,u_2,t}\right) = \sigma\left(z_2^{u_1,u_2,t}\right)$ for all $(u_1, u_2) \in U_{\text{ad}}$ and for all $0 < t \le T$.

In particular, the condition $(\mathbf{G}_5)$ holds if

$$\begin{cases} Z_1 = Z_2 = Z, \ C_1 = C_2 = C \in B_\infty(\mathbf{T}, \mathcal{L}(X, Z)), \\ c_1 = c_2 = c \in L_2(\mathbf{T} \times \Omega, Z), \ n_1 = n_2 = n \in M_2(\mathbf{T}, Z). \end{cases}$$

Obviously, the condition $(\mathbf{G}_5)$ implies $\mathbf{E}_{1,t}^{u_1^*,u_2^*} = \mathbf{E}_{2,t}^{u_1^*,u_2^*}$.

Theorem 5.19. *Under the conditions $(\mathbf{G}_1)$, $(\mathbf{G}_2)$, $(\mathbf{G}_3')$, $(\mathbf{RE})$, $(\mathbf{G}_4)$ and $(\mathbf{G}_5)$, let $u^* = (u_1^*, u_2^*) \in U_{\text{ad}}$ be a saddle point in the game* (5.15)–(5.18). *Denote*

$$u^* = \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix}, \ x^* = x^{u_1^*,u_2^*}, \ \mathbf{E}_t^* = \mathbf{E}_{1,t}^{u_1^*,u_2^*} = \mathbf{E}_{2,t}^{u_1^*,u_2^*}.$$

Then u^* has the form

$$u_t^* = -G_t^{-1}(B_t^* Q_t + L_t)\mathbf{E}_t^* x_t^* - G_t^{-1} B_t^* \mathbf{E}_t^* \alpha_t - G_t^{-1} \mathbf{E}_t^* g_t, \quad a.e.\ t \in \mathbf{T}, \tag{5.30}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q is a solution of the equation (3.9) *and α is defined by* (5.11). *A saddle point is unique if the equation* (3.9) *has a unique solution.*

Proof. By Theorem 5.18, we have (5.29). Under the condition $(\mathbf{G}_5)$, Theorem 5.18 can be applied to the second player. Therefore, we also have

$$u_{2,t}^* = -\big(G_{2,t} - R_t G_{1,t}^{-1} R_t^*\big)^{-1} \mathbf{E}_{2,t}^{u_1^*,u_2^*} \big(\big((B_{2,t}^* - R_t G_{1,t}^{-1} B_{1,t}^*)Q_t + L_{2,t} - R_t G_{1,t}^{-1} L_{1,t}\big)x_t^* + \big(B_{2,t}^* - R_t G_{1,t}^{-1} B_{1,t}^*\big)\alpha_t + g_{2,t} - R_t G_{1,t}^{-1} g_{1,t}\big), \quad \text{a.e. } t \in \mathbf{T}. \tag{5.31}$$

Combining (5.29) and (5.31) and using Proposition 1.24, we obtain (5.30). If Q is a unique solution of the equation (3.9), then, obviously, (5.30) defines a unique saddle point (if it exists). □

Finally, note that in the game considered in this section the noise processes of the state system (5.15) are arbitrarily related with the noise processes of the observation systems of the players, which are defined by (5.16). But, relation between the noise processes of the observation systems of the players is such that the condition $(\mathbf{G_5'})$ or $(\mathbf{G_5})$ holds. A general case, when the noise processes of the observation systems of the players are arbitrarily related, needs a further investigation.

5.4 Minimizing Sequence

In this section a simple idea of constructing a minimizing sequence in the problem (5.1)–(5.4) will be realized. In this section it is assumed that the conditions $(\mathbf{C_1})$, $(\mathbf{C_2})$, $(\mathbf{C_4})$ and the following hold:

$(\mathbf{C_3})$ $Q_T \in \mathcal{L}(X)$, $Q_T \geq 0$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$, $L \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \geq 0$ for a.e. $t \in \mathbf{T}$.

Note that the condition $(\mathbf{C_3})$ is stronger than the condition $(\mathbf{C_3'})$. In particular, by Theorem 3.22, the conditions $(\mathbf{C_1})$ and $(\mathbf{C_3})$ imply the condition $(\mathbf{RE})$, i.e., the existence and, moreover, the uniqueness of a solution Q of the equation (3.9) in $B(\mathbf{T}, \mathcal{L}(X))$ satisfying $Q_t \geq 0$, $0 \leq t \leq T$. In this section the problem (5.1)–(5.4) will be considered under the conditions $(\mathbf{C_1})$–$(\mathbf{C_4})$.

5.4.1 Properties of Cost Functional

Lemma 5.20. *The functional J defined by* (5.4) *is strictly convex on $\tilde{U}^0$.*

Proof. Suppose $v, w \in \tilde{U}^0$, $v \neq w$ and $0 < \lambda < 1$. From (5.1), we have

$$x^{\lambda v + (1-\lambda) w} = \lambda x^v + (1 - \lambda) x^w.$$

Using this equality, one can compute

$$\begin{aligned} J(\lambda v + (1-\lambda) w) = \lambda J(v) + (1-\lambda) J(w) \\ - \lambda(1-\lambda)\mathbf{E}\Big(\langle x_T^v - x_T^w, Q_T(x_T^v - x_T^w)\rangle \\ + \int_0^T \left\langle \begin{bmatrix} x_t^v - x_t^w \\ v_t - w_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix} \begin{bmatrix} x_t^v - x_t^w \\ v_t - w_t \end{bmatrix} \right\rangle dt \Big). \end{aligned}$$

Since $Q_T \geq 0$ and $F_t - L_t^* G_t^{-1} L_t \geq 0$ for a.e. $t \in \mathbf{T}$, we obtain

$$J(\lambda v + (1-\lambda) w) = \lambda J(v) + (1-\lambda) J(w) - \lambda(1-\lambda)\mathbf{E}\int_0^T \langle y_t^{v,w}, G_t y_t^{v,w}\rangle \, dt, \quad (5.32)$$

where

$$y_t^{v,w} = v_t - w_t + G_t^{-1}L_t(x_t^v - x_t^w), \text{ a.e. } t \in \mathbf{T}.$$

Let us show that $y_t^{v,w} = 0$ for a.e. $t \in \mathbf{T}$ can not be a case. Suppose the converse. Then from (5.1), we have

$$x_t^v - x_t^w = \int_0^t \mathcal{U}_{t,s} B_s (v_s - w_s)\, ds = -\int_0^t \mathcal{U}_{t,s} B_s G_s^{-1} L_s (x_s^v - x_s^w)\, ds.$$

Therefore, for some $c > 0$,

$$\mathbf{E}\|x_t^v - x_t^w\| \le c \int_0^t \mathbf{E}\|x_s^v - x_s^w\|\, ds.$$

Applying Theorem 2.45, we obtain $x_t^v - x_t^w = 0$ for all $0 \le t \le T$ and consequently, $v_t - w_t = 0$ for a.e. $t \in \mathbf{T}$. The last equality is contrary to the assumption $v \ne w$. Hence, $v_t - w_t + G_t^{-1}L_t(x_t^v - x_t^w) \ne 0$ on some subset of $\mathbf{T}$ which has a positive measure. Using this conclusion and $G_t > 0$, a.e. $t \in \mathbf{T}$, from (5.32), we obtain

$$J(\lambda v + (1-\lambda)w) < \lambda J(v) + (1-\lambda)J(w),$$

i.e., J is strictly convex. □

Lemma 5.21. *If $\{u^n\}$ is a sequence in $\tilde{U}^0$ satisfying $\|u^n\|_{L_2} \to \infty$ as $n \to \infty$, then $J(u^n) \to \infty$ as $n \to \infty$.*

Proof. Without loss of generality we will take $q = 0$, $f = 0$ and $g = 0$. First, let us show that $\|u^n\|_{L_2} \to \infty$ implies

$$\mathbf{E}\int_0^T \left\|u_t^n + G_t^{-1}L_t x_t^{u^n}\right\|^2 dt \to \infty. \tag{5.33}$$

Let

$$h_t^n = u_t^n + G_t^{-1}L_t x_t^{u^n}, \text{ a.e. } t \in \mathbf{T}.$$

Substituting this in (5.1), we have

$$x_t^{u^n} = \mathcal{U}_{t,0}l + \int_0^t \mathcal{U}_{t,s}\big(B_s h_s^n - B_s G_s^{-1} L_s x_s^{u^n} + b_s\big)\, ds + \int_0^t \mathcal{U}_{t,s}\, dm_s.$$

Let $\mathcal{R} = \mathcal{P}_{-BG^{-1}L}(\mathcal{U})$. By Proposition 4.28,

$$x_t^{u^n} = \mathcal{R}_{t,0}l + \int_0^t \mathcal{R}_{t,s}(B_s h_s^n + b_s)\, ds + \int_0^t \mathcal{R}_{t,s}\, dm_s, \ 0 \le t \le T. \tag{5.34}$$

(5.34) defines a continuous operator $h^n \to x^{u^n}$ from the space $L_2(\mathbf{T}, L_2(\Omega, U))$ to the space $L_2(\mathbf{T}, L_2(\Omega, X))$. Therefore, if a subsequence of $\{h^n\}$ is bounded in $L_2(\mathbf{T}, L_2(\Omega, U))$, then the corresponding subsequence of $\{x^{u^n}\}$ is bounded in

$L_2(\mathbf{T}, L_2(\Omega, X))$ too. This means the boundedness of the corresponding subsequence of $\{u^n\}$ in $L_2(\mathbf{T}, L_2(\Omega, U))$. This is contrary to $\|u^n\|_{L_2} \to \infty$. Therefore, (5.33) holds. Now let us write (5.4) in the form

$$J(u) = \mathbf{E}\bigg(\langle x_T^u, Q_T x_T^u \rangle + \int_0^T \left(\langle x_t^u, \left(F_t - L_t^* G_t^{-1} L_t\right) x_t^u \rangle \, dt \right. \\ + \int_0^T \langle u_t + G_t^{-1} L_t x_t^u, G_t \left(u_t + G_t^{-1} L_t x_t^u\right) \rangle \, dt \bigg).$$

By the condition $(\mathbf{C_3})$, $Q_T \geq 0$, $F_t - L_t^* G_t^{-1} L_t \geq 0$, $G_t > 0$ for a.e. $t \in \mathbf{T}$ and $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$. Hence, by Proposition 2.35,

$$J(u^n) \geq k^{-1} \mathbf{E} \int_0^T \left\| u_t^n + G_t^{-1} L_t x_t^{u^n} \right\|^2 dt,$$

where $k = \text{ess sup} \left\| G_t^{-1} \right\| > 0$. This shows that $J(u^n) \to \infty$. □

Proposition 5.22. *There exists a unique function $u^* \in \tilde{U}^0$ such that the functional (5.4) takes its minimum value on $\tilde{U}^0$ at u^* and*

$$u_t^* = -G_t^{-1}(B_t^* Q_t + L_t)\, \mathbf{E}_t^0 x_t^* - G_t^{-1} B_t^* \, \mathbf{E}_t^0 \alpha_t - G_t^{-1}\, \mathbf{E}_t^0 g_t, \ \ a.e.\ t \in \mathbf{T}, \tag{5.35}$$

where $x^* = x^{u^*}$, $\mathbf{E}_t^0 = \mathbf{E}(\cdot\,|z^{0,t})$, $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q is a solution of the equation (3.9) *and α is defined by* (5.11).

Proof. Since $\tilde{U}^0$ is a subspace of $L_2(\mathbf{T}, L_2(\Omega, U))$, $\tilde{U}^0$ is closed and convex. By Lemmas 5.20 and 5.21, the functional (5.4) satisfies the conditions of Proposition 2.3. So, there exists a unique $u^* \in \tilde{U}^0$ at which the functional (5.4) takes its minimum value on $\tilde{U}^0$. The formula (5.35) follows from Lemmas 5.14 and 5.15. □

5.4.2 Minimizing Sequence

Recall that a sequence $\{u^n\}$ in U_{ad} is called a *minimizing sequence* in the problem (5.1)–(5.4) if

$$J(u^n) \to \inf_{U_{\text{ad}}} J(u), \ \ n \to \infty.$$

By definition of infimum, a minimizing sequence exists always and, obviously, is not unique.

Theorem 5.23. *Under the conditions $(\mathbf{C_1})$–$(\mathbf{C_4})$, let*

$$u_t^\lambda = -\, G_t^{-1}(B_t^* Q_t + L_t) \mathbf{E}_{t-\lambda}^0 x_t^\lambda \\ -\, G_t^{-1} B_t^* \mathbf{E}_{t-\lambda}^0 \alpha_t - G_t^{-1} \mathbf{E}_{t-\lambda}^0 g_t, \ \ a.e.\ t \in \mathbf{T}, \ \lambda > 0, \tag{5.36}$$

where $x^\lambda = x^{u^\lambda}$, $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q is a solution of the equation (3.9), *α is defined by* (5.11) *and $\mathbf{E}_{t-\lambda}^0 = \mathbf{E}$ for $0 < t \leq \lambda$. Then $\{u^{\lambda_n}\}$ is a minimizing* *sequence in the problem* (5.1)–(5.4) *for any sequence* $\{\lambda_n\}$ *of real numbers with* $\lambda_n \to 0$. *If there exists an optimal control* $u^* \in U_{\text{ad}}$ *in the problem* (5.1)–(5.4), *then* u^λ *converges to* u^* *in the norm of* $L_2(\mathbf{T} \times \Omega, U)$ *as* $\lambda \to 0$.

Proof. By Proposition 5.11, $u^\lambda \in U_{\text{ad}}$ for all $0 < \lambda \le T$. On the other hand, u^λ converges to u^* defined by (5.35) in the norm of $L_2(\mathbf{T} \times \Omega, U)$ when $\lambda \to 0$. So, by Propositions 5.13 and 5.22, we have

$$J(u^\lambda) \to J(u^*) = \inf_{\tilde{U}^0} J(u) = \inf_{U_{\text{ad}}} J(u), \ \lambda \to 0.$$

If an optimal control $u^* \in U_{\text{ad}}$ in the problem (5.1)–(5.4) exists, then it has the form (5.12) as well as (5.35). Hence, $u^\lambda \to u^*$ as $\lambda \to 0$ in the norm of $L_2(\mathbf{T} \times \Omega, U)$. □

5.5 Linear Regulator Problem

A nonrandom version of the problem (5.1)–(5.4) is called a (deterministic) *linear regulator problem.*

5.5.1 Setting of Linear Regulator Problem

One can pass from the problem (5.1)–(5.4) to a linear regulator problem by taking $\mathcal{F} = \{\Omega, \varnothing\}$. In this case, all martingales in $M_2(\mathbf{T}, X)$ reduce to the zero function. Also, in this case l, b, c, q, f, g become nonrandom, the observation system (5.2) has no meaning since the nonrandom state process need not be estimated and

$$U_{\text{ad}} = \tilde{U}^1 = \tilde{U}^0 = L_2(\mathbf{T}, U).$$

Below we consider a linear regulator problem on $[\tau, T]$, where $0 \le \tau < T$, instead of $[0, T]$ and select l, b, q, f and g in a special form. Realizing the above, we set a linear regulator problem in the form

$$x_t^u = \mathcal{U}_{t,\tau} l + \int_\tau^t \mathcal{U}_{t,s} B_s u_s ds, \ 0 \le \tau \le t \le T, \tag{5.37}$$

$$\begin{aligned} J(u) &= \langle x_T^u + \rho_T, Q_T(x_T^u + \rho_T)\rangle \\ &\quad + \int_\tau^T \left\langle \begin{bmatrix} x_t^u + \rho_t \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix} \begin{bmatrix} x_t^u + \rho_t \\ u_t \end{bmatrix} \right\rangle dt, \end{aligned} \tag{5.38}$$

$$U_{\text{ad}} = L_2(\tau, T; U). \tag{5.39}$$

In this section it is supposed that the following conditions hold:

($\mathbf{R_1}$) $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $B \in B_\infty(\mathbf{T}, \mathcal{L}(U, X))$, $l \in X$, $0 \le \tau < T$;

($\mathbf{R_2}$) $Q_T \in \mathcal{L}(X)$, $Q_T \ge 0$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$, $L \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \ge 0$ for a.e. $t \in \mathbf{T}$, $\rho_T \in X$, $\rho \in L_2(\mathbf{T}, X)$.

A linear regulator problem consists of minimizing the functional (5.38) on the set of admissible controls (5.39) over the state system (5.37). This problem briefly will be called the problem (5.37)–(5.39). We prefer to call the feedback rule of a given control u (the function φ satisfying $u_t = \varphi_t(x_t^u)$) a *regulator*. A regulator is *optimal* if the respective control is optimal.

5.5.2 Optimal Regulator

The following theorem completely solves the problem (5.37)–(5.39).

Theorem 5.24. *Under the conditions* ($\mathbf{R_1}$) *and* ($\mathbf{R_2}$), *there exists a unique optimal regulator in the problem* (5.37)–(5.39) *and the respective optimal control has the form*

$$u_t^* = -G_t^{-1}(B_t^* Q_t + L_t)x_t^* - G_t^{-1}B_t^*\alpha_t - G_t^{-1}L_t\rho_t, \text{ a.e. } t \in [\tau, T], \qquad (5.40)$$

where $x^* = x^{u^*}$, $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, Q *is a solution of the equation* (3.9) *and*

$$\alpha_t = \mathcal{Y}_{T,t}^* Q_T \rho_T + \int_t^T \mathcal{Y}_{s,t}^* \big(F_s - (Q_s B_s + L_s^*)G_s^{-1}L_s\big)\rho_s \, ds, \ \tau \le t \le T, \qquad (5.41)$$

$$x_t^* = \mathcal{Y}_{t,\tau} l - \int_\tau^t \mathcal{Y}_{t,s} B_s G_s^{-1}(B_s^*\alpha_s + L_s\rho_s)ds, \ \tau \le t \le T. \qquad (5.42)$$

If in addition $\rho = 0$ *and* $\rho_T = 0$, *then* $J(u^*) = \langle l, Q_\tau l\rangle$.

Proof. Let $\tau = 0$ in the problem (5.37)–(5.39). Also, let $\mathcal{F} = \{\Omega, \varnothing\}$ and $b = 0$ in the problem (5.1)–(5.4). Then as it was mentioned above, the sets of admissible controls in both of these problems are equal to $L_2(\mathbf{T}, U)$. One can observe that for

$$q = Q_T\rho_T, \ f_t = F_t\rho_t, \ g_t = L_t\rho_t, \ \text{a.e. } t \in \mathbf{T}, \qquad (5.43)$$

the difference of the functionals (5.38) and (5.4) is equal to

$$\langle \rho_T, Q_T\rho_T\rangle + \int_0^T \langle \rho_t, F_t\rho_t\rangle \, dt$$

and it is independent of control actions. So, under the above assumptions, Theorem 5.17 could be applied to the problem (5.37)–(5.39). This leads to the formulae (5.40)–(5.42) for the optimal control (when $\tau = 0$). The existence and the uniqueness of the optimal regulator follows from Proposition 5.22 because $\mathcal{F} = \{\Omega, \varnothing\}$ implies $U_{\text{ad}} = \tilde{U}^0 = \tilde{U}^1 = L_2(\mathbf{T}, U)$. Now let $0 < \tau < T$. In this case by an easy manipulation, the problem (5.37)–(5.39) reduces to the case $\tau = 0$ and the formulae (5.40)–(5.42) can be obtained. Finally, if $\rho = 0$ and $\rho_\tau = 0$, then $\alpha = 0$. Substituting them in (5.38), one can obtain $J(u^*) = \langle l, Q_\tau l\rangle$. □

5.6 Existence of Optimal Control

In this section, using properties of controls in linear feedback form, the existence of an optimal control in the problem (5.1)–(5.4) will be reduced to a certain filtering problem. In this section it is supposed that the conditions $(\mathbf{C_1})$–$(\mathbf{C_3})$ and $(\mathbf{C_4'})$ hold and n is a Wiener martingale, i.e.,

$(\mathbf{E_1})$ n has the representation

$$n_t = \int_0^t \Psi_s \, dv_s, \ 0 \le t \le T,$$

where $\Psi \in B_\infty(\mathbf{T}, \mathcal{L}(Z))$ and v is a Z-valued Wiener process on $\mathbf{T}$.

Also, we use the notation of Section 5.1.

5.6.1 Controls in Linear Feedback Form

Consider the stochastic optimal control problem (5.1)–(5.4). A function in the form

$$u_t = u_t^0 + \int_0^t M_{t,s} \, dz_s^u, \text{ a.e. } t \in \mathbf{T}, \tag{5.44}$$

where $u^0 \in L_2(\mathbf{T}, U)$ and $M \in B_2(\Delta_T, \mathcal{L}(Z, U))$, is called a *control in linear feedback form.* Note that by $(\mathbf{E_1})$, the integral in (5.44) is well defined.

Lemma 5.25. *If u has the representation (5.44) with $u^0 \in L_2(\mathbf{T}, U)$ and with $M \in B_2(\Delta_T, \mathcal{L}(Z, U))$, then it also has the representation*

$$u_t = u_t^1 + \int_0^t N_{t,s} \, dz_s^0, \ \textit{a.e. } t \in \mathbf{T}, \tag{5.45}$$

for some $u^1 \in L_2(\mathbf{T}, U)$ and $N \in B_2(\Delta_T, \mathcal{L}(Z, U))$ and vice versa, where z^0 is the observation process (5.2) corresponding to zero control.

Proof. Suppose (5.44) holds. By (5.1)–(5.2), we have

$$z_t^u = z_t^0 + \int_0^t \int_0^s C_s \mathcal{U}_{s,r} B_r u_r \, dr \, ds, \ 0 \le t \le T. \tag{5.46}$$

Using (5.46) in (5.44), we obtain

$$u_t = u_t^0 + \int_0^t M_{t,s} \, dz_s^0 + \int_0^t \int_0^s M_{t,s} C_s \mathcal{U}_{s,r} B_r u_r \, dr \, ds, \text{ a.e. } t \in \mathbf{T}. \tag{5.47}$$

The equation (5.47) is a Volterra integral equation in u. Therefore, by Theorem 2.38, there exists $G \in B_2(\Delta_T, \mathcal{L}(U))$ such that

$$u_t = u_t^0 + \int_0^t M_{t,s} \, dz_s^0 + \int_0^t G_{t,s} \left(u_s^0 + \int_0^s M_{s,r} \, dz_r^0 \right) ds, \text{ a.e. } t \in \mathbf{T}.$$

Hence by Theorem 4.23, for

$$u_t^1 = u_t^0 + \int_0^t G_{t,s} u_s^0 \, ds, \text{ a.e. } t \in \mathbf{T},$$

and for

$$N_{t,s} = M_{t,s} + \int_s^t G_{t,r} M_{r,s} \, dr, \ 0 \le s \le t \le T,$$

we obtain the representation (5.45) for u. In a similar manner, starting from (5.45), the representation (5.44) can be obtained. □

Proposition 5.26. *The set* U_{ad}*, defined by* (5.3)*, contains all controls which are in the linear feedback form* (5.44) *with* $u^0 \in L_2(\mathbf{T}, U)$ *and* $M \in B_2(\Delta_T, \mathcal{L}(Z, U))$.

Proof. If u has the form (5.44), then $u \in \tilde{U}$. By Lemma 5.25, u has the form (5.45) as well. Hence, $u \in \tilde{U}^0$. Thus, $u \in \tilde{U}^0 \cap \tilde{U} = U_{\text{ad}}$. □

Proposition 5.27. *The set* U_{ad}*, defined by* (5.3)*, contains all controls in the form* (5.45) *where* $u^1 \in L_2(\mathbf{T}, U)$ *and* $N \in B_2(\Delta_T, \mathcal{L}(Z, U))$.

Proof. This is similar to that of Proposition 5.26. □

5.6.2 Existence of Optimal Control

By Proposition 5.22, the functional (5.4) takes its minimum value on $\tilde{U}^0$ at a unique control u^* defined by (5.35). Let us find another representation for this u^*. Denote

$$x_t = \mathcal{U}_{t,0}(l - \mathbf{E}l) + \int_0^t \mathcal{U}_{t,s}(b_s - \mathbf{E}b_s) \, ds + \int_0^t \mathcal{U}_{t,s} \, dm_s, \ 0 \le t \le T, \tag{5.48}$$

and

$$\eta_t = \mathcal{U}_{t,0}\mathbf{E}l + \int_0^t \mathcal{U}_{t,s}(B_s u_s^* + \mathbf{E}b_s) \, ds, \ 0 \le t \le T. \tag{5.49}$$

Then

$$\mathbf{E}_t^0 x_t^* = \mathbf{E}_t^0 x_t + \eta_t, \ 0 < t \le T, \tag{5.50}$$

where $x^* = x^{u^*}$. Using (5.50) in (5.35), we have

$$u^* = -G_t^{-1}(B_t^* Q_t + L_t)\left(\mathbf{E}_t^0 x_t + \eta_t\right) - G_t^{-1} g_t - G_t^{-1} B_t^* \mathbf{E}_t^0 \alpha_t, \text{ a.e. } t \in \mathbf{T}. \tag{5.51}$$

Let $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$. By Proposition 4.28,

$$\begin{aligned}\eta_t &= \mathcal{U}_{t,0}\mathbf{E}l - \int_0^t \mathcal{U}_{t,s}B_sG_s^{-1}(B_s^*Q_s + L_s)\big(\mathbf{E}_t^0 x_t + \eta_s\big)ds \\ &\quad + \int_0^t \mathcal{U}_{t,s}\big(\mathbf{E}b_s - B_sG_s^{-1}g_s - B_sG_s^{-1}B_s^*\,\mathbf{E}_s^0\alpha_s\big)ds \\ &= \mathcal{Y}_{t,0}\mathbf{E}l - \int_0^t \mathcal{Y}_{t,s}B_sG_s^{-1}(B_s^*Q_s + L_s)\mathbf{E}_t^0 x_t\,ds \\ &\quad + \int_0^t \mathcal{Y}_{t,s}\big(\mathbf{E}b_s - B_sG_s^{-1}g_s - B_sG_s^{-1}B_s^*\mathbf{E}_s^0\alpha_s\big)ds. \end{aligned} \tag{5.52}$$

Let

$$\begin{aligned} y_t &= (B_t^*Q_t + L_t)x_t + B_t^* \int_t^T \mathcal{Y}_{s,t}^* Q_s(b_s - \mathbf{E}b_s)\,ds \\ &\quad + B_t^* \int_t^T \mathcal{Y}_{s,t}^* Q_s\,dm_s, \ \ 0 \le t \le T, \end{aligned} \tag{5.53}$$

and let

$$\begin{aligned} u_t^1 =& - G_t^{-1}g_t - G_t^{-1}B_t^*\mathcal{Y}_{T,t}^*q - G_t^{-1}(B_t^*Q_t + L_t)\mathcal{Y}_{t,0}\mathbf{E}l \\ &+ G_t^{-1}B_t^* \int_t^T \mathcal{Y}_{s,t}^*\big((Q_sB_s + L_s^*)G_s^{-1}g_s - Q_s\mathbf{E}b_s - f_s\big)\,ds \\ &- G_t^{-1}(B_t^*Q_t + L_t)\int_0^t \mathcal{Y}_{t,s}\big(\mathbf{E}b_s - B_sG_s^{-1}g_s\big)\,ds \\ &+ G_t^{-1}(B_t^*Q_t + L_t)\int_0^t \mathcal{Y}_{t,s}B_sG_s^{-1}B_s^*\mathcal{Y}_{T,s}^*q\,ds \\ &- G_t^{-1}(B_t^*Q_t + L_t)\int_0^t\int_s^T \mathcal{Y}_{t,s}B_sG_s^{-1}B_s^*\mathcal{Y}_{r,s}^* \\ &\qquad \times \big((Q_rB_r + L_r^*)G_r^{-1}g_r - Q_r\mathbf{E}b_r - f_r\big)\,dr\,ds, \ \text{a.e. } t \in \mathbf{T}. \end{aligned} \tag{5.54}$$

Substituting (5.52) and (5.11) in (5.51) and using (5.53) and (5.54), we obtain

$$u_t^* = u_t^1 - G_t^{-1}\mathbf{E}_t^0 y_t + G_t^{-1}(B_t^*Q_t + L_t)\int_0^t \mathcal{Y}_{t,s}B_sG_s^{-1}\mathbf{E}_s^0 y_s\,ds, \ \text{a.e. } t \in \mathbf{T}. \tag{5.55}$$

Let z_t be defined by

$$z_t = \int_0^t (C_sx_s + c_s - \mathbf{E}c_s)\,ds + \int_0^t \Psi_s\,dv_s, \ \ 0 \le t \le T, \tag{5.56}$$

where x is defined by (5.48). Obviously, the difference $z_t^0 - z_t$ is nonrandom for all $0 \le t \le T$. Also by $(\mathbf{E_1})$, z and z^0 have continuous paths. So by Proposition 2.30, we can write

$$\mathbf{E}_t^0 y_t = \mathbf{E}(y_t|z_s^0;\, 0 < s \le t) = \mathbf{E}(y_t|z_s;\, 0 < s \le t), \ \ 0 < t \le T. \tag{5.57}$$

Below we will refer to the condition:

($\mathbf{E_2}$) For y and z, defined by (5.48), (5.53) and (5.56),

$$\mathbf{E}(y_t|z_s;\,0 < s \le t) = \int_0^t K_{t,s}\,dz_s,\ 0 < t \le T,$$

where $K \in B_2(\Delta_T, \mathcal{L}(Z,U))$.

Theorem 5.28. *Under the conditions* ($\mathbf{C_1}$)–($\mathbf{C_3}$), ($\mathbf{C_4'}$) *and* ($\mathbf{E_1}$)–($\mathbf{E_2}$), *there exists a unique optimal control in the stochastic optimal control problem* (5.1)–(5.4).

Proof. By (5.57) and ($\mathbf{E_2}$), the formula (5.55) defines a control u^* in the form (5.45) for some $u^1 \in L_2(\mathbf{T}, U)$ and $N \in B_2(\Delta_T, \mathcal{L}(Z,U))$. Therefore, by Proposition 5.27, u^* belongs to $\tilde{U}_{\mathrm{ad}}$. On the other hand, by Proposition 5.22, the functional (5.4) takes its minimum value on $\tilde{U}^0$ at u^*, defined by (5.35), which is equal to u^*, defined by (5.55). Therefore, by Proposition 5.13, u^* is an optimal control in the problem (5.1)–(5.4). The uniqueness of the optimal control follows from Theorem 5.16 because in view of Theorem 3.22, the conditions ($\mathbf{C_1}$) and ($\mathbf{C_3}$) imply the existence of a unique solution of the equation (3.9). □

5.6.3 Application to Existence of Saddle Points

Theorem 5.28 can be used to study the existence of saddle points. To demonstrate this, in this section we consider two games involving heat and wave equations.

Example 5.29. Assume that a string of unit length is given. Two players have access to every point on the string at any time moment. For the time period T, the first player wants to make the temperature of the string close to a certain value, say, to zero and the other one wants to make it much away from zero, both of them with minimum effort. At both end points of the string a constant temperature equal to zero is kept. Obviously, the fluctuation of temperature at points along the string will be modelled as wide band noise. Both players observe the average temperature over the string up to the current time moment with measurement noise that will be modelled as white noise. Mathematically, this "heating-cooling" game can be formulated as

$$\frac{\partial}{\partial t}x_{t,\theta} = \frac{\partial^2}{\partial\theta^2}x_{t,\theta} + u_{1,t,\theta} + u_{2,t,\theta} + \varphi_t,\ 0 < t \le T,$$

$$x_{0,\theta} = f_\theta,\ x_{t,0} = x_{t,1} = 0,\ 0 \le \theta \le 1,\ 0 \le t \le T, \tag{5.58}$$

$$z_t = \int_0^t \int_0^1 x_{s,\theta}\,d\theta\,ds + w_t,\ 0 \le t \le T, \tag{5.59}$$

$$J(u_1,u_2) = \mathbf{E}\int_0^1 x_{T,\theta}^2\,d\theta + \mathbf{E}\int_0^T\int_0^1 \left(u_{1,t,\theta}^2 - u_{2,t,\theta}^2\right)d\theta\,dt, \tag{5.60}$$

where the function x expresses the temperature of the string, u_1 and u_2 are controls of the first and second players, respectively, and z is the common observation process of the players. Here w is a standard Wiener process and φ is a wide band noise process represented in the form (see Section 4.6.2)

$$\varphi_t = \int_{\max(0,t-\varepsilon)}^{t} \phi_{\theta-t}\, dw_\theta,\ 0 \le t \le T, \tag{5.61}$$

with $0 < \varepsilon < T$ and $\phi \in L_2(-\varepsilon, 0; \mathbb{R})$.

Let $X = L_2(0,1;\mathbb{R})$. Recall that the second order differential operator $d^2/d\theta^2$ defined on $D\left(d^2/d\theta^2\right) = \left\{h \in X : \left(d^2/d\theta^2\right)h \in X,\ h_0 = h_1 = 0\right\}$ generates the strongly continuous semigroup $\mathcal{U}$ as defined in Example 3.5 with $\mathcal{U}_t = \mathcal{U}_t^*$, $t \ge 0$. Since $Q_T = I$, $F_s = 0$, $L_s = 0$ and

$$B_s G_s^{-1} B_s^* = \begin{bmatrix} I & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} I \\ I \end{bmatrix} = 0,$$

for the game (5.58)–(5.60), one can easily observe that $\mathcal{Y}_t = \mathcal{U}_t$, $t \ge 0$, and the Riccati equation (3.9) is trivial, expressing its solution explicitly in the form $Q_t = \mathcal{U}_{2(T-t)}$. Hence, by Theorem 5.19, if there exists a saddle point in this game, then it is unique and has the components

$$u_{1,t,\cdot}^* = -\mathcal{U}_{2(T-t)} \mathbf{E}_t^* x_{t,\cdot}^* - \mathbf{E}_t^* \int_t^T \mathcal{U}_{2T-t-s} D\varphi_s\, ds,\ \text{a.e. } t \in \mathbf{T}, \tag{5.62}$$

$$u_{2,t,\cdot}^* = \mathcal{U}_{2(T-t)} \mathbf{E}_t^* x_{t,\cdot}^* + \mathbf{E}_t^* \int_t^T \mathcal{U}_{2T-t-s} D\varphi_s\, ds,\ \text{a.e. } t \in \mathbf{T}, \tag{5.63}$$

where D is defined by $[Dh]_\theta = h$, $-\infty < h < \infty$, $0 \le \theta \le 1$. Using (5.62)–(5.63) in (5.58), we obtain

$$x_{t,\cdot}^* = \mathcal{U}_t f + \int_0^t \mathcal{U}_{t-s} D\varphi_s\, ds,\ 0 \le t \le T. \tag{5.64}$$

Therefore, (5.62)–(5.63) can be rewritten as

$$u_{1,t,\cdot}^* = -\mathcal{U}_{2T-t} f - \mathbf{E}_t^* \int_0^T \mathcal{U}_{2T-t-s} D\varphi_s\, ds,\ \text{a.e. } t \in \mathbf{T}, \tag{5.65}$$

$$u_{2,t,\cdot}^* = \mathcal{U}_{2T-t} f + \mathbf{E}_t^* \int_0^T \mathcal{U}_{2T-t-s} D\varphi_s\, ds,\ \text{a.e. } t \in \mathbf{T}. \tag{5.66}$$

In order to show the existence of the saddle point, we will verify it on the pair (u_1^*, u_2^*). At first, note that from the results of Section 8.1 it will follow that the conditions $(\mathbf{E_1})$–$(\mathbf{E_2})$ for the game (5.58)–(5.60) hold. Hence, $(u_1^*, u_2^*) \in U_{\text{ad}}$. From Theorem 5.28, it is clear that

$$J(u_1^*, u_2^*) \le J(u_1, u_2^*),\ (u_1, u_2^*) \in U_{\text{ad}}. \tag{5.67}$$

Let us show that

$$J(u_1^*, u_2) \le J(u_1^*, u_2^*), \ (u_1^*, u_2) \in U_{\text{ad}}. \tag{5.68}$$

For, take any u_2 with $(u_1^*, u_2^* + u_2) \in U_{\text{ad}}$. Then

$$\begin{aligned}
&J(u_1^*, u_2^* + u_2) - J(u_1^*, u_2^*)\\
&\quad = 2\mathbf{E}\Big\langle \mathcal{U}_T f + \int_0^T \mathcal{U}_{T-s} D\varphi_s\, ds, \int_0^T \mathcal{U}_{T-t} u_{2,t,\cdot}\, dt\Big\rangle_X\\
&\qquad + \mathbf{E}\Big\| \int_0^T \mathcal{U}_{T-t} u_{2,t,\cdot}\, dt\Big\|_X^2 - 2\mathbf{E}\int_0^T \langle u_{2,t,\cdot}^*, u_{2,t,\cdot}\rangle_X\, dt - \mathbf{E}\int_0^T \|u_{2,t,\cdot}\|_X^2\, dt\\
&\quad \le 2\mathbf{E}\int_0^T \Big\langle \mathcal{U}_{2T-t} f + \int_0^T \mathcal{U}_{2T-t-s} D\varphi_s\, ds, u_{2,t,\cdot}\Big\rangle_X dt\\
&\qquad - 2\mathbf{E}\int_0^T \Big\langle \mathcal{U}_{2T-t} f + \mathbf{E}_t^* \int_0^T \mathcal{U}_{2T-t-s} D\varphi_s\, ds, u_{2,t,\cdot}\Big\rangle_X dt\\
&\qquad + \mathbf{E}\int_0^T \|u_{2,t,\cdot}\|_X^2\, dt\Big(\int_0^T \|\mathcal{U}_{T-t}\|_{\mathcal{L}(X)}^2\, dt - 1\Big)\\
&\quad = \mathbf{E}\int_0^T \|u_{2,t,\cdot}\|_X^2\, dt\Big(\int_0^T \|\mathcal{U}_t\|_{\mathcal{L}(X)}^2\, dt - 1\Big).
\end{aligned}$$

Since

$$\begin{aligned}
\int_0^T \|\mathcal{U}_t\|_{\mathcal{L}(X)}^2\, dt &= \int_0^T \sup_{\|h\|=1} \sum_{n=1}^\infty 2e^{-2n^2\pi^2 t}\Big(\int_0^1 h_\theta \sin(n\pi\theta)\, d\theta\Big)^2 dt\\
&\le \sum_{n=1}^\infty \int_0^T e^{-2n^2\pi^2 t}\, dt \le \sum_{n=1}^\infty \frac{1}{2n^2\pi^2}\\
&\le \frac{1}{2\pi^2}\Big(1 + \int_1^\infty \frac{1}{z^2}\, dz\Big) = \frac{1}{\pi^2} < 1
\end{aligned}$$

we obtain that a saddle point in the game (5.58)–(5.60) exists. Thus, the pair (u_1^*, u_2^*) defined by (5.62)–(5.63) or (5.65)–(5.66) is a unique saddle point and $J(u_1^*, u_2^*) = \mathbf{E}\|x_{T,\cdot}^*\|_X^2$ where x^* is defined by (5.64).

Example 5.30. Now consider a game involving a vibrating string of unit length: let the state system be given by

$$\begin{aligned}
&\frac{\partial^2}{\partial t^2} x_{t,\theta} = \frac{\partial^2}{\partial \theta^2} x_{t,\theta} + u_{1,t,\theta} + u_{2,t,\theta} + \varphi_t, \ 0 < t \le T,\\
&\qquad x_{0,\theta} = f_\theta, \ \frac{\partial}{\partial t} x_{t,\theta}\Big|_{t=0} = g_\theta, x_{t,0} = x_{t,1} = 0, \ 0 \le \theta \le 1, \ 0 \le t \le T,
\end{aligned} \tag{5.69}$$

the observation system by

$$z_t = \int_0^t \int_0^1 x_{s,\theta}\, d\theta\, ds + w_t,\ 0 \le t \le T, \tag{5.70}$$

and the functional by

$$J(u_1, u_2) = \mathbf{E} \int_0^1 \left(\left(\frac{\partial}{\partial t} x_{t,\theta} \Big|_{t=T} \right)^2 + \left(\frac{\partial}{\partial \theta} x_{t,\theta} \Big|_{t=T} \right)^2 \right) d\theta$$
$$+ \mathbf{E} \int_0^T \int_0^1 \left(u_{1,t,\theta}^2 - u_{2,t,\theta}^2 \right) d\theta\, dt. \tag{5.71}$$

Here $x_{t,\theta}$ expresses the displacement of the string at the point θ and at the time moment t, u_1 is a control of the first player which tries to minimize the functional (5.71) and u_2 is a control of the second player maximizing (5.71). Both players observe the process z defined by (5.70). We assume that w is a standard Wiener process and φ is a wide band noise process defined by (5.61) with $0 < \varepsilon < T$ and $\phi \in L_2(-\varepsilon, 0; \mathbb{R})$.

Consider the Hilbert space X and the semigroup $\mathcal{U}$ on X from Example 3.6. Recall that $\mathcal{U}$ is a contraction semigroup and its natural extension to the real line forms a group.

Let

$$y_{t,\theta} = \begin{bmatrix} x_{t,\theta} \\ (\partial/\partial t) x_{t,\theta} \end{bmatrix},\ y_{0,\cdot} = \begin{bmatrix} f \\ g \end{bmatrix} \in X.$$

Since for $\xi \in L_2(0,1;\mathbb{R})$ with $(d/d\theta)\xi \in L_2(0,1;\mathbb{R})$, we have

$$\xi_\theta = \sum_{n=1}^\infty \check{\xi}_n \sqrt{2} \sin(n\pi\theta) \Rightarrow \frac{d}{d\theta}\xi_\theta = \sum_{n=1}^\infty n\pi \check{\xi}_n \sqrt{2} \cos(n\pi\theta),$$

by Parseval identity,

$$\int_0^1 \xi_\theta^2\, d\theta = \sum_{n=1}^\infty \check{\xi}_n^2 \quad \text{and} \quad \int_0^1 \left(\frac{\partial}{\partial \theta} \xi_\theta \right)^2 d\theta = \sum_{n=1}^\infty n^2 \pi^2 \check{\xi}_n^2,$$

where $\{\check{\xi}_n\}$ is the sequence of Fourier coefficients in the half-range Fourier sine expansion of ξ. Therefore, the game (5.69)–(5.71) can be written as

$$y_{t,\cdot} = \mathcal{U}_t y_{0,\cdot} + \int_0^t \mathcal{U}_{t-s} \check{I}(u_{1,s,\cdot} + u_{2,s,\cdot})\, ds + \int_0^t \mathcal{U}_{t-s} \check{I} D \varphi_s\, ds,\ 0 \le t \le T, \tag{5.72}$$

$$z_t = \int_0^t \int_0^1 \begin{bmatrix} I & 0 \end{bmatrix} y_{s,\theta}\, d\theta\, ds + w_t,\ 0 \le t \le T, \tag{5.73}$$

$$J(u_1, u_2) = \mathbf{E} \|y_{T,\cdot}\|_X^2 + \mathbf{E} \int_0^T \left(\|u_{1,t,\cdot}\|_{L_2}^2 - \|u_{2,t,\cdot}\|_{L_2}^2 \right) dt, \tag{5.74}$$

where D is defined in Example 5.29 and

$$\check{I} = \begin{bmatrix} 0 \\ I \end{bmatrix} \in \mathcal{L}(L_2(0,1;\mathbb{R}), X).$$

Since $Q_T = I$, $F_s = 0$, $L_s = 0$ and $B_s G_s^{-1} B_s^* = 0$ for the game (5.72)–(5.74), we obtain $\mathcal{Y}_t = \mathcal{U}_t$ and $Q_t = I$, $0 \le t \le T$. Hence, by Theorem 5.19, if there exists a saddle point in this game, then it is unique and has the components

$$u_{1,t,\cdot}^* = -\check{I}^* \mathbf{E}_t^* y_{t,\cdot}^* - \check{I}^* \mathbf{E}_t^* \int_t^T \mathcal{U}_{t-s} \check{I} D \varphi_s \, ds, \text{ a.e. } t \in \mathbf{T}, \tag{5.75}$$

$$u_{2,t,\cdot}^* = \check{I}^* \mathbf{E}_t^* y_{t,\cdot}^* + \check{I}^* \mathbf{E}_t^* \int_t^T \mathcal{U}_{t-s} \check{I} D \varphi_s \, ds, \text{ a.e. } t \in \mathbf{T}. \tag{5.76}$$

Substituting (5.75)–(5.76) in (5.72), we obtain

$$y_{t,\cdot}^* = \mathcal{U}_t y_{0,\cdot} + \int_0^t \mathcal{U}_{t-s} \check{I} D \varphi_s \, ds, \; 0 \le t \le T. \tag{5.77}$$

Therefore, (5.75)–(5.76) can be rewritten as

$$u_{1,t,\cdot}^* = -\check{I}^* \mathcal{U}_t y_{0,\cdot} - \mathbf{E}_t^* \int_0^T \check{I}^* \mathcal{U}_{t-s} \check{I} D \varphi_s \, ds, \text{ a.e. } t \in \mathbf{T}, \tag{5.78}$$

$$u_{2,t,\cdot}^* = \check{I}^* \mathcal{U}_t y_{0,\cdot} + \mathbf{E}_t^* \int_0^T \check{I}^* \mathcal{U}_{t-s} \check{I} D \varphi_s \, ds, \text{ a.e. } t \in \mathbf{T}. \tag{5.79}$$

In order to show the existence of the saddle point, we will verify it on the pair (u_1^*, u_2^*). From the results of Section 8.1 it will follow that the conditions $(\mathbf{E_1})$–$(\mathbf{E_2})$ for the game (5.72)–(5.74) hold. Hence, $(u_1^*, u_2^*) \in U_{\text{ad}}$. From Theorem 5.28, it is clear that the inequality (5.67) holds. Let us show the inequality (5.68) for the game (5.72)–(5.74). For, take any u_2 with $(u_1^*, u_2^* + u_2) \in U_{\text{ad}}$. Then

$$\begin{aligned}
&J(u_1^*, u_2^* + u_2) - J(u_1^*, u_2^*) \\
&\quad = 2\mathbf{E}\left\langle \mathcal{U}_T y_{0,\cdot} + \int_0^T \mathcal{U}_{T-s} \check{I} D \varphi_s \, ds, \int_0^T \mathcal{U}_{T-t} \check{I} u_{2,t,\cdot} \, dt \right\rangle_X \\
&\qquad + \mathbf{E}\left\| \int_0^T \mathcal{U}_{T-t} \check{I} u_{2,t,\cdot} \, dt \right\|_X^2 - 2\mathbf{E}\int_0^T \langle u_{2,t,\cdot}^*, u_{2,t,\cdot} \rangle_{L_2} \, dt - \mathbf{E}\int_0^T \|u_{2,t,\cdot}\|_{L_2}^2 \, dt \\
&\quad \le 2\mathbf{E}\int_0^T \left\langle \check{I}^* \mathcal{U}_T \mathcal{U}_{T-t}^* y_{0,\cdot} + \int_0^T \check{I}^* \mathcal{U}_{T-s} \mathcal{U}_{T-t}^* \check{I} D \varphi_s \, ds, u_{2,t,\cdot} \right\rangle_{L_2} dt \\
&\qquad - 2\mathbf{E}\int_0^T \left\langle \check{I}^* \mathcal{U}_t y_{0,\cdot} + \mathbf{E}_t^* \int_0^T \check{I}^* \mathcal{U}_{t-s} \check{I} D \varphi_s \, ds, u_{2,t,\cdot} \right\rangle_{L_2} dt \\
&\qquad + \mathbf{E}\int_0^T \|u_{2,t,\cdot}\|_{L_2}^2 \, dt \cdot \left(\int_0^T \|\mathcal{U}_{T-t} \check{I}\|_{\mathcal{L}}^2 \, dt - 1 \right) \le 0
\end{aligned}$$

if $0 < T \leq 1$. Hence, the inequality (5.68) holds for $0 < T \leq 1$ proving that a saddle point in the game (5.72)–(5.74) or (5.69)–(5.71) exists if $0 < T \leq 1$. Thus, the pair (u_1^*, u_2^*) defined by (5.75)–(5.76) or (5.78)–(5.79) is a unique saddle point and $J(u_1^*, u_2^*) = \mathbf{E}\|y_{T,\cdot}^*\|_X^2$ where y^* is defined by (5.77) and $0 < T \leq 1$.

5.7 Concluding Remarks

Let

$$x_t = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\varphi_s^1\, ds + \int_0^t \mathcal{U}_{t,s}\, dm_s,\ 0 \leq t \leq T, \tag{5.80}$$

$$y_t = M_t x_t + \int_t^T N_{s,t}\varphi_s^1\, ds + \int_t^T N_{s,t}\, dm_s,\ 0 \leq t \leq T, \tag{5.81}$$

$$z_t = \int_0^t (C_s x_s + \varphi_s^2)\, ds + \int_0^t \Psi_s\, dv_s,\ 0 \leq t \leq T. \tag{5.82}$$

One can observe that the random processes x, y and z defined by (5.80)–(5.82) are the same as x, y and z defined by the formulae (5.48), (5.53) and (5.56) if

$$x_0 = l - \mathbf{E}l,\ \varphi_t^1 = b_t - \mathbf{E}b_t,\ \varphi_t^2 = c_t - \mathbf{E}c_t,\ 0 \leq t \leq T,$$
$$N_{s,t} = B_t^*\mathcal{Y}_{s,t}^*Q_s,\ (s,t) \in \Delta_T,\ M_t = B_t^*Q_t + L_t,\ 0 \leq t \leq T.$$

Indeed the condition $(\mathbf{E_2})$ means finding the best estimate of y_t based on observations z_s, $0 \leq s \leq t$, in a linear feedback form for all $0 < t \leq T$. With the notation introduced above, this problem will be called the filtering problem (5.80)–(5.82) and will be studied in the following chapters under the condition that the covariance operator of the Wiener process v is coercive, which in term implies that Z has a finite dimension, i.e., $Z = \mathbb{R}^n$. Also, the processes φ^1 and φ^2 will be taken to be colored or wide band noise processes.

Two cases in the problem (5.80)–(5.82) must be specified. The first case assumes that the past of observations and the future of signal noise are independent, i.e., $\sigma(z_s;\, 0 \leq s \leq t)$ and $\sigma(\varphi_s^1, m_s - m_t;\, t < s \leq T)$ are independent for all $0 < t \leq T$. In this case,

$$\mathbf{E}_t y_t = \mathbf{E}(y_t | z_s;\, 0 \leq s \leq t) = M_t\mathbf{E}(x_t | z_s;\, 0 \leq s \leq t),\ 0 < t \leq T,$$

and the problem (5.80)–(5.82) reduces to estimating x_t based on observations z_s, $0 \leq s \leq t$. This case includes the Kalman-Bucy filtering for independent or correlated white noises, discussed in Chapter 6, and the linear filtering problems when the signal noise anticipates the observation noise.

The second case is more complicated and requires the dependence of the σ-algebras $\sigma(z_s;\, 0 \leq s \leq t)$ and $\sigma(\varphi_s^1, m_s - m_t;\, t < s \leq T)$ for t in some nontrivial subinterval of $\mathbf{T}$. In particular, this case includes filtering problems with the signal noise delaying the observation noise.

Chapter 6

Control and Estimation under Correlated White Noises

In this chapter the Kalman–Bucy estimation theory concerning correlated white noise processes and its application to designing optimal stochastic regulators are considered.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $X, U, H \in \mathcal{H}$, $T > 0$, $\mathbf{T} = [0, T]$ is a finite time interval and $\Delta_t = \{(s,r) : 0 \le r \le s \le t\}$ for $t > 0$.

6.1 Estimation: Preliminaries

6.1.1 Setting of Estimation Problems

Let (x, z) be a pair of random processes where

$$x_t = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\Phi_s\, dw_s,\ 0 \le t \le T, \tag{6.1}$$

is a signal process and

$$z_t = \int_0^t C_s x_s ds + \int_0^t \Psi_s\, dv_s,\ 0 \le t \le T, \tag{6.2}$$

is an observation process. We will suppose that the following conditions hold:

$(\mathbf{E_1^w})$ $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$;

$(\mathbf{E_2^w})$ $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with $\operatorname{cov} v_T > 0$, x_0 is a an X-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0 and (w, v) are independent.

Also, throughout this chapter we will denote

$$\begin{cases} P_0 = \mathrm{cov} x_0, \begin{bmatrix} \bar{W} & \bar{R} \\ \bar{R}^* & \bar{V} \end{bmatrix} = T^{-1}\mathrm{cov} \begin{bmatrix} w_T \\ v_T \end{bmatrix}, \\ W_t = \Phi_t \bar{W} \Phi_t^*, \ V_t = \Psi_t \bar{V} \Psi_t^*, \ R_t = \Phi_t \bar{R} \Psi_t^*, \ 0 \le t \le T. \end{cases} \tag{6.3}$$

Let t and τ be two time moments in $\mathbf{T}$. Since x_0, w and v define a Gaussian system, according to Proposition 4.34, in case $\tau > 0$ we can expect that the best estimate $\hat{x}_t^\tau$ of x_t based on z_s, $0 \le s \le \tau$, has a linear feedback form, i.e.,

$$\hat{x}_t^\tau = \mathbf{E}(x_t | z_s;\ 0 \le s \le \tau) = \int_0^\tau K_s \, dz_s \tag{6.4}$$

for some $K \in B_2(0, \tau; \mathcal{L}(\mathbb{R}^n, X))$. One can see that (6.4) includes the case $\tau = 0$ as well in view of

$$\hat{x}_t^0 = \mathbf{E}(x_t | z_0) = \mathbf{E} x_t = \mathcal{U}_{t,0} \mathbf{E} x_0 + \mathbf{E} \int_0^t \mathcal{U}_{t,s} \Phi_s \, dw_s = 0.$$

Designing the conditional expectation $\mathbf{E}(x_t | z_s;\ 0 \le s \le \tau)$ in the linear feedback form (6.4), where x and z are defined by (6.1) and (6.2), respectively, will be called the estimation problem (6.1)–(6.2). When $0 \le t = \tau \le T$ (respectively, $0 \le \tau < t \le T$ or $0 \le t < \tau \le T$), the estimation problem (6.1)–(6.2) will be called the filtering (respectively, prediction or smoothing) problem (6.1)–(6.2). In (6.4) we write $\hat{x}_t = \hat{x}_t^\tau$ in the case $t = \tau$.

Note that in (6.4) K depends on s as well as on t and τ. For brevity, we indicate only the dependence of K on s. We say that an optimal linear feedback estimator in the estimation problem (6.1)–(6.2) is unique if the function K satisfying (6.4) is unique a.e. on $[0, \tau]$ for t and τ considered as fixed.

6.1.2 Wiener–Hopf Equation

By use of Proposition 4.34, one can derive an equation for K satisfying (6.4).

Lemma 6.1. *For $0 < \tau \le T$ and for $0 \le t \le T$, the equality (6.4) holds if and only if K satisfies*

$$K_s V_s + \int_0^\tau K_r \tilde{\Lambda}_{r,s} \, dr = \Lambda_{t,s} C_s^* + \chi_{[0,t]}(s) \mathcal{U}_{t,s} R_s, \ \text{a.e. } s \in [0, \tau], \tag{6.5}$$

where Λ and $\tilde{\Lambda}$ are defined by

$$\Lambda_{r,s} = \mathcal{U}_{r,0} P_0 \mathcal{U}_{s,0}^* + \int_0^{\min(s,r)} \mathcal{U}_{r,\sigma} W_\sigma \mathcal{U}_{s,\sigma}^* \, d\sigma, \ s, r \in \mathbf{T}, \tag{6.6}$$

$$\tilde{\Lambda}_{r,s} = C_r \Lambda_{r,s} C_s^* + \begin{Bmatrix} C_r \mathcal{U}_{r,s} R_s, & r > s \\ R_r^* \mathcal{U}_{s,r}^* C_s^*, & r \le s \end{Bmatrix}, \ s, r \in \mathbf{T}. \tag{6.7}$$

Proof. First, note that by Proposition 4.26(e),

$$\begin{aligned}\operatorname{cov}(x_r, x_s) &= \operatorname{cov}(\mathcal{U}_{r,0}x_0, \mathcal{U}_{s,0}x_0) + \operatorname{cov}\left(\int_0^r \mathcal{U}_{r,\sigma}\Phi_\sigma\, dw_\sigma, \int_0^s \mathcal{U}_{s,\sigma}\Phi_\sigma\, dw_\sigma\right)\\ &= \mathcal{U}_{r,0}P_0\mathcal{U}_{s,0}^* + \int_0^{\min(s,r)} \mathcal{U}_{r,\sigma}W_\sigma\mathcal{U}_{s,\sigma}^*\, d\sigma = \Lambda_{r,s}.\end{aligned}$$

By Proposition 4.34, the equality (6.4) holds if and only if

$$\operatorname{cov}\left(\int_0^\tau K_s\, dz_s, \int_0^\tau G_s\, dz_s\right) = \operatorname{cov}\left(x_t, \int_0^\tau G_s\, dz_s\right) \tag{6.8}$$

for all $G \in B_2(0, \tau; \mathcal{L}(\mathbb{R}^n))$. Computing the left-hand side of (6.8), we obtain

$$\begin{aligned}&\operatorname{cov}\left(\int_0^\tau K_s\, dz_s, \int_0^\tau G_s\, dz_s\right)\\ &\quad= \operatorname{cov}\left(\int_0^\tau K_sC_sx_s\, ds + \int_0^\tau K_s\Psi_s\, dv_s, \int_0^\tau G_sC_sx_s\, ds + \int_0^\tau G_s\Psi_s\, dv_s\right)\\ &\quad= \int_0^\tau\int_0^\tau K_rC_r\Lambda_{r,s}C_s^*G_s^*\, dr\, ds + \int_0^\tau K_sV_sG_s^*\, ds\\ &\qquad+ \operatorname{cov}\left(\int_0^\tau\int_0^r K_rC_r\mathcal{U}_{r,s}\Phi_s\, dw_s\, dr, \int_0^\tau G_s\Psi_s\, dv_s\right)\\ &\qquad+ \operatorname{cov}\left(\int_0^\tau K_r\Psi_r\, dv_r, \int_0^\tau\int_0^s G_sC_s\mathcal{U}_{s,r}\Phi_r\, dw_r\, ds\right)\\ &\quad= \int_0^\tau \left(K_sV_s + \int_0^\tau K_r\left(C_r\Lambda_{r,s}C_s^* + \begin{Bmatrix} C_r\mathcal{U}_{r,s}R_s, & r > s\\ R_r^*\mathcal{U}_{s,r}^*C_s^*, & r \le s\end{Bmatrix}\right) dr\right)G_s^*\, ds\\ &\quad= \int_0^\tau \left(K_sV_s + \int_0^\tau K_r\tilde{\Lambda}_{r,s}\, dr\right) G_s^*\, ds.\end{aligned}$$

Similarly, for the right-hand side of (6.8),

$$\begin{aligned}\operatorname{cov}\left(x_t, \int_0^\tau G_s\, dz_s\right) &= \operatorname{cov}\left(x_t, \int_0^\tau G_sC_sx_s\, ds\right)\\ &\qquad+ \operatorname{cov}\left(\int_0^t \mathcal{U}_{t,s}\Phi_s\, dw_s, \int_0^\tau G_s\Psi_s\, dv_s\right)\\ &= \int_0^\tau \left(\Lambda_{t,s}C_s^* + \chi_{[0,t]}(s)\mathcal{U}_{t,s}R_s\right) G_s^*\, ds.\end{aligned}$$

Equating the expressions for the two sides of (6.8) and using the arbitrariness of G, we obtain (6.5). So, (6.8) implies (6.5). Running back in the above calculations, one can show that (6.5) implies (6.8) too. □

The equation (6.5) for K is called a *Wiener–Hopf equation.*

6.2 Filtering

In this section we suppose that the conditions $(\mathbf{E}_1^{\mathbf{w}})$ and $(\mathbf{E}_2^{\mathbf{w}})$ hold and we let $0 \le t = \tau \le T$.

6.2.1 Dual Linear Regulator Problem

Denote

$$\begin{cases} \mathcal{R} = \mathcal{D}_t(\mathcal{U}), \ B = D_t(C), \ F = D_t(W), \\ G = D_t(V), \ L = D_t(R), \ Q_t = P_0, \end{cases} \tag{6.9}$$

where the transformations $\mathcal{D}_t$ and D_t are defined in Section 3.2.2. Consider the linear regulator problem of minimizing the functional

$$J(\eta) = \langle \xi_t^\eta, Q_t \xi_t^\eta \rangle + \int_0^t (\langle \xi_s^\eta, F_s \xi_s^\eta \rangle + \langle \eta_s, G_s \eta_s \rangle + 2\langle \eta_s, L_s \xi_s^\eta \rangle)\, ds, \tag{6.10}$$

where

$$\xi_s^\eta = -\mathcal{R}_{s,0} l + \int_0^s \mathcal{R}_{s,r} B_r \eta_r \, dr, \ 0 \le s \le t, \tag{6.11}$$

and η is an admissible control taken from $U_{\text{ad}} = L_2(0,t;\mathbb{R}^n)$.

Lemma 6.2. *Suppose that* $0 < t \le T$ *and* $l \in X$. *Then the control* $\eta^* \in L_2(0,t;\mathbb{R}^n)$ *is optimal in the linear regulator problem* (6.10)–(6.11) *if and only if it satisfies*

$$G_s \eta_s^* + \int_0^t \tilde{\Sigma}_{s,r}^t \eta_r^* \, dr = B_s^* \Sigma_{s,0}^t l + L_s \mathcal{R}_{s,0} l, \ \text{a.e. } s \in [0,t], \tag{6.12}$$

where

$$\Sigma_{s,r}^t = \mathcal{R}_{t,s}^* Q_t \mathcal{R}_{t,r} + \int_{\max(s,r)}^t \mathcal{R}_{\sigma,s}^* F_\sigma \mathcal{R}_{\sigma,r} \, d\sigma, \ s,r \in [0,t], \tag{6.13}$$

$$\tilde{\Sigma}_{s,r}^t = B_s^* \Sigma_{s,r}^t B_r + \begin{Bmatrix} L_s \mathcal{R}_{s,r} B_r, & r < s \\ B_s^* \mathcal{R}_{r,s}^* L_r^*, & r \ge s \end{Bmatrix} \ s,r \in [0,t]. \tag{6.14}$$

Proof. Let η^* be an optimal control in the problem (6.10)–(6.11) and let

$$\xi^* = \xi^{\eta^*}, \ \eta \in L_2(0,t;\mathbb{R}^n), \ \lambda \in \mathbb{R}, \ \nu_s = \int_0^s \mathcal{R}_{s,r} B_r \eta_r \, dr, \ 0 \le s \le t.$$

We have

$$\begin{aligned} 0 &\le J(\eta^* + \lambda\eta) - J(\eta^*) \\ &= 2\lambda\Big(\langle \nu_t, Q_t \xi_t^* \rangle + \int_0^t (\langle \nu_s, F_s \xi_s^* \rangle + \langle \eta_s, G_s \eta_s^* \rangle + \langle \eta_s, L_s \xi_s^* \rangle + \langle \eta_s^*, L_s \nu_s \rangle)\, ds \Big) \\ &\quad + \lambda^2 \Big(\langle \nu_t, Q_t \nu_t \rangle + \int_0^t (\langle \nu_s, F_s \nu_s \rangle + \langle \eta_s, G_s \eta_s \rangle + 2\langle \eta_s, L_s \nu_s \rangle)\, ds \Big). \end{aligned}$$

Dividing both sides of the obtained inequality consequently by $\lambda > 0$ and $\lambda < 0$ and then tending λ to 0, we obtain

$$\langle \nu_t, Q_t \xi_t^* \rangle + \int_0^t (\langle \nu_s, F_s \xi_s^* \rangle + \langle \eta_s, G_s \eta_s^* \rangle + \langle \eta_s, L_s \xi_s^* \rangle + \langle \eta_s^*, L_s \nu_s \rangle)\, ds = 0.$$

Substituting ν in the obtained equality and using the arbitrariness of η in the space $L_2(0,t;\mathbb{R}^n)$, we obtain that for a.e. $s \in [0,t]$,

$$G_s \eta_s^* + B_s^* \mathcal{R}_{t,s}^* Q_t \xi_t^* + \int_s^t B_s^* \mathcal{R}_{r,s}^* F_r \xi_r^*\, dr + L_s \xi_s^* + \int_s^t B_s^* \mathcal{R}_{r,s}^* L_r^* \eta_r^*\, dr = 0.$$

Now, substituting ξ^* from (6.11) in the obtained equality and using (6.13) and (6.14), we obtain (6.12). Also, running back in the above calculations, we obtain that if η^* satisfies (6.12), then it is an optimal control in the problem (6.10)–(6.11). □

Theorem 6.3. *Let* $0 < t \le T$. *Then under the conditions* ($\mathbf{E_1^w}$), ($\mathbf{E_2^w}$) *and* (6.9), *the best estimate* $\hat{x}_t$ *of* x_t *based on* z_s, $0 \le s \le t$, *in the filtering problem* (6.1)–(6.2) *is equal to* (6.4) (*when* $\tau = t$) *if and only if the function, defined by* $\eta_s^* = K_{t-s}^* l$, *a.e.* $s \in [0,t]$, *is an optimal control in the linear regulator problem* (6.10)–(6.11) *for all* $l \in X$.

Proof. By Lemmas 6.1 and 6.2, it is sufficient to show that K is a solution of the equation (6.5) (when $\tau = t$) if and only if $\eta_s^* = K_{t-s}^* l$, a.e. $s \in [0,t]$, is a solution of the equation (6.12). We will use the relation (6.9) between the functions included in the equations (6.5) and (6.12). First, let us show that

$$\Sigma_{s,r}^t = \Lambda_{t-r,t-s}^* \quad \text{and} \quad \tilde{\Sigma}_{s,r}^t = \tilde{\Lambda}_{t-r,t-s}^*, \tag{6.15}$$

where Σ^t, $\tilde{\Sigma}^t$, Λ and $\tilde{\Lambda}$ are defined by (6.13)–(6.14) and (6.6)–(6.7). These equalities follow from

$$\begin{aligned}
\Sigma_{s,r}^t &= \mathcal{U}_{t-s,0} P_0 \mathcal{U}_{t-r,0}^* + \int_{\max(s,r)}^t \mathcal{U}_{t-s,t-\sigma} W_{t-\sigma} \mathcal{U}_{t-r,t-\sigma}^*\, d\sigma \\
&= \mathcal{U}_{t-s,0} P_0 \mathcal{U}_{t-r,0}^* + \int_0^{\min(t-s,t-r)} \mathcal{U}_{t-s,\sigma} W_\sigma \mathcal{U}_{t-r,\sigma}^*\, d\sigma \\
&= \Lambda_{t-r,t-s}^*, \\
\tilde{\Sigma}_{s,r}^t &= B_s^* \Sigma_{s,r}^t B_r + \begin{Bmatrix} L_s \mathcal{R}_{s,r} B_r, & r < s \\ B_s^* \mathcal{R}_{r,s}^* L_r^*, & r \ge s \end{Bmatrix} \\
&= C_{t-s} \Lambda_{t-r,t-s}^* C_{t-r}^* + \begin{Bmatrix} R_{t-s}^* \mathcal{U}_{t-r,t-s}^* C_{t-r}^*, & r < s \\ C_{t-s} \mathcal{U}_{t-s,t-r} R_{t-r}, & r \ge s \end{Bmatrix} \\
&= \tilde{\Lambda}_{t-r,t-s}^*.
\end{aligned}$$

Now let $\eta_s^* = K_{t-s}^* l$, $0 \le s \le t$, be a solution of the equation (6.12). By (6.9) and (6.15), we have

$$V_{t-s}K_{t-s}^* l + \int_0^t \tilde{\Lambda}_{t-r,t-s}^* K_{t-r}^* l \, dr = C_{t-s}\Lambda_{t,t-s}^* l + R_{t-s}^* \mathcal{U}_{t,t-s}^* l.$$

Using the arbitrariness of l and replacing s by $t-s$, we obtain

$$V_s K_s^* + \int_0^t \tilde{\Lambda}_{t-r,s}^* K_{t-r}^* \, dr = C_s \Lambda_{t,s}^* + R_s^* \mathcal{U}_{t,s}^*$$

or

$$V_s K_s^* + \int_0^t \tilde{\Lambda}_{r,s}^* K_r^* \, dr = C_s \Lambda_{t,s}^* + R_s^* \mathcal{U}_{t,s}^*.$$

Taking adjoints of the left- and right-hand sides of the last equality, we conclude that K is a solution of the equation (6.5) (when $\tau = t$). In a similar way, considering K as a solution of the equation (6.5) (when $\tau = t$), one can show that $\eta_s^* = K_{t-s}^* l$, a.e. $s \in [0, t]$, is a solution of the equation (6.12) for all $l \in X$. □

Theorem 6.3 expresses the duality between the filtering problem (6.1)–(6.2) and the linear regulator problem (6.10)–(6.11). According to this duality, the optimal filter in the problem (6.1)–(6.2) and the optimal control in the problem (6.10)–(6.11) are related as the functions used in the setting of these problems. This relation includes (a) reversing time, (b) taking adjoints and (c) corresponding rearrangement.

6.2.2 Optimal Linear Feedback Filter

Theorems 6.3 and 5.24 lead to the following result.

Theorem 6.4. *Under the conditions* $(\mathbf{E_1^w})$ *and* $(\mathbf{E_2^w})$, *there exists a unique optimal linear feedback filter in the filtering problem* (6.1)–(6.2) *and the best estimate* $\hat{x}_t$ *of* x_t *based on* z_s, $0 \le s \le t$, *is equal to*

$$\hat{x}_t = \int_0^t \mathcal{Y}_{t,s}(P_s C_s^* + R_s) V_s^{-1} \, dz_s, \ \ 0 \le t \le T, \tag{6.16}$$

where $\mathcal{Y} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$ *and* P *is a unique solution of the dual Riccati equation*

$$P_s = \mathcal{U}_{s,0} P_0 \mathcal{U}_{s,0}^* + \int_0^s \mathcal{U}_{s,r}\big(W_r - (P_r C_r^* + R_r) V_r^{-1}(C_r P_r + R_r^*)\big)\mathcal{U}_{s,r}^* \, dr, \ \ 0 \le s \le T. \tag{6.17}$$

Proof. The case $t = 0$ is trivial. Let $0 < t \le T$. Introduce the notation (6.9) and consider the linear regulator problem (6.10)–(6.11). We have

$$\begin{aligned} F_s - L_s^* G_s^{-1} L_s &= W_{t-s} - R_{t-s} V_{t-s}^{-1} R_{t-s}^* \\ &= \Phi_{t-s} \bar{W} \Phi_{t-s}^* - \Phi_{t-s} \bar{R} \Psi_{t-s}^* \big(\Psi_{t-s} \bar{V} \Psi_{t-s}^*\big)^{-1} \Psi_{t-s} \bar{R}^* \Phi_{t-s}^* \\ &= \Phi_{t-s} \big(\bar{W} - \bar{R} \bar{V}^{-1} \bar{R}^*\big) \Phi_{t-s}^* \\ &= T^{-1} \Phi_{t-s} \mathrm{cov}\big(w_T - \bar{R} \bar{V}^{-1} v_T\big) \Phi_{t-s}^* \\ &\ge 0, \text{ a.e. } s \in [0,t]. \end{aligned}$$

Also, in view of $\bar{V} > 0$ and $\|h\|^2 = \big\|\Psi_{t-s}^{-1*} \Psi_{t-s}^* h\big\|^2 \le \big\|\Psi_{t-s}^{-1*}\big\|^2 \|\Psi_{t-s}^* h\|^2$,

$$\begin{aligned} \langle G_s h, h\rangle &= \big\langle \Psi_{t-s} \bar{V} \Psi_{t-s}^* h, h\big\rangle \\ &\ge c\|\Psi_{t-s}^* h\|^2 \ge c\big\|\Psi_{t-s}^{-1*}\big\|^{-2} \|h\|^2, \text{ a.e. } s \in [0,t], \end{aligned}$$

where $c > 0$ is a constant. Therefore, Theorem 5.24 can be applied to the problem (6.10)–(6.11) according to which an optimal control in the problem (6.10)–(6.11) is unique and has the form

$$\eta_s^* = G_s^{-1}(B_s^* Q_s + L_s)\mathcal{K}_{s,0} l, \text{ a.e. } s \in [0,t], \tag{6.18}$$

where $\mathcal{K} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{R})$ and Q is a unique solution of the equation

$$\begin{aligned} Q_s = \mathcal{R}_{t,s}^* Q_t \mathcal{R}_{t,s} + \int_s^t \mathcal{R}_{r,s}^* \big(F_r \\ - (Q_r B_r + L_r^*) G_r^{-1} (B_r^* Q_r + L_r)\big) \mathcal{R}_{r,s}\, dr, \ 0 \le s \le t. \end{aligned} \tag{6.19}$$

Let $P_s = Q_{t-s}$. Then by (6.19), we have

$$\begin{aligned} P_s &= \mathcal{R}_{t,t-s}^* P_0 \mathcal{R}_{t,t-s} + \int_{t-s}^t \mathcal{R}_{r,t-s}^* \big(F_r \\ &\qquad -(P_{t-r} B_r + L_r^*) G_r^{-1} (B_r^* P_{t-r} + L_r)\big) \mathcal{R}_{r,t-s}\, dr \\ &= \mathcal{R}_{t,t-s}^* P_0 \mathcal{R}_{t,t-s} + \int_0^s \mathcal{R}_{t-r,t-s}^* \big(F_{t-r} \\ &\qquad -(P_r B_{t-r} + L_{t-r}^*) G_{t-r}^{-1} (B_{t-r}^* P_r + L_{t-r})\big) \mathcal{R}_{t-r,t-s}\, dr. \end{aligned}$$

So by (6.9), we obtain that P satisfies the equation (6.17). Note that the equation (6.17) defines P on $\mathbf{T}$ while (6.19) defines Q only on $[0,t]$. By Theorem 3.24, P is a unique solution of the equation (6.17) in $B(\mathbf{T}, \mathcal{L}(X))$ satisfying $P_s \ge 0$ for $0 \le s \le T$. Now consider the mild evolution operator $\mathcal{K}$. It is defined on the triangular set $\Delta_t = \{(s,r) : 0 \le r \le s \le t\}$. By Proposition 3.18(a) and (6.9), we have

$$\mathcal{D}_t(\mathcal{K}) = \mathcal{D}_t\left(\mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{D}_t(\mathcal{U}))\right) = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})|_{\Delta_t} = \mathcal{Y}|_{\Delta_t}.$$

Therefore, $\mathcal{D}_t(\mathcal{K})$ is the restriction of $\mathcal{Y}$ from Δ_T to the narrower triangular set Δ_t. Finally, using (6.9), the relation between $(\mathcal{K}, Q)$ and $(\mathcal{Y}, P)$ and Theorem 6.3, we conclude that there exists a unique optimal linear feedback filter in the filtering problem (6.1)–(6.2) and it is determined by

$$K_s = K_{t,s} = \mathcal{Y}_{t,s}(P_s C_s^* + R_s)V_s^{-1}, \text{ a.e. } s \in [0, t]. \tag{6.20}$$

This implies (6.16). □

6.2.3 Error Process

The difference $e_t = x_t - \hat{x}_t$, $0 \le t \le T$, is called the *error process* in the filtering problem (6.1)–(6.2).

Proposition 6.5. *For the error process in the filtering problem* (6.1)–(6.2), *the equality* $\text{cov} e_t = P_t$, $0 \le t \le T$, *holds where* P *is a solution of the equation* (6.17).

Proof. Obviously,

$$\text{cov} e_0 = \text{cov} x_0 = P_0.$$

Let $0 < t \le T$. Since $\hat{x}_t$ is the projection of x_t from $L_2(\Omega, X)$ onto its subspace $L_2(\Omega, \sigma(z_s;\ 0 \le s \le t), \mathbf{P}, X)$,

$$\text{cov}(x_t, \hat{x}_t) = \text{cov}\hat{x}_t + \text{cov}(x_t - \hat{x}_t, \hat{x}_t) = \text{cov}\hat{x}_t.$$

Therefore,

$$\text{cov} e_t = \text{cov}(x_t - \hat{x}_t) = \text{cov} x_t - \text{cov}\hat{x}_t. \tag{6.21}$$

Recall that

$$\text{cov} x_t = \Lambda_{t,t} = \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,r} W_r \mathcal{U}_{t,r}^* \, dr. \tag{6.22}$$

Let us compute $\text{cov}\hat{x}_t$. Similar to the proof of Lemma 6.1, we have

$$\text{cov}\hat{x}_t = \text{cov} \int_0^t K_s \, dz_s = \int_0^t \left(K_s V_s + \int_0^t K_r \tilde{\Lambda}_{r,s} \, dr \right) K_s^* \, ds,$$

where K is defined by (6.20). Since K is a solution of the Wiener–Hopf equation (6.5) when $\tau = t$,

$$\begin{aligned} \text{cov}\hat{x}_t &= \int_0^t (\Lambda_{t,s} C_s^* + \mathcal{U}_{t,s} R_s) K_s^* \, ds \\ &= \int_0^t \mathcal{U}_{t,s} \left(\left(\mathcal{U}_{s,0} P_0 \mathcal{U}_{s,0}^* + \int_0^s \mathcal{U}_{s,r} W_r \mathcal{U}_{s,r}^* \, dr \right) C_s^* + R_s \right) K_s^* \, ds. \end{aligned}$$

By use of (6.17) and (6.20), we have

$$\begin{aligned}\operatorname{cov}\hat{x}_t = &\int_0^t \mathcal{U}_{t,s} R_s V_s^{-1}(C_s P_s + R_s^*)\mathcal{Y}_{t,s}^* \, ds \\ &+ \int_0^t \mathcal{U}_{t,s} P_s C_s^* V_s^{-1}(C_s P_s + R_s^*)\mathcal{Y}_{t,s}^* \, ds \\ &+ \int_0^t \int_0^s \mathcal{U}_{t,r}(P_r C_r^* + R_r)V_r^{-1}(C_r P_r + R_r^*)\mathcal{U}_{s,r}^* C_s^* V_s^{-1} \\ &\qquad \times (C_s P_s + R_s^*)\mathcal{Y}_{t,s}^* \, dr\, ds.\end{aligned}$$

Since $\mathcal{Y} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$, we obtain

$$\begin{aligned}\operatorname{cov}\hat{x}_t &= \int_0^t \mathcal{U}_{t,s}(P_s C_s^* + R_s)V_s^{-1}(C_s P_s + R_s^*)\mathcal{Y}_{t,s}^* \, ds \\ &\quad + \int_0^t \mathcal{U}_{t,r}(P_r C_r^* + R_r)V_r^{-1}(C_r P_r + R_r^*)(\mathcal{U}_{t,r}^* - \mathcal{Y}_{t,r}^*)\, dr \\ &= \int_0^t \mathcal{U}_{t,s}(P_s C_s^* + R_s)V_s^{-1}(C_s P_s + R_s^*)\mathcal{U}_{t,s}^* \, ds. \qquad (6.23)\end{aligned}$$

Finally, combining (6.21), (6.22) and (6.23) and using (6.17), we conclude that $\operatorname{cov} e_t = P_t$. □

Proposition 6.6. *For the error process in the filtering problem* (6.1)–(6.2), *the equality*

$$\operatorname{cov}(e_t, e_s) = \begin{cases} \mathcal{Y}_{t,s} P_s, & 0 \le s < t \le T, \\ P_t \mathcal{Y}_{s,t}^*, & 0 \le t < s \le T, \end{cases}$$

holds, where $\mathcal{Y} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$ *and* P *is a solution of the equation* (6.17).

Proof. Let $0 \le s < t \le T$. Since $\mathcal{Y} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$, by Propositions 4.28 and 4.29,

$$x_t = \mathcal{Y}_{t,s} x_s + \int_s^t \mathcal{Y}_{t,r}(P_r C_r^* + R_r)V_r^{-1} C_r x_r \, dr + \int_s^t \mathcal{Y}_{t,r} \Phi_r \, dw_r$$

and

$$\hat{x}_t = \mathcal{Y}_{t,s} \hat{x}_s + \int_s^t \mathcal{Y}_{t,r}(P_r C_r^* + R_r)V_r^{-1}(C_r x_r \, dr + \Psi_r \, dv_r).$$

Therefore,

$$e_t = \mathcal{Y}_{t,s} e_s - \int_s^t \mathcal{Y}_{t,r}(P_r C_r^* + R_r)V_r^{-1}\Psi_r \, dv_r + \int_s^t \mathcal{Y}_{t,r}\Phi_r \, dw_r.$$

Since $\sigma(x_0, w_r, v_r;\ 0 \le r \le s)$ and $\sigma(w_r - w_s, v_r - v_s;\ s < r \le T)$ are independent,

$$\operatorname{cov}(e_t, e_s) = \operatorname{cov}(\mathcal{Y}_{t,s} e_s, e_s) = \mathcal{Y}_{t,s} P_s,$$

where Proposition 6.5 is used. If $0 \le t < s \le T$, then in a similar way one can show that $\operatorname{cov}(e_t, e_s) = P_t \mathcal{Y}_{s,t}^*$. □

6.2.4 Innovation Process

The random process

$$\bar{z}_t = z_t - \int_0^t C_s \hat{x}_s \, ds, \ 0 \le t \le T, \tag{6.24}$$

is called the *innovation process* associated with the estimation problem (6.1)–(6.2).

Proposition 6.7. *Let $\mathcal{F}_t^z$ be the natural filtration of z and let $\bar{z}$ be the innovation process defined by* (6.24)*. Then $\{\bar{z}_t, \mathcal{F}_t^z\} \in M_2^c(\mathbf{T}, \mathbb{R}^n)$.*

Proof. Obviously, $\bar{z}_t \in L_2(\Omega, \mathcal{F}_t^z, \mathbf{P}, \mathbb{R}^n)$ for all $0 \le t \le T$, $\bar{z}$ has continuous paths and $\bar{z}_0 = 0$. Let us show that $\bar{z}$ is a martingale. By (6.2),

$$\begin{aligned}\bar{z}_t &= \int_0^t C_r x_r \, dr + \int_0^t \Psi_r \, dv_r - \int_0^t C_r \hat{x}_r \, dr \\ &= \int_0^t C_r (x_r - \hat{x}_r) \, dr + \int_0^t \Psi_r \, dv_r.\end{aligned}$$

Hence, for $0 \le s < t \le T$, we have

$$\mathbf{E}(\bar{z}_t - \bar{z}_s | \mathcal{F}_s^z) = \int_s^t C_r \mathbf{E}(x_r - \mathbf{E}(x_r|\mathcal{F}_r^z)|\mathcal{F}_s^z) \, dr - \mathbf{E}\left(\int_s^t \Psi_r \, dv_r \middle| \mathcal{F}_s^z \right) = 0.$$

This implies $\bar{z}_s = \mathbf{E}(\bar{z}_t|\mathcal{F}_s^z)$. Thus, $\{\bar{z}_t, \mathcal{F}_t^z\} \in M_2^c(\mathbf{T}, \mathbb{R}^n)$. □

Proposition 6.8. *The innovation process* (6.24) *has the representation*

$$\bar{z}_t = \int_0^t \Psi_s \, d\gamma_s, \ 0 \le t \le T, \tag{6.25}$$

where $\{\gamma_t, \mathcal{F}_t^z\}$ is an $\mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with the covariance operator $\bar{V}$.

Proof. It is sufficient to show that the random process γ, defined by

$$\gamma_t = \int_0^t \Psi_s^{-1} \, d\bar{z}_s, \ 0 \le t \le T, \tag{6.26}$$

is a Wiener process with respect to the filtration $\{\mathcal{F}_t^z\}$ and its covariance operator is $\bar{V}$. By Proposition 6.7 and Theorem 4.21(b), $\{\gamma_t, \mathcal{F}_t^z\} \in M_2^c(\mathbf{T}, \mathbb{R}^n)$. In order to study the Dolean measure and the covariance function of γ, at first, note that, if

$0 \le s \le t$, then

$$\begin{aligned}
\operatorname{cov}(v_t, e_s) &= \operatorname{cov}\left(v_t, \int_0^s \mathcal{U}_{s,r}\Phi_r\, dw_r - \int_0^s K_{s,r}(C_r x_r\, dr + \Psi_r\, dv_r)\right) \\
&= \operatorname{cov}\left(v_t, \int_0^s \mathcal{U}_{s,r}\Phi_r\, dw_r\right) - \operatorname{cov}\left(v_t, \int_0^s \int_0^r K_{s,r}C_r\mathcal{U}_{r,\alpha}\Phi_\alpha\, dw_\alpha\, dr\right) \\
&\quad - \operatorname{cov}\left(v_t, \int_0^s K_{s,r}\Psi_r\, dv_r\right) \\
&= \int_0^s \bar{R}^*\Phi_r^*\mathcal{U}_{s,r}^*\, dr - \int_0^s \int_0^r \bar{R}^*\Phi_\alpha^*\mathcal{U}_{r,\alpha}^* C_r^* K_{s,r}^*\, d\alpha\, dr - \int_0^s \bar{V}\Psi_r^* K_{s,r}^*\, dr \\
&= \int_0^s \bar{R}^*\Phi_\alpha^*\left(\mathcal{U}_{s,\alpha}^* - \int_\alpha^s \mathcal{U}_{r,\alpha}^* C_r^* K_{s,r}^*\, dr\right) d\alpha - \int_0^s \bar{V}\Psi_r^* K_{s,r}^*\, dr,
\end{aligned}$$

where e is the error process (see Section 6.2.3) and K is defined by (6.20). From $\mathcal{Y} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$ and (6.20), we obtain that $0 \le s \le t$ implies

$$\begin{aligned}
\operatorname{cov}(v_t, e_s) &= \int_0^s \bar{R}^*\Phi_\alpha^*\mathcal{Y}_{s,\alpha}^*\, d\alpha - \int_0^s \bar{V}\Psi_r^*\left(\Psi_r\bar{V}\Psi_r^*\right)^{-1}\left(C_r P_r + \Psi_r\bar{R}^*\Phi_r^*\right)\mathcal{Y}_{s,r}^*\, dr \\
&= \int_0^s \bar{R}^*\Phi_r^*\mathcal{Y}_{s,r}^*\, dr - \int_0^s \Psi_r^{-1}\left(C_r P_r + \Psi_r\bar{R}^*\Phi_r^*\right)\mathcal{Y}_{s,r}^*\, dr \\
&= -\int_0^s \Psi_r^{-1} C_r P_r \mathcal{Y}_{s,r}^*\, dr.
\end{aligned} \tag{6.27}$$

Thus, using (6.27) and Proposition 6.6, we can calculate

$$\begin{aligned}
\operatorname{cov}\gamma_t &= \operatorname{cov}\int_0^t \Psi_s^{-1}\, d\bar{z}_s = \operatorname{cov}\left(\int_0^t \Psi_s^{-1}\, dz_s - \int_0^t \Psi_s^{-1} C_s\hat{x}_s\, ds\right) \\
&= \operatorname{cov}\left(\int_0^t \Psi_s^{-1}\Psi_s\, dv_s + \int_0^t \Psi_s^{-1} C_s(x_s - \hat{x}_s)\, ds\right) \\
&= \operatorname{cov}\left(v_t + \int_0^t \Psi_s^{-1} C_s e_s\, ds\right) \\
&= \bar{V}t + \int_0^t \operatorname{cov}(v_t, e_s) C_s^*\Psi_s^{-1*}\, ds + \int_0^t \Psi_s^{-1} C_s \operatorname{cov}(e_s, v_t)\, ds \\
&\quad + \int_0^t \int_0^t \Psi_r^{-1} C_r \operatorname{cov}(e_r, e_s) C_s^*\Psi_s^{-1*}\, ds\, dr \\
&= \bar{V}t - \int_0^t \int_0^s \Psi_r^{-1} C_r P_r \mathcal{Y}_{s,r}^* C_s^*\Psi_s^{-1*}\, dr\, ds \\
&\quad - \int_0^t \int_0^s \Psi_s^{-1} C_s \mathcal{Y}_{s,r} P_r C_r^*\Psi_r^{-1*}\, dr\, ds + \int_0^t \int_0^r \Psi_r^{-1} C_r \mathcal{Y}_{r,s} P_s C_s^*\Psi_s^{-1*}\, ds\, dr \\
&\quad + \int_0^t \int_r^t \Psi_r^{-1} C_r P_r \mathcal{Y}_{s,r}^* C_s^*\Psi_s^{-1*}\, ds\, dr = \bar{V}t.
\end{aligned}$$

Now let us study the Dolean measure λ and the covariance function M of γ (see Section 4.1.4). Let $(s,t] \times F$ be any predictable rectangle with respect to the filtration $\{\mathcal{F}_t^z\}$. It is easy to see that γ, $\bar{z}$ and z define a Gaussian system. Therefore, $\mathbf{E}(\bar{z}_t - \bar{z}_s)|\mathcal{F}_s^z) = 0$ implies the independence of $\bar{z}_t - \bar{z}_s$ and $\mathcal{F}_s^z$ and, hence, the independence of $\gamma_t - \gamma_s$ and $\mathcal{F}_s^z$. Using this fact and Proposition 4.2(f), we obtain

$$\begin{aligned}
\lambda((s,t] \times F) &= \mathbf{E}\left(\chi_F \|\gamma_t - \gamma_s\|^2\right) \\
&= \mathbf{P}(F)\mathbf{E}\|\gamma_t - \gamma_s\|^2 \\
&= \mathbf{P}(F)\,\mathrm{tr}(\mathrm{cov}(\gamma_t - \gamma_s)) \\
&= \mathbf{P}(F)(\mathrm{tr}(\mathrm{cov}\gamma_t) - \mathrm{tr}(\mathrm{cov}\gamma_s)) \\
&= \mathbf{P}(F)\big(\mathrm{tr}\bar{V}\big)(t-s) \\
&= \big(\mathrm{tr}\bar{V}\big)\mathbf{P}(F)\ell((s,t]).
\end{aligned}$$

So, $\lambda = \big(\mathrm{tr}\bar{V}\big)(\mathbf{P} \otimes \ell)$. Respectively, for the covariance function M of γ, we have

$$\begin{aligned}
\int_{(s,t]\times F} M\, d\lambda &= \mathbf{E}(\chi_F((\gamma_t - \gamma_s) \otimes (\gamma_t - \gamma_s))) \\
&= \mathbf{P}(F)\,\mathrm{cov}(\gamma_t - \gamma_s) = \mathbf{P}(F)\bar{V}(t-s) \\
&= \int_{(s,t]} \int_F \bar{V}\, d\mathbf{P}\, dt = \big(\mathrm{tr}\bar{V}\big)^{-1} \int_{(s,t]\times F} \bar{V}\, d\lambda.
\end{aligned}$$

We conclude that $M_{t,\omega} = \big(\mathrm{tr}\bar{V}\big)^{-1}\bar{V}$, $(t,\omega) \in \mathbf{T} \times \Omega$. Thus, by Definition 4.10, $\{\gamma_t, \mathcal{F}_t^z\}$ is an $\mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with the covariance operator $\bar{V}$. □

Theorem 6.9. *The best estimate in the filtering problem* (6.1)–(6.2) *has the following representation in the form of a stochastic integral with respect to the innovation process:*

$$\hat{x}_t = \int_0^t \mathcal{U}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\, d\bar{z}_s, \quad 0 \le t \le T. \tag{6.28}$$

Proof. By Theorem 6.4 and Proposition 4.28, we have

$$\begin{aligned}
\hat{x}_t &= \int_0^t \mathcal{Y}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\, dz_s \\
&= \int_0^t \mathcal{U}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\, dz_s \\
&\quad - \int_0^t \mathcal{U}_{t,s}(P_s C_s^* + R_s)V_s^{-1}C_s\hat{x}_s\, ds \\
&= \int_0^t \mathcal{U}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\, d\bar{z}_s,
\end{aligned}$$

where $\mathcal{Y}$ and $\mathcal{U}$ are related as in Theorem 6.4. □

Example 6.10. Let $\mathcal{U}_t$, $t \geq 0$, be a strongly continuous semigroup (see Examples 3.3–3.9 and 3.42–3.45 for the definitions of useful specific semigroups) and let A be its infinitesimal generator. Then by Theorem 6.9, the best estimate in the filtering problem (6.1)–(6.2) is a mild solution of the linear stochastic differential equation

$$d\hat{x}_t = A\hat{x}_t dt + (P_t C_t^* + R_t)V_t^{-1}(dz_t - C_t\hat{x}_t dt), \; 0 < t \leq T, \; \hat{x}_0 = 0.$$

6.3 Prediction

In this section we suppose that the conditions $(\mathbf{E}_1^w)$ and $(\mathbf{E}_2^w)$ hold and, additionally, $0 \leq \tau < t \leq T$.

6.3.1 Dual Linear Regulator Problem

Introduce the notation (6.9) and consider the linear regulator problem of minimizing the functional

$$J(\eta) = \langle \xi_t^\eta, Q_t \xi_t^\eta \rangle + \int_{t-\tau}^t (\langle \xi_s^\eta, F_s \xi_s^\eta \rangle + \langle \eta_s, G_s \eta_s \rangle + 2\langle \eta_s, L_s \xi_s^\eta \rangle)\, ds, \tag{6.29}$$

where

$$\xi_s^\eta = -\mathcal{R}_{s,0} l + \int_{t-\tau}^s \mathcal{R}_{s,r} B_r \eta_r \, dr, \; t - \tau \leq s \leq t, \tag{6.30}$$

and η is an admissible control taken from $U_{\mathrm{ad}} = L_2(t-\tau, t; \mathbb{R}^n)$.

Lemma 6.11. *Suppose that* $0 < \tau < t \leq T$ *and* $l \in X$. *Then a control* $\eta^* \in L_2(t-\tau, t; \mathbb{R}^n)$ *is optimal in the linear regulator problem* (6.29)–(6.30) *if and only if it satisfies*

$$G_s \eta_s^* + \int_{t-\tau}^t \tilde{\Sigma}_{s,r}^t \eta_r^* \, dr = B_s^* \Sigma_{s,0}^t l + L_s \mathcal{R}_{s,0} l, \; a.e. \; s \in [t-\tau, t], \tag{6.31}$$

where Σ^t *and* $\tilde{\Sigma}^t$ *are defined by* (6.13) *and* (6.14).

Proof. Let η^* be an optimal control in the problem (6.29)–(6.30) and let

$$\xi^* = \xi^{\eta^*}, \; \eta \in L_2(t-\tau, t; \mathbb{R}^n), \; \nu_s = \int_{t-\tau}^s \mathcal{R}_{s,r} B_r \eta_r \, dr, \; t-\tau \leq s \leq t. \tag{6.32}$$

Similar to the proof of Lemma 6.2, one can obtain

$$\langle \nu_t, Q_t \xi_t^* \rangle + \int_{t-\tau}^t (\langle \nu_s, F_s \xi_s^* \rangle + \langle \eta_s, G_s \eta_s^* \rangle + \langle \eta_s, L_s \xi_s^* \rangle + \langle \eta_s^*, L_s \nu_s \rangle)\, ds = 0.$$

Substituting ν from (6.32) in the obtained equality and using arbitrariness of η in $L_2(t-\tau,t;\mathbb{R}^n)$, we obtain that for a.e. $s \in [t-\tau,t]$,

$$G_s\eta_s^* + B_s^*\mathcal{R}_{t,s}^*Q_t\xi_t^* + \int_s^t B_s^*\mathcal{R}_{r,s}^*F_r\xi_r^*\,dr + L_s\xi_s^* + \int_s^t B_s^*\mathcal{R}_{r,s}^*L_r^*\eta_r^*\,dr = 0.$$

Now substituting ξ^* from (6.30) in the obtained equality and using (6.13) and (6.14), we obtain (6.31). Also, running back in the above calculations, we conclude that if η^* satisfies (6.31), then it is an optimal control in the problem (6.29)–(6.30). □

Theorem 6.12. *Let $0 < \tau < t \le T$. Then under the conditions $(\mathbf{E_1^w})$, $(\mathbf{E_2^w})$ and (6.9), the best estimate $\hat{x}_t^\tau$ of x_t based on z_s, $0 \le s \le \tau$, in the prediction problem (6.1)–(6.2) is equal to (6.4) if and only if the function, defined by $\eta_s^* = K_{t-s}^*l$, a.e. $s \in [t-\tau,t]$, is an optimal control in the linear regulator problem (6.29)–(6.30) for all $l \in X$.*

Proof. This can be proved in a similar way as Theorem 6.3 by use of Lemmas 6.1 and 6.11. □

By Theorem 6.12, the linear regulator problem (6.29)–(6.30) is dual to the prediction problem (6.1)–(6.2).

6.3.2 Optimal Linear Feedback Predictor

Theorems 6.12 and 5.24 lead to the following result.

Theorem 6.13. *Under the conditions $(\mathbf{E_1^w})$ and $(\mathbf{E_2^w})$, there exists a unique optimal linear feedback predictor in the prediction problem (6.1)–(6.2) and the best estimate $\hat{x}_t^\tau$ of x_t based on z_s, $0 \le s \le \tau$, is equal to*

$$\hat{x}_t^\tau = \mathcal{U}_{t,\tau}\hat{x}_\tau, \quad 0 \le \tau < t \le T,$$

where $\hat{x}_\tau$ is the best estimate in the filtering problem (6.1)–(6.2).

Proof. The case $\tau = 0$ is trivial. Let $0 < \tau < t \le T$. Introduce the notation (6.9) and consider the linear regulator problem (6.29)–(6.30). By Theorem 5.24, the optimal control in the problem (6.29)–(6.30) is unique and has the form

$$\eta_s^* = G_s^{-1}(B_s^*Q_s + L_s)\mathcal{K}_{s,t-\tau}\mathcal{R}_{t-\tau,0}l, \text{ a.e. } s \in [t-\tau,t],$$

where $\mathcal{K} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{R})$ and Q is a solution of the equation (6.19) on interval $[t-\tau,t]$. In proving Theorem 6.4, it was shown that the function $P_s = Q_{t-s}$, $0 \le s \le t$, is the unique solution of the Riccati equation (6.17) and $\mathcal{D}_t(\mathcal{K}) = \mathcal{Y}|_{\Delta_t}$, where $\mathcal{Y} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$. Thus, using the notation (6.9), by Theorem 6.12, we easily obtain that there exists a unique optimal linear feedback predictor in the prediction problem (6.1)–(6.2) which is determined by

$$K_s = K_{t,\tau,s} = \mathcal{U}_{t,\tau}\mathcal{Y}_{\tau,s}(P_sC_s^* + R_s)V_s^{-1}, \text{ a.e. } s \in [0,\tau].$$

Consequently, the best estimate $\hat{x}_t^\tau$ in the prediction problem (6.1)–(6.2) is equal to

$$\hat{x}_t^\tau = \mathcal{U}_{t,\tau}\int_0^\tau \mathcal{Y}_{\tau,s}(P_sC_s^* + R_s)V_s^{-1}\,dz_s = \mathcal{U}_{t,\tau}\hat{x}_\tau.$$

The proof is completed. □

6.4 Smoothing

In this section we suppose that the conditions ($\mathbf{E_1^w}$) and ($\mathbf{E_2^w}$) hold and, additionally, $0 \le t < \tau \le T$.

6.4.1 Dual Linear Regulator Problem

Introduce the notation (6.9) in which t is replaced by τ, i.e.,

$$\begin{cases} \mathcal{R} = \mathcal{D}_\tau(\mathcal{U}),\ B = D_\tau(C),\ F = D_\tau(W), \\ G = D_\tau(V),\ L = D_\tau(R),\ Q_\tau = P_0. \end{cases} \tag{6.33}$$

Consider the linear regulator problem of minimizing the functional

$$J(\eta) = \langle \xi_\tau^\eta, Q_\tau\xi_\tau^\eta\rangle + \int_0^\tau (\langle \xi_s^\eta, F_s\xi_s^\eta\rangle + \langle \eta_s, G_s\eta_s\rangle + 2\langle \eta_s, L_s\xi_s^\eta\rangle)\,ds, \tag{6.34}$$

where

$$\xi_s^\eta = \begin{Bmatrix} -\mathcal{R}_{s,\tau-t}l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix} + \int_0^s \mathcal{R}_{s,r}B_r\eta_r\,dr,\ 0 \le s \le \tau, \tag{6.35}$$

and η is an admissible control taken from $U_{\text{ad}} = L_2(0,\tau;\mathbb{R}^n)$.

Lemma 6.14. *Suppose that $0 \le t < \tau \le T$ and $l \in X$. Then the control $\eta^* \in L_2(0,\tau;\mathbb{R}^n)$ is optimal in the linear regulator problem* (6.34)–(6.35) *if and only if it satisfies*

$$G_s\eta_s^* + \int_0^\tau \tilde{\Sigma}_{s,r}^\tau\eta_r^*\,dr = B_s^*\Sigma_{s,\tau-t}^\tau l + \chi_{[\tau-t,\tau]}(s)L_s\mathcal{R}_{s,\tau-t}l,\ \textit{a.e. } s \in [0,\tau], \tag{6.36}$$

where Σ^τ and $\tilde{\Sigma}^\tau$ are defined by (6.13) *and* (6.14).

Proof. Let η^* be an optimal control in the problem (6.34)–(6.35) and let

$$\xi^* = \xi^{\eta^*},\ \eta \in L_2(0,\tau;\mathbb{R}^n),\ \nu_s = \int_0^s \mathcal{R}_{s,r} B_r \eta_r \, dr,\ 0 \le s \le \tau. \tag{6.37}$$

Similar to the proof of Lemma 6.2, we have

$$\langle \nu_\tau, Q_\tau \xi_\tau^* \rangle + \int_0^\tau (\langle \nu_s, F_s \xi_s^* \rangle + \langle \eta_s, G_s \eta_s^* \rangle + \langle \eta_s, L_s \xi_s^* \rangle + \langle \eta_s^*, L_s \nu_s \rangle) ds = 0.$$

Substituting ν from (6.37) in the obtained equality and using the arbitrariness of η in $L_2(0,\tau;\mathbb{R}^n)$, we obtain that for a.e. $s \in [0,\tau]$,

$$G_s \eta_s^* + B_s^* \mathcal{R}_{\tau,s}^* Q_\tau \xi_\tau^* + \int_s^\tau B_s^* \mathcal{R}_{r,s}^* F_r \xi_r^* \, dr + L_s \xi_s^* + \int_s^\tau B_s^* \mathcal{R}_{r,s}^* L_r^* \eta_r^* \, dr = 0.$$

Now substituting ξ^* from (6.35) in the obtained equality and using (6.13) and (6.14), we obtain (6.36). Also, running back in the above calculations, we obtain that if η^* satisfies (6.36), then it is an optimal control in the problem (6.34)–(6.35). □

Theorem 6.15. *Let* $0 \le t < \tau \le T$. *Then under the conditions* ($\mathbf{E_1^w}$), ($\mathbf{E_2^w}$) *and* (6.33), *the best estimate* $\hat{x}_t^\tau$ *of* x_t *based on* z_s, $0 \le s \le \tau$, *in the smoothing problem* (6.1)–(6.2) *is equal to* (6.4) *if and only if the function, defined by* $\eta_s^* = K_{\tau-s}^* l$, *a.e.* $s \in [0,\tau]$, *is an optimal control in the linear regulator problem* (6.34)–(6.35) *for all* $l \in X$.

Proof. This can be proved in a similar way as Theorem 6.3 by use of Lemmas 6.1 and 6.14. □

Thus, the smoothing problem (6.1)–(6.2) and the linear regulator problem (6.34)–(6.35) are dual.

6.4.2 Optimal Linear Feedback Smoother

To find the formulae for the optimal linear feedback smoother, the problem (6.34)–(6.35) will be written in the following equivalent form:

$$\begin{aligned} J(\eta) = \langle \zeta_\tau^\eta - \rho_\tau, Q_\tau(\zeta_\tau^\eta - \rho_\tau) \rangle + \int_0^\tau (\langle \zeta_s^\eta - \rho_s, F_s(\zeta_s^\eta - \rho_s) \rangle \\ + \langle \eta_s, G_s \eta_s \rangle + 2\langle \eta_s, L_s(\zeta_s^\eta - \rho_s) \rangle) \, ds, \end{aligned} \tag{6.38}$$

where

$$\zeta_s^\eta = \int_0^s \mathcal{R}_{s,r} B_r \eta_r \, dr,\ 0 \le s \le \tau, \tag{6.39}$$

η is an admissible control in $U_{\text{ad}} = L_2(0,\tau;\mathbb{R}^n)$ and

$$\rho_s = \begin{Bmatrix} \mathcal{R}_{s,\tau-t} l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix},\ 0 \le s \le \tau. \tag{6.40}$$

One can observe that the functionals (6.34) and (6.38) are equal. According to Theorem 5.24, the optimal control η^* in the linear regulator problem (6.38)–(6.40) (and (6.34)–(6.35) as well) can be represented as

$$\eta_s^* = -G_s^{-1}\big((B_s^* Q_s + L_s)\zeta_s^* - B_s^*\alpha_s - L_s\rho_s\big), \text{ a.e. } s \in [0,\tau], \tag{6.41}$$

where

$$\zeta_s^* = \zeta_s^{\eta^*} = \int_0^s \mathcal{K}_{s,r} B_r G_r^{-1}(B_r^*\alpha_r + L_r\rho_r)\, dr, \ 0 \le s \le \tau, \tag{6.42}$$

$$\alpha_s = \mathcal{K}_{\tau,s}^* Q_\tau \rho_\tau + \int_s^\tau \mathcal{K}_{r,s}^*\big(F_r - (Q_r B_r + L_r^*)G_r^{-1}L_r\big)\rho_r\, dr, \ 0 \le s \le \tau, \tag{6.43}$$

$\mathcal{K} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{R})$ and Q is a solution of the Riccati equation

$$\begin{aligned} Q_s = {} & \mathcal{R}_{\tau,s}^* Q_\tau \mathcal{R}_{\tau,s} + \int_s^\tau \mathcal{R}_{r,s}^*\big(F_r \\ & - (Q_r B_r + L_r^*)G_r^{-1}(B_r^* Q_r + L_r)\big)\mathcal{R}_{r,s}\, dr, \ 0 \le s \le \tau. \end{aligned} \tag{6.44}$$

Lemma 6.16. *Under the above notation,*

$$\alpha_s = \left\{ \begin{array}{ll} Q_s \mathcal{R}_{s,\tau-t} l, & s \ge \tau - t \\ \mathcal{K}_{\tau-t,s}^* Q_{\tau-t} l, & s < \tau - t \end{array} \right\}, \ 0 \le s \le \tau. \tag{6.45}$$

Proof. By Proposition 3.23, the equation (6.44) is equivalent to

$$Q_s = \mathcal{K}_{\tau,s}^* Q_\tau \mathcal{R}_{\tau,s} + \int_s^\tau \mathcal{K}_{r,s}^*\big(F_r - (Q_r B_r + L_r^*)G_r^{-1}L_r\big)\mathcal{R}_{r,s}\, dr, \ 0 \le s \le \tau. \tag{6.46}$$

Let $\tau - t \le s \le \tau$. Using (6.40) and (6.46) in (6.43), we have

$$\begin{aligned} \alpha_s &= \mathcal{K}_{\tau,s}^* Q_\tau \mathcal{R}_{\tau,\tau-t} l + \int_s^\tau \mathcal{K}_{r,s}^*\big(F_r - (Q_r B_r + L_r^*)G_r^{-1}L_r\big)\mathcal{R}_{r,\tau-t} l\, dr \\ &= \left(\mathcal{K}_{\tau,s}^* Q_\tau \mathcal{R}_{\tau,s} + \int_s^\tau \mathcal{K}_{r,s}^*\big(F_r - (Q_r B_r + L_r^*)G_r^{-1}L_r\big)\mathcal{R}_{r,s}\, dr\right)\mathcal{R}_{s,\tau-t} l \\ &= Q_s \mathcal{R}_{s,\tau-t} l. \end{aligned}$$

If $0 \le s < \tau - t$, then in a similar way, we have

$$\begin{aligned} \alpha_s &= \mathcal{K}_{\tau,s}^* Q_\tau \mathcal{R}_{\tau,\tau-t} l + \int_{\tau-t}^\tau \mathcal{K}_{r,s}^*\big(F_r - (Q_r B_r + L_r^*)G_r^{-1}L_r\big)\mathcal{R}_{r,\tau-t} l\, dr \\ &= \mathcal{K}_{\tau-t,s}^*\Bigg(\mathcal{K}_{\tau,\tau-t}^* Q_\tau \mathcal{R}_{\tau,\tau-t} \\ &\qquad + \int_{\tau-t}^\tau \mathcal{K}_{r,\tau-t}^*\big(F_r - (Q_r B_r + L_r^*)G_r^{-1}L_r\big)\mathcal{R}_{r,\tau-t}\, dr\Bigg) l \\ &= \mathcal{K}_{\tau-t,s}^* Q_{\tau-t} l. \end{aligned}$$

Thus, the proof of the lemma is completed. □

Lemma 6.17. *Under the above notation,*

$$\zeta_s^* = \int_0^{\min(s,\tau-t)} \mathcal{K}_{s,r} B_r G_r^{-1} B_r^* \mathcal{K}_{\tau-t,r}^* Q_{\tau-t} l \, dr + \begin{Bmatrix} \mathcal{R}_{s,\tau-t} l - \mathcal{K}_{s,\tau-t} l, & s \geq \tau - t \\ 0, & s < \tau - t \end{Bmatrix}, \ 0 \leq s \leq \tau. \tag{6.47}$$

Proof. By (6.40) and (6.45), we have

$$B_r^* \alpha_r + L_r \rho_r = \begin{Bmatrix} (B_r^* Q_r + L_r) \mathcal{R}_{r,\tau-t} l, & r \geq \tau - t \\ B_r^* \mathcal{K}_{\tau-t,r}^* Q_{\tau-t} l, & r < \tau - t \end{Bmatrix}, \text{ a.e. } r \in [0, \tau]. \tag{6.48}$$

Substituting this expression in (6.42), we obtain

$$\zeta_s^* = \int_0^{\min(s,\tau-t)} \mathcal{K}_{s,r} B_r G_r^{-1} B_r^* \mathcal{K}_{\tau-t,r}^* Q_{\tau-t} l \, dr + \int_{\min(s,\tau-t)}^{s} \mathcal{K}_{s,r} B_r G_r^{-1} (B_r^* Q_r + L_r) \mathcal{R}_{r,\tau-t} l \, dr.$$

Since $\mathcal{K} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{R})$, the last equality implies (6.47). □

Theorem 6.18. *Under the conditions* $(\mathbf{E_1^w})$ *and* $(\mathbf{E_2^w})$, *there exists a unique optimal linear feedback smoother in the smoothing problem* (6.1)–(6.2) *and the best estimate* $\hat{x}_t^\tau$ *of* x_t *based on* z_s, $0 \leq s \leq \tau$, *is equal to*

$$\hat{x}_t^\tau = \hat{x}_t + P_t \int_t^\tau \mathcal{Y}_{s,t}^* C_s^* V_s^{-1} \, d\bar{z}_s, \ 0 \leq t < \tau \leq T, \tag{6.49}$$

where $\bar{z}$ *is the innovation process defined by* (6.24), $\hat{x}_t$ *is the best estimate in the filtering problem* (6.1)–(6.2), $\mathcal{Y} = \mathcal{P}_{-(PC^*+R)V^{-1}C}(\mathcal{U})$ *and* P *is a unique solution of the equation* (6.17).

Proof. Introduce the notation (6.33) and consider the linear regulator problem (6.34)–(6.35). By (6.41), (6.47) and (6.48), the optimal control in the problem (6.34)–(6.35) is unique and is equal to

$$\eta_s^* = -G_s^{-1}(B_s^* Q_s + L_s) \int_0^{\min(s,\tau-t)} \mathcal{K}_{s,r} B_r G_r^{-1} B_r^* \mathcal{K}_{\tau-t,r}^* Q_{\tau-t} l \, dr + G_s^{-1} \begin{Bmatrix} (B_s^* Q_s + L_s) \mathcal{K}_{s,\tau-t} l, & s \geq \tau - t \\ B_s^* \mathcal{K}_{\tau-t,s}^* Q_{\tau-t} l, & s < \tau - t \end{Bmatrix}, \text{ a.e. } s \in [0, \tau], \tag{6.50}$$

where $\mathcal{K} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{R})$ and Q is a solution of the equation (6.44). In proving Theorem 6.4 it was shown that $P_s = Q_{\tau-s}$, $0 \leq s \leq \tau$, is a unique solution of the equation (6.17) and $\mathcal{D}_\tau(\mathcal{K}) = \mathcal{Y}|_{\Delta_\tau}$. Therefore, applying Theorem

6.15, we conclude that there is a unique optimal linear feedback smoother in the smoothing problem (6.1)–(6.2) and it is determined by

$$K_s = K_{t,\tau,s} = -\int_{\max(s,t)}^{\tau} P_t \mathcal{Y}_{r,t}^* C_r^* V_r^{-1} C_r \mathcal{Y}_{r,s}(P_s C_s^* + R_s) V_s^{-1}\, dr + \begin{cases} \mathcal{Y}_{t,s}(P_s C_s^* + R_s)V_s^{-1}, & s \le t \\ P_t \mathcal{Y}_{s,t}^* C_s^* V_s^{-1}, & s > t \end{cases}, \text{ a.e. } s \in [0,\tau].$$

Finally by Theorem 6.4 and (6.24), we obtain

$$\begin{aligned}
\hat{x}_t^\tau &= \int_0^\tau K_s\, dz_s \\
&= \int_0^t \mathcal{Y}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\, dz_s + \int_t^\tau P_t \mathcal{Y}_{s,t}^* C_s^* V_s^{-1}\, dz_s \\
&\quad - \int_t^\tau \int_0^r P_t \mathcal{Y}_{r,t}^* C_r^* V_r^{-1} C_r \mathcal{Y}_{r,s}(P_s C_s^* + R_s)V_s^{-1} dz_s\, dr \\
&= \hat{x}_t + P_t \int_t^\tau \mathcal{Y}_{r,t}^* C_r^* V_r^{-1}(dz_r - C_r \hat{x}_r\, dr) \\
&= \hat{x}_t + P_t \int_t^\tau \mathcal{Y}_{r,t}^* C_r^* V_r^{-1}\, d\bar{z}_r.
\end{aligned}$$

Thus, the proof of the theorem is completed. □

6.5 Stochastic Regulator Problem

In this section the control problem (5.1)–(5.4) from Chapter 5 will be called a *linear stochastic regulator problem* and it will be considered under correlated white noise processes.

6.5.1 Setting of the Problem

Consider the problem (5.1)–(5.4) in which the state-observation system (5.1)–(5.2) and the functional (5.4) are defined in the form

$$x_t^u = \mathcal{U}_{t,0} x_0 + \int_0^t \mathcal{U}_{t,s} B_s u_s\, ds + \int_0^t \mathcal{U}_{t,s} \Phi_s\, dw_s, \; 0 \le t \le T, \tag{6.51}$$

$$z_t^u = \int_0^t C_s x_s^u\, ds + \int_0^t \Psi_s\, dv_s, \; 0 \le t \le T, \tag{6.52}$$

$$J(u) = \mathbf{E}\Big(\langle x_T^u, Q_T x_T^u \rangle + \int_0^T \Big\langle \begin{bmatrix} x_t^u \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix} \begin{bmatrix} x_t^u \\ u_t \end{bmatrix} \Big\rangle\, dt \Big) \tag{6.53}$$

and a control u is taken from the set of admissible controls U_{ad} as defined by (5.3) in Section 5.1.2. This problem will be called the linear stochastic regulator problem (6.51)–(6.53). The deterministic function φ is called a *stochastic regulator* if $u_t = \varphi_t(x_t^u)$, a.e. $t \in \mathbf{T}$, belongs to U_{ad}. A stochastic regulator is *optimal* if the respective control is optimal.

In this section the following conditions are supposed to hold:

$(\mathbf{R_1^w})$ $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $B \in B_\infty(\mathbf{T}, \mathcal{L}(U, X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$;

$(\mathbf{R_2^w})$ $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with $\operatorname{cov} v_T > 0$, x_0 is an X-valued Gaussian random variable with $\mathbf{E} x_0 = 0$, x_0 and (w, v) are independent;

$(\mathbf{R_3^w})$ $Q_T \in \mathcal{L}(X)$, $Q_T \geq 0$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$, $L \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \geq 0$ for a.e. $t \in \mathbf{T}$.

Note that $(\mathbf{R_1^w})$ is the same as $(\mathbf{E_1^w})$ completed with the condition about B, and $(\mathbf{R_2^w})$ and $(\mathbf{R_3^w})$ are the same as $(\mathbf{E_2^w})$ and $(\mathbf{C_3})$, respectively.

In this section we will use the operators $\bar{W}$, $\bar{V}$, $\bar{R}$ and P_0, as well as the functions W, V and R defined by (6.3). Also, for given $u \in U_{\text{ad}}$, we denote

$$\mathbf{E}_t^u = \mathbf{E}(\cdot \,|z_s^u, 0 \leq s \leq t) \text{ and } \hat{x}_t^u = \mathbf{E}_t^u x_t^u, \ 0 \leq t \leq T.$$

If $u_t = u_t^*$, then $x_t^* = x_t^{u^*}$, $z_t^* = z_t^{u^*}$, $\mathbf{E}_t^* = \mathbf{E}_t^{u^*}$ and $\hat{x}_t^* = \hat{x}_t^{u^*}$, $0 \leq t \leq T$. Similarly, we use the notation $x_t^0 = x_t^u$, $z_t^0 = z_t^u$, $\mathbf{E}_t^0 = \mathbf{E}_t^u$ and $\hat{x}_t^0 = \hat{x}_t^u$ when $u_t = 0$, $0 \leq t \leq T$.

6.5.2 Optimal Stochastic Regulator

The following results completely solve the problem (6.51)–(6.53).

Proposition 6.19. *For the system* (5.51)–(5.52), *the error process*

$$e_t = x_t^u - \hat{x}_t^u, \ 0 \leq t \leq T, \ u \in U_{\text{ad}},$$

and the innovation process

$$\bar{z}_t = z_t^u - \int_0^t C_s \hat{x}_s^u \, ds, \ 0 \leq t \leq T, \ u \in U_{\text{ad}},$$

are independent of selection of $u \in U_{\text{ad}}$. *Furthermore,*

$$\operatorname{cov} e_t = P_t \text{ and } \bar{z}_t = \int_0^t \Psi_s \, d\gamma_s, \ 0 \leq t \leq T, \tag{6.54}$$

where P *is a unique solution of the equation* (6.17), $\{\gamma_t, \mathcal{F}_t^{z^0}\}$ *is an* $\mathbb{R}^n$*-valued Wiener process on* $\mathbf{T}$ *with the covariance operator* $\bar{V}$.

Proof. For $u \in U_{\mathrm{ad}}$,

$$\hat{x}_t^u = \mathbf{E}_t^u\Big(\mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\Phi_s\, dw_s\Big) + \int_0^t \mathcal{U}_{t,s}B_s u_s\, ds.$$

Therefore, by Proposition 5.5,

$$e_t = x_t^u - \hat{x}_t^u = x_t^0 - \mathbf{E}_t^u x_t^0 = x_t^0 - \mathbf{E}_t^0 x_t^0 = x_t^0 - \hat{x}_t^0,$$

i.e., the error process e is independent of selection of $u \in U_{\mathrm{ad}}$. Thus, we have

$$\begin{aligned}\bar{z}_t &= z_t^u - \int_0^t C_s \hat{x}_s^u\, ds \\ &= \int_0^t C_s(x_s^u - \hat{x}_s^u)\, ds + \int_0^t \Psi_s\, dv_s \\ &= \int_0^t C_s\big(x_s^0 - \hat{x}_s^0\big)\, ds + \int_0^t \Psi_s\, dv_s \\ &= z_t^0 - \int_0^t C_s \hat{x}_s^0\, ds,\end{aligned}$$

i.e., the innovation process $\bar{z}$ is independent of selection of $u \in U_{\mathrm{ad}}$ too. Finally, (6.54) follows from Propositions 6.5 and 6.8. □

Theorem 6.20. *Under the conditions* $(\mathbf{R}_1^{\mathbf{w}})$–$(\mathbf{R}_3^{\mathbf{w}})$, *there exists a unique optimal stochastic regulator in the problem* (6.51)–(6.53) *and the respective optimal control has the form*

$$u_t^* = -G_t^{-1}(B_t^* Q_t + L_t)\hat{x}_t^*, \ \text{a.e. } t \in \mathbf{T}, \tag{6.55}$$

where

$$\begin{aligned}\hat{x}_t^* &= \int_0^t \mathcal{R}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\, dz_s^* \\ &= \int_0^t \mathcal{Y}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\, d\bar{z}_s, \\ &= \int_0^t \mathcal{Y}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\Psi_s\, d\gamma_s,\ 0 \le t \le T,\end{aligned} \tag{6.56}$$

$\mathcal{R} = \mathcal{P}_{-BG^{-1}(B^*Q+L)-(PC^*+R)V^{-1}C}(\mathcal{U})$, $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, *$Q$ and P are unique solutions of the Riccati equations* (3.9) *and* (6.17), *respectively,* $\{\gamma_t, \mathcal{F}_t^{z^0}\}$ *is an* $\mathbb{R}^n$*-valued Wiener process on* $\mathbf{T}$ *with the covariance operator* $\bar{V}$.

Proof. The existence and the uniqueness of an optimal stochastic regulator in the problem (6.51)–(6.53) follow from Theorems 5.28 and 6.4. Moreover, by Theorem 5.17, the representation (6.55) holds for u^*. Using this in (6.51), we obtain

$$x_t^* = \mathcal{U}_{t,0}x_0 - \int_0^t \mathcal{U}_{t,s}B_sG_s^{-1}(B_s^*Q_s + L_s)\hat{x}_s^*\,ds + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s.$$

By Proposition 5.5, $\mathbf{E}_t^* = \mathbf{E}_t^0$. Therefore, applying Theorem 6.9,

$$\begin{aligned}\hat{x}_t^* &= \mathbf{E}_t^*\Big(\mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s\Big) - \int_0^t \mathcal{U}_{t,s}B_sG_s^{-1}(B_s^*Q_s + L_s)\hat{x}_s^*\,ds \\ &= \mathbf{E}_t^0\Big(\mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s\Big) - \int_0^t \mathcal{U}_{t,s}B_sG_s^{-1}(B_s^*Q_s + L_s)\hat{x}_s^*\,ds \\ &= \int_0^t \mathcal{U}_{t,s}(P_sC_s^* + R_s)V_s^{-1}(dz_s^0 - C_s\hat{x}_s^0\,ds) - \int_0^t \mathcal{U}_{t,s}B_sG_s^{-1}(B_s^*Q_s + L_s)\hat{x}_s^*\,ds.\end{aligned}$$

Hence, by Proposition 6.19,

$$\hat{x}_t^* = \int_0^t \mathcal{U}_{t,s}(P_sC_s^* + R_s)V_s^{-1}(dz_s^* - C_s\hat{x}_s^*\,ds) - \int_0^t \mathcal{U}_{t,s}B_sG_s^{-1}(B_s^*Q_s + L_s)\hat{x}_s^*\,ds.$$

Thus, using Propositions 4.28 and 6.19, we obtain the equations in (6.56) too. □

Lemma 6.21. *For the functional*

$$\hat{J}(u) = \mathbf{E}\bigg(\langle \hat{x}_T^u, Q_T\hat{x}_T^u\rangle + \int_0^T \left\langle \begin{bmatrix}\hat{x}_t^u \\ u_t\end{bmatrix}, \begin{bmatrix}F_t & L_t^* \\ L_t & G_t\end{bmatrix}\begin{bmatrix}\hat{x}_t^u \\ u_t\end{bmatrix}\right\rangle dt\bigg) \tag{6.57}$$

subjected to the system (6.51)–(6.52),

$$\hat{J}(u^*) = \operatorname{tr}\int_0^T \bar{V}^{-1}\Psi_t^{-1}(C_tP_t + R_t^*)Q_t(P_tC_t^* + R_t)\Psi_t^{-1*}\,dt,$$

where u^* is an optimal control in the problem (6.51)–(6.53) *and Q and P are unique solutions of the Riccati equations* (3.9) *and* (6.17), *respectively.*

Proof. Using (6.55) and Proposition 4.2(f), we obtain

$$\begin{aligned}\hat{J}(u^*) &= \mathbf{E}\langle \hat{x}_T^*, Q_T\hat{x}_T^*\rangle \\ &\quad + \mathbf{E}\int_0^T \langle \hat{x}_t^*, (F_t - L_t^*G_t^{-1}L_t + Q_tB_tG_t^{-1}B_t^*Q_t)\hat{x}_t^*\rangle\,dt \\ &= \operatorname{tr}\Big(\operatorname{cov}\Big(Q_T^{1/2}\hat{x}_T^*\Big)\Big) \\ &\quad + \operatorname{tr}\int_0^T \operatorname{cov}\Big(\big(F_t - L_t^*G_t^{-1}L_t + Q_tB_tG_t^{-1}B_t^*Q_t\big)^{1/2}\hat{x}_t^*\Big)\,dt.\end{aligned}$$

We will use the representation (6.56) for $\hat{x}^*$. By Proposition 4.26(f) and Theorem 3.22,

$$\begin{aligned}
\hat{J}(u^*) &= \operatorname{tr}\int_0^T \bar{V}^{-1}\Psi_s^{-1}(C_sP_s + R_s^*)\mathcal{Y}_{T,s}^* Q_T \mathcal{Y}_{T,s}(P_sC_s^* + R_s)\Psi_s^{-1*}\,ds \\
&\quad + \operatorname{tr}\int_0^T\int_0^t \bar{V}^{-1}\Psi_s^{-1}(C_sP_s + R_s^*)\mathcal{Y}_{t,s}^* F_t \mathcal{Y}_{t,s}(P_sC_s^* + R_s)\Psi_s^{-1*}\,ds\,dt \\
&\quad - \operatorname{tr}\int_0^T\int_0^t \bar{V}^{-1}\Psi_s^{-1}(C_sP_s + R_s^*)\mathcal{Y}_{t,s}^* L_t^* G_t^{-1} L_t \mathcal{Y}_{t,s} \\
&\qquad\qquad \times (P_sC_s^* + R_s)\Psi_s^{-1*}\,ds\,dt \\
&\quad + \operatorname{tr}\int_0^T\int_0^t \bar{V}^{-1}\Psi_s^{-1}(C_sP_s + R_s^*)\mathcal{Y}_{t,s}^* Q_t B_t G_t^{-1} B_t^* Q_t \mathcal{Y}_{t,s} \\
&\qquad\qquad \times (P_sC_s^* + R_s)\Psi_s^{-1*}\,ds\,dt \\
&= \operatorname{tr}\int_0^T \bar{V}^{-1}\Psi_s^{-1}(C_sP_s + R_s^*)Q_s(P_sC_s^* + R_s)\Psi_s^{-1*}\,ds.
\end{aligned}$$

The proof is completed. □

Proposition 6.22. *The minimum of the functional J in the problem* (6.51)–(6.53) *is equal to*

$$\begin{aligned}
J(u^*) &= \operatorname{tr}(Q_TP_T) + \operatorname{tr}\int_0^T F_tP_t\,dt \\
&\quad + \operatorname{tr}\int_0^T \bar{V}^{-1}\Psi_t^{-1}(C_tP_t + R_t^*)Q_t(P_tC_t^* + R_t)\Psi_t^{-1*}\,dt,
\end{aligned}$$

where Q and P are unique solutions of the Riccati equations (3.9) *and* (6.17), *respectively.*

Proof. Let $e_t = x_t^* - \hat{x}_t^*$, $0 \le t \le T$, be the error process. We have

$$\begin{aligned}
\mathbf{E}\langle e_T, Q_Te_T\rangle &= \mathbf{E}\langle x_T^*, Q_Tx_T^*\rangle - 2\mathbf{E}\langle x_T^*, Q_T\hat{x}_T^*\rangle + \mathbf{E}\langle \hat{x}_T^*, Q_T\hat{x}_T^*\rangle \\
&= \mathbf{E}\langle x_T^*, Q_Tx_T^*\rangle - \mathbf{E}\langle \hat{x}_T^*, Q_T\hat{x}_T^*\rangle.
\end{aligned}$$

Similarly,

$$\mathbf{E}\int_0^T \langle e_t, F_te_t\rangle\,dt = \mathbf{E}\int_0^T \langle x_t^*, F_tx_t^*\rangle\,dt - \mathbf{E}\int_0^T \langle \hat{x}_t^*, F_t\hat{x}_t^*\rangle\,dt.$$

Also,

$$\mathbf{E}\int_0^T \langle u_t^*, L_tx_t^*\rangle\,dt = \mathbf{E}\int_0^T \langle u_t^*, L_t\mathbf{E}_t^*x_t^*\rangle\,dt = \mathbf{E}\int_0^T \langle u_t^*, L_t\hat{x}_t^*\rangle\,dt.$$

Therefore, by Proposition 6.5,

$$\begin{aligned}
J(u^*) &= \mathbf{E}\left(\langle x_T^*, Q_T x_T^*\rangle + \int_0^T (\langle x_t^*, F_t x_t^*\rangle + 2\langle u_t^*, L_t x_t^*\rangle + \langle u_t^*, G_T u_t^*\rangle)\, dt\right) \\
&= \mathbf{E}\left(\langle \hat{x}_T^*, Q_T \hat{x}_T^*\rangle + \int_0^T (\langle \hat{x}_t^*, F_t \hat{x}_t^*\rangle + 2\langle u_t^*, L_t \hat{x}_t^*\rangle + \langle u_t^*, G_T u_t^*\rangle)\, dt\right) \\
&\quad + \mathbf{E}\left(\langle e_T, Q_T e_T\rangle + \int_0^T \langle e_t, F_t e_t\rangle\, dt\right) \\
&= \hat{J}(u^*) + \mathrm{tr}\left(\mathrm{cov}\left(Q_T^{1/2} e_T\right)\right) + \mathrm{tr}\int_0^T \mathrm{cov}\left(F_t^{1/2} e_t\right) dt \\
&= \hat{J}(u^*) + \mathrm{tr}(Q_T P_T) + \mathrm{tr}\int_0^T F_t P_t\, dt.
\end{aligned}$$

Thus, by Lemma 6.21, the statement is obtained. □

Example 6.23. In order to present the optimal stochastic regulator in differential form, assume that the conditions $(\mathbf{R_1^w})$–$(\mathbf{R_3^w})$ hold so that $\mathcal{U} \in \mathcal{S}(X)$ and A is the infinitesimal generator of $\mathcal{U}$. Then the state-observation system (6.51)–(6.52) under $u = u^*$ can be written in the differential form

$$\begin{cases} dx_t^* = (Ax_t^* + B_t u_t^*)dt + \Phi_t dw_t, \ 0 < t \le T, \ x_0^* = x_0, \\ dz_t^* = C_t x_t^* dt + \Psi_t dv_t, \ 0 < t \le T, \ z_t^* = 0. \end{cases} \tag{6.58}$$

By (6.55), the optimal control u^* in the problem (6.51)–(6.53) has the form

$$u_t^* = -G_t^{-1}(B_t^* Q_t + L_t)\hat{x}_t^*, \ \text{a.e. } t \in \mathbf{T}, \tag{6.59}$$

where $\hat{x}^*$ is a mild solution of the linear stochastic differential equation

$$\begin{aligned} d\hat{x}_t^* &= (A\hat{x}_t^* + B_t u_t^*)dt + (P_t C_t^* + \Phi_t \bar{R}\Psi_t^*) \\ &\quad \times (\Psi_t \bar{V}\Psi_t^*)^{-1}(dz_t^* - C_t \hat{x}_t^* dt), \ 0 < t \le T, \ \hat{x}_0^* = 0, \end{aligned} \tag{6.60}$$

with P and Q being, respectively, unique scalar product solutions of the differential Riccati equations

$$\begin{aligned} &\frac{d}{dt}P_t - P_t A^* - AP_t - \Phi_t \bar{W}\Phi_t^* + \left(P_t C_t^* + \Phi_t \bar{R}\Psi_t^*\right) \\ &\quad \times \left(\Psi_t \bar{V}\Psi_t^*\right)^{-1}\left(C_t P_t + \Psi_t \bar{R}^* \Phi_t^*\right) = 0, \ 0 < t \le T, \ P_0 = \mathrm{cov} x_0, \end{aligned} \tag{6.61}$$

$$\begin{aligned} &\frac{d}{dt}Q_t + Q_t A + A^* Q_t + F_t \\ &\quad - (Q_t B_t + L_t^*)G_t^{-1}(B_t^* Q_t + L_t) = 0, \ 0 \le t < T, \ Q_T \text{ is given.} \end{aligned} \tag{6.62}$$

Chapter 7

Control and Estimation under Colored Noises

In this chapter the control and estimation results of Chapter 6 are modified to the colored noise processes.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $X, U, H \in \mathcal{H}$, $T > 0$, $\mathbf{T} = [0, T]$ is a finite time interval and $\Delta_t = \{(s, r) : 0 \le r \le s \le t\}$ for $t > 0$.

7.1 Estimation

7.1.1 Setting of Estimation Problems

Let (x, z) be a partially observable system so that

$$x_t = \mathcal{U}_{t,0} x_0 + \int_0^t \mathcal{U}_{t,s} \varphi_s^1 \, ds + \int_0^t \mathcal{U}_{t,s} \Phi_s \, dw_s, \ 0 \le t \le T, \tag{7.1}$$

$$z_t = \int_0^t (C_s x_s + \varphi_s^2) \, ds + \int_0^t \Psi_s \, dv_s, \ 0 \le t \le T, \tag{7.2}$$

where

$$\varphi_t^1 = \int_0^t \mathcal{U}_{t,s}^1 \Phi_s^1 \, dw_s, \ \varphi_t^2 = \int_0^t \mathcal{U}_{t,s}^2 \Phi_s^2 \, dw_s, \ 0 \le t \le T, \tag{7.3}$$

Suppose that the following conditions hold:

$(\mathbf{E_1^c})$ $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$;

$(\mathbf{E_2^c})$ $\mathcal{U}^1 \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $\mathcal{U}^2 \in \mathcal{E}(\Delta_T, \mathcal{L}(\mathbb{R}^n))$, $\Phi, \Phi^1 \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Phi^2 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \mathbb{R}^n))$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with $\mathrm{cov} v_T > 0$, x_0 is a an X-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0 and (w, v) are independent.

Also, we will use the notation

$$P_0 = \mathrm{cov} x_0, \begin{bmatrix} \bar{W} & \bar{R} \\ \bar{R}^* & \bar{V} \end{bmatrix} = T^{-1}\mathrm{cov}\begin{bmatrix} w_T \\ v_T \end{bmatrix}, \; V_t = \Psi_t \bar{V} \Psi_t^*, \; 0 \le t \le T, \tag{7.4}$$

and the decompositions

$$\tilde{W}_t = \begin{bmatrix} W_t & W_t^{01} & W_t^{02} \\ W_t^{01*} & W_t^{11} & W_t^{12} \\ W_t^{02*} & W_t^{12*} & W_t^{22} \end{bmatrix} = \begin{bmatrix} \Phi_t \bar{W} \Phi_t^* & \Phi_t \bar{W} \Phi_t^{1*} & \Phi_t \bar{W} \Phi_t^{2*} \\ \Phi_t^1 \bar{W} \Phi_t^* & \Phi_t^1 \bar{W} \Phi_t^{1*} & \Phi_t^1 \bar{W} \Phi_t^{2*} \\ \Phi_t^2 \bar{W} \Phi_t^* & \Phi_t^2 \bar{W} \Phi_t^{1*} & \Phi_t^2 \bar{W} \Phi_t^{2*} \end{bmatrix}, \tag{7.5}$$

$$\tilde{R}_t = \begin{bmatrix} R_t \\ R_t^1 \\ R_t^2 \end{bmatrix} = \begin{bmatrix} \Phi_t \bar{R} \Psi_t^* \\ \Phi_t^1 \bar{R} \Psi_t^* \\ \Phi_t^2 \bar{R} \Psi_t^* \end{bmatrix}, \; 0 \le t \le T. \tag{7.6}$$

One can observe that the difference between the signal-observation systems (6.1)–(6.2) and (7.1)–(7.2) is the presence of the colored noise processes φ^1 and φ^2 in (7.1)–(7.2). We will study the estimation problem of estimating x_t based on z_s, $0 \le s \le \tau$, where $t, \tau \in \mathbf{T}$ and (x, z) is defined by (7.1)–(7.3). This problem will be called the estimation (filtering, prediction, smoothing) problem (7.1)–(7.3).

7.1.2 Reduction

Let

$$\tilde{\mathcal{U}} = \mathcal{P}_D(\mathcal{U} \odot \mathcal{U}^1 \odot \mathcal{U}^2), \; D = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times X \times \mathbb{R}^n).$$

Using Theorem 3.13, one can obtain that

$$\tilde{\mathcal{U}}_{t,s} = \begin{bmatrix} \mathcal{U}_{t,s} & \mathcal{E}_{t,s} & 0 \\ 0 & \mathcal{U}_{t,s}^1 & 0 \\ 0 & 0 & \mathcal{U}_{t,s}^2 \end{bmatrix}, \; 0 \le s \le t \le T, \tag{7.7}$$

where

$$\mathcal{E}_{t,s} = \int_s^t \mathcal{U}_{t,r} \mathcal{U}_{r,s}^1 \, dr, \; 0 \le s \le t \le T. \tag{7.8}$$

Also, let

$$\tilde{\Phi} = \begin{bmatrix} \Phi \\ \Phi^1 \\ \Phi^2 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(H, X \times X \times \mathbb{R}^n)). \tag{7.9}$$

With this notation, the evolution of the process

$$\tilde{x}_t = \begin{bmatrix} x_t \\ \varphi_t^1 \\ \varphi_t^2 \end{bmatrix}, \; 0 \le t \le T, \tag{7.10}$$

can be expressed as

$$\tilde{x}_t = (\mathcal{U} \odot \mathcal{U}^1 \odot \mathcal{U}^2)_{t,0}\tilde{x}_0 + \int_0^t (\mathcal{U} \odot \mathcal{U}^1 \odot \mathcal{U}^2)_{t,s} D\tilde{x}_s \, ds$$
$$+ \int_0^t (\mathcal{U} \odot \mathcal{U}^1 \odot \mathcal{U}^2)_{t,s}\tilde{\Phi}_s \, dw_s, \ 0 \le t \le T,$$

or as (see Proposition 4.28)

$$\tilde{x}_t = \tilde{\mathcal{U}}_{t,0}\tilde{x}_0 + \int_0^t \tilde{\mathcal{U}}_{t,s}\tilde{\Phi}_s \, dw_s, \ 0 \le t \le T. \tag{7.11}$$

Similarly, for the observation system (7.2), we have

$$z_t = \int_0^t \tilde{C}_s \tilde{x}_s \, ds + \int_0^t \Psi_s \, dv_s, \ 0 \le t \le T, \tag{7.12}$$

where

$$\tilde{C} = [C \ \ 0 \ \ I] \in B_\infty(\mathbf{T}, \mathcal{L}(X \times X \times \mathbb{R}^n, \mathbb{R}^n)). \tag{7.13}$$

Thus, we can state the following result.

Lemma 7.1. *The best estimates in the estimation problems* (7.1)–(7.3) *and* (7.11)–(7.12) *are related as in*

$$\mathbf{E}(x_t|z_s;\ 0 \le s \le \tau) = \tilde{I}\,\mathbf{E}(\tilde{x}_t|z_s;\ 0 \le s \le \tau), \ t, \tau \in \mathbf{T},$$

where

$$\tilde{I} = [I \ \ 0 \ \ 0] \in \mathcal{L}(X \times X \times \mathbb{R}^n, X). \tag{7.14}$$

Proof. This follows from the equality $x_t = \tilde{I}\tilde{x}_t$, $0 \le t \le T$. □

Thus, the estimation problem (7.1)–(7.3) is reduced to the estimation problem (7.11)–(7.12).

7.1.3 Optimal Linear Feedback Estimators

Theorem 7.2. *Suppose the conditions* ($\mathbf{E_1^c}$) *and* ($\mathbf{E_2^c}$) *hold, let*

$$\hat{\tilde{x}}_t = \begin{bmatrix} \hat{x}_t \\ \psi_t^1 \\ \psi_t^2 \end{bmatrix} = \int_0^t \tilde{\mathcal{Y}}_{t,s}(\tilde{P}_s\tilde{C}_s^* + \tilde{R}_s)V_s^{-1} \, dz_s, \ 0 \le t \le T,$$

and let

$$\bar{z}_t = z_t - \int_0^t \tilde{C}_s \hat{\tilde{x}}_s \, ds = z_t - \int_0^t \left(C_s\hat{x}_s + \psi_s^2\right) ds, \ 0 \le t \le T,$$

where $\tilde{\mathcal{U}}$, $\tilde{C}$, $\tilde{\Phi}$, $\tilde{W}$ *and* $\tilde{R}$ *are defined by* (7.7)–(7.8), (7.13), (7.9), (7.5) *and* (7.6), $\tilde{\mathcal{Y}} = \mathcal{P}_{-(\tilde{P}\tilde{C}^*+\tilde{R})V^{-1}\tilde{C}}(\tilde{\mathcal{U}})$, *and* $\tilde{P}$ *is a unique solution of the Riccati equation*

$$\begin{aligned} \tilde{P}_t = \tilde{\mathcal{U}}_{t,0}\tilde{P}_0\tilde{\mathcal{U}}_{t,0}^* + \int_0^t \tilde{\mathcal{U}}_{t,s}(\tilde{W}_s \\ - (\tilde{P}_s\tilde{C}_s^* + \tilde{R}_s)V_s^{-1}(\tilde{C}_s\tilde{P}_s + \tilde{R}_s^*))\tilde{\mathcal{U}}_{t,s}^*\,ds,\ 0 \le t \le T, \end{aligned} \tag{7.15}$$

with $\tilde{P}_0 = \operatorname{cov}\tilde{x}_0$. *Then there exists a unique optimal linear feedback filter (predictor and smoother) in the estimation problem* (7.1)–(7.3) *and depending on* t *and* τ, *the best estimate* $\hat{x}_t^\tau$ *of* x_t *based on* z_s, $0 \le s \le \tau$, *in the estimation problem* (7.1)–(7.3) *is equal to*

$$\begin{aligned} \hat{x}_t &= \hat{x}_t^t = \int_0^t \tilde{I}\tilde{\mathcal{Y}}_{t,s}(\tilde{P}_s\tilde{C}_s^* + \tilde{R}_s)V_s^{-1}\,dz_s,\ 0 \le t \le T, \\ \hat{x}_t^\tau &= \tilde{I}\tilde{\mathcal{U}}_{t,\tau}\hat{\tilde{x}}_\tau = \mathcal{U}_{t,\tau}\hat{x}_\tau + \mathcal{E}_{t,\tau}\psi_\tau^1,\ 0 \le \tau < t \le T, \\ \hat{x}_t^\tau &= \hat{x}_t + \tilde{I}\tilde{P}_t\int_t^\tau \tilde{\mathcal{Y}}_{s,t}^*\tilde{C}_s^*V_s^{-1}\,d\tilde{z}_s,\ 0 \le t < \tau \le T, \end{aligned}$$

where $\tilde{I}$ *and* $\mathcal{E}$ *are as defined by* (7.14) *and* (7.8), *respectively.*

Proof. This is a consequence of Theorems 6.4, 6.13, 6.18 and Lemma 7.1. □

7.1.4 About the Riccati Equation (7.15)

The solution $\tilde{P}$ of the Riccati equation (7.15) can be decomposed as

$$\tilde{P}_t = \begin{bmatrix} P_t^{00} & P_t^{01} & P_t^{02} \\ P_t^{01*} & P_t^{11} & P_t^{12} \\ P_t^{02*} & P_t^{12*} & P_t^{22} \end{bmatrix} \in \mathcal{L}(X \times X \times \mathbb{R}^n),\ 0 \le t \le T. \tag{7.16}$$

Below we derive an equation for each component of $\tilde{P}$ in this decomposition.

Proposition 7.3. *Suppose the conditions* $(\mathbf{E}_1^c)$ *and* $(\mathbf{E}_2^c)$ *hold, let the solution* $\tilde{P}$ *of the equation* (7.15) *be decomposed as* (7.16) *and let*

$$\begin{cases} M_t = P_t^{00}C_t^* + P_t^{02} + R_t,\ 0 \le t \le T, \\ M_t^1 = P_t^{01*}C_t^* + P_t^{12} + R_t^1,\ 0 \le t \le T, \\ M_t^2 = P_t^{02*}C_t^* + P_t^{22} + R_t^2,\ 0 \le t \le T. \end{cases} \tag{7.17}$$

Then $(P^{00}, P^{01}, P^{02}, P^{11}, P^{12}, P^{22})$ *is a unique solution of the following system of* *equations:*

$$P_t^{00} = \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s}(W_s + P_s^{01} + P_s^{01*} - M_s V_s^{-1} M_s^*)\mathcal{U}_{t,s}^* \, ds, \ 0 \le t \le T, \tag{7.18}$$

$$P_t^{01} = \int_0^t \mathcal{U}_{t,s}(W_s^{01} + P_s^{11} - M_s V_s^{-1} M_s^{1*})\mathcal{U}_{t,s}^{1*} \, ds, \ 0 \le t \le T, \tag{7.19}$$

$$P_t^{02} = \int_0^t \mathcal{U}_{t,s}(W_s^{02} + P_s^{12} - M_s V_s^{-1} M_s^{2*})\mathcal{U}_{t,s}^{2*} \, ds, \ 0 \le t \le T, \tag{7.20}$$

$$P_t^{11} = \int_0^t \mathcal{U}_{t,s}^1(W_s^{11} - M_s^1 V_s^{-1} M_s^{1*})\mathcal{U}_{t,s}^{1*} \, ds, \ 0 \le t \le T, \tag{7.21}$$

$$P_t^{12} = \int_0^t \mathcal{U}_{t,s}^1(W_s^{12} - M_s^1 V_s^{-1} M_s^{2*})\mathcal{U}_{t,s}^{2*} \, ds, \ 0 \le t \le T, \tag{7.22}$$

$$P_t^{22} = \int_0^t \mathcal{U}_{t,s}^2(W_s^{22} - M_s^2 V_s^{-1} M_s^{2*})\mathcal{U}_{t,s}^{2*} \, ds, \ 0 \le t \le T. \tag{7.23}$$

Proof. Using (7.5)–(7.7), (7.13) and

$$\tilde{P}_0 = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times X \times \mathbb{R}^n) \tag{7.24}$$

in the equation (7.15), one can easily verify that the components P^{11}, P^{12} and P^{22} of $\tilde{P}$ satisfy the equations (7.21)–(7.23). For the component P^{01},

$$P_t^{01} = \int_0^t \mathcal{U}_{t,s}(W_s^{01} - M_s V_s^{-1} M_s^{1*})\mathcal{U}_{t,s}^{1*} \, ds + \int_0^t \mathcal{E}_{t,s}(W_s^{11} - M_s^1 V_s^{-1} M_s^{1*})\mathcal{U}_{t,s}^{1*} \, ds. \tag{7.25}$$

By (7.8) and (7.21),

$$\begin{aligned} P_t^{01} &= \int_0^t \mathcal{U}_{t,s}(W_s^{01} - M_s V_s^{-1} M_s^{1*})\mathcal{U}_{t,s}^{1*} \, ds \\ &\quad + \int_0^t \int_s^t \mathcal{U}_{t,r}\mathcal{U}_{r,s}^1(W_s^{11} - M_s^1 V_s^{-1} M_s^{1*})\mathcal{U}_{r,s}^{1*}\mathcal{U}_{t,r}^{1*} \, dr \, ds \\ &= \int_0^t \mathcal{U}_{t,s}(W_s^{01} - M_s V_s^{-1} M_s^{1*})\mathcal{U}_{t,s}^{1*} \, ds \\ &\quad + \int_0^t \mathcal{U}_{t,r}\left(\int_0^r \mathcal{U}_{r,s}^1(W_s^{11} - M_s^1 V_s^{-1} M_s^{1*})\mathcal{U}_{r,s}^{1*} \, ds\right)\mathcal{U}_{t,r}^{1*} \, dr \\ &= \int_0^t \mathcal{U}_{t,s}(W_s^{01} + P_s^{11} - M_s V_s^{-1} M_s^{1*})\mathcal{U}_{t,s}^{1*} \, ds. \end{aligned}$$

This proves (7.19). For the component P^{00} of $\tilde{P}$,

$$
\begin{aligned}
P_t^{00} = &\, \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s}(W_s - M_s V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds \\
&+ \int_0^t \mathcal{U}_{t,s}(W_s^{01} - M_s V_s^{-1} M_s^{1*}) \mathcal{E}_{t,s}^* \, ds \\
&+ \int_0^t \mathcal{E}_{t,s}(W_s^{01*} - M_s^1 V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds \\
&+ \int_0^t \mathcal{E}_{t,s}(W_s^{11} - M_s^1 V_s^{-1} M_s^{1*}) \mathcal{E}_{t,s}^* \, ds.
\end{aligned}
$$

Note that, if $0 \le s \le r \le t \le T$, then

$$
\begin{aligned}
\mathcal{E}_{t,s} &= \int_s^t \mathcal{U}_{t,\sigma} \mathcal{U}_{\sigma,s}^1 \, d\sigma = \int_s^r \mathcal{U}_{t,\sigma} \mathcal{U}_{\sigma,s}^1 \, d\sigma + \int_r^t \mathcal{U}_{t,\sigma} \mathcal{U}_{\sigma,s}^1 \, d\sigma \\
&= \mathcal{U}_{t,r} \mathcal{E}_{r,s} + \mathcal{E}_{t,r} \mathcal{U}_{r,s}^1.
\end{aligned}
\tag{7.26}
$$

Hence, by (7.26), (7.25), (7.21), (7.19) and (7.8),

$$
\begin{aligned}
P_t^{00} = &\, \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s}(W_s - M_s V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds \\
&+ \int_0^t \int_s^t \mathcal{U}_{t,r} \mathcal{U}_{r,s}(W_s^{01} - M_s V_s^{-1} M_s^{1*}) \mathcal{U}_{r,s}^{1*} \mathcal{U}_{t,r}^* \, dr \, ds \\
&+ \int_0^t \mathcal{E}_{t,s}(W_s^{01*} - M_s^1 V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds \\
&+ \int_0^t \int_s^t \mathcal{U}_{t,r} \mathcal{E}_{r,s}(W_s^{11} - M_s^1 V_s^{-1} M_s^{1*}) \mathcal{U}_{r,s}^{1*} \mathcal{U}_{t,r}^* \, dr \, ds \\
&+ \int_0^t \int_s^t \mathcal{E}_{t,r} \mathcal{U}_{r,s}^1 (W_s^{11} - M_s^1 V_s^{-1} M_s^{1*}) \mathcal{U}_{r,s}^{1*} \mathcal{U}_{t,r}^* \, dr \, ds \\
= &\, \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s}(W_s - M_s V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds \\
&+ \int_0^t \mathcal{U}_{t,r} P_r^{01} \mathcal{U}_{t,r}^* \, dr + \int_0^t \mathcal{E}_{t,r} P_r^{11} \mathcal{U}_{t,r}^* \, dr \\
&+ \int_0^t \mathcal{E}_{t,s}(W_s^{01*} - M_s^1 V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds \\
= &\, \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s}(W_s + P_s^{01} - M_s V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds \\
&+ \int_0^t \int_s^t \mathcal{U}_{t,r} \mathcal{U}_{r,s}^1 (W_s^{01*} + P_s^{11} - M_s^1 V_s^{-1} M_s^*) \mathcal{U}_{r,s}^* \mathcal{U}_{t,r}^* \, dr \, ds \\
= &\, \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s}(W_s + P_s^{01} + P_s^{01*} - M_s V_s^{-1} M_s^*) \mathcal{U}_{t,s}^* \, ds.
\end{aligned}
$$

This proves (7.18). Finally, for the component P^{02},

$$P_t^{02} = \int_0^t \mathcal{U}_{t,s}(W_s^{02} - M_s V_s^{-1} M_s^{2*})\mathcal{U}_{t,s}^{2*}\, ds + \int_0^t \mathcal{E}_{t,s}(W_s^{12} - M_s^1 V_s^{-1} M_s^{2*})\mathcal{U}_{t,s}^{2*}\, ds.$$

By (7.8) and (7.22),

$$\begin{aligned} P_t^{02} &= \int_0^t \mathcal{U}_{t,s}(W_s^{02} - M_s V_s^{-1} M_s^{2*})\mathcal{U}_{t,s}^{2*}\, ds \\ &\quad + \int_0^t \int_s^t \mathcal{U}_{t,r}\mathcal{U}_{r,s}^1 (W_s^{12} - M_s^1 V_s^{-1} M_s^{2*})\mathcal{U}_{r,s}^{2*}\mathcal{U}_{t,r}^{2*}\, dr\, ds \\ &= \int_0^t \mathcal{U}_{t,s}(W_s^{02} + P_s^{12} - M_s V_s^{-1} M_s^{2*})\mathcal{U}_{t,s}^{2*}\, ds. \end{aligned}$$

This proves (7.20). The uniqueness of solution of the system (7.18)–(7.23) follows from the uniqueness of solution of the equation (7.15). □

Proposition 7.4. *Suppose the conditions* ($\mathbf{E_1^c}$) *and* ($\mathbf{E_2^c}$) *hold. Then for the error process* $e_t = x_t - \hat{x}_t$, $0 \le t \le T$, *in the filtering problem* (7.1)–(7.3), *the equality* $\operatorname{cov} e_t = P_t^{00}$, $0 \le t \le T$, *holds where* P^{00} *is defined by* (7.18)–(7.23).

Proof. This follows from Proposition 6.5. □

7.1.5 Example: Optimal Filter in Differential Form

Example 7.5. In order to obtain the equations of the best estimate in the filtering problem (7.1)–(7.3) in differential form, assume that the conditions ($\mathbf{E_1^c}$)–($\mathbf{E_2^c}$) hold so that $\mathcal{U}, \mathcal{U}^1 \in \mathcal{S}(X)$, $\mathcal{U}^2 \in \mathcal{S}(\mathbb{R}^n)$ and A, A_1 and A_2 are the infinitesimal generators of $\mathcal{U}$, $\mathcal{U}^1$ and $\mathcal{U}^2$, respectively. Then the best estimate $\hat{x}$ in the filtering problem (7.1)–(7.3) together with ψ^1 and ψ^2 is a mild solution of the simultaneous linear stochastic differential equations

$$\begin{cases} d\hat{x}_t = A\hat{x}_t dt + \psi_t^1 dt + \left(P_t^{00} C_t^* + P_t^{02} + \Phi_t \bar{R} \Psi_t^*\right)\left(\Psi_t \bar{V} \Psi_t^*\right)^{-1} \\ \qquad \times \left(dz_t - C_t \hat{x}_t dt - \psi_t^2 dt\right), \ 0 < t \le T, \ \hat{x}_0 = 0, \\ d\psi_t^1 = A_1 \psi_t^1 dt + \left(P_t^{01*} C_t^* + P_t^{12} + \Phi_t^1 \bar{R} \Psi_t^*\right)\left(\Psi_t \bar{V} \Psi_t^*\right)^{-1} \\ \qquad \times \left(dz_t - C_t \hat{x}_t dt - \psi_t^2 dt\right), \ 0 < t \le T, \ \psi_0^1 = 0, \\ d\psi_t^2 = A_2 \psi_t^2 dt + \left(P_t^{02*} C_t^* + P_t^{22} + \Phi_t^2 \bar{R} \Psi_t^*\right)\left(\Psi_t \bar{V} \Psi_t^*\right)^{-1} \\ \qquad \times \left(dz_t - C_t \hat{x}_t dt - \psi_t^2 dt\right), \ 0 < t \le T, \ \psi_0^2 = 0, \end{cases} \tag{7.27}$$

where $(P^{00}, P^{01}, P^{02}, P^{11}, P^{12}, P^{22})$ is a unique scalar product solution of the system of differential equations

$$\frac{d}{dt}P_t^{00} - P_t^{00}A^* - AP_t^{00} - P_t^{01} - P_t^{01*} - \Phi_t\bar{W}\Phi_t^* \\ + M_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^* = 0,\ 0 < t \le T,\ P_0^{00} = P_0, \tag{7.28}$$

$$\frac{d}{dt}P_t^{01} - P_t^{01}A_1^* - AP_t^{01} - P_t^{11} - \Phi_t\bar{W}\Phi_t^{1*} \\ + M_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^{1*} = 0,\ 0 < t \le T,\ P_0^{01} = 0, \tag{7.29}$$

$$\frac{d}{dt}P_t^{02} - P_t^{02}A_2^* - AP_t^{02} - P_t^{12} - \Phi_t\bar{W}\Phi_t^{2*} \\ + M_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^{2*} = 0,\ 0 < t \le T,\ P_0^{02} = 0, \tag{7.30}$$

$$\frac{d}{dt}P_t^{11} - P_t^{11}A_1^* - A_1P_t^{11} - \Phi_t^1\bar{W}\Phi_t^{1*} \\ + M_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^{1*} = 0,\ 0 < t \le T,\ P_0^{11} = 0, \tag{7.31}$$

$$\frac{d}{dt}P_t^{12} - P_t^{12}A_2^* - A_1P_t^{12} - \Phi_t^1\bar{W}\Phi_t^{2*} \\ + M_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^{2*} = 0,\ 0 < t \le T,\ P_0^{12} = 0, \tag{7.32}$$

$$\frac{d}{dt}P_t^{22} - P_t^{22}A_2^* - A_2P_t^{22} - \Phi_t^2\bar{W}\Phi_t^{2*} \\ + M_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^{2*} = 0,\ 0 < t \le T,\ P_0^{22} = 0, \tag{7.33}$$

with the functions M, M^1 and M^2 as defined by (7.17). Indeed, under the above conditions, we have

$$\mathcal{U} \odot \mathcal{U}^1 \odot \mathcal{U}^2 \in \mathcal{S}(X \times X \times \mathbb{R}^n).$$

Since $\tilde{\mathcal{U}}$ is the perturbation of the semigroup $\mathcal{U} \odot \mathcal{U}^1 \odot \mathcal{U}^2$ by the bounded operator

$$D = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

by Theorem 3.15, we conclude that $\tilde{\mathcal{U}} \in \mathcal{S}(X \times X \times \mathbb{R}^n)$ and its infinitesimal generator is

$$\tilde{A} = \begin{bmatrix} A & I & 0 \\ 0 & A_1 & 0 \\ 0 & 0 & A_2 \end{bmatrix}. \tag{7.34}$$

Thus, applying the results of Example 6.10 to the filtering problem (7.11)–(7.12), we obtain the system (7.27) for the optimal filter. The equations (7.28)–(7.33) are exactly the equations (7.18)–(7.23) written in differential form which follow from Theorem 3.28 and (7.15).

7.2 Stochastic Regulator Problem

7.2.1 Setting of the Problem

Consider the problem (5.1)–(5.4) in which the state-observation system (5.1)–(5.2) and the functional (5.4) are defined in the form

$$x_t^u = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}B_s u_s\,ds + \int_0^t \mathcal{U}_{t,s}\varphi_s^1\,ds + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s,\ 0 \le t \le T, \tag{7.35}$$

$$z_t^u = \int_0^t (C_s x_s^u + \varphi_s^2)ds + \int_0^t \Psi_s\,dv_s,\ 0 \le t \le T, \tag{7.36}$$

$$J(u) = \mathbf{E}\left(\langle x_T^u, Q_T x_T^u\rangle + \int_0^T \left\langle \begin{bmatrix} x_t^u \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix} \begin{bmatrix} x_t^u \\ u_t \end{bmatrix} \right\rangle dt\right), \tag{7.37}$$

where φ^1 and φ^2 are defined by (7.3) and a control u is taken from the set of admissible controls U_{ad} as defined by (5.3) in Section 5.1.2. This problem will be called the linear stochastic regulator problem (7.35)–(7.37).

In this section the following conditions are supposed to hold:

($\mathbf{R_1^c}$) $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $B \in B_\infty(\mathbf{T}, \mathcal{L}(U, X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$;

($\mathbf{R_2^c}$) $\mathcal{U}^1 \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $\mathcal{U}^2 \in \mathcal{E}(\Delta_T, \mathcal{L}(\mathbb{R}^n))$, $\Phi, \Phi^1 \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Phi^2 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \mathbb{R}^n))$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with $\operatorname{cov} v_T > 0$, x_0 is a an X-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0 and (w, v) are independent;

($\mathbf{R_3^c}$) $Q_T \in \mathcal{L}(X)$, $Q_T \ge 0$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$, $L \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \ge 0$ for a.e. $t \in \mathbf{T}$.

Note that ($\mathbf{R_1^c}$) is the same as ($\mathbf{E_1^c}$) completed with the condition about B, and ($\mathbf{R_2^c}$) and ($\mathbf{R_3^c}$) are the same as ($\mathbf{E_2^c}$) and ($\mathbf{C_3}$), respectively.

7.2.2 Reduction

Let

$$\tilde{x}_t^u = \begin{bmatrix} x_t^u \\ \varphi_t^1 \\ \varphi_t^2 \end{bmatrix},\ 0 \le t \le T. \tag{7.38}$$

Similar to Section 7.1.2, for $\tilde{\mathcal{U}}$, $\tilde{C}$ and $\tilde{\Phi}$, defined by (7.7)–(7.8), (7.13) and (7.9), and for

$$\tilde{B} = \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(U, X \times X \times \mathbb{R}^n)), \tag{7.39}$$

the processes $\tilde{x}^u$ and z^u can be expressed by

$$\tilde{x}_t^u = \tilde{\mathcal{U}}_{t,0}\tilde{x}_0 + \int_0^T \tilde{\mathcal{U}}_{t,s}\tilde{B}_s u_s\, ds + \int_0^t \tilde{\mathcal{U}}_{t,s}\tilde{\Phi}_s\, dw_s,\ 0 \le t \le T, \tag{7.40}$$

$$z_t^u = \int_0^t \tilde{C}_s \tilde{x}_s^u\, ds + \int_0^t \Psi_s\, dv_s,\ 0 \le t \le T. \tag{7.41}$$

Also, the functional (7.37) can be written as

$$J(u) = \mathbf{E}\Bigg(\langle \tilde{x}_T^u, \tilde{Q}_T \tilde{x}_T^u\rangle + \int_0^T \left\langle \begin{bmatrix} \tilde{x}_t^u \\ u_t \end{bmatrix}, \begin{bmatrix} \tilde{F}_t & \tilde{L}_t^* \\ \tilde{L}_t & G_t \end{bmatrix} \begin{bmatrix} \tilde{x}_t^u \\ u_t \end{bmatrix} \right\rangle dt\Bigg), \tag{7.42}$$

where

$$\tilde{Q}_T = \begin{bmatrix} Q_T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times X \times \mathbb{R}^n), \tag{7.43}$$

$$\tilde{F} = \begin{bmatrix} F & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(X \times X \times \mathbb{R}^n)), \tag{7.44}$$

$$\tilde{L} = \begin{bmatrix} L & 0 & 0 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(X \times X \times \mathbb{R}^n, U)). \tag{7.45}$$

Lemma 7.6. *The functional* (7.37), *subject to* (7.35)–(7.36), *and the functional* (7.42), *subject to* (7.40)–(7.41), *are the same on* U_{ad} *as defined by* (5.3).

Proof. This follows from (7.38). □

Thus, the problem (7.35)–(7.37) is reduced to the linear stochastic regulator problem (7.40)–(7.42).

7.2.3 Optimal Stochastic Regulator

Theorem 7.7. *Under the conditions* $(\mathbf{R_1^c})$–$(\mathbf{R_3^c})$, *there exists a unique optimal stochastic regulator in the problem* (7.35)–(7.37) *and the respective optimal control has the form*

$$u_t^* = -G_t^{-1}(\tilde{B}_t^* \tilde{Q}_t + \tilde{L}_t)\hat{\tilde{x}}_t^*,\ a.e.\ t \in \mathbf{T}, \tag{7.46}$$

where

$$\hat{\tilde{x}}_t^* = \int_0^t \tilde{\mathcal{R}}_{t,s}(\tilde{P}_s \tilde{C}_s^* + \tilde{R}_s)V_s^{-1}\, dz_s^*,\ 0 \le t \le T, \tag{7.47}$$

$\tilde{P}$ *is a unique solution of the Riccati equation* (7.15), $\tilde{Q}$ *is a unique solution of the Riccati equation*

$$\begin{aligned} \tilde{Q}_t &= \tilde{\mathcal{U}}_{T,t}^* \tilde{Q}_T \tilde{\mathcal{U}}_{T,t} + \int_t^T \tilde{\mathcal{U}}_{s,t}^*(\tilde{F}_s \\ &\quad -(\tilde{Q}_s\tilde{B}_s + \tilde{L}_s^*)G_s^{-1}(\tilde{B}_s^*\tilde{Q}_s + \tilde{L}_s))\tilde{\mathcal{U}}_{s,t}\, ds,\ 0 \le t \le T, \end{aligned} \tag{7.48}$$

$\tilde{\mathcal{U}}$, $\tilde{B}$, $\tilde{C}$, $\tilde{\Phi}$, $\tilde{Q}_T$, $\tilde{F}$ and $\tilde{L}$ are defined by (7.7)–(7.8), (7.39), (7.13), (7.9) and (7.43)–(7.45) and $\tilde{\mathcal{R}} = \mathcal{P}_{-\tilde{B}G^{-1}(\tilde{B}^*\tilde{Q}+\tilde{L})-(\tilde{P}\tilde{C}^*+\tilde{R})V^{-1}\tilde{C}}(\tilde{\mathcal{U}})$.

Proof. This follows from Theorem 6.20 and Lemma 7.6. □

7.2.4 About the Riccati Equation (7.48)

Proposition 7.8. *Suppose the conditions ($\mathbf{R_1^c}$) and ($\mathbf{R_3^c}$) hold. Then the solution $\tilde{Q}$ of the equation (7.48) can be decomposed as*

$$\tilde{Q}_t = \begin{bmatrix} Q_t^{00} & Q_t^{01} & 0 \\ Q_t^{01*} & Q_t^{11} & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times X \times \mathbb{R}^n), \ 0 \le t \le T, \tag{7.49}$$

where (Q^{00}, Q^{01}, Q^{11}) is a unique solution of the system of equations

$$Q_t^{00} = \mathcal{U}_{T,t}^* Q_T \mathcal{U}_{T,t} + \int_t^T \mathcal{U}_{s,t}^*(F_s - N_s G_s^{-1} N_s^*)\mathcal{U}_{s,t}\, ds, \ 0 \le t \le T, \tag{7.50}$$

$$Q_t^{01} = \int_t^T \mathcal{U}_{s,t}^*(Q_s^{00} - N_s G_s^{-1} B_s^* Q_s^{01})\mathcal{U}_{s,t}^1\, ds, \ 0 \le t \le T, \tag{7.51}$$

$$Q_t^{11} = \int_t^T \mathcal{U}_{s,t}^{1*}(Q_s^{01} + Q_s^{01*} - Q_s^{01*} B_s G_s^{-1} B_s^* Q_s^{01})\mathcal{U}_{s,t}^1\, ds, \ 0 \le t \le T, \tag{7.52}$$

and $N = Q^{00}B + L^*$.

Proof. Using (7.7), (7.39), (7.43)–(7.45) and writing

$$\tilde{Q}_t = \begin{bmatrix} Q_t^{00} & Q_t^{01} & Q_t^{02} \\ Q_t^{01*} & Q_t^{11} & Q_t^{12} \\ Q_t^{02*} & Q_t^{12*} & Q_t^{22} \end{bmatrix}$$

in the equation (7.48), one can easily verify that

$$Q_t^{02} = 0, \ Q_t^{12} = 0 \text{ and } Q_t^{22} = 0$$

and the component Q^{00} satisfies the equation (7.50). For the component Q^{01},

$$\begin{aligned} Q_t^{01} &= \mathcal{U}_{T,t}^* Q_T \mathcal{E}_{T,t} + \int_t^T \mathcal{U}_{s,t}^*(F_s - N_s G_s^{-1} N_s^*)\mathcal{E}_{s,t}\, ds \\ &\quad - \int_t^T \mathcal{U}_{s,t}^* N_s G_s^{-1} B_s^* Q_s^{01} \mathcal{U}_{s,t}^1\, ds. \end{aligned} \tag{7.53}$$

By (7.8) and (7.50),

$$\begin{aligned}
Q_t^{01} &= \int_t^T \mathcal{U}_{r,t}^* \mathcal{U}_{T,r}^* Q_T \mathcal{U}_{T,r} \mathcal{U}_{r,t}^1 \, dr \\
&\quad + \int_t^T \int_t^s \mathcal{U}_{r,t}^* \mathcal{U}_{s,r}^* (F_s - N_s G_s^{-1} N_s^*) \mathcal{U}_{s,r} \mathcal{U}_{r,t}^1 \, dr \, ds \\
&\quad - \int_t^T \mathcal{U}_{s,t}^* N_s G_s^{-1} B_s^* Q_s^{01} \mathcal{U}_{s,t}^1 \, ds \\
&= \int_t^T \mathcal{U}_{s,t}^* (Q_s^{00} - N_s G_s^{-1} B_s^* Q_s^{01}) \mathcal{U}_{s,t}^1 \, ds.
\end{aligned}$$

This proves (7.51). For the component Q^{11} of $\tilde{Q}$,

$$\begin{aligned}
Q_t^{11} &= \mathcal{E}_{T,t}^* Q_T \mathcal{E}_{T,t} + \int_t^T \mathcal{E}_{s,t}^* (F_s - N_s G_s^{-1} N_s^*) \mathcal{E}_{s,t} \, ds \\
&\quad - \int_t^T \mathcal{E}_{s,t}^* N_s G_s^{-1} B_s^* Q_s^{01} \mathcal{U}_{s,t}^1 \, ds - \int_t^T \mathcal{U}_{s,t}^{1*} Q_s^{01*} B_s G_s^{-1} N_s^* \mathcal{E}_{s,t} \, ds \\
&\quad - \int_t^T \mathcal{U}_{s,t}^{1*} Q^{01*} B_s G_s^{-1} B_s^* Q_s^{01} \mathcal{U}_{s,t}^1 \, ds.
\end{aligned}$$

Using (7.8), (7.26), (7.50), (7.51) and (7.53),

$$\begin{aligned}
Q_t^{11} &= \int_t^T \mathcal{U}_{r,t}^{1*} \mathcal{U}_{T,r}^* Q_T (\mathcal{U}_{T,r} \mathcal{E}_{r,t} + \mathcal{E}_{T,r} \mathcal{U}_{r,t}^1) \, dr \\
&\quad + \int_t^T \int_t^s \mathcal{U}_{r,t}^{1*} \mathcal{U}_{s,r}^* (F_s - N_s G_s^{-1} N_s^*) (\mathcal{U}_{s,r} \mathcal{E}_{r,t} + \mathcal{E}_{s,r}^1 \mathcal{U}_{r,t}^1) \, dr \, ds \\
&\quad - \int_t^T \int_t^s \mathcal{U}_{r,t}^{1*} \mathcal{U}_{s,r}^* N_s G_s^{-1} B_s^* Q_s^{01} \mathcal{U}_{s,r}^1 \mathcal{U}_{r,t}^1 \, dr \, ds \\
&\quad - \int_t^T \mathcal{U}_{s,t}^{1*} Q_s^{01*} B_s G_s^{-1} N_s^* \mathcal{E}_{s,t} \, ds - \int_t^T \mathcal{U}_{s,t}^{1*} Q_s^{01*} B_s G_s^{-1} B_s^* Q_s^{01} \mathcal{U}_{s,t}^1 \, ds \\
&= \int_t^T \mathcal{U}_{r,t}^{1*} Q_r^{00} \mathcal{E}_{r,t} \, dr + \int_t^T \mathcal{U}_{r,t}^{1*} Q_r^{01} \mathcal{U}_{r,t}^1 \, dr \\
&\quad - \int_t^T \mathcal{U}_{s,t}^{1*} Q_s^{01*} B_s G_s^{-1} N_s^* \mathcal{E}_{s,t} \, ds - \int_t^T \mathcal{U}_{s,t}^{1*} Q_s^{01*} B_s G_s^{-1} B_s^* Q_s^{01} \mathcal{U}_{s,t}^1 \, ds \\
&= \int_t^T \int_t^s \mathcal{U}_{r,t}^{1*} \mathcal{U}_{s,r}^{1*} (Q_s^{00} - Q_s^{01*} B_s G_s^{-1} N_s^*) \mathcal{U}_{s,r} \mathcal{U}_{r,t}^1 \, dr \, ds \\
&\quad + \int_t^T \mathcal{U}_{s,t}^{1*} (Q_s^{01} - Q_s^{01*} B_s G_s^{-1} B_s^* Q_s^{01}) \mathcal{U}_{s,t}^1 \, ds \\
&= \int_t^T \mathcal{U}_{s,t}^{1*} (Q_s^{01} + Q_s^{01*} - Q_s^{01*} B_s G_s^{-1} B_s^* Q_s^{01}) \mathcal{U}_{s,t}^1 \, ds.
\end{aligned}$$

This proves (7.51). The uniqueness of solution of the system (7.50)–(7.52) follows from the uniqueness of solution of the equation (7.48). □

Proposition 7.9. *Suppose the conditions* ($\mathbf{R}_1^c$) *and* ($\mathbf{R}_3^c$) *hold. Then the solution* Q^{01} *of the equation* (7.51) *has the representation*

$$Q_t^{01} = \int_t^T \mathcal{Y}_{s,t}^* Q_s^{00} \mathcal{U}_{s,t}^1 \, ds, \ 0 \le t \le T, \tag{7.54}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q^{00}+L)}(\mathcal{U})$.

Proof. Let $N = Q^{00}B + L$. Then we have

$$\mathcal{Y}_{s,t}^* = \mathcal{U}_{s,t}^* - \int_t^s \mathcal{Y}_{r,t}^* N_r G_r^{-1} B_r^* \mathcal{U}_{s,r}^* \, dr, \ 0 \le t \le s \le T.$$

Using this in (7.51), we obtain

$$\begin{aligned} Q_t^{01} &= \int_t^T \mathcal{Y}_{s,t}^* (Q_s^{00} - N_s G_s^{-1} B_s^* Q_s^{01}) \mathcal{U}_{s,t}^1 \, ds \\ &\quad + \int_t^T \int_t^s \mathcal{Y}_{r,t}^* N_r G_r^{-1} B_r^* \mathcal{U}_{s,r}^* (Q_s^{00} - N_s G_s^{-1} B_s^* Q_s^{01}) \mathcal{U}_{s,t}^1 \, dr \, ds \\ &= \int_t^T \mathcal{Y}_{s,t}^* (Q_s^{00} - N_s G_s^{-1} B_s^* Q_s^{01}) \mathcal{U}_{s,t}^1 \, ds \\ &\quad + \int_t^T \int_r^T \mathcal{Y}_{r,t}^* N_r G_r^{-1} B_r^* \mathcal{U}_{s,r}^* (Q_s^{00} - N_s G_s^{-1} B_s^* Q_s^{01}) \mathcal{U}_{s,r}^1 \mathcal{U}_{r,t}^1 \, ds \, dr \\ &= \int_t^T \mathcal{Y}_{s,t}^* Q_s^{00} \mathcal{U}_{s,t}^1 \, ds. \end{aligned}$$

The proof is completed. □

Proposition 7.10. *Suppose the conditions* ($\mathbf{R}_1^c$)–($\mathbf{R}_3^c$) *hold. Then the minimum of the functional* J *in the problem* (7.35)–(7.37) *is equal to*

$$\begin{aligned} J(u^*) &= \operatorname{tr}(Q_T P_T) + \operatorname{tr} \int_0^T F_t P_t^{00} \, dt \\ &\quad + \operatorname{tr} \int_0^T \bar{V}^{-1} \Psi_t^{-1} \begin{bmatrix} M_t \\ M_t^1 \end{bmatrix}^* \begin{bmatrix} Q_t^{00} & Q_t^{01} \\ Q_t^{01*} & Q_t^{11} \end{bmatrix} \begin{bmatrix} M_t \\ M_t^1 \end{bmatrix} \Psi_t^{-1*} \, dt, \end{aligned}$$

where $(P^{00}, P^{01}, P^{02}, P^{11}, P^{12}, P^{22})$ *is a solution of the system of equations* (7.18)–(7.23), (Q^{00}, Q^{01}, Q^{11}) *is a solution of the system of equations* (7.50)–(7.52) *and* M *and* M^1 *are defined by* (7.17).

Proof. This follows from Proposition 6.22. □

7.2.5 Example: Optimal Stochastic Regulator in Differential Form

Example 7.11. Assume that the conditions $(\mathbf{R}_1^c)$–$(\mathbf{R}_3^c)$ hold so that $\mathcal{U},\mathcal{U}^1 \in \mathcal{S}(X)$, $\mathcal{U}^2 \in \mathcal{S}(\mathbb{R}^n)$ and A, A_1 and A_2 are the infinitesimal generators of $\mathcal{U}$, $\mathcal{U}^1$ and $\mathcal{U}^2$, respectively. Then the system (7.35)–(7.36), under $u = u^*$, can be written as

$$\begin{cases} dx_t^* = \big(Ax_t^* + \varphi_t^1 + B_t u_t^*\big)dt + \Phi_t dw_t, \ 0 < t \le T, \ x_0^* = x_0, \\ d\varphi_t^1 = A_1\varphi_t^1 dt + \Phi_t^1 dw_t, \ 0 < t \le T, \ \varphi_0^1 = 0, \\ d\varphi_t^2 = A_2\varphi_t^2 dt + \Phi_t^2 dw_t, \ 0 < t \le T, \ \varphi_0^2 = 0, \\ dz_t^* = \big(C_t x_t^* + \varphi_t^2\big)dt + \Psi_t dv_t, \ 0 < t \le T, \ z_0^* = 0. \end{cases} \tag{7.55}$$

By (7.46), the optimal control u^* has the form

$$u_t^* = -G_t^{-1}\big(B_t^* Q_t^{00} + L_t\big)\hat{x}_t^* - G_t^{-1} B_t^* Q_t^{01}\psi_t^1, \text{ a.e. } t \in \mathbf{T}. \tag{7.56}$$

Note that by Proposition 7.9, u^* can also be represented in the form

$$u_t^* = -G_t^{-1}\big(B_t^* Q_t^{00} + L_t\big)\hat{x}_t^* - G_t^{-1} B_t^* \int_t^T \mathcal{Y}_{s,t}^* Q_s^{00} \mathcal{U}_{s,t}^1 \psi_t^1 \, ds, \text{ a.e. } t \in \mathbf{T}, \tag{7.57}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q^{00}+L)}(\mathcal{U})$, which agrees with the extended separation principle (see Theorem 5.16). Here, $(\hat{x}^*, \psi^1, \psi^2)$ is a mild solution of the system of linear stochastic differential equations (see Example 7.5)

$$\begin{cases} d\hat{x}_t^* = \big(A\hat{x}_t^* + \psi_t^1 + B_t u_t^*\big)dt + \big(P_t^{00}C_t^* + P_t^{02} + \Phi_t\bar{R}\Psi_t^*\big)\big(\Psi_t\bar{V}\Psi_t^*\big)^{-1} \\ \qquad \times \big(dz_t^* - C_t\hat{x}_t^* dt - \psi_t^2 dt\big), \ 0 < t \le T, \ \hat{x}_0^* = 0, \\ d\psi_t^1 = A_1\psi_t^1 dt + \big(P_t^{01*}C_t^* + P_t^{12} + \Phi_t^1\bar{R}\Psi_t^*\big)\big(\Psi_t\bar{V}\Psi_t^*\big)^{-1} \\ \qquad \times \big(dz_t^* - C_t\hat{x}_t^* dt - \psi_t^2 dt\big), \ 0 < t \le T, \ \psi_0^1 = 0, \\ d\psi_t^2 = A_2\psi_t^2 dt + \big(P_t^{02*}C_t^* + P_t^{22} + \Phi_t^2\bar{R}\Psi_t^*\big)\big(\Psi_t\bar{V}\Psi_t^*\big)^{-1} \\ \qquad \times \big(dz_t^* - C_t\hat{x}_t^* dt - \psi_t^2 dt\big), \ 0 < t \le T, \ \psi_0^2 = 0, \end{cases} \tag{7.58}$$

with $\big(P^{00}, P^{01}, P^{02}, P^{11}, P^{12}, P^{22}\big)$ satisfying (7.28)–(7.33) in the scalar product sense. Also, using Theorem 3.27, one can show that $\big(Q^{00}, Q^{01}, Q^{11}\big)$ is a unique scalar product solution of the system of differential equations

$$\begin{aligned} &\frac{d}{dt}Q_t^{00} + Q_t^{00}A + A^*Q_t^{00} + F_t \\ &\qquad - \big(Q_t^{00}B_t + L_t^*\big)G_t^{-1}\big(B_t^*Q_t^{00} + L_t\big) = 0, \ 0 \le t < T, \ Q_T^{00} = Q_T, \\ &\frac{d}{dt}Q_t^{01} + Q_t^{01}A_1 + A^*Q_t^{01} + Q_t^{00} \\ &\qquad - \big(Q_t^{00}B_t + L_t^*\big)G_t^{-1}B_t^*Q_t^{01} = 0, \ 0 \le t < T, \ Q_T^{01} = 0, \\ &\frac{d}{dt}Q_t^{11} + Q_t^{11}A_1 + A_1^*Q_t^{11} + Q_t^{01} + Q_t^{01*} \\ &\qquad - Q_t^{01*}B_tG_t^{-1}B_t^*Q_t^{01} = 0, \ 0 \le t < T, \ Q_T^{11} = 0. \end{aligned}$$

Chapter 8

Control and Estimation under Wide Band Noises

In this chapter the control and estimation results of Chapter 6 are modified to the wide band noise processes.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $X, U, H \in \mathcal{H}$, $T > 0$, $\mathbf{T} = [0, T]$ is a finite time interval and $\Delta_t = \{(s, r) : 0 \le r \le s \le t\}$ for $t > 0$.

8.1 Estimation

8.1.1 Setting of Estimation Problems

In this section we will study the estimation problems for the partially observable system (7.1)–(7.2) in which φ^1 and φ^2 are wide band noise processes. So, let

$$x_t = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\varphi_s^1\, ds + \int_0^t \mathcal{U}_{t,s}\Phi_s\, dw_s,\ 0 \le t \le T, \tag{8.1}$$

$$z_t = \int_0^t \left(C_s x_s + \varphi_s^2\right) ds + \int_0^t \Psi_s\, dv_s,\ 0 \le t \le T, \tag{8.2}$$

where

$$\varphi_t^1 = \int_{\max(0,t-\varepsilon)}^t \Phi_{t,\theta-t}^1\, dw_\theta,\ \varphi_t^2 = \int_{\max(0,t-\delta)}^t \Phi_{t,\alpha-t}^2\, dw_\alpha,\ 0 \le t \le T. \tag{8.3}$$

Recall that the integral representation for wide band noise processes was discussed in Section 4.6.2 according to which (8.3) defines a pair of wide band noise processes φ^1 and φ^2.

Assume that the following conditions hold:

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

($\mathbf{E_1^b}$) $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$;

($\mathbf{E_2^b}$) $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Phi^1 \in L_\infty(\mathbf{T}, B_2(-\varepsilon, 0; \mathcal{L}(H, X)))$, $0 < \varepsilon < T$, $\Phi^2 \in L_\infty(\mathbf{T}, B_2(-\delta, 0; \mathcal{L}(H, \mathbb{R}^n)))$, $0 < \delta < T$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with $\operatorname{cov} v_T > 0$, x_0 is a an X-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0 and (w, v) are independent.

The problem of estimating x_t based on the observations z_s, $0 \le s \le \tau$, where $t, \tau \in \mathbf{T}$ and (x, z) is defined by (8.1)–(8.3), will be called the estimation (filtering, prediction, smoothing) problem (8.1)–(8.3). This problem will be solved by use of the product space method, successfully applied in Chapter 7 dealing with colored noises.

In this section we will use the notation P_0, $\bar{W}$, $\bar{R}$, $\bar{V}$ and V from (7.4). Also, we will denote

$$\begin{cases} \tilde{X} = L_2(-\varepsilon, 0; X), \ \tilde{\mathbb{R}}^n = L_2(-\delta, 0; \mathbb{R}^n), \\ \check{X} = W^{1,2}(-\varepsilon, 0; X), \ \check{\mathbb{R}}^n = W^{1,2}(-\delta, 0; \mathbb{R}^n), \end{cases} \tag{8.4}$$

and consider the semigroups of right translation $\mathcal{T}^1$ and $\mathcal{T}^2$, defined by

$$[\mathcal{T}_t^1 f]_\theta = \begin{Bmatrix} f_{\theta - t}, & \theta - t \ge -\varepsilon \\ 0, & \theta - t < -\varepsilon \end{Bmatrix}, \ -\varepsilon \le \theta \le 0, \ t \ge 0, \ f \in \tilde{X}, \tag{8.5}$$

$$[\mathcal{T}_t^2 g]_\alpha = \begin{Bmatrix} g_{\alpha - t}, & \alpha - t \ge -\delta \\ 0, & \alpha - t < -\delta \end{Bmatrix}, \ -\delta \le \alpha \le 0, \ t \ge 0, \ g \in \tilde{\mathbb{R}}^n. \tag{8.6}$$

8.1.2 The First Reduction

Define the linear operators $\check{\Gamma}^1$ and $\check{\Gamma}^2$ by

$$\check{\Gamma}^1 \in \mathcal{L}(\check{X}, X) : \ \check{\Gamma}^1 f = f_0, \ f \in \check{X}, \tag{8.7}$$

$$\check{\Gamma}^2 \in \mathcal{L}(\check{\mathbb{R}}^n, \mathbb{R}^n) : \ \check{\Gamma}^2 g = g_0, \ g \in \check{\mathbb{R}}^n. \tag{8.8}$$

Recall that $\check{\Gamma}^1$ and $\check{\Gamma}^2$ are unbounded operators from $\tilde{X}$ and $\tilde{\mathbb{R}}^n$ to X and $\mathbb{R}^n$, respectively.

Let

$$\check{x}_t = \begin{bmatrix} x_t \\ \check{\varphi}_t^1 \\ \check{\varphi}_t^2 \end{bmatrix} \in X \times \tilde{X} \times \tilde{\mathbb{R}}^n, \ 0 \le t \le T, \tag{8.9}$$

where x is the random process, defined by (8.1), and $\check{\varphi}^1$ and $\check{\varphi}^2$ are the $\tilde{X}$- and $\tilde{\mathbb{R}}^n$-valued random processes defined by

$$[\check{\varphi}_t^1]_\theta = \int_{\max(0, t-\varepsilon-\theta)}^t \bar{\Phi}_{s, s-t+\theta}^1 \, dw_s, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \tag{8.10}$$

$$[\check{\varphi}_t^2]_\alpha = \int_{\max(0, t-\delta-\alpha)}^t \bar{\Phi}_{s, s-t+\alpha}^2 \, dw_s, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \tag{8.11}$$

with

$$\bar{\Phi}^1_{t,\theta} = \begin{Bmatrix} \Phi^1_{t-\theta,\theta}, & t-\theta \le T \\ 0, & t-\theta > T \end{Bmatrix}, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \tag{8.12}$$

$$\bar{\Phi}^2_{t,\alpha} = \begin{Bmatrix} \Phi^2_{t-\alpha,\alpha}, & t-\alpha \le T \\ 0, & t-\alpha > T \end{Bmatrix}, \ -\delta \le \alpha \le 0, \ 0 \le t \le T. \tag{8.13}$$

In this section we will derive equations for the processes $\check{x}$ and z, defined by (8.9) and (8.2) that will use the unbounded operators $\check{\Gamma}^1$ and $\check{\Gamma}^2$.

Lemma 8.1. *For $\check{\varphi}^1$ and $\check{\varphi}^2$, defined by* (8.10) *and* (8.11), *respectively,*

$$\check{\varphi}^1_t = \int_0^t \mathcal{T}^1_{t-s}\check{\Phi}^1_s \, dw_s, \ 0 \le t \le T, \tag{8.14}$$

$$\check{\varphi}^2_t = \int_0^t \mathcal{T}^2_{t-s}\check{\Phi}^2_s \, dw_s, \ 0 \le t \le T, \tag{8.15}$$

where $\mathcal{T}^1$ and $\mathcal{T}^2$ are the semigroups of right translation defined by (8.5) *and* (8.6), *$\check{\Phi}^1 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \tilde{X}))$ and $\check{\Phi}^2 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \tilde{\mathbb{R}}^n))$ are defined by*

$$\left[\check{\Phi}^1_t h\right]_\theta = \bar{\Phi}^1_{t,\theta} h, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ h \in H, \tag{8.16}$$

$$\left[\check{\Phi}^2_t h\right]_\alpha = \bar{\Phi}^2_{t,\alpha} h, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \ h \in H, \tag{8.17}$$

with $\bar{\Phi}^1$ and $\bar{\Phi}^2$ from (8.12)–(8.13).

Proof. Clearly, $\check{\Phi}^1 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \tilde{X}))$ and $\check{\Phi}^2 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \tilde{\mathbb{R}}^n))$. Let us prove the representation (8.14) for $\check{\varphi}^1$. By (8.5) and (8.16), for $h \in H$, we have

$$\begin{aligned} \left[\mathcal{T}^1_{t-s}\check{\Phi}^1_s h\right]_\theta &= \begin{Bmatrix} \left[\check{\Phi}^1_s h\right]_{s-t+\theta}, & s-t+\theta \ge -\varepsilon \\ 0, & s-t+\theta < -\varepsilon \end{Bmatrix} \\ &= \begin{Bmatrix} \bar{\Phi}^1_{s,s-t+\theta} h, & s-t+\theta \ge -\varepsilon \\ 0, & s-t+\theta < -\varepsilon \end{Bmatrix}. \end{aligned}$$

Therefore, by (8.10),

$$\left[\check{\varphi}^1_t\right]_\theta = \int_{\max(0,t-\varepsilon-\theta)}^t \bar{\Phi}^1_{s,s-t+\theta} \, dw_s = \left[\int_0^t \mathcal{T}^1_{t-s}\check{\Phi}^1_s \, dw_s\right]_\theta$$

and, consequently, (8.14) holds. In a similar way the representation (8.15) for $\check{\varphi}^2$ can be proved. □

Lemma 8.2. *Let*

$$\check{\Phi} = \begin{bmatrix} \Phi \\ \check{\Phi}^1 \\ \check{\Phi}^2 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(H, X \times \tilde{X} \times \tilde{\mathbb{R}}^n)), \tag{8.18}$$

where the functions $\check{\Phi}^1$ *and* $\check{\Phi}^2$ *are defined by* (8.16) *and* (8.17), *respectively, and let*

$$\check{\mathcal{U}}_{t,s} = \begin{bmatrix} \mathcal{U}_{t,s} & \check{\mathcal{E}}_{t,s} & 0 \\ 0 & \mathcal{T}^1_{t-s} & 0 \\ 0 & 0 & \mathcal{T}^2_{t-s} \end{bmatrix} \in \mathcal{L}\big(X \times \tilde{X} \times \tilde{\mathbb{R}}^n\big),\ 0 \le s \le t \le T, \tag{8.19}$$

where

$$\check{\mathcal{E}}_{t,s} f = \int_{\max(-\varepsilon, s-t)}^{0} \mathcal{U}_{t,s-r} f_r \, dr,\ 0 \le s \le t \le T,\ f \in \tilde{X}. \tag{8.20}$$

Then the random process $\check{x}$, *defined by* (8.9), *has the representation*

$$\check{x}_t = \check{\mathcal{U}}_{t,0}\check{x}_0 + \int_0^t \check{\mathcal{U}}_{t,s}\check{\Phi}_s \, dw_s,\ 0 \le t \le T. \tag{8.21}$$

Furthermore, the observation system (8.2) *can be written in the form*

$$z_t = \int_0^t \check{C}_s \check{x}_s \, ds + \int_0^t \Psi_s \, dv_s,\ 0 \le t \le T, \tag{8.22}$$

where

$$\check{C} = \begin{bmatrix} C & 0 & \check{\Gamma}^2 \end{bmatrix} \in B_\infty\big(\mathbf{T}, \mathcal{L}\big(X \times \tilde{X} \times \check{\mathbb{R}}^n, \mathbb{R}^n\big)\big). \tag{8.23}$$

Proof. First, note that by Example 3.12, the function $\check{\mathcal{U}}$, defined by (8.19)–(8.20), belongs to $\mathcal{E}\big(\Delta_T, \mathcal{L}\big(X \times \tilde{X} \times \tilde{\mathbb{R}}^n\big)\big)$. Since $\big[\check{\varphi}^2_t\big]_0 = \varphi^2_t$, $0 \le t \le T$, the formula (8.22) is obvious. Let us prove the representation (8.21). For the random process x, defined by (8.1) and (8.3), we have

$$\begin{aligned} x_t &= \mathcal{U}_{t,0}x_0 + \int_0^t \int_{\max(0,s-\varepsilon)}^{s} \mathcal{U}_{t,s}\Phi^1_{s,\theta-s} \, dw_\theta \, ds + \int_0^t \mathcal{U}_{t,s}\Phi_s \, dw_s \\ &= \mathcal{U}_{t,0}x_0 + \int_0^t \int_r^{\min(r+\varepsilon,t)} \mathcal{U}_{t,s}\Phi^1_{s,r-s} \, ds \, dw_r + \int_0^t \mathcal{U}_{t,s}\Phi_s \, dw_s. \end{aligned}$$

Since for $h \in H$,

$$\begin{aligned} \left(\int_r^{\min(r+\varepsilon,t)} \mathcal{U}_{t,s}\Phi^1_{s,r-s} \, ds\right) h &= \int_r^{\min(r+\varepsilon,t)} \mathcal{U}_{t,s}\Phi^1_{s,r-s} h \, ds \\ &= \int_r^{\min(r+\varepsilon,t)} \mathcal{U}_{t,s}\bar{\Phi}^1_{r,r-s} h \, ds \\ &= \int_r^{\min(r+\varepsilon,t)} \mathcal{U}_{t,s}\big[\check{\Phi}^1_r h\big]_{r-s} \, ds \\ &= \int_{\max(-\varepsilon,r-t)}^{0} \mathcal{U}_{t,r-s}\big[\check{\Phi}^1_r h\big]_s \, ds = \check{\mathcal{E}}_{t,r}\check{\Phi}^1_r h, \end{aligned}$$

where (8.12), (8.16) and (8.20) are used, we obtain

$$x_t = \mathcal{U}_{t,0}x_0 + \int_0^t \check{\mathcal{E}}_{t,r}\check{\Phi}_r^1 \, dw_r + \int_0^t \mathcal{U}_{t,r}\Phi_r \, dw_r, \ \ 0 \le t \le T.$$

Combining this equality with (8.14) and (8.15) and using (8.19), we obtain the representation (8.21). □

Lemma 8.3. *The best estimates in the estimation problems* (8.1)–(8.3) *and* (8.21)–(8.22) *are related as in*

$$\mathbf{E}(x_t|z_s;\, 0 \le s \le \tau) = \tilde{I}\,\mathbf{E}(\check{x}_t|z_s;\, 0 \le s \le \tau), \ \ t, \tau \in \mathbf{T},$$

where

$$\tilde{I} = [I \ \ 0 \ \ 0] \in \mathcal{L}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n, X). \tag{8.24}$$

Proof. This follows from the equality $x_t = \tilde{I}\check{x}_t$, $0 \le t \le T$. □

Thus, the estimation problem (8.1)–(8.3) is reduced to the estimation problem (8.21)–(8.22), where in the latter the noise disturbances are white noises.

8.1.3 The Second Reduction

Although the partially observable system (8.21)–(8.22) is driven only by white noise, the estimation results stated in Chapter 6 can be directly applied to this system only if $\Phi^2 = 0$ which compensates for the unboundedness of $\check{\Gamma}^2$ (as an operator from $\tilde{\mathbb{R}}^n$ to $\mathbb{R}^n$). This in turn means that in the original estimation problem (8.1)–(8.3) the observation system (8.2) must be free of wide band noise disturbance. If $\Phi^2 \ne 0$, then to handle the unboundedness of $\check{\Gamma}^2$, one can approximate $\check{\Gamma}^2$ by bounded operators. This approach is used by Bashirov [9] under an additional condition of continuity on Φ^2 and it is closely related with a linear regulator problem under delays in control studied by Ichikawa [55]. The realization of this approach meets many difficulties and requires routine calculations. Instead, under some more restrictive conditions of differentiability on Φ^2, the difficulties of the above mentioned approach can be easily avoided and a reduced filtering problem can be obtained in terms of only bounded linear operators.

Additionally, note that the operator $\check{\Gamma}^1$, defined by (8.7), is implicitly involved in the system (8.21)–(8.22) such that the unboundedness of $\check{\Gamma}^1$ (as an operator from $\tilde{X}$ to X) is absorbed by $\check{\mathcal{E}}$, defined by (8.20). In studying the components of the Riccati equation associated with the system (8.21)–(8.22) this unboundedness creates difficulties. To avoid them, we will impose a differentiability condition on Φ^1 too.

Thus, we will suppose that the conditions $(\mathbf{E_1^b})$ and $(\mathbf{E_2^b})$ hold and

$(\mathbf{E_3^b})$ the operator-valued functions Φ^1, Φ^{1*}, Φ^2 and Φ^{2*} are strongly differentiable in each of the variables with

$$\frac{\partial}{\partial t}\Phi^1,\ \frac{\partial}{\partial \theta}\Phi^1 \in L_\infty(\mathbf{T}, B_2(-\varepsilon, 0; \mathcal{L}(H, X))),$$
$$\frac{\partial}{\partial t}\Phi^2, \frac{\partial}{\partial \alpha}\Phi^2 \in L_\infty(\mathbf{T}, B_2(-\delta, 0; \mathcal{L}(H, \mathbb{R}^n)))$$

and $\Phi^1_{t,-\varepsilon} = 0$ and $\Phi^2_{t,-\delta} = 0$ for all $0 \le t \le T$.

Note that by Proposition 2.8(d), $(\mathbf{E_3^b})$ implies

$$\frac{\partial}{\partial t}\Phi^{1*},\ \frac{\partial}{\partial \theta}\Phi^{1*} \in L_\infty(\mathbf{T}, B_2(-\varepsilon, 0; \mathcal{L}(X, H))),$$
$$\frac{\partial}{\partial t}\Phi^{2*}, \frac{\partial}{\partial \alpha}\Phi^{2*} \in L_\infty(\mathbf{T}, B_2(-\delta, 0; \mathcal{L}(\mathbb{R}^n, H))),$$

$\Phi^{1*}_{t,-\varepsilon} = 0$ and $\Phi^{2*}_{t,-\delta} = 0$ for all $0 \le t \le T$.

Remark 8.4. By the physical interpretation of wide band noises discussed in Section 4.6.2, the functions Φ^1 and Φ^2 are coefficients of relaxing. Therefore, in the one-dimensional case, they are naturally expected to be increasing (and, hence, a.e. differentiable) functions of the second variable satisfying $\Phi^1_{t,-\varepsilon} = 0$ and $\Phi^2_{t,-\delta} = 0$.

Let

$$\tilde{x}_t = \begin{bmatrix} x_t \\ \tilde{\varphi}^1_t \\ \tilde{\varphi}^2_t \end{bmatrix} \in X \times \tilde{X} \times \tilde{\mathbb{R}}^n,\ 0 \le t \le T, \tag{8.25}$$

where x is defined by (8.1) and (8.3) and $\tilde{\varphi}^1$ and $\tilde{\varphi}^2$ are defined by

$$[\tilde{\varphi}^1_t]_\theta = \int_{\max(0,t-\varepsilon-\theta)}^t \frac{\partial}{\partial \theta}\bar{\Phi}^1_{s,s-t+\theta}\, dw_s,\ -\varepsilon \le \theta \le 0,\ 0 \le t \le T, \tag{8.26}$$

$$[\tilde{\varphi}^2_t]_\alpha = \int_{\max(0,t-\delta-\alpha)}^t \frac{\partial}{\partial \alpha}\bar{\Phi}^2_{s,s-t+\alpha}\, dw_s,\ -\delta \le \alpha \le 0,\ 0 \le t \le T, \tag{8.27}$$

with $\bar{\Phi}^1$ and $\bar{\Phi}^2$ from (8.12)–(8.13).

Lemma 8.5. *For $\tilde{\varphi}^1$ and $\tilde{\varphi}^2$, defined by (8.26) and (8.27), respectively,*

$$\tilde{\varphi}^1_t = \int_0^t \mathcal{T}^1_{t-s}\tilde{\Phi}^1_s\, dw_s,\ 0 \le t \le T, \tag{8.28}$$

$$\tilde{\varphi}^2_t = \int_0^t \mathcal{T}^2_{t-s}\tilde{\Phi}^2_s\, dw_s,\ 0 \le t \le T, \tag{8.29}$$

where $\mathcal{T}^1$ *and* $\mathcal{T}^2$ *are defined by* (8.5) *and* (8.6) *and* $\tilde{\Phi}^1 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \tilde{X}))$ *and* $\tilde{\Phi}^2 \in B_\infty(\mathbf{T}, \mathcal{L}(H, \tilde{\mathbb{R}}^n))$ *are defined by*

$$\left[\tilde{\Phi}_t^1 h\right]_\theta = \frac{\partial}{\partial\theta}\bar{\Phi}_{t,\theta}^1 h, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ h \in H, \tag{8.30}$$

$$\left[\tilde{\Phi}_t^2 h\right]_\alpha = \frac{\partial}{\partial\alpha}\bar{\Phi}_{t,\alpha}^2 h, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \ h \in H, \tag{8.31}$$

with $\bar{\Phi}^1$ *and* $\bar{\Phi}^2$ *from* (8.12)–(8.13).

Proof. By the condition $(\mathbf{E_3^b})$, we have

$$\tilde{\Phi}^1 \in B_\infty\big(\mathbf{T}, \mathcal{L}(H, \tilde{X})\big) \text{ and } \tilde{\Phi}^2 \in B_\infty\big(\mathbf{T}, \mathcal{L}(H, \tilde{\mathbb{R}}^n)\big).$$

Let us prove the representation (8.28) for $\tilde{\varphi}^1$. By (8.5) and (8.30), for $h \in H$, we have

$$\begin{aligned}\left[\mathcal{T}_{t-s}^1 \tilde{\Phi}_s^1 h\right]_\theta &= \begin{Bmatrix} \left[\tilde{\Phi}_s^1 h\right]_{s-t+\theta}, & s-t+\theta \ge -\varepsilon \\ 0, & s-t+\theta < -\varepsilon \end{Bmatrix} \\ &= \begin{Bmatrix} (\partial/\partial\theta)\bar{\Phi}_{s,s-t+\theta}^1 h, & s-t+\theta \ge -\varepsilon \\ 0, & s-t+\theta < -\varepsilon \end{Bmatrix}.\end{aligned}$$

Therefore, by (8.26),

$$\left[\tilde{\varphi}_t^1\right]_\theta = \int_{\max(0,t-\varepsilon-\theta)}^t \frac{\partial}{\partial\theta}\bar{\Phi}_{s,s-t+\theta}^1 \, dw_s = \left[\int_0^t \mathcal{T}_{t-s}^1 \tilde{\Phi}_s^1 \, dw_s\right]_\theta$$

and, consequently, (8.28) holds. In a similar way the representation (8.29) for $\tilde{\varphi}^2$ can be proved. □

Let Γ^1 and Γ^2 be the integral operators from $\tilde{X}$ and $\tilde{\mathbb{R}}^n$ to X and $\mathbb{R}^n$, respectively, defined by

$$\Gamma^1 f = \int_{-\varepsilon}^0 f_\theta \, d\theta, \ f \in \tilde{X}, \tag{8.32}$$

$$\Gamma^2 g = \int_{-\delta}^0 g_\alpha \, d\alpha, \ g \in \tilde{\mathbb{R}}^n. \tag{8.33}$$

In the following lemma we will use the mild evolution operator $\tilde{\mathcal{U}}$ which we define by

$$\tilde{\mathcal{U}} = \mathcal{P}_D\big(\mathcal{U} \odot \mathcal{T}^1 \odot \mathcal{T}^2\big), \ D = \begin{bmatrix} 0 & \Gamma^1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}\big(X \times \tilde{X} \times \tilde{\mathbb{R}}^n\big).$$

Using Theorem 3.13, one can obtain

$$\tilde{\mathcal{U}}_{t,s} = \begin{bmatrix} \mathcal{U}_{t,s} & \mathcal{E}_{t,s} & 0 \\ 0 & \mathcal{T}_{t-s}^1 & 0 \\ 0 & 0 & \mathcal{T}_{t-s}^2 \end{bmatrix}, \ 0 \le s \le t \le T, \tag{8.34}$$

where

$$\mathcal{E}_{t,s} = \int_s^t \mathcal{U}_{t,r}\Gamma^1\mathcal{T}^1_{r-s}\,dr,\ 0 \le s \le t \le T. \tag{8.35}$$

Lemma 8.6. *Let*

$$\tilde{\Phi} = \begin{bmatrix} \Phi \\ \tilde{\Phi}^1 \\ \tilde{\Phi}^2 \end{bmatrix} \in B_\infty\big(\mathbf{T}, \mathcal{L}\big(H, X \times \tilde{X} \times \tilde{\mathbb{R}}^n\big)\big), \tag{8.36}$$

where the functions $\tilde{\Phi}^1$ and $\tilde{\Phi}^2$ are defined by (8.30)–(8.31). *Then the random process $\tilde{x}$ defined by* (8.25) *has the representation*

$$\tilde{x}_t = \tilde{\mathcal{U}}_{t,0}\tilde{x}_0 + \int_0^t \tilde{\mathcal{U}}_{t,s}\tilde{\Phi}_s\,dw_s,\ 0 \le t \le T, \tag{8.37}$$

where $\tilde{\mathcal{U}}$ is defined by (8.34)–(8.35).

Proof. Let $h \in H$ and let $0 \le r \le s \le t \le T$. Using the expression for $\big[\mathcal{T}^1_{s-r}\tilde{\Phi}^1_r h\big]_\theta$, obtained in proving Lemma 8.5, we have

$$\Gamma^1\mathcal{T}^1_{s-r}\tilde{\Phi}^1_r h = \int_{\min(0,s-r-\varepsilon)}^0 \frac{\partial}{\partial\theta}\bar{\Phi}^1_{r,r-s+\theta}h\,d\theta.$$

Hence, by $\big(\mathbf{E}^{\mathbf{b}}_{\mathbf{3}}\big)$,

$$\begin{aligned}\int_r^t \mathcal{U}_{t,s}\Gamma^1\mathcal{T}^1_{s-r}\tilde{\Phi}^1_r h\,ds &= \int_r^t \mathcal{U}_{t,s}\int_{\min(0,s-r-\varepsilon)}^0 \frac{\partial}{\partial\theta}\bar{\Phi}^1_{r,r-s+\theta}h\,d\theta\,ds \\ &= \int_r^{\min(t,r+\varepsilon)} \mathcal{U}_{t,s}\int_{s-r-\varepsilon}^0 \frac{\partial}{\partial\theta}\bar{\Phi}^1_{r,r-s+\theta}h\,d\theta\,ds \\ &= \int_r^{\min(t,r+\varepsilon)} \mathcal{U}_{t,s}\big(\bar{\Phi}^1_{r,r-s} - \bar{\Phi}^1_{r,-\varepsilon}\big)h\,ds \\ &= \int_r^{\min(t,r+\varepsilon)} \mathcal{U}_{t,s}\Phi^1_{s,r-s}h\,ds.\end{aligned}$$

Using this equality and (8.35), for the random process x defined by (8.1) and (8.3), we obtain

$$\begin{aligned}x_t &= \mathcal{U}_{t,0}x_0 + \int_0^t\int_{\max(0,s-\varepsilon)}^s \mathcal{U}_{t,s}\Phi^1_{s,\theta-s}\,dw_\theta\,ds + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s \\ &= \mathcal{U}_{t,0}x_0 + \int_0^t\int_r^{\min(r+\varepsilon,t)} \mathcal{U}_{t,s}\Phi^1_{s,r-s}\,ds\,dw_r + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s \\ &= \mathcal{U}_{t,0}x_0 + \int_0^t\int_r^t \mathcal{U}_{t,s}\Gamma^1\mathcal{T}^1_{s-r}\tilde{\Phi}^1_r\,ds\,dw_r + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s \\ &= \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{E}_{t,r}\tilde{\Phi}^1_r\,dw_r + \int_0^t \mathcal{U}_{t,r}\Phi_r\,dw_r.\end{aligned}$$

Combining the last equality together with the equalities (8.28) and (8.29), we obtain (8.37). □

Lemma 8.7. *The observation process* z, *defined by* (8.2) *and* (8.3), *has the representation*

$$z_t = \int_0^t \tilde{C}_s \tilde{x}_s \, ds + \int_0^t \Psi_s \, dv_s, \ \ 0 \le t \le T, \tag{8.38}$$

where $\tilde{x}$ *is defined by* (8.25),

$$\tilde{C} = \begin{bmatrix} C & 0 & \Gamma^2 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n, \mathbb{R}^n)), \tag{8.39}$$

and Γ^2 *is defined by* (8.33).

Proof. It must be shown that

$$\varphi_t^2 = \Gamma^2 \tilde{\varphi}_t^2, \ \ 0 \le t \le T,$$

where φ^2 and $\tilde{\varphi}^2$ are defined by (8.3) and (8.27), respectively. This follows from

$$\begin{aligned}
\Gamma^2 \tilde{\varphi}_t^2 &= \int_{-\delta}^0 \left[\tilde{\varphi}_t^2\right]_\alpha d\alpha \\
&= \int_{-\delta}^0 \int_{\max(0,t-\delta-\alpha)}^t \frac{\partial}{\partial \alpha} \bar{\Phi}^2_{s,s-t+\alpha} \, dw_s \, d\alpha \\
&= \int_{\max(0,t-\delta)}^t \int_{t-s-\delta}^0 \frac{\partial}{\partial \alpha} \bar{\Phi}^2_{s,s-t+\alpha} \, d\alpha \, dw_s \\
&= \int_{\max(0,t-\delta)}^t \left(\bar{\Phi}^2_{s,s-t} - \bar{\Phi}^2_{s,-\delta}\right) dw_s \\
&= \int_{\max(0,t-\delta)}^t \Phi^2_{t,s-t} \, dw_s = \varphi_t^2.
\end{aligned}$$

Thus, (8.38) is proved. □

Lemma 8.8. *The best estimates in the estimation problems* (8.1)–(8.3) *and* (8.37)–(8.38) *are related as in*

$$\mathbf{E}(x_t | z_s; \ 0 \le s \le \tau) = \tilde{I} \mathbf{E}(\tilde{x}_t | z_s, 0 \le s \le \tau), \ \ t, \tau \in \mathbf{T},$$

where $\tilde{I}$ *is defined by* (8.24).

Proof. In fact, this is Lemma 8.3 in which the system (8.21)–(8.22) is replaced by the system (8.37)–(8.38). □

Thus, the estimation problem (8.1)–(8.3) is reduced to the estimation problem (8.37)–(8.38) in which the noise disturbances are white noises and all operators are bounded.

8.1.4 Optimal Linear Feedback Estimators

The theorem, stated below, formally looks similar to Theorem 7.2, but the same symbols used in these theorems may have different meanings.

Theorem 8.9. *Suppose the conditions* $(\mathbf{E}_1^b)$–$(\mathbf{E}_3^b)$ *hold, let*

$$\hat{\tilde{x}}_t = \begin{bmatrix} \hat{x}_t \\ \tilde{\psi}_t^1 \\ \tilde{\psi}_t^2 \end{bmatrix} = \int_0^t \tilde{\mathcal{Y}}_{t,s}(\tilde{P}_s\tilde{C}_s^* + \tilde{\Phi}_s\bar{R}\Psi_s^*)V_s^{-1}\,dz_s,\ 0 \le t \le T,$$

and let

$$\tilde{z}_t = z_t - \int_0^t \tilde{C}_s\hat{\tilde{x}}_s\,ds = z_t - \int_0^t (C_s\hat{x}_s + \Gamma^2\tilde{\psi}_s^2)ds,\ 0 \le t \le T,$$

where $\tilde{\mathcal{U}}$, Γ^1, Γ^2, $\tilde{C}$, $\tilde{\Phi}$, V, $\bar{W}$ *and* $\bar{R}$ *are defined by* (8.34)–(8.35), (8.32)–(8.33), (8.39), (8.36), (8.30)–(8.31) *and* (7.4), $\tilde{\mathcal{Y}} = \mathcal{P}_{-(\tilde{P}\tilde{C}^*+\tilde{\Phi}\bar{R}\Psi^*)V^{-1}\tilde{C}}(\tilde{\mathcal{U}})$ *and* $\tilde{P}$ *is a unique solution of the Riccati equation*

$$\begin{aligned}\tilde{P}_t = \tilde{\mathcal{U}}_{t,0}\tilde{P}_0\tilde{\mathcal{U}}_{t,0}^* + \int_0^t \tilde{\mathcal{U}}_{t,s}\Big(\tilde{\Phi}_s\bar{W}\tilde{\Phi}_s^* \\ - \Big(\tilde{P}_s\tilde{C}_s^* + \tilde{\Phi}_s\bar{R}\Psi_s^*\Big)V_s^{-1}\Big(\tilde{C}_s\tilde{P}_s + \Psi_s\bar{R}^*\tilde{\Phi}_s^*\Big)\Big)\tilde{\mathcal{U}}_{t,s}^*\,ds,\ 0 \le t \le T,\end{aligned} \tag{8.40}$$

with

$$\tilde{P}_0 = \operatorname{cov}\begin{bmatrix} x_0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n). \tag{8.41}$$

Then there exists a unique optimal linear feedback filter (predictor, smoother) in the estimation problem (8.1)–(8.3) *and depending on* t *and* τ, *the best estimate* $\hat{x}_t^\tau$ *of* x_t *based on* z_s, $0 \le s \le \tau$, *in the estimation problem* (8.1)–(8.3) *is equal to*

$$\begin{aligned}\hat{x}_t = \hat{x}_t^t &= \int_0^t \tilde{I}\tilde{\mathcal{Y}}_{t,s}(\tilde{P}_s\tilde{C}_s^* + \tilde{\Phi}_s\bar{R}\Psi_s^*)V_s^{-1}\,dz_s,\ 0 \le t \le T,\\ \hat{x}_t^\tau &= \tilde{I}\tilde{\mathcal{U}}_{t,\tau}\hat{\tilde{x}}_\tau = \mathcal{U}_{t,\tau}\hat{x}_\tau + \mathcal{E}_{t,\tau}\tilde{\psi}_\tau^1,\ 0 \le \tau < t \le T,\\ \hat{x}_t^\tau &= \hat{x}_t + \tilde{I}\tilde{P}_t\int_t^\tau \tilde{\mathcal{Y}}_{s,t}^*\tilde{C}_s^*V_s^{-1}\,d\tilde{z}_s,\ 0 \le t < \tau \le T,\end{aligned}$$

where $\tilde{I}$ *and* $\mathcal{E}$ *are as defined by* (8.24) *and* (8.35).

Proof. This is a consequence of Theorems 6.4, 6.13, 6.18 and Lemma 8.8. □

8.1.5 About the Riccati Equation (8.40)

The solution $\tilde{P}$ of the Riccati equation (8.40) can be decomposed as

$$\tilde{P}_t = \begin{bmatrix} P_t^{00} & \tilde{P}_t^{01} & \tilde{P}_t^{02} \\ \tilde{P}_t^{01*} & \tilde{P}_t^{11} & \tilde{P}_t^{12} \\ \tilde{P}_t^{02*} & \tilde{P}_t^{12*} & \tilde{P}_t^{22} \end{bmatrix} \in \mathcal{L}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n),\ 0 \le t \le T. \tag{8.42}$$

The following proposition presents a system of equations for the components of $\tilde{P}$ in this decomposition.

Proposition 8.10. *Suppose the conditions* $(\mathbf{E_1^b})$–$(\mathbf{E_3^b})$ *hold, let the solution* $\tilde{P}$ *of the equation* (8.40) *be decomposed as* (8.42) *and let*

$$\begin{cases} \tilde{M}_t = P_t^{00} C_t^* + \tilde{P}_t^{02} \Gamma^{2*} + \Phi_t \bar{R} \Psi_t^*,\ 0 \le t \le T, \\ \tilde{M}_t^1 = \tilde{P}_t^{01*} C_t^* + \tilde{P}_t^{12} \Gamma^{2*} + \tilde{\Phi}_t^1 \bar{R} \Psi_t^*,\ 0 \le t \le T, \\ \tilde{M}_t^2 = \tilde{P}_t^{02*} C_t^* + \tilde{P}_t^{22} \Gamma^{2*} + \tilde{\Phi}_t^2 \bar{R} \Psi_t^*,\ 0 \le t \le T. \end{cases} \tag{8.43}$$

Then $(P^{00}, \tilde{P}^{01}, \tilde{P}^{02}, \tilde{P}^{11}, \tilde{P}^{12}, \tilde{P}^{22})$ *is a unique solution of the system of equations*

$$P_t^{00} = \mathcal{U}_{t,0} P_0 \mathcal{U}_{t,0}^* + \int_0^t \mathcal{U}_{t,s} (\Phi_s \bar{W} \Phi_s^* + \tilde{P}_s^{01} \Gamma^{1*} + \Gamma^1 \tilde{P}_s^{01*} - \tilde{M}_s V_s^{-1} \tilde{M}_s^*) \mathcal{U}_{t,s}^*\, ds,\ 0 \le t \le T, \tag{8.44}$$

$$\tilde{P}_t^{01} = \int_0^t \mathcal{U}_{t,s} (\Phi_s \bar{W} \tilde{\Phi}_s^{1*} + \Gamma^1 \tilde{P}_s^{11} - \tilde{M}_s V_s^{-1} \tilde{M}_s^{1*}) \mathcal{T}_{t-s}^{1*}\, ds,\ 0 \le t \le T, \tag{8.45}$$

$$\tilde{P}_t^{02} = \int_0^t \mathcal{U}_{t,s} (\Phi_s \bar{W} \tilde{\Phi}_s^{2*} + \Gamma^1 \tilde{P}_s^{12} - \tilde{M}_s V_s^{-1} \tilde{M}_s^{2*}) \mathcal{T}_{t-s}^{2*}\, ds,\ 0 \le t \le T, \tag{8.46}$$

$$\tilde{P}_t^{11} = \int_0^t \mathcal{T}_{t-s}^1 (\tilde{\Phi}_s^1 \bar{W} \tilde{\Phi}_s^{1*} - \tilde{M}_s^1 V_s^{-1} \tilde{M}_s^{1*}) \mathcal{T}_{t-s}^{1*}\, ds,\ 0 \le t \le T, \tag{8.47}$$

$$\tilde{P}_t^{12} = \int_0^t \mathcal{T}_{t-s}^1 (\tilde{\Phi}_s^1 \bar{W} \tilde{\Phi}_s^{2*} - \tilde{M}_s^1 V_s^{-1} \tilde{M}_s^{2*}) \mathcal{T}_{t-s}^{2*}\, ds,\ 0 \le t \le T, \tag{8.48}$$

$$\tilde{P}_t^{22} = \int_0^t \mathcal{T}_{t-s}^2 (\tilde{\Phi}_s^2 \bar{W} \tilde{\Phi}_s^{2*} - \tilde{M}_s^2 V_s^{-1} \tilde{M}_s^{2*}) \mathcal{T}_{t-s}^{2*}\, ds,\ 0 \le t \le T. \tag{8.49}$$

Proof. This can be proved in a similar way as Proposition 7.3. □

Proposition 8.11. *Suppose the conditions* $(\mathbf{E_1^b})$–$(\mathbf{E_3^b})$ *hold. Then for the error process* $e_t = x_t - \hat{x}_t$, $0 \le t \le T$, *in the filtering problem* (8.1)–(8.3), *the equality* $\operatorname{cov} e_t = P_t^{00}$, $0 \le t \le T$, *holds where* P^{00} *is defined by the system of equations* (8.44)–(8.49).

Proof. This follows from Proposition 6.5. □

8.1.6 Example: Optimal Filter in Differential Form

The following example concerning the filtering problem (8.1)–(8.3) is similar to Example 7.5.

Example 8.12. Assume that the conditions $(\mathbf{E}_1^\flat)$–$(\mathbf{E}_3^\flat)$ hold so that $\mathcal{U} \in \mathcal{S}(X)$ and A is the infinitesimal generator of $\mathcal{U}$. Then the best estimate $\hat{x}$ in the filtering problem (8.1)–(8.3) together with $\tilde{\psi}^1$ and $\tilde{\psi}^2$ is a mild solution of the system of equations

$$\begin{cases} d\hat{x}_t = A\hat{x}_t dt + \Gamma^1\tilde{\psi}_t^1 dt + \left(P_t^{00}C_t^* + \tilde{P}_t^{02}\Gamma^{2*} + \Phi_t\bar{R}\Psi_t^*\right) \\ \qquad \times\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\left(dz_t - C_t\hat{x}_t dt - \Gamma^2\tilde{\psi}_t^2 dt\right),\ 0 < t \le T,\ \hat{x}_0 = 0, \\ d\tilde{\psi}_t^1 = (-d/d\theta)\tilde{\psi}_t^1 dt + \left(\tilde{P}_t^{01*}C_t^* + \tilde{P}_t^{12}\Gamma^{2*} + \tilde{\Phi}_t^1\bar{R}\Psi_t^*\right) \\ \qquad \times\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\left(dz_t - C_t\hat{x}_t dt - \Gamma^2\tilde{\psi}_t^2 dt\right),\ 0 < t \le T,\ \tilde{\psi}_0^1 = 0, \\ d\tilde{\psi}_t^2 = (-d/d\alpha)\tilde{\psi}_t^2 dt + \left(\tilde{P}_t^{02*}C_t^* + \tilde{P}_t^{22}\Gamma^{2*} + \tilde{\Phi}_t^2\bar{R}\Psi_t^*\right) \\ \qquad \times\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\left(dz_t - C_t\hat{x}_t dt - \Gamma^2\tilde{\psi}_t^2 dt\right),\ 0 < t \le T,\ \tilde{\psi}_0^2 = 0, \end{cases} \tag{8.50}$$

where $\left(P^{00}, \tilde{P}^{01}, \tilde{P}^{02}, \tilde{P}^{11}, \tilde{P}^{12}, \tilde{P}^{22}\right)$ is a unique scalar product solution of the system of differential equations

$$\frac{d}{dt}P_t^{00} - P_t^{00}A^* - AP_t^{00} - \tilde{P}_t^{01}\Gamma^{1*} - \Gamma^1\tilde{P}_t^{01*} - \Phi_t\bar{W}\Phi_t^* \\ + \tilde{M}_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\tilde{M}_t^* = 0,\ 0 < t \le T,\ P_0^{00} = P_0, \tag{8.51}$$

$$\frac{d}{dt}\tilde{P}_t^{01} - \tilde{P}_t^{01}\frac{d}{d\theta} - A\tilde{P}_t^{01} - \Gamma^1\tilde{P}_t^{11} - \Phi_t\bar{W}\tilde{\Phi}_t^{1*} \\ + \tilde{M}_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\tilde{M}_t^{1*} = 0,\ 0 < t \le T,\ \tilde{P}_0^{01} = 0, \tag{8.52}$$

$$\frac{d}{dt}\tilde{P}_t^{02} - \tilde{P}_t^{02}\frac{d}{d\alpha} - A\tilde{P}_t^{02} - \Gamma^1\tilde{P}_t^{12} - \Phi_t\bar{W}\tilde{\Phi}_t^{2*} \\ + \tilde{M}_t\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\tilde{M}_t^{2*} = 0,\ 0 < t \le T,\ \tilde{P}_0^{02} = 0, \tag{8.53}$$

$$\frac{d}{dt}\tilde{P}_t^{11} - \tilde{P}_t^{11}\frac{d}{d\theta} - \left(-\frac{d}{d\theta}\right)\tilde{P}_t^{11} - \tilde{\Phi}_t^1\bar{W}\tilde{\Phi}_t^{1*} \\ + \tilde{M}_t^1\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\tilde{M}_t^{1*} = 0,\ 0 < t \le T,\ \tilde{P}_0^{11} = 0, \tag{8.54}$$

$$\frac{d}{dt}\tilde{P}_t^{12} - \tilde{P}_t^{12}\frac{d}{d\alpha} - \left(-\frac{d}{d\theta}\right)\tilde{P}_t^{12} - \tilde{\Phi}_t^1\bar{W}\tilde{\Phi}_t^{2*} \\ + \tilde{M}_t^1\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\tilde{M}_t^{2*} = 0,\ 0 < t \le T,\ \tilde{P}_0^{12} = 0, \tag{8.55}$$

$$\frac{d}{dt}\tilde{P}_t^{22} - \tilde{P}_t^{22}\frac{d}{d\alpha} - \left(-\frac{d}{d\alpha}\right)\tilde{P}_t^{22} - \tilde{\Phi}_t^2\bar{W}\tilde{\Phi}_t^{2*} \\ + \tilde{M}_t^2\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}\tilde{M}_t^{2*} = 0,\ 0 < t \le T,\ \tilde{P}_0^{22} = 0, \tag{8.56}$$

$d/d\theta$ is a differential operator from

$$D(d/d\theta) = \{f \in \check{X} : f_0 = 0\}$$

to $\tilde{X}$ and $-d/d\theta$ is its adjoint from

$$D(-d/d\theta) = \{f \in \check{X} : f_{-\varepsilon} = 0\}$$

to $\tilde{X}$, $d/d\alpha$ is a differential operator from

$$D(d/d\alpha) = \{g \in \check{\mathbb{R}}^n : g_0 = 0\}$$

to $\tilde{\mathbb{R}}^n$ and $-d/d\alpha$ is its adjoint from

$$D(-d/d\alpha) = \{g \in \check{\mathbb{R}}^n : g_{-\delta} = 0\}$$

to $\tilde{\mathbb{R}}^n$, $\tilde{M}$, $\tilde{M}^1$ and $\tilde{M}^2$ are defined by (8.43).

Indeed, by Theorem 3.15,

$$\tilde{\mathcal{U}} \in \mathcal{S}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n).$$

Since the infinitesimal generators of $\mathcal{T}^1$ and $\mathcal{T}^2$ are $-d/d\theta$ and $-d/d\alpha$, we obtain that the infinitesimal generator of $\tilde{\mathcal{U}}$ is

$$\tilde{A} = \begin{bmatrix} A & \Gamma^1 & 0 \\ 0 & -d/d\theta & 0 \\ 0 & 0 & -d/d\alpha \end{bmatrix}. \tag{8.57}$$

Also,

$$\tilde{A}^* = \begin{bmatrix} A^* & 0 & 0 \\ \Gamma^{1*} & d/d\theta & 0 \\ 0 & 0 & d/d\alpha \end{bmatrix}. \tag{8.58}$$

Thus, applying the results of Example 6.10 to the filtering problem (8.37)–(8.38), we obtain the system (8.50) for the optimal filter in the filtering problem (8.1)–(8.3). The equations (8.51)–(8.56) are exactly the equations (8.44)–(8.49) written in differential form which follow from Theorem 3.28 and (8.40).

8.2 More About the Optimal Filter

In this section we develop the results from Example 8.12 and find more convenient equations for the optimal filter than in (8.50)–(8.56) .

8.2.1 More About the Riccati Equation (8.40)

By Proposition 3.26, the solution $\tilde{P}$ of the Riccati equation (8.40) belongs to the space $L_\infty(\mathbf{T}, \mathcal{L}_1(X \times \tilde{X} \times \tilde{\mathbb{R}}^n))$ and, hence, to $L_\infty(\mathbf{T}, \mathcal{L}_2(X \times \tilde{X} \times \tilde{\mathbb{R}}^n))$. Thus, each component of $\tilde{P}$ in the decomposition (8.42) is an $\mathcal{L}_2$-valued function. By Proposition 2.39, we can represent these components (except P^{00}) in the form:

$$\tilde{P}_t^{01} f = \int_{-\varepsilon}^{0} \bar{P}_{t,\theta}^{01} f_\theta \, d\theta, \ 0 \le t \le T, \ f \in \tilde{X},$$

$$\big[\tilde{P}_t^{01*} h\big]_\theta = \bar{P}_{t,\theta}^{01*} h, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ h \in X,$$

$$\tilde{P}_t^{02} g = \int_{-\delta}^{0} \bar{P}_{t,\alpha}^{02} g_\alpha \, d\alpha, \ 0 \le t \le T, \ g \in \tilde{\mathbb{R}}^n,$$

$$\big[\tilde{P}_t^{02*} h\big]_\alpha = \bar{P}_{t,\alpha}^{02*} h, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \ h \in X,$$

$$\big[\tilde{P}_t^{11} f\big]_\theta = \int_{-\varepsilon}^{0} \bar{P}_{t,\theta,\tau}^{11} f_\tau \, d\tau, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ f \in \tilde{X},$$

$$\big[\tilde{P}_t^{12} g\big]_\theta = \int_{-\delta}^{0} \bar{P}_{t,\theta,\alpha}^{12} g_\alpha \, d\alpha, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ g \in \tilde{\mathbb{R}}^n,$$

$$\big[\tilde{P}_t^{12*} f\big]_\alpha = \int_{-\varepsilon}^{0} \bar{P}_{t,\theta,\alpha}^{12*} f_\theta \, d\theta, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \ f \in \tilde{X},$$

$$\big[\tilde{P}_t^{22} g\big]_\alpha = \int_{-\delta}^{0} \bar{P}_{t,\alpha,\sigma}^{22} g_\sigma \, d\sigma, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \ g \in \tilde{\mathbb{R}}^n,$$

where $\bar{P}^{01} \in L_\infty(\mathbf{T}, L_2(-\varepsilon, 0; \mathcal{L}_2(X)))$, $\bar{P}^{02} \in L_\infty(\mathbf{T}, L_2(-\delta, 0; \mathcal{L}_2(\mathbb{R}^n, X)))$, $\bar{P}^{11} \in L_\infty(\mathbf{T}, L_2([-\varepsilon, 0] \times [-\varepsilon, 0], \mathcal{L}_2(X)))$, $\bar{P}^{12} \in L_\infty(\mathbf{T}, L_2([-\varepsilon, 0] \times [-\delta, 0], \mathcal{L}_2(\mathbb{R}^n, X)))$ and $\bar{P}^{22} \in L_\infty(\mathbf{T}, L_2([-\delta, 0] \times [-\delta, 0], \mathcal{L}_2(\mathbb{R}^n)))$. We define the functions:

$$P_{t,\theta}^{01} = \int_{-\varepsilon}^{\theta} \bar{P}_{t,\theta_1}^{01} \, d\theta_1, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T,$$

$$P_{t,\alpha}^{02} = \int_{-\delta}^{\alpha} \bar{P}_{t,\alpha_1}^{02} \, d\alpha_1, \ -\delta \le \alpha \le 0, \ 0 \le t \le T,$$

$$P_{t,\theta,\tau}^{11} = \int_{-\varepsilon}^{\theta} \int_{-\varepsilon}^{\tau} \bar{P}_{t,\theta_1,\tau_1}^{11} \, d\tau_1 \, d\theta_1, \ -\varepsilon \le \theta \le 0, \ -\varepsilon \le \tau \le 0, \ 0 \le t \le T,$$

$$P_{t,\theta,\alpha}^{12} = \int_{-\varepsilon}^{\theta} \int_{-\delta}^{\alpha} \bar{P}_{t,\theta_1,\alpha_1}^{12} \, d\alpha_1 \, d\theta_1, \ -\varepsilon \le \theta \le 0, \ -\delta \le \alpha \le 0, \ 0 \le t \le T,$$

$$P_{t,\alpha,\sigma}^{22} = \int_{-\delta}^{\alpha} \int_{-\delta}^{\sigma} \bar{P}_{t,\alpha_1,\sigma_1}^{22} \, d\sigma_1 \, d\alpha_1, \ -\delta \le \alpha \le 0, \ -\delta \le \sigma \le 0, \ 0 \le t \le T.$$

Note that the above integrals are in the strong sense for operator-valued functions (see Section 2.4.4). Also, since $\tilde{P}_t^{11} \ge 0$ and $\tilde{P}_t^{22} \ge 0$, one can show that $\bar{P}_{t,\theta,\tau}^{11} = \bar{P}_{t,\tau,\theta}^{11*}$ and $\bar{P}_{t,\alpha,\sigma}^{22} = \bar{P}_{t,\sigma,\alpha}^{22*}$, and, consequently, $P_{t,\theta,\tau}^{11} = P_{t,\tau,\theta}^{11*}$ and $P_{t,\alpha,\sigma}^{22} = P_{t,\sigma,\alpha}^{22*}$.

In this section we will derive equations for the operator-valued functions P^{00}, P^{01}, P^{02}, P^{11}, P^{12} and P^{22}.

Lemma 8.13. *With the above notation,*

$$\begin{aligned}
\Gamma^1 \tilde{P}_t^{01*} h &= P_{t,0}^{01*} h, \ 0 \le t \le T, \ h \in X, \\
\tilde{P}_t^{01} \Gamma^{1*} h &= P_{t,0}^{01} h, \ 0 \le t \le T, \ h \in X, \\
\Gamma^2 \tilde{P}_t^{02*} h &= P_{t,0}^{02*} h, \ 0 \le t \le T, \ h \in X, \\
\tilde{P}_t^{02} \Gamma^{2*} h &= P_{t,0}^{02} h, \ 0 \le t \le T, \ h \in \mathbb{R}^n.
\end{aligned}$$

Proof. The first equality follows from

$$\Gamma^1 \tilde{P}_t^{01*} h = \int_{-\varepsilon}^{0} \bar{P}_{t,\theta}^{01*} h \, d\theta = \int_{-\varepsilon}^{0} \frac{\partial}{\partial \theta} P_{t,\theta}^{01*} h \, d\theta = P_{t,0}^{01*} h - P_{t,-\varepsilon}^{01*} h = P_{t,0}^{01*} h.$$

In a similar way the third equality can be proved. The second and fourth equalities are consequences of the first and third equalities. ☐

Lemma 8.14. *With the above notation,*

$$\begin{aligned}
\left[\tilde{P}_t^{11} \Gamma^{1*} h\right]_\theta &= \frac{\partial}{\partial \theta} P_{t,\theta,0}^{11} h, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ h \in X, \\
\left[\tilde{P}_t^{12*} \Gamma^{1*} h\right]_\alpha &= \frac{\partial}{\partial \alpha} P_{t,0,\alpha}^{12*} h, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \ h \in X, \\
\left[\tilde{P}_t^{12} \Gamma^{2*} h\right]_\theta &= \frac{\partial}{\partial \theta} P_{t,\theta,0}^{12} h, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ h \in \mathbb{R}^n, \\
\left[\tilde{P}_t^{22} \Gamma^{2*} h\right]_\alpha &= \frac{\partial}{\partial \alpha} P_{t,\alpha,0}^{22} h, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \ h \in \mathbb{R}^n.
\end{aligned}$$

Proof. We have

$$\Gamma^1 \tilde{P}_t^{11} f = \int_{-\varepsilon}^{0} \int_{-\varepsilon}^{0} \bar{P}_{t,\tau,\theta}^{11} f_\theta \, d\theta \, d\tau = \int_{-\varepsilon}^{0} \left(\frac{\partial}{\partial \theta} P_{t,0,\theta}^{11} \right) f_\theta \, d\theta,$$

where $f \in \tilde{X}$. Therefore, by Proposition 2.8(d), for $h \in X$, we obtain

$$\left[\tilde{P}_t^{11} \Gamma^{1*} h\right]_\theta = \left(\frac{\partial}{\partial \theta} P_{t,0,\theta}^{11} \right)^* h = \frac{\partial}{\partial \theta} P_{t,0,\theta}^{11*} h = \frac{\partial}{\partial \theta} P_{t,\theta,0}^{11} h,$$

which proves the first equality. Similarly, the other equalities can be proved. ☐

Lemma 8.15. *Let*

$$\begin{cases}
M_t = P_t^{00} C_t^* + P_{t,0}^{02} + \Phi_t \bar{R} \Psi_t^*, \ 0 \le t \le T, \\
M_{t,\theta}^1 = P_{t,\theta}^{01*} C_t^* + P_{t,\theta,0}^{12} + \bar{\Phi}_{t,\theta}^1 \bar{R} \Psi_t^*, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \\
M_{t,\alpha}^2 = P_{t,\alpha}^{02*} C_t^* + P_{t,\alpha,0}^{22} + \bar{\Phi}_{t,\alpha}^2 \bar{R} \Psi_t^*, \ -\delta \le \alpha \le 0, \ 0 \le t \le T.
\end{cases} \tag{8.59}$$

Then $M^1_{t,-\varepsilon} = 0$ *and* $M^2_{t,-\delta} = 0$ *for* $0 \le t \le T$, *and*

$$\tilde{M}_t h = M_t h, \text{ a.e. } t \in \mathbf{T},\ h \in \mathbb{R}^n,$$
$$\left[\tilde{M}^1_t h\right]_\theta = \frac{\partial}{\partial \theta} M^1_{t,\theta} h, \text{ a.e. } \theta \in [-\varepsilon, 0], \text{ a.e. } t \in \mathbf{T},\ h \in \mathbb{R}^n,$$
$$\left[\tilde{M}^2_t h\right]_\alpha = \frac{\partial}{\partial \alpha} M^2_{t,\alpha} h, \text{ a.e. } \alpha \in [-\delta, 0], \text{ a.e. } t \in \mathbf{T},\ h \in \mathbb{R}^n.$$

Proof. The first two equalities are clear. The third equality follows from (8.43) and Lemma 8.13. For the fourth equality, from (8.43), (8.30) and Lemma 8.14,

$$\begin{aligned}\left[\tilde{M}^1_t h\right]_\theta &= \left[\left(\tilde{P}^{01*}_t C^*_t + \tilde{P}^{12}_t \tilde{\Gamma}^{2*} + \tilde{\Phi}^1_t \bar{R} \Psi^*_t\right) h\right]_\theta \\ &= \frac{\partial}{\partial \theta}\left(P^{01*}_{t,\theta} C^*_t + P^{12}_{t,\theta,0} + \bar{\Phi}^1_{t,\theta} \bar{R} \Psi^*_t\right) h = \frac{\partial}{\partial \theta} M^1_{t,\theta} h.\end{aligned}$$

In a similar way the fifth equality can be proved. □

Theorem 8.16. *Assume that the conditions* $(\mathbf{E}^{\mathbf{b}}_{\mathbf{1}})$–$(\mathbf{E}^{\mathbf{b}}_{\mathbf{3}})$ *hold so that* $\mathcal{U} \in \mathcal{S}(X)$ *and* A *is the infinitesimal generator of* $\mathcal{U}$. *Let* M, M^1 *and* M^2 *be defined by* (8.59). *Then* $(P^{00}, P^{01}, P^{02}, P^{11}, P^{12}, P^{22})$ *is a unique solution of the system of equations*

$$\begin{aligned}&\frac{d}{dt} P^{00}_t - P^{00}_t A^* - A P^{00}_t - P^{01}_{t,0} - P^{01*}_{t,0} - \Phi_t \bar{W} \Phi^*_t \\ &\qquad + M_t \left(\Psi_t \bar{V} \Psi^*_t\right)^{-1} M^*_t = 0,\ P^{00}_0 = P_0,\ 0 < t \le T, \qquad (8.60)\\
&\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta}\right) P^{01}_{t,\theta} - A P^{01}_{t,\theta} - P^{11}_{t,0,\theta} - \Phi_t \bar{W} \bar{\Phi}^{1*}_{t,\theta} + M_t \left(\Psi_t \bar{V} \Psi^*_t\right)^{-1} M^{1*}_{t,\theta} = 0,\\ &\qquad P^{01}_{0,\theta} = P^{01}_{t,-\varepsilon} = 0,\ -\varepsilon \le \theta \le 0,\ 0 < t \le T, \qquad (8.61)\\
&\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \alpha}\right) P^{02}_{t,\alpha} - A P^{02}_{t,\alpha} - P^{12}_{t,0,\alpha} - \Phi_t \bar{W} \bar{\Phi}^{2*}_{t,\alpha} + M_t \left(\Psi_t \bar{V} \Psi^*_t\right)^{-1} M^{2*}_{t,\alpha} = 0,\\ &\qquad P^{02}_{0,\alpha} = P^{02}_{t,-\delta} = 0,\ -\delta \le \alpha \le 0,\ 0 < t \le T, \qquad (8.62)\\
&\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta} + \frac{\partial}{\partial \tau}\right) P^{11}_{t,\theta,\tau} - \bar{\Phi}^1_{t,\theta} \bar{W} \bar{\Phi}^{1*}_{t,\tau} + M^1_{t,\theta} \left(\Psi_t \bar{V} \Psi^*_t\right)^{-1} M^{1*}_{t,\tau} = 0,\\ &\qquad P^{11}_{0,\theta,\tau} = P^{11}_{t,-\varepsilon,\tau} = P^{11}_{t,\theta,-\varepsilon} = 0,\ -\varepsilon \le \theta \le 0,\ -\varepsilon \le \tau \le 0,\ 0 < t \le T, \qquad (8.63)\\
&\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta} + \frac{\partial}{\partial \alpha}\right) P^{12}_{t,\theta,\alpha} - \bar{\Phi}^1_{t,\theta} \bar{W} \bar{\Phi}^{2*}_{t,\alpha} + M^1_{t,\theta} \left(\Psi_t \bar{V} \Psi^*_t\right)^{-1} M^{2*}_{t,\alpha} = 0,\\ &\qquad P^{12}_{0,\theta,\alpha} = P^{12}_{t,-\varepsilon,\alpha} = P^{12}_{t,\theta,-\delta} = 0,\ -\varepsilon \le \theta \le 0,\ -\delta \le \alpha \le 0,\ 0 < t \le T, \qquad (8.64)\\
&\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \alpha} + \frac{\partial}{\partial \sigma}\right) P^{22}_{t,\alpha,\sigma} - \bar{\Phi}^2_{t,\alpha} \bar{W} \bar{\Phi}^{2*}_{t,\sigma} + M^2_{t,\alpha} \left(\Psi_t \bar{V} \Psi^*_t\right)^{-1} M^{2*}_{t,\sigma} = 0,\\ &\qquad P^{22}_{0,\alpha,\sigma} = P^{22}_{t,-\delta,\sigma} = P^{22}_{t,\alpha,-\delta} = 0,\ -\delta \le \alpha \le 0,\ -\delta \le \sigma \le 0,\ 0 < t \le T, \qquad (8.65)\end{aligned}$$

in the following sense: P^{00} *is a solution of the equation* (8.60) *in the scalar product sense; for all* $h^* \in D(A^*)$, P^{01} *and* P^{02} *satisfy, respectively,*

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial \theta}\right) P_{t,\theta}^{01*}h^* - P_{t,\theta}^{01*}A^*h^* - P_{t,0,\theta}^{11*}h^*$$
$$- \bar{\Phi}_{t,\theta}^{1}\bar{W}\Phi_t^*h^* + M_{t,\theta}^{1}\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^*h^* = 0,\ P_{0,\theta}^{01} = P_{t,-\varepsilon}^{01} = 0,$$

for a.e. $\theta \in [-\varepsilon, 0]$ *and a.e.* $t \in (0,T]$, *and*

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial \alpha}\right) P_{t,\alpha}^{02*}h^* - P_{t,\alpha}^{02*}A^*h^* - P_{t,0,\alpha}^{12*}h^*$$
$$- \bar{\Phi}_{t,\alpha}^{2}\bar{W}\Phi_t^*h^* + M_{t,\alpha}^{2}\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}M_t^*h^* = 0,\ P_{0,\alpha}^{02} = P_{t,-\delta}^{02} = 0,$$

for a.e $\alpha \in [-\delta, 0]$ *and a.e.* $t \in (0,T]$; P^{11}, P^{12} *and* P^{22} *satisfy* (8.63), (8.64) *and* (8.65), *respectively, for a.e.* $\theta, \tau \in [-\varepsilon, 0]$, *a.e.* $\alpha, \sigma \in [-\delta, 0]$ *and a.e.* $t \in (0,T]$.

Proof. We will use the brief notation $V_t = \Psi_t\bar{V}\Psi_t^*$. Derivation of the equations (8.63), (8.64) and (8.65) are similar. So, we will derive one of them, say, (8.64). For that, consider the equation (8.55). $\tilde{P}^{12}$ is a scalar product solution of this equation. Let $f \in D(d/d\theta)$ and let $g \in D(d/d\alpha)$. This means that $f \in \check{X}$ with $f_0 = 0$ and $g \in \check{\mathbb{R}}^n$ with $g_0 = 0$. Using Lemma 8.15, Propositions 2.8(d), 2.43 and (8.30)–(8.31), one can evaluate each term in (8.55) in the scalar product and obtain for the first term,

$$\begin{aligned}
\left\langle \frac{d}{dt}\tilde{P}_t^{12}g, f\right\rangle_{\check{X}} &= \int_{-\varepsilon}^{0}\left\langle \frac{\partial}{\partial t}\int_{-\delta}^{0}\bar{P}_{t,\theta,\alpha}^{12}g_\alpha\,d\alpha, f_\theta\right\rangle_X d\theta\\
&= \int_{-\varepsilon}^{0}\left\langle \frac{\partial}{\partial t}\int_{-\delta}^{0}\left(\frac{\partial^2}{\partial\theta\,\partial\alpha}P_{t,\theta,\alpha}^{12}\right)g_\alpha\,d\alpha, f_\theta\right\rangle_X d\theta\\
&= \int_{-\varepsilon}^{0}\left\langle \frac{\partial}{\partial t}\int_{-\delta}^{0}\frac{\partial}{\partial\theta}P_{t,\theta,\alpha}^{12}g'_\alpha\,d\alpha, f_\theta\right\rangle_X d\theta\\
&= \int_{-\varepsilon}^{0}\int_{-\delta}^{0}\left\langle \frac{\partial}{\partial t}P_{t,\theta,\alpha}^{12}g'_\alpha, f'_\theta\right\rangle_X d\alpha\,d\theta;
\end{aligned}$$

for the second term,

$$\begin{aligned}
-\left\langle \tilde{P}_t^{12}\frac{d}{d\alpha}g, f\right\rangle_{\check{X}} &= -\int_{-\varepsilon}^{0}\left\langle \int_{-\delta}^{0}\bar{P}_{t,\theta,\alpha}^{12}g'_\alpha\,d\alpha, f_\theta\right\rangle_X d\theta\\
&= -\int_{-\varepsilon}^{0}\int_{-\delta}^{0}\left\langle \left(\frac{\partial^2}{\partial\theta\,\partial\alpha}P_{t,\theta,\alpha}^{12}\right)g'_\alpha, f_\theta\right\rangle_X d\alpha\,d\theta\\
&= \int_{-\varepsilon}^{0}\int_{-\delta}^{0}\left\langle \left(\frac{\partial}{\partial\alpha}P_{t,\theta,\alpha}^{12}\right)g'_\alpha, f'_\theta\right\rangle_X d\alpha\,d\theta;
\end{aligned}$$

for the third term,

$$
\begin{aligned}
-\left\langle \tilde{P}_t^{12} g, \frac{d}{d\theta} f \right\rangle_{\tilde{X}} &= -\int_{-\varepsilon}^{0} \left\langle \int_{-\delta}^{0} \bar{P}_{t,\theta,\alpha}^{12} g_\alpha \, d\alpha, f'_\theta \right\rangle_X d\theta \\
&= -\int_{-\varepsilon}^{0} \int_{-\delta}^{0} \left\langle \left(\frac{\partial^2}{\partial\theta\,\partial\alpha} P_{t,\theta,\alpha}^{12} \right) g_\alpha, f'_\theta \right\rangle_X d\alpha\, d\theta \\
&= \int_{-\varepsilon}^{0} \int_{-\delta}^{0} \left\langle \frac{\partial}{\partial\theta} P_{t,\theta,\alpha}^{12} g'_\alpha, f'_\theta \right\rangle_X d\alpha\, d\theta;
\end{aligned}
$$

for the fourth term,

$$
\begin{aligned}
-\left\langle \tilde{\Phi}_t^1 \bar{W} \tilde{\Phi}_t^{2*} g, f \right\rangle_{\tilde{X}} &= -\int_{-\varepsilon}^{0} \left\langle \frac{\partial}{\partial\theta} \bar{\Phi}_{t,\theta}^1 \bar{W} \int_{-\delta}^{0} \left(\frac{\partial}{\partial\alpha} \bar{\Phi}_{t,\alpha}^{2*} \right) g_\alpha \, d\alpha, f_\theta \right\rangle_X d\theta \\
&= -\int_{-\varepsilon}^{0} \int_{-\delta}^{0} \langle \bar{\Phi}_{t,\theta}^1 \bar{W} \bar{\Phi}_{t,\alpha}^{2*} g'_\alpha, f'_\theta \rangle_X \, d\alpha\, d\theta;
\end{aligned}
$$

and, finally, for the last term,

$$
\begin{aligned}
\left\langle \tilde{M}_t^1 V^{-1} \tilde{M}_t^{2*} g, f \right\rangle_{\tilde{X}} &= \int_{-\varepsilon}^{0} \left\langle \frac{\partial}{\partial\theta} M_{t,\theta}^1 V_t^{-1} \int_{-\delta}^{0} \left(\frac{\partial}{\partial\alpha} M_{t,\alpha}^{2*} \right) g_\alpha \, d\alpha, f_\theta \right\rangle_X d\theta \\
&= \int_{-\varepsilon}^{0} \int_{-\delta}^{0} \langle M_{t,\theta}^1 V_t^{-1} M_{t,\alpha}^{2*} g'_\alpha, f'_\theta \rangle_X \, d\alpha\, d\theta.
\end{aligned}
$$

Substituting these expressions in the equation (8.55) and using the arbitrariness of h' and g' in $\tilde{X}$ and $\tilde{\mathbb{R}}^n$, respectively, we obtain that P^{12} satisfies the equation (8.64) for a.e. $\theta, \tau \in [-\varepsilon, 0]$ and for a.e. $t \in (0, T]$. The initial and boundary conditions for the equation (8.64) follow from the initial condition for the equation (8.55) and the definition of P^{12}.

The equations (8.61) and (8.62) can be derived in a similar manner as well and, hence, we will derive one of them, say, the equation (8.61). $\tilde{P}^{01}$ is a scalar product solution of the equation (8.52). Let $h^* \in D(A^*)$ and $f \in \check{X}$ with $f_0 = 0$. Using Lemmas 8.14 and 8.15, Propositions 2.8(d) and 2.43 and (8.30), one can evaluate each term in (8.52) in the scalar product and obtain for the first term,

$$
\begin{aligned}
\left\langle \frac{d}{dt} \tilde{P}_t^{01} f, h^* \right\rangle &= \left\langle \frac{\partial}{\partial t} \int_{-\varepsilon}^{0} \bar{P}_{t,\theta}^{01} f_\theta \, d\theta, h^* \right\rangle \\
&= \int_{-\varepsilon}^{0} \left\langle \left(\frac{\partial^2}{\partial t \partial\theta} P_{t,\theta}^{01} \right) f_\theta, h^* \right\rangle d\theta \\
&= \int_{-\varepsilon}^{0} \left\langle \frac{\partial}{\partial t} P_{t,\theta}^{01} f'_\theta, h^* \right\rangle d\theta \\
&= -\int_{-\varepsilon}^{0} \left\langle f'_\theta, \frac{\partial}{\partial t} P_{t,\theta}^{01*} h^* \right\rangle d\theta;
\end{aligned}
$$

for the second term,

$$\begin{aligned}-\left\langle \tilde{P}_t^{01}\frac{d}{d\theta}f,h^*\right\rangle &= -\left\langle \int_{-\varepsilon}^{0} \bar{P}_{t,\theta}^{01} f'_\theta\, d\theta, h^*\right\rangle \\ &= -\int_{-\varepsilon}^{0}\left\langle \left(\frac{\partial}{\partial\theta}P_{t,\theta}^{01}\right) f'_\theta, h^*\right\rangle d\theta \\ &= -\int_{-\varepsilon}^{0}\left\langle f'_\theta, \frac{\partial}{\partial\theta}P_{t,\theta}^{01*}h^*\right\rangle d\theta;\end{aligned}$$

for the third term,

$$\begin{aligned}-\left\langle \tilde{P}_t^{01}f, A^*h^*\right\rangle &= -\left\langle \int_{-\varepsilon}^{0} \bar{P}_{t,\theta}^{01} f_\theta\, d\theta, A^*h^*\right\rangle \\ &= -\int_{-\varepsilon}^{0}\left\langle \left(\frac{\partial}{\partial\theta}P_{t,\theta}^{01}\right) f_\theta, A^*h^*\right\rangle d\theta \\ &= \int_{-\varepsilon}^{0}\langle f'_\theta, P_{t,\theta}^{01*}A^*h^*\rangle\, d\theta;\end{aligned}$$

for the fourth term,

$$\begin{aligned}-\left\langle \tilde{\Gamma}^1\tilde{P}_t^{11}f, h^*\right\rangle &= -\left\langle \int_{-\varepsilon}^{0} \left(\frac{\partial}{\partial\theta}P_{t,\theta,0}^{11}\right)^* f_\theta\, d\theta, h^*\right\rangle \\ &= -\int_{-\varepsilon}^{0}\left\langle \left(\frac{\partial}{\partial\theta}P_{t,0,\theta}^{11}\right) f_\theta, h^*\right\rangle d\theta \\ &= \int_{-\varepsilon}^{0}\langle f'_\theta, P_{t,0,\theta}^{11*}h^*\rangle\, d\theta;\end{aligned}$$

for the fifth term,

$$\begin{aligned}-\left\langle \Phi\bar{W}\tilde{\Phi}^{1*}f, h^*\right\rangle &= -\left\langle \int_{-\varepsilon}^{0} \Phi_t\bar{W}\left(\frac{\partial}{\partial\theta}\bar{\Phi}_{t,\theta}^{1}\right)^* f_\theta\, d\theta, h^*\right\rangle \\ &= \int_{-\varepsilon}^{0}\langle f'_\theta, \bar{\Phi}_{t,\theta}^{1}\bar{W}\Phi_t^*h^*\rangle\, d\theta;\end{aligned}$$

and, finally, for the last term,

$$\begin{aligned}\left\langle \tilde{M}_tV_t^{-1}\tilde{M}_t^{1*}f, h^*\right\rangle &= \left\langle \int_{-\varepsilon}^{0} M_tV_t^{-1}\left(\frac{\partial}{\partial\theta}M_{t,\theta}^{1}\right)^* f_\theta\, d\theta, h^*\right\rangle \\ &= -\int_{-\varepsilon}^{0}\langle f'_\theta, M_{t,\theta}^{1}V_t^{-1}M_t^*h^*\rangle\, d\theta.\end{aligned}$$

Substituting the obtained expressions in the equation (8.52) and using the arbitrariness of f' in $\tilde{X}$, we obtain that the equation (8.61) is satisfied by P^{01} in the above mentioned sense. The initial and boundary conditions for the equation (8.61) follow from the initial condition for the equation (8.52) and the definition of P^{01}.

Finally, the equation (8.60) follows from the equation (8.51) and Lemmas 8.13 and 8.15. The uniqueness of a solution of the system (8.60)–(8.65) follows from the uniqueness of a solution of the system (8.51)–(8.56). □

8.2.2 Equations for the Optimal Filter

Below we present the equations for the optimal filter in the problem (8.1)–(8.3) in practically useful form.

Theorem 8.17. *Assume that the conditions* $(\mathbf{E}_1^b)$–$(\mathbf{E}_3^b)$ *hold so that* $\mathcal{U} \in \mathcal{S}(X)$ *and* A *is the infinitesimal generator of* $\mathcal{U}$. *Let* $(P^{00}, P^{01}, P^{02}, P^{11}, P^{12}, P^{22})$ *be the solution (in the sense that is defined in* Theorem 8.16) *of the system* (8.60)–(8.65) *and let* $\bar{\Phi}^1$ *and* $\bar{\Phi}^2$ *be defined by* (8.12)–(8.13). *Then the best estimate* $\hat{x}$ *in the filtering problem* (8.1)–(8.3) *together with* ψ^1 *and* ψ^2 *is a unique solution of the system of equations*

$$\begin{cases} d\hat{x}_t = A\hat{x}_t dt + \psi_{t,0}^1 dt + \left(P_t^{00}C_t^* + P_{t,0}^{02} + \Phi_t \bar{R}\Psi_t^*\right)\left(\Psi_t \bar{V}\Psi_t^*\right)^{-1} \\ \quad \times \left(dz_t - C_t\hat{x}_t dt - \psi_{t,0}^2 dt\right), \ \hat{x}_0 = 0, \ 0 < t \le T, \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta}^1 dt = \left(P_{t,\theta}^{01*}C_t^* + P_{t,\theta,0}^{12} + \bar{\Phi}_{t,\theta}^1 \bar{R}\Psi_t^*\right)\left(\Psi_t \bar{V}\Psi_t^*\right)^{-1} \\ \quad \times \left(dz_t - C_t\hat{x}_t dt - \psi_{t,0}^2 dt\right), \ \psi_{0,\theta}^1 = \psi_{t,-\varepsilon}^1 = 0, \ -\varepsilon \le \theta \le 0, \ 0 < t \le T, \\ (\partial/\partial t + \partial/\partial\alpha)\psi_{t,\alpha}^2 dt = \left(P_{t,\alpha}^{02*}C_t^* + P_{t,\alpha,0}^{22} + \bar{\Phi}_{t,\alpha}^2 \bar{R}\Psi_t^*\right)\left(\Psi_t \bar{V}\Psi_t^*\right)^{-1} \\ \quad \times \left(dz_t - C_t\hat{x}_t dt - \psi_{t,0}^2 dt\right), \ \psi_{0,\alpha}^2 = \psi_{t,-\delta}^2 = 0, \ -\delta \le \alpha \le 0, \ 0 < t \le T, \end{cases} \tag{8.66}$$

in the following sense: $\hat{x}$ *is a mild solution of the first equation in* (8.66) *and* ψ^1 *and* ψ^2 *are ordinary solutions of the second and third equations in* (8.66), *respectively.*

Proof. We will derive the equations in (8.66) from the equations in (8.50). For this, we define the random processes

$$\psi_{t,\theta}^1 = \int_{-\varepsilon}^{\theta} [\tilde{\psi}_t^1]_\tau d\tau, \ \theta \in [-\varepsilon, 0], \ 0 < t \le T,$$

$$\psi_{t,\alpha}^2 = \int_{-\delta}^{\alpha} [\tilde{\psi}_t^2]_\sigma d\sigma, \ \alpha \in [-\delta, 0], \ 0 < t \le T,$$

where $\tilde{\psi}^1$ and $\tilde{\psi}^2$ are solutions of the second and third equations in (8.50), and show that $(\hat{x}, \psi^1, \psi^2)$ is a solution of the system (8.66) in the above mentioned sense. It is clear that

$$\Gamma^1\tilde{\psi}_t^1 = \int_{-\varepsilon}^{0} \frac{\partial}{\partial\theta}\psi_{t,\theta}^1 \, d\theta = \psi_{t,0}^1 - \psi_{t,-\varepsilon}^1 = \psi_{t,0}^1.$$

In a similar way, $\Gamma^2\tilde{\psi}_t^2 = \psi_{t,0}^2$. So, using Lemma 8.13, we obtain the first equation in (8.66) from the first equation in (8.50). Derivation of the second and third equations in (8.66) are similar and we will derive only one of them, say, the second equation in (8.66). For brevity, we will use the innovation process $\bar{z}$, defined in Theorem 8.9, and $\tilde{M}^1$, defined by (8.43). By (8.5) and Lemma 8.15, for $h \in \mathbb{R}^n$,

$$\left[\mathcal{T}^1_{t-s}\tilde{M}^1_s h\right]_\theta = \left\{\begin{matrix} \left[\tilde{M}^1_s h\right]_{s-t+\theta}, & s-t+\theta \geq -\varepsilon \\ 0, & s-t+\theta < -\varepsilon \end{matrix}\right\}$$
$$= \left\{\begin{matrix} (\partial/\partial\theta) M^1_{s,s-t+\theta} h, & s-t+\theta \geq -\varepsilon \\ 0, & s-t+\theta < -\varepsilon \end{matrix}\right\}.$$

Since $\tilde{\psi}^1$ is a mild solution of the second equation in (8.50),

$$\left[\tilde{\psi}^1_t\right]_\theta = \left[\int_0^t \mathcal{T}^1_{t-s}\tilde{M}^1_s \left(\Psi_s \bar{V}\Psi^*_s\right)^{-1} d\bar{z}_s\right]_\theta$$
$$= \int_{\max(0,t-\varepsilon-\theta)}^t \frac{\partial}{\partial\theta} M^1_{s,s-t+\theta}\left(\Psi_s \bar{V}\Psi^*_s\right)^{-1} d\bar{z}_s.$$

Hence,

$$\psi^1_{t,\theta} = \int_{-\varepsilon}^\theta \left[\tilde{\psi}^1_t\right]_\tau d\tau = \int_{-\varepsilon}^\theta \int_{\max(0,t-\varepsilon-\tau)}^t \frac{\partial}{\partial\tau} M^1_{s,s-t+\tau}\left(\Psi_s \bar{V}\Psi^*_s\right)^{-1} d\bar{z}_s\, d\tau$$
$$= \int_{\max(0,t-\varepsilon-\theta)}^t \int_{t-s-\varepsilon}^\theta \frac{\partial}{\partial\tau} M^1_{s,s-t+\tau}\left(\Psi_s \bar{V}\Psi^*_s\right)^{-1} d\tau\, d\bar{z}_s$$
$$= \int_{\max(0,t-\varepsilon-\theta)}^t M^1_{s,s-t+\theta}\left(\Psi_s \bar{V}\Psi^*_s\right)^{-1} d\bar{z}_s$$
$$= \left[\int_0^t \mathcal{T}^1_{t-s}\left[M^1_s\right]\left(\Psi_s \bar{V}\Psi^*_s\right)^{-1} d\bar{z}_s\right]_\theta,$$

where we refer to Remark 2.21 for the notation $\left[M^1_s\right]$. We see that ψ^1 is a mild solution of the equation

$$d\left[\psi^1_t\right] = (-d/d\theta)\left[\psi^1_t\right] dt + \left[M^1_t\right]\left(\Psi_t \bar{V}\Psi^*_t\right)^{-1} d\bar{z}_t, \tag{8.67}$$

which in turn is the same as the second equation in (8.66). Also, one can see that for the equation (8.67), the conditions of Theorem 4.27 hold. Indeed, by Lemma 8.15, $M^1_{t,-\varepsilon} = 0$ and $(\partial/\partial\theta)M^1_{t,\theta}h = \left[\tilde{M}^1_t h\right]_\theta$ for all $h \in \mathbb{R}^n$. So, $\left[\tilde{M}^1\right]$, $0 \leq t \leq T$, is a $D(-d/d\theta)$-valued function (the space $D(-d/d\theta)$ is defined in Theorem 1.22). Using Proposition 1.15(a), one can obtain that $M \in B_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n, D(-d/d\theta)))$. Thus, by Theorem 4.27, the equation (8.67) (and, respectively, the second equation in (8.66)) has ψ^1 as its unique ordinary solution. The initial and boundary conditions in (8.66) follow from the initial conditions in (8.50) and the definitions of ψ^1 and ψ^2. □

Remark 8.18. If A is taken from $\mathcal{L}(X)$, then one can easily show that all the solutions of the equations (8.60)–(8.65) and (8.66) are in the ordinary sense.

8.3 Stochastic Regulator Problem

8.3.1 Setting of the Problem

Consider the problem (5.1)–(5.4) in which the state-observation system (5.1)–(5.2) and the functional (5.4) are defined in the form

$$x_t^u = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}(B_s u_s + \varphi_s^1)ds + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s,\ 0 \le t \le T, \tag{8.68}$$

$$z_t^u = \int_0^t (C_s x_s^u + \varphi_s^2)ds + \int_0^t \Psi_s\,dv_s,\ 0 \le t \le T, \tag{8.69}$$

$$J(u) = \mathbf{E}\left(\langle x_T^u, Q_T x_T^u\rangle + \int_0^T \left\langle \begin{bmatrix} x_t^u \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix}\begin{bmatrix} x_t^u \\ u_t \end{bmatrix}\right\rangle dt\right), \tag{8.70}$$

where φ^1 and φ^2 are defined by (8.3) and a control u is taken from the set of admissible controls U_{ad} as defined by (5.3) in Section 5.1.2. This problem will be called the linear stochastic regulator problem (8.68)–(8.70). The problem (8.68)–(8.70) differs from the similar problems (6.51)–(6.53) and (7.35)–(7.37) because the noise processes φ^1 and φ^2 are wide band.

In this section the following conditions are supposed to hold:

$(\mathbf{R_1^b})$ $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$, $B \in B_\infty(\mathbf{T}, \mathcal{L}(U, X))$;

$(\mathbf{R_2^b})$ $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Phi^1 \in L_\infty(\mathbf{T}, B_2(-\varepsilon, 0; \mathcal{L}(H, X)))$, $0 < \varepsilon < T$, $\Phi^2 \in L_\infty(\mathbf{T}, B_2(-\delta, 0; \mathcal{L}(H, \mathbb{R}^n)))$, $0 < \delta < T$, the operator-valued functions Φ^1, Φ^{1*}, Φ^2 and Φ^{2*} are strongly differentiable in each of the variables with

$$\frac{\partial}{\partial t}\Phi^1,\ \frac{\partial}{\partial \theta}\Phi^1 \in L_\infty(\mathbf{T}, B_2(-\varepsilon, 0; \mathcal{L}(H, X))),$$
$$\frac{\partial}{\partial t}\Phi^2,\ \frac{\partial}{\partial \alpha}\Phi^2 \in L_\infty(\mathbf{T}, B_2(-\delta, 0; \mathcal{L}(H, \mathbb{R}^n)))$$

and $\Phi^1_{t,-\varepsilon} = 0$ and $\Phi^2_{t,-\delta} = 0$ for all $0 \le t \le T$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with $\operatorname{cov} v_T > 0$, x_0 is a an X-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0 and (w, v) are independent;

$(\mathbf{R_3^b})$ $Q_T \in \mathcal{L}(X)$, $Q_T \ge 0$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$, $L \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \ge 0$ for a.e. $t \in \mathbf{T}$.

Note that $(\mathbf{R_1^b})$ is the same as $(\mathbf{E_1^b})$ completed with the condition about B, $(\mathbf{R_2^b})$ is the combination of $(\mathbf{E_2^b})$ and $(\mathbf{E_3^b})$, and $(\mathbf{R_3^b})$ is the same as $(\mathbf{C_3})$. We will also use the notation of Sections 8.1 and 8.2.

8.3.2 Reduction

Let

$$\tilde{x}_t^u = \begin{bmatrix} x_t^u \\ \tilde{\varphi}_t^1 \\ \tilde{\varphi}_t^2 \end{bmatrix}, \quad 0 \le t \le T, \tag{8.71}$$

where $\tilde{\varphi}^1$ and $\tilde{\varphi}^2$ are defined by (8.26) and (8.27), respectively. Similar to Section 8.1.3, for $\tilde{\mathcal{U}}$, $\tilde{\Phi}$, $\tilde{C}$ defined by (8.30)–(8.36), (8.39), and for

$$\tilde{B} = \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(U, X \times \tilde{X} \times \tilde{\mathbb{R}}^n)), \tag{8.72}$$

where $\tilde{X}$ and $\tilde{\mathbb{R}}^n$ are defined by (8.4), the processes $\tilde{x}^u$ and z^u can be represented as

$$\tilde{x}_t^u = \tilde{\mathcal{U}}_{t,0}\tilde{x}_0 + \int_0^T \tilde{\mathcal{U}}_{t,s}\tilde{B}_s u_s\, ds + \int_0^t \tilde{\mathcal{U}}_{t,s}\tilde{\Phi}_s\, dw_s, \quad 0 \le t \le T, \tag{8.73}$$

$$z_t^u = \int_0^t \tilde{C}_s \tilde{x}_s^u\, ds + \int_0^t \Psi_s\, dv_s, \quad 0 \le t \le T. \tag{8.74}$$

Also, the functional (8.70) can be written as

$$J(u) = \mathbf{E}\left(\langle \tilde{x}_T^u, \tilde{Q}_T \tilde{x}_T^u \rangle + \int_0^T \left\langle \begin{bmatrix} \tilde{x}_t^u \\ u_t \end{bmatrix}, \begin{bmatrix} \tilde{F}_t & \tilde{L}_t^* \\ \tilde{L}_t & G_t \end{bmatrix} \begin{bmatrix} \tilde{x}_t^u \\ u_t \end{bmatrix} \right\rangle dt \right), \tag{8.75}$$

where

$$\tilde{Q}_T = \begin{bmatrix} Q_T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n), \tag{8.76}$$

$$\tilde{F} = \begin{bmatrix} F & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n)), \tag{8.77}$$

$$\tilde{L} = \begin{bmatrix} L & 0 & 0 \end{bmatrix} \in B_\infty(\mathbf{T}, \mathcal{L}(X \times \tilde{X} \times \tilde{\mathbb{R}}^n, U)). \tag{8.78}$$

Thus, the problem (8.68)–(8.70) is reduced to the linear stochastic regulator problem (8.73)–(8.75). This we state in the following form.

Lemma 8.19. *The functional* (8.70), *subject to* (8.68)–(8.69), *and the functional* (8.75), *subject to* (8.73)–(8.74), *are the same on* U_{ad} *as defined by* (5.3).

Proof. This follows from (8.71). □

8.3.3 Optimal Stochastic Regulator

Theorem 8.20. *Under the conditions* $(\mathbf{R_1^b})$–$(\mathbf{R_3^b})$, *there exists a unique optimal stochastic regulator in the problem* (8.73)–(8.75) *and the respective optimal control has the form*

$$u_t^* = -G_t^{-1}\big(\tilde{B}_t^*\tilde{Q}_t + \tilde{L}_t\big)\hat{\tilde{x}}_t^*, \ \text{a.e. } t \in \mathbf{T}, \tag{8.79}$$

where

$$\hat{\tilde{x}}_t^* = \int_0^t \tilde{\mathcal{R}}_{t,s}\big(\tilde{P}_s\tilde{C}_s^* + \tilde{\Phi}_s R\Psi_s^*\big)\big(\Psi_s\bar{V}\Psi_s^*\big)^{-1}\,dz_s^*, \ 0 \le t \le T, \tag{8.80}$$

$\tilde{\mathcal{R}} = \mathcal{P}_{-\tilde{B}G^{-1}(\tilde{B}^*\tilde{Q}+\tilde{L})-(\tilde{P}\tilde{C}^*+\tilde{\Phi}R\Psi^*)(\Psi\bar{V}\Psi^*)^{-1}\tilde{C}}(\tilde{\mathcal{U}})$, $\tilde{P}$ *is a unique solution of the Riccati equation* (8.40), $\tilde{Q}$ *is a unique solution of the Riccati equation*

$$\tilde{Q}_t = \tilde{\mathcal{U}}_{T,t}^*\tilde{Q}_T\tilde{\mathcal{U}}_{T,t} + \int_t^T \tilde{\mathcal{U}}_{s,t}^*\Big(\tilde{F}_s - \Big(\tilde{Q}_s\tilde{B}_s + \tilde{L}_s^*\Big)G_s^{-1}\Big(\tilde{B}_s^*\tilde{Q}_s + \tilde{L}_s\Big)\Big)\tilde{\mathcal{U}}_{s,t}\,ds, \ 0 \le t \le T, \tag{8.81}$$

$\tilde{\mathcal{U}}$, $\tilde{B}$, $\tilde{C}$, $\tilde{\Phi}$, $\tilde{Q}_T$, $\tilde{F}$ *and* $\tilde{L}$ *are defined by* (8.30)–(8.36), (8.72), (8.39) *and* (8.76)–(8.78).

Proof. This follows from Theorem 6.20 and Lemma 8.19. □

8.3.4 About the Riccati Equation (8.81)

Proposition 8.21. *Suppose the conditions* $(\mathbf{R_1^b})$ *and* $(\mathbf{R_3^b})$ *hold. Then the solution* $\tilde{Q}$ *of the equation* (8.81) *can be decomposed as*

$$\tilde{Q}_t = \begin{bmatrix} Q_t^{00} & \tilde{Q}_t^{01} & 0 \\ \tilde{Q}_t^{01*} & \tilde{Q}_t^{11} & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{L}\big(X \times \tilde{X} \times \tilde{\mathbb{R}}^n\big), \ 0 \le t \le T, \tag{8.82}$$

where $\big(Q^{00}, \tilde{Q}^{01}, \tilde{Q}^{11}\big)$ *is a unique solution of the system of equations*

$$Q_t^{00} = \mathcal{U}_{T,t}^*Q_T\mathcal{U}_{T,t} + \int_t^T \mathcal{U}_{s,t}^*\big(F_s - (Q_s^{00}B_s + L_s^*)G_s^{-1}(B_s^*Q_s^{00} + L_s)\mathcal{U}_{s,t}\,ds, \ 0 \le t \le T, \tag{8.83}$$

$$\tilde{Q}_t^{01} = \int_t^T \mathcal{U}_{s,t}^*(Q_s^{00}\Gamma^1 - (Q_s^{00}B_s + L_s^*)G_s^{-1}B_s^*\tilde{Q}_s^{01})\mathcal{T}_{s-t}^1\,ds, \ 0 \le t \le T, \tag{8.84}$$

$$\tilde{Q}_t^{11} = \int_t^T \mathcal{T}_{s-t}^{1*}(\Gamma^{1*}\tilde{Q}_s^{01} + \tilde{Q}_s^{01*}\Gamma^1 - \tilde{Q}_s^{01*}B_sG_s^{-1}B_s^*\tilde{Q}_s^{01})\mathcal{T}_{s-t}^1\,ds, \ 0 \le t \le T. \tag{8.85}$$

Proof. This can be proved in a similar way as Proposition 7.8. □

It turns out that the solutions $\tilde{Q}^{01}$ and $\tilde{Q}^{11}$ of the equations (8.84) and (8.85) can be explicitly expressed through the solution Q^{00} of the equation (8.83). This is shown in the following two propositions.

Proposition 8.22. *Suppose the conditions* $(\mathbf{R_1^b})$ *and* $(\mathbf{R_3^b})$ *hold. Then the component* $\tilde{Q}^{01}$ *of* $\tilde{Q}$ *in the decomposition* (8.82) *has the representation*

$$\tilde{Q}_t^{01} f = \int_{-\varepsilon}^{0} \bar{Q}_{t,\theta}^{01} f_\theta \, d\theta, \ 0 \le t \le T, \ f \in \tilde{X},$$

where

$$\bar{Q}_{t,\theta}^{01} = \int_t^{\min(T,t-\theta)} \mathcal{Y}_{s,t}^* Q_s^{00} \, ds, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T. \tag{8.86}$$

Furthermore,

$$\frac{\partial}{\partial \theta} \bar{Q}_{t,\theta}^{01} = -Q_{t,\theta}^{01} \text{ and } \bar{Q}_{t,0}^{01} = 0, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T,$$

where

$$Q_{t,\theta}^{01} = \left\{ \begin{matrix} \mathcal{Y}_{t-\theta,t}^* Q_{t-\theta}^{00}, & \theta \ge t - T \\ 0, & \theta < t - T \end{matrix} \right\}, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \tag{8.87}$$

$\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^* Q^{00} + L)}(\mathcal{U})$ *and* Q^{00} *is a solution of the equation* (8.83).

Proof. In a similar way as in the proof of Proposition 7.9 one can obtain

$$\tilde{Q}_t^{01} = \int_t^T \mathcal{Y}_{s,t}^* Q_s^{00} \Gamma^1 \mathcal{T}_{s-t}^1 \, ds.$$

If $f \in \tilde{X}$, then

$$\begin{aligned} \tilde{Q}_t^{01} f &= \int_t^T \mathcal{Y}_{s,t}^* Q_s^{00} \Gamma^1 \mathcal{T}_{s-t}^1 f \, ds \\ &= \int_t^T \mathcal{Y}_{s,t}^* Q_s^{00} \int_{\min(0,s-t-\varepsilon)}^{0} f_{\theta-s+t} \, d\theta \, ds \\ &= \int_t^T \mathcal{Y}_{s,t}^* Q_s^{00} \int_{\min(-\varepsilon,t-s)}^{t-s} f_\theta \, d\theta \, ds \\ &= \int_{-\varepsilon}^{0} \left(\int_t^{\min(T,t-\theta)} \mathcal{Y}_{s,t}^* Q_s^{00} \, ds \right) f_\theta \, d\theta, \end{aligned}$$

which implies the conclusions of the proposition. □

Proposition 8.23. *Suppose the conditions* $(\mathbf{R_1^b})$ *and* $(\mathbf{R_3^b})$ *hold. Then the component* $\tilde{Q}^{11}$ *of* $\tilde{Q}$ *in the decomposition* (8.82) *has the representation*

$$[\tilde{Q}_t^{11} f]_\theta = \int_{-\varepsilon}^{0} \bar{Q}_{t,\theta,\tau}^{11} f_\tau \, d\tau, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \ f \in \tilde{X},$$

where

$$\bar{Q}_{t,\theta,\tau}^{11} = \int_t^{\min(T,t-\theta,t-\tau)} \left(\bar{Q}_{s,\tau+s-t}^{01} + \bar{Q}_{s,\theta+s-t}^{01*} - \bar{Q}_{s,\theta+s-t}^{01*} B_s G_s^{-1} B_s^* \bar{Q}_{s,\tau+s-t}^{01}\right) ds,$$
$$-\varepsilon \le \theta \le 0, \ -\varepsilon \le \tau \le 0, \ 0 \le t \le T,$$

and $\bar{Q}^{01}$ *is as defined by* (8.86). *Furthermore,*

$$\frac{\partial^2}{\partial\theta\partial\tau} \bar{Q}_{t,\theta,\tau}^{11} = Q_{t,\theta,\tau}^{11} \ \text{and} \ \bar{Q}_{t,\theta,0}^{11} = \bar{Q}_{t,0,\tau}^{11} = 0, \ -\varepsilon \le \theta \le 0, \ -\varepsilon \le \tau \le 0, \ 0 \le t \le T,$$

where

$$Q_{t,\theta,\tau}^{11} = \begin{Bmatrix} Q_{t-\theta,\tau-\theta}^{01}, & \theta \ge \tau \\ Q_{t-\tau,\theta-\tau}^{01*}, & \theta < \tau \end{Bmatrix} - \int_t^{\min(t-\theta,t-\tau)} Q_{s,\theta+s-t}^{01*} B_s G_s^{-1} B_s^* Q_{s,\tau+s-t}^{01} \, ds,$$
$$-\varepsilon \le \theta \le 0, \ -\varepsilon \le \tau \le 0, \ 0 \le t \le T, \tag{8.88}$$

and Q^{01} *is defined by* (8.87).

Proof. Using Proposition 8.22, for $f \in \tilde{X}$, we obtain

$$\begin{aligned}
[\tilde{Q}_s^{01} \mathcal{T}_{s-t}^1 f]_\tau &= \int_{-\varepsilon}^{0} \bar{Q}_{s,\tau}^{01} [\mathcal{T}_{s-t}^1 f]_\tau d\tau \\
&= \int_{\min(0,s-t-\varepsilon)}^{0} \bar{Q}_{s,\tau}^{01} f_{\tau-s+t} \, d\tau \\
&= \int_{\min(-\varepsilon,t-s)}^{t-s} \bar{Q}_{s,\tau+s-t}^{01} f_\tau \, d\tau.
\end{aligned}$$

Hence,

$$\begin{aligned}
\left[\int_t^T \mathcal{T}_{s-t}^{1*} \Gamma^{1*} \tilde{Q}_s^{01} \mathcal{T}_{s-t}^1 f \, ds\right]_\theta &= \int_t^{\min(T,t-\theta)} \int_{\min(-\varepsilon,t-s)}^{t-s} \bar{Q}_{s,\tau+s-t}^{01} f_\tau \, d\tau \, ds \\
&= \int_{-\varepsilon}^{0} \int_t^{\min(T,t-\theta,t-\tau)} \bar{Q}_{s,\tau+s-t}^{01} f_\tau \, ds \, d\tau.
\end{aligned}$$

In a similar way one can obtain

$$\left[\int_t^T \mathcal{T}_{s-t}^{1*} \tilde{Q}_s^{01*} \Gamma^1 \mathcal{T}_{s-t}^1 f \, ds\right]_\theta = \int_{-\varepsilon}^{0} \int_t^{\min(T,t-\theta,t-\tau)} \bar{Q}_{s,\theta+s-t}^{01*} f_\tau \, ds \, d\tau$$

and

$$\left[\int_t^T \mathcal{T}_{s-t}^{1*}\tilde{Q}_s^{01*}B_sG_s^{-1}B_s^*\tilde{Q}_s^{01}\mathcal{T}_{s-t}^1 f\,ds\right]_\theta$$
$$= \int_{-\varepsilon}^0 \int_t^{\min(T,t-\theta,t-\tau)} \bar{Q}_{s,\theta+s-t}^{01*}B_sG_s^{-1}B_s^*\bar{Q}_{s,\tau+s-t}^{01}f_\tau\,ds\,d\tau.$$

Thus, using these equalities in (8.85), we obtain the integral representation for $\tilde{Q}^{11}$ with $\bar{Q}^{11}$ as defined in the proposition. Clearly, $\bar{Q}_{t,\theta,0}^{11} = \bar{Q}_{t,0,\tau}^{11} = 0$. By Propositions 2.41 and 2.42, we have

$$\frac{\partial}{\partial\theta}\bar{Q}_{t,\theta,\tau}^{11} = \int_t^{\min(T,t-\theta,t-\tau)} \frac{\partial}{\partial\theta}\bar{Q}_{s,\theta+s-t}^{01*}\left(I - B_sG_s^{-1}B_s^*\bar{Q}_{s,\tau+s-t}^{01}\right)ds$$
$$-\begin{Bmatrix} \bar{Q}_{t-\theta,\tau-\theta}^{01}, & \theta\geq\tau \text{ and } \theta\geq t-T \\ 0, & \text{otherwise} \end{Bmatrix}$$

and

$$\frac{\partial^2}{\partial\theta\partial\tau}\bar{Q}_{t,\theta,\tau}^{11} = -\int_t^{\min(T,t-\theta,t-\tau)} \frac{\partial}{\partial\theta}\bar{Q}_{s,\theta+s-t}^{01*}B_sG_s^{-1}B_s^*\frac{\partial}{\partial\tau}\bar{Q}_{s,\tau+s-t}^{01}\,ds$$
$$-\begin{Bmatrix} (\partial/\partial\tau)\bar{Q}_{t-\theta,\tau-\theta}^{01} & \theta\geq\tau \text{ and } \theta\geq t-T \\ (\partial/\partial\theta)\bar{Q}_{t-\tau,\theta-\tau}^{01*} & \theta<\tau \text{ and } \tau\geq t-T \\ 0, & \text{otherwise} \end{Bmatrix}.$$

Finally, by Proposition 8.22, we obtain that $(\partial^2/\partial\theta\partial\tau)\bar{Q}_{t,\theta,\tau}^{11} = Q_{t,\theta,\tau}^{11}$. □

Proposition 8.24. *Suppose the conditions* $(\mathbf{R}_1^\mathbf{b})$–$(\mathbf{R}_3^\mathbf{b})$ *hold. Then the minimum of the functional J in the problem* (8.68)–(8.70) *is equal to*

$$J(u^*) = \operatorname{tr}(Q_T^{00}P_T^{00}) + \operatorname{tr}\int_0^T \left(F_tP_t^{00} + \bar{V}^{-1}\Psi_t^{-1}M_t^*Q_t^{00}M_t\Psi_t^{-1*}\right)dt$$
$$+ \operatorname{tr}\int_0^T\int_{-\varepsilon}^0 \bar{V}^{-1}\Psi_t^{-1}\left(M_t^*Q_{t,\theta}^{01}M_{t,\theta}^1 + M_{t,\theta}^{1*}Q_{t,\theta}^{01*}M_t\right)\Psi_t^{-1*}\,d\theta\,dt$$
$$+ \operatorname{tr}\int_0^T\int_{-\varepsilon}^0\int_{-\varepsilon}^0 \bar{V}^{-1}\Psi_t^{-1}M_{t,\theta}^{1*}Q_{t,\theta,\tau}^{11}M_{t,\tau}^1\Psi_t^{-1*}\,d\tau\,d\theta\,dt,$$

where P^{00} *is defined by* (8.51)–(8.56), Q^{00}, Q^{01} *and* Q^{11} *are defined by* (8.83), (8.87) *and* (8.88), *respectively, and* M *and* M^1 *are defined by* (8.59).

Proof. From Proposition 6.22, one can obtain

$$J(u^*) = \operatorname{tr}(Q_T^{00}P_T^{00}) + \operatorname{tr}\int_0^T F_tP_t^{00}\,dt$$
$$+ \operatorname{tr}\int_0^T \bar{V}^{-1}\Psi_t^{-1}\begin{bmatrix}\tilde{M}_t \\ \tilde{M}_t^1\end{bmatrix}^*\begin{bmatrix}Q_t^{00} & \tilde{Q}_t^{01} \\ \tilde{Q}_t^{01*} & \tilde{Q}_t^{11}\end{bmatrix}\begin{bmatrix}\tilde{M}_t \\ \tilde{M}_t^1\end{bmatrix}\Psi_t^{-1*}\,dt.$$

In view of Propositions 8.22, 8.23, 2.39 and 2.43 and Lemma 8.15, the above equality implies the required expression for $J(u^*)$. □

8.3.5 Example: Optimal Stochastic Regulator in Differential Form

Example 8.25. Assume that the conditions $(\mathbf{R}_1^b)$–$(\mathbf{R}_3^b)$ hold so that $\mathcal{U} \in \mathcal{S}(X)$ and A is the infinitesimal generator of $\mathcal{U}$. Then the state-observation system (8.68)–(8.69), under $u = u^*$, can be written in the differential form

$$\begin{cases} dx_t^* = \big(Ax_t^* + \varphi_t^1 + B_t u_t^*\big)dt + \Phi_t dw_t, \ x_0^* = x_0, \ 0 < t \le T, \\ dz_t^* = \big(C_t x_t^* + \varphi_t^2\big)dt + \Psi_t dv_t, \ z_0^* = 0, \ 0 < t \le T. \end{cases}$$

Denote

$$\begin{cases} \varphi_{t,\theta}^1 = \big[\breve{\varphi}_t^1\big]_\theta, \ -\varepsilon \le \theta \le 0, \ 0 \le t \le T, \\ \varphi_{t,\alpha}^2 = \big[\breve{\varphi}_t^2\big]_\alpha, \ -\delta \le \alpha \le 0, \ 0 \le t \le T, \end{cases}$$

where $\breve{\varphi}^1$ and $\breve{\varphi}^2$ are defined by (8.10) and (8.11), respectively. Then by Lemmas 8.1 and 8.2, this system can be written as

$$\begin{cases} dx_t^* = \big(Ax_t^* + \varphi_{t,0}^1 + B_t u_t^*\big)dt + \Phi_t dw_t, \ x_0^* = x_0, \ 0 < t \le T, \\ dz_t^* = \big(C_t x_t^* + \varphi_{t,0}^2\big)dt + \Psi_t dv_t, \ z_0^* = 0, \ 0 < t \le T, \\ (\partial/\partial t + \partial/\partial\theta)\varphi_{t,\theta}^1 dt = \bar{\Phi}_{t,\theta}^1 dw_t, \ \varphi_{0,\theta}^1 = \varphi_{t,-\varepsilon}^1 = 0, \ -\varepsilon \le \theta \le 0, \ 0 < t \le T, \\ (\partial/\partial t + \partial/\partial\alpha)\varphi_{t,\alpha}^2 dt = \bar{\Phi}_{t,\alpha}^2 dw_t, \ \varphi_{0,\alpha}^2 = \varphi_{t,-\delta}^1 = 0, \ -\delta \le \alpha \le 0, \ 0 < t \le T, \end{cases} \tag{8.89}$$

where $\bar{\Phi}^1$ and $\bar{\Phi}^2$ are defined by (8.12)–(8.13). By Theorem 8.20, the optimal control u^* in the problem (8.68)–(8.70) has the form

$$u_t^* = -G_t^{-1}\big(B_t^* Q_t^{00} + L_t\big)\hat{x}_t^* - G_t^{-1} B_t^* \tilde{Q}_t^{01} \tilde{\psi}_t^1, \ \text{a.e. } t \in \mathbf{T}. \tag{8.90}$$

Note that, by Proposition 8.22,

$$\begin{aligned} \tilde{Q}_t^{01}\tilde{\psi}_t^1 &= \int_{-\varepsilon}^0 \bar{Q}_{t,\theta}^{01}\big[\tilde{\psi}_t^1\big]_\theta \, d\theta \\ &= \int_{-\varepsilon}^0 \bar{Q}_{t,\theta}^{01} \frac{\partial}{\partial\theta}\psi_{t,\theta}^1 \, d\theta \\ &= -\int_{-\varepsilon}^0 \left(\frac{\partial}{\partial\theta}\bar{Q}_{t,\theta}^{01}\right)\psi_{t,\theta}^1 \, d\theta \\ &= \int_{\max(-\varepsilon, t-T)}^0 \mathcal{Y}_{t-\theta,t}^* Q_{t-\theta}^{00} \psi_{t,\theta}^1 \, d\theta, \end{aligned}$$

where $\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q^{00}+L)}(\mathcal{U})$. So, (8.90) can also be written in the form

$$\begin{aligned} u_t^* = &- G_t^{-1}\big(B_t^* Q_t^{00} + L_t\big)\hat{x}_t^* \\ &- G_t^{-1} B_t^* \int_{\max(-\varepsilon, t-T)}^0 \mathcal{Y}_{t-\theta,t}^* Q_{t-\theta}^{00} \psi_{t,\theta}^1 \, d\theta, \ \text{a.e. } t \in \mathbf{T}, \end{aligned} \tag{8.91}$$

which agrees with the extended separation principle (see Theorem 5.16). Here $(\hat{x}^*, \psi^1, \psi^2)$ is a solution of the system of linear stochastic differential equations

$$\begin{cases} d\hat{x}_t^* = \left(A\hat{x}_t^* + \psi_{t,0}^1 + B_t u_t^*\right)dt + \left(P_t^{00}C_t^* + P_{t,0}^{02} + \Phi_t\bar{R}\Psi_t^*\right)\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1} \\ \quad \times \left(dz_t^* - C_t\hat{x}_t^*dt - \psi_{t,0}^2 dt\right),\ \hat{x}_0^* = 0,\ 0 < t \le T, \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta}^1 dt = \left(P_{t,\theta}^{01*}C_t^* + P_{t,\theta,0}^{12} + \bar{\Phi}_{t,\theta}^1\bar{R}\Psi_t^*\right)\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1} \\ \quad \times \left(dz_t^* - C_t\hat{x}_t^*dt - \psi_{t,0}^2 dt\right),\ \psi_{0,\theta}^1 = \psi_{t,-\varepsilon}^1 = 0,\ -\varepsilon \le \theta \le 0,\ 0 < t \le T, \\ (\partial/\partial t + \partial/\partial\alpha)\psi_{t,\alpha}^2 dt = \left(P_{t,\alpha}^{02*}C_t^* + P_{t,\alpha,0}^{22} + \bar{\Phi}_{t,\alpha}^2\bar{R}\Psi_t^*\right)\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1} \\ \quad \times \left(dz_t^* - C_t\hat{x}_t^*dt - \psi_{t,0}^2 dt\right),\ \psi_{0,\alpha}^2 = \psi_{t,-\delta}^2 = 0,\ -\delta \le \alpha \le 0,\ 0 < t \le T, \end{cases} \tag{8.92}$$

in the sense as defined in Theorem 8.17 with $\left(P^{00}, P^{01}, P^{02}, P^{11}, P^{12}, P^{22}\right)$ being a unique solution of the system (8.60)–(8.65) in the sense defined in Theorem 8.16. Also, using Theorem 3.27, one can show that Q^{00} is a unique scalar product solution of the differential equation

$$\begin{aligned} &\frac{d}{dt}Q_t^{00} + Q_t^{00}A + A^*Q_t^{00} + F_t \\ &\quad - \left(Q_t^{00}B_t + L_t^*\right)G_t^{-1}\left(B_t^*Q_t^{00} + L_t\right) = 0,\ 0 \le t < T,\ Q_T^{00} = Q_T. \end{aligned} \tag{8.93}$$

We point out the similarity of the equations obtained for the optimal stochastic regulators in the problems (8.68)–(8.70) and (7.35)–(7.37). The formulae (8.89) and (8.92) include the functions $\bar{\Phi}^1$ and $\bar{\Phi}^1$ defined by (8.12) and (8.13), respectively. We can give the following partial differential equations for them together with the final and boundary conditions that can be easily verified:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial\theta}\right)\bar{\Phi}_{t,\theta}^1 = \frac{\partial}{\partial 2}\Phi_{t-\theta,\theta}^1,\ \bar{\Phi}_{T,\theta}^1 = \bar{\Phi}_{t,-\varepsilon}^1 = 0,\ -\varepsilon \le \theta \le 0,\ 0 \le t < T, \tag{8.94}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial\alpha}\right)\bar{\Phi}_{t,\alpha}^2 = \frac{\partial}{\partial 2}\Phi_{t-\alpha,\alpha}^2,\ \bar{\Phi}_{T,\alpha}^2 = \bar{\Phi}_{t,-\delta}^1 = 0,\ -\delta \le \alpha \le 0,\ 0 \le t < T. \tag{8.95}$$

Here $(\partial/\partial 2)\Phi_{t,\theta}^1$ and $(\partial/\partial 2)\Phi_{t,\alpha}^2$ mean the partial differentiation with respect to the second variable.

8.4 Concluding Remarks

The wide band noises φ^1 and φ^2, defined by (8.3) and used in the estimation and stochastic regulator problems (8.1)–(8.3) and (8.68)–(8.70), respectively, are in the form of (4.30). A slight modification of this form may be

$$\int_{t-\varepsilon}^{t} \Phi_{t,\theta-t}\, dw_\theta,\ 0 \le t \le T, \tag{8.96}$$

where w is a Wiener process with the time variable varying in $[-\varepsilon, T]$ instead of $\mathbf{T} = [0, T]$ and Φ is a deterministic operator-valued function on $\mathbf{T} \times [-\varepsilon, 0]$.

The problems (8.1)–(8.3) and (8.68)–(8.70) with the wide band noises φ^1 and φ^2, defined in the form of (8.96), can be solved in a similar way that was demonstrated in this chapter by making slight changes.

The idea of the formulae (4.30) and (8.96) is that they define a wide band noise as a distributed left translation of a white noise. One can try a distributed right translation and define a wide band noise in the form

$$\int_t^{t+\varepsilon} \Phi_{t,\theta-t}\, dw_\theta,\ 0 \le t \le T, \tag{8.97}$$

or, more generally,

$$\int_{t-\varepsilon}^{t+\varepsilon} \Phi_{t,\theta-t}\, dw_\theta, d\theta,\ 0 \le t \le T, \tag{8.98}$$

where w is again a Wiener process and Φ is a deterministic operator-valued function. The wide band noise (8.97) (or (8.98)) could be rewritten in the form of (8.96) by considering the shifted (translated) Wiener process $\bar{w}_t = w_{t+\varepsilon} - w_\varepsilon$ (see Proposition 4.16(a)). Therefore, the problems (8.1)–(8.2) and (8.68)–(8.70) with the wide band noises φ^1 and φ^2, given in the form of (8.97) (or (8.98)), can be reduced to the respective problems with the shifted Wiener processes.

Another interesting problem, originally stated in [16, 26], is as follows. In applications a wide band noise is given by its autocovariance function. By Theorem 4.38, each autocovariance function generates an infinite collection of wide band noise processes represented in the integral form (4.30). Consequently, the error of estimation (see Proposition 8.11) and the minimum of the cost functional (see Proposition 8.24) are for the sample wide band noises φ^1 and φ^2 corresponding to the given relaxing functions Φ^1 and Φ^2. Generally, we must talk about their distributions. What are the least upper bounds and the greatest lower bounds of these distributions? Which φ^1 and φ^2 are the most preferable samples for these distributions? What are the shapes of the relaxing functions of most preferable φ^1 and φ^2?

One can try an analytic method to study this problem. Another way may be developing numerical methods for the respective Riccati equations and using simulations.

Chapter 9

Control and Estimation under Shifted White Noises

In this chapter the control and estimation results of Chapter 6 are modified to the shifted white noise processes in case of the state or signal noise delaying the observation noise. As a method of study we use the duality principle for estimation problems and the extended separation principle for control problem.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $X, U, H \in \mathcal{H}$, $T > 0$, $\mathbf{T} = [0, T]$ is a finite time interval and $\Delta_t = \{(s, r) : 0 \le r \le s \le t\}$ for $t > 0$.

9.1 Preliminaries

Chapters 7 and 8 contain the control and estimation results for the two specific kinds of dependence of noise processes acting on state and observation systems. The method of study used in these chapters is based on a reduction of the originally given system to a system with correlated white noises. This reduction allows us to extend (in a certain way) the control and estimation results of Chapter 6 to the colored and wide band noise driven systems. Even, in Chapters 7 and 8 we did not refer to the extended separation principle stated in Theorem 5.16 since it was sufficient to use its particular case from Theorem 6.20.

Another kind of dependence of noise processes can be exposed if one of two correlated white noise processes is a pointwise delay of the other one. We regard such noise processes as shifted white noises. Especially, the case when the state or signal noise is a delay of the observation noise can have significant implications in engineering. Here are some illustrations.

Mapping the ocean floor. Getting a correct map of the ocean floor is important for installation of fixed mobile drilling platforms, locating pipelines in the ocean

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

etc. This job is assisted by a device called a sonar. A sound signal radiates into the water through the sonar transducer that normally is mounted near the keel of a surface ship. Echoes are reflected from the ocean bottom to the sonar which detects them and determines water depth. However, the ocean waves affect the calculated water depth. If ε is the difference of the detecting and radiating times of the sound signal and y is the actual water depth (corresponding to the ocean level), then the water depth is

$$z = y + w'_t$$

at the detecting time moment t of the sound signal and it is

$$x = y + w'_{t-\varepsilon}$$

at the radiating time moment $t - \varepsilon$ of the same sound signal. Here w'_t is the displacement in a surface wave at the time t and can be characterized as the sum of wind-generated waves at previous times over a large area in conjunction with the Earth's gravity. Considering w' as a white noise, we see that x and z are random perturbations of y by shifted white noises.

Space navigation and guidance. In the previous illustration, ε is negligible. For instant, if a sound propagates in water at a speed about 1500 m/s, for an ordinary water depth of 750 m, one can calculate $\varepsilon = 1\ s$. The change $w'(t) - w'(t - \varepsilon)$ of the ocean-wave height for the time of 1 s is much smaller than the depth of 750 m. But, the previous illustration exposes well a mechanism that forms shifted white noises. Should the ocean bottom be replaced with a spacecraft, a sonar with a ground radar, a sound signal with an electromagnetic signal and a white noise caused by ocean waves with a white noise caused by atmospheric propagation, then the corresponding value of ε would be significant. It is nearly constant for Earth orbiting satellites and time dependent for space probes having interplanetary missions.

To understand the nature of the shift arising in space navigation, fix some time moment t and let ε be the time needed for electromagnetic signals to run the distance from the ground radar to the spacecraft and then to turn back. Assume that the control action u changes the position x of the spacecraft in accordance with the linear equation

$$x' = Ax + Bu$$

if noise effects and the distance to the spacecraft are neglected. Then at the time t the ground radar detects the signal

$$z_t = x_{t-\varepsilon/2} + w'_t$$

consisting of the useful information $x_{t-\varepsilon/2}$ about the position of the spacecraft at $t - \varepsilon/2$ corrupted by white noise w'_t caused by atmospheric propagation. Furthermore, the position of the spacecraft at $t - \varepsilon/2$ is changed by the control action $u_{t-\varepsilon}$ that is sent by the ground radar at the time moment $t - \varepsilon$. This control passing through the atmosphere is corrupted by the noise $w'_{t-\varepsilon}$. Hence, the equation for the position of the spacecraft must be written as

$$x'_{t-\varepsilon/2} = Ax_{t-\varepsilon/2} + B(u_{t-\varepsilon} + w'_{t-\varepsilon}).$$

Substituting $\tilde{x}_t = x_{t-\varepsilon/2}$ and $\tilde{u}_t = u_{t-\varepsilon}$, we obtain the partially observable system

$$\begin{cases} \tilde{x}'_t = A\tilde{x}_t + B\tilde{u}_t + Bw'_{t-\varepsilon}, \\ z_t = \tilde{x}_t + w'_t, \end{cases}$$

or, if $\tilde{w}'_t = w'_{t-\varepsilon}$,

$$\begin{cases} \tilde{x}'_t = A\tilde{x}_t + B\tilde{u}_t + B\tilde{w}'_t, \\ z_t = \tilde{x}_t + \tilde{w}'_{t+\varepsilon}, \end{cases}$$

disturbed by shifted white noises with the state noise delaying the observation noise.

In this chapter we will discuss linear stochastic regulator and estimation problems for partially observable linear systems under shifted noises with the state (signal) noise delaying the observation noise. As a method of study we use the duality principle for estimation problems and the extended separation principle for the stochastic regulator problem. For this, we will assume that $0 < \varepsilon < T$ and will consider two correlated Wiener processes w and v on the interval $[0, T + \varepsilon]$ and the function $\mu \in W^{1,\infty}(\mathbf{T}, \mathbb{R})$ satisfying

$$t \le \mu_t \le t + \varepsilon, \ 0 \le t \le T, \text{ and } \mu_s < \mu_t, \ 0 \le s < t \le T.$$

Since μ is increasing and continuous, its inverse μ^{-1} exists on $[\mu_0, \mu_T]$ and it is increasing and continuous as well. The noises of the state (signal) and the observations will be formed by the random processes w_t and v_{μ_t}, $0 \le t \le T$, respectively. Thus, the state (signal) noise will be a delay of the observation noise in time.

Three particular cases of the function μ are as follows:

(a) $\mu_t = t$ (identity);

(b) $\mu_t = t + \varepsilon$ with $\varepsilon > 0$ (right translation);

(c) $\mu_t = ct$ with $c > 1$ (rotation).

These cases will form our examples to demonstrate the theory in this and succeeding chapters. Note that the case (a) means there is no shift, and in fact it was discussed in Chapter 6. The case (b) describes the shift arising in navigation of Earth orbiting satellites since they have nearly constant distance from the Earth, but the case (c) describes the shift arising in navigation of space probes flying away from the Earth since their distance from the Earth increases with nearly constant rate of change.

9.2 State Noise Delaying Observation Noise: Filtering

9.2.1 Setting of the Problem

As it was mentioned above, we are going to apply the extended separation principle to study the control problem for the partially observable linear system under the state noise delaying the observation noise. Therefore, the related filtering problem will be set in the form mentioned in Section 5.7.

Let

$$x_t = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}\Phi_s\,dw_s,\ 0 \le t \le T, \tag{9.1}$$

$$y_t = M_t x_t + \int_t^T N_{s,t}\Phi_s\,dw_s,\ 0 \le t \le T, \tag{9.2}$$

$$z_t = \int_0^t C_s x_s\,ds + \int_0^t \Psi_s\,dv_{\mu_s},\ 0 \le t \le T. \tag{9.3}$$

Assume that the following conditions hold:

($\mathbf{E_1^s}$) $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$;

($\mathbf{E_2^s}$) $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $[0, T+\varepsilon]$ with $\mathrm{cov} v_T > 0$, $0 < \varepsilon < T$, $\mu \in W^{1,\infty}(\mathbf{T}, \mathbb{R})$ is a function satisfying $t \le \mu_t \le t + \varepsilon$ for $0 \le t \le T$, $\mu_s < \mu_t$ for $0 \le s < t \le T$ and $\mu_t' \ge c$ for a.e. $t \in \mathbf{T}$ and for some $c > 0$, x_0 is a an X-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0 and (w, v) are independent;

($\mathbf{E_3^s}$) $M \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $N \in B_\infty(\Delta_T, \mathcal{L}(X, U))$.

We will use the operators P_0, $\bar{W}$, $\bar{V}$ and $\bar{R}$ as defined by

$$P_0 = \mathrm{cov} x_0,\ \begin{bmatrix} \bar{W} & \bar{R} \\ \bar{R}^* & \bar{V} \end{bmatrix} = T^{-1}\mathrm{cov}\begin{bmatrix} w_T \\ v_T \end{bmatrix},$$

and always assume that $N_{t,s} = 0$ if $t > T$. For brevity, we will also use the notation

$$W_t = \Phi_t \bar{W} \Phi_t^*,\ V_t = \Psi_t \bar{V} \Psi_t^* \mu_t',\ R_t = \begin{Bmatrix} \Phi_{\mu_t} \bar{R} \Psi_t^* \mu_t', & \mu_t \le T \\ 0, & \mu_t > T \end{Bmatrix},\ \text{a.e. } t \in \mathbf{T}. \tag{9.4}$$

Obviously,

$$W \in B_\infty(\mathbf{T}, \mathcal{L}(X)),\ V, V^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n)) \text{ and } R \in B_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n, X)).$$

Since x_0, w and v define a Gaussian system, according to Proposition 4.34, we can expect that

$$\hat{y}_t = \mathbf{E}(y_t | z_s;\ 0 \le s \le t) = \int_0^t K_s\,dz_s,\ 0 \le t \le T, \tag{9.5}$$

for some function $K \in B_2(0,t;\mathcal{L}(\mathbb{R}^n,U))$. Estimating the random variable y_t based on the observations z_s, $0 \le s \le t$, in the linear feedback form (9.5), where x, y and z are defined by (9.1)–(9.3), will be called the filtering problem (9.1)–(9.3).

9.2.2 Dual Linear Regulator Problem

Lemma 9.1. *For $0 < t \le T$, the equality (9.5) holds if and only if K satisfies*

$$K_s V_s + \int_0^t K_r \tilde{\Lambda}_{r,s}\,dr = M_t \Lambda_{t,s} C_s^* + \left\{ \begin{array}{ll} M_t \mathcal{U}_{t,\mu_s}, & t \ge \mu_s \\ N_{\mu_s,t}, & t < \mu_s \end{array} \right\} R_s, \ \ a.e.\ s \in [0,t],$$

where Λ and $\tilde{\Lambda}$ are defined by

$$\Lambda_{r,s} = \mathcal{U}_{r,0} P_0 \mathcal{U}_{s,0}^* + \int_0^{\min(s,r)} \mathcal{U}_{r,\sigma} W_\sigma \mathcal{U}_{s,\sigma}^*\,d\sigma, \ \ s,r \in \mathbf{T},$$

$$\tilde{\Lambda}_{r,s} = C_r \Lambda_{r,s} C_s^* + \left\{ \begin{array}{ll} C_r \mathcal{U}_{r,\mu_s} R_s, & r > \mu_s \\ R_r^* \mathcal{U}_{s,\mu_r}^* C_s^*, & s \ge \mu_r \\ 0, & otherwise \end{array} \right\}, \ \ s,r \in \mathbf{T}.$$

Proof. This can by proved in a similar way as Lemma 6.1 by use of Proposition 4.34. □

Now fix $0 < t \le T$ and denote

$$\left\{ \begin{array}{l} \mathcal{R} = \mathcal{D}_t(\mathcal{U}),\ B = D_t(C),\ F = D_t(W), \\ G = D_t(V),\ L = D_t(R),\ Q_t = P_0,\ \nu = D_t(\mu), \end{array} \right. \tag{9.6}$$

where the transformations $\mathcal{D}_t$ and D_t are defined in Section 3.2.2. Consider the linear regulator problem of minimizing the functional

$$\begin{aligned} J(\eta) = \langle \xi_t^\eta, Q_t \xi_t^\eta \rangle &+ \int_0^t (\langle \xi_s^\eta, F_s \xi_s^\eta \rangle + \langle \eta_s, G_s \eta_s \rangle)\,ds \\ &+ 2\int_0^t \left\langle \eta_s, L_s \left\{ \begin{array}{ll} \xi_{\nu_s}^\eta, & \nu_s > 0 \\ -N_{t-\nu_s,t}^* l, & \nu_s \le 0 \end{array} \right\} \right\rangle ds, \end{aligned} \tag{9.7}$$

where

$$\xi_s^\eta = -\mathcal{R}_{s,0} M_t^* l + \int_0^s \mathcal{R}_{s,r} B_r \eta_r\,dr, \ \ 0 \le s \le t, \tag{9.8}$$

$l \in U$ and η is a control from the set of admissible controls $L_2(0,t;\mathbb{R}^n)$.

Lemma 9.2. *Suppose that $0 < t \le T$ and $l \in U$. Then a control $\eta^* \in L_2(0,t;\mathbb{R}^n)$ is optimal in the linear regulator problem (9.7)–(9.8) if and only if it satisfies*

$$G_s \eta_s^* + \int_0^t \tilde{\Sigma}_{s,r}^t \eta_r^*\,dr = B_s^* \Sigma_{s,0}^t M_t^* l + L_s \left\{ \begin{array}{ll} \mathcal{R}_{\nu_s,0} M_t^* l, & \nu_s \ge 0 \\ N_{t-\nu_s,t}^* l, & \nu_s < 0 \end{array} \right\}, \ \ a.e.\ s \in [0,t],$$

where

$$\Sigma^t_{s,r} = \mathcal{R}^*_{t,s} Q_t \mathcal{R}_{t,r} + \int_{\max(s,r)}^{t} \mathcal{R}^*_{\sigma,s} F_\sigma \mathcal{R}_{\sigma,r}\, d\sigma, \ \ s,r \in [0,t],$$

$$\tilde{\Sigma}^t_{s,r} = B^*_s \Sigma^t_{s,r} B_r + \left\{ \begin{array}{ll} L_s \mathcal{R}_{\nu_s,r} B_r, & r < \nu_s \\ B^*_s \mathcal{R}^*_{\nu_r,s} L^*_r, & s \le \nu_r \\ 0, & \textit{otherwise} \end{array} \right\}, \ \ s,r \in [0,t].$$

Proof. This can by proved in a similar way as Lemma 6.2. □

Theorem 9.3. *Let $0 < t \le T$. Then under the conditions $(\mathbf{E^s_1})$–$(\mathbf{E^s_3})$ and (9.6), the best estimate $\hat{y}_t$ of y_t based on z_s, $0 \le s \le t$, in the filtering problem (9.1)–(9.3) is equal to (9.5) if and only if the function, defined by $\eta^*_s = K^*_{t-s} l$, a.e. $s \in [0,t]$, is an optimal control in the linear regulator problem (9.7)–(9.8) for all $l \in U$.*

Proof. This can by proved in a similar way as Theorem 6.3 by use of Lemmas 9.1 and 9.2. □

By Theorem 9.3, the linear regulator problem (9.7)–(9.8) is dual to the filtering problem (9.1)–(9.3).

9.2.3 Optimal Linear Feedback Filter

We will write a control $\eta \in L_2(0,t;\mathbb{R}^n)$ in the problem (9.7)–(9.8) in the form

$$\eta_s = \zeta_s - G_s^{-1} L_s \left\{ \begin{array}{ll} \xi^\eta_{\nu_s}, & \nu_s > 0 \\ -N^*_{t-\nu_s,t} l, & \nu_s \le 0 \end{array} \right\}, \ \text{a.e. } s \in [0,t], \tag{9.9}$$

where $\zeta \in L_2(0,t;\mathbb{R}^n)$. Substituting (9.9) in (9.7)–(9.8), we obtain that the function $\eta = \eta^*$ is an optimal control in the problem (9.7)–(9.8) if and only if the function $\zeta = \zeta^*$, which is related with $\eta = \eta^*$ as in (9.9), is an optimal control in the linear regulator problem of minimizing the functional

$$J_1(\zeta) = \langle \rho^\zeta_t, Q_t \rho^\zeta_t \rangle + \int_0^t \left(\langle \rho^\zeta_s, \check{F}_s \rho^\zeta_s \rangle + \langle \zeta_s, G_s \zeta_s \rangle \right) ds \tag{9.10}$$

with

$$\begin{aligned} \rho^\zeta_s = & -\mathcal{R}_{s,0} M^*_t l + \int_0^s \mathcal{R}_{s,r} B_r \zeta_r \, dr \\ & - \int_0^s \mathcal{R}_{s,r} B_r G_r^{-1} L_r \left\{ \begin{array}{ll} \rho^\zeta_{\nu_r}, & \nu_r > 0 \\ -N^*_{t-\nu_r,t} l, & \nu_r \le 0 \end{array} \right\} dr, \ \ 0 \le s \le t, \end{aligned} \tag{9.11}$$

where $\rho^\zeta = \xi^\eta$ if ζ and η are related as in (9.9) and

$$\check{F}_s = \left\{ \begin{array}{ll} F_s - L^*_{\nu_s^{-1}} G^{-1}_{\nu_s^{-1}} L_{\nu_s^{-1}} \left(\nu^{-1}\right)'_s, & s < \nu_t \\ F_s, & s \ge \nu_t \end{array} \right\}, \ \text{a.e. } s \in [0,t]. \tag{9.12}$$

Note that, since $s \le \mu_s \le s+\varepsilon$, we have $s-\varepsilon \le \nu_s \le s$. Hence, (9.11) is a delay equation. Using Proposition 3.18(f), one can compute that

$$\begin{aligned}
\check{F}_s &= \begin{Bmatrix} W_{t-s} - R_{t-\nu_s^{-1}} V_{t-\nu_s^{-1}}^{-1} R_{t-\nu_s^{-1}}^{*} (\nu^{-1})'_s, & s < \nu_t \\ W_{t-s}, & s \ge \nu_t \end{Bmatrix} \\
&= \begin{Bmatrix} W_{t-s} - R_{\mu_{t-s}^{-1}} V_{\mu_{t-s}^{-1}}^{-1} R_{\mu_{t-s}^{-1}}^{*} (\mu^{-1})'_{t-s}, & t-s > \mu_0 \\ W_{t-s}, & t-s \le \mu_0 \end{Bmatrix} \\
&= \Phi_{t-s} \begin{Bmatrix} \bar{W} - \bar{R}\bar{V}^{-1}\bar{R}^{*}\mu'_{\mu_{t-s}^{-1}} (\mu^{-1})'_{t-s}, & t-s > \mu_0 \\ \bar{W}, & t-s \le \mu_0 \end{Bmatrix} \Phi_{t-s}^{*} \\
&= \Phi_{t-s} \begin{Bmatrix} T^{-1}\mathrm{cov}\big(w_T - \bar{R}\bar{V}^{-1}v_T\big), & t-s > \mu_0 \\ T^{-1}\mathrm{cov}\, w_T, & t-s \le \mu_0 \end{Bmatrix} \Phi_{t-s}^{*} \ge 0.
\end{aligned}$$

Thus, $\check{F}$ is a function the values of which are nonnegative operators.

Let $\tilde{X} = L_2(-\varepsilon, 0; X)$ and let $\check{X} = W^{12}(-\varepsilon, 0; X)$. Define the semigroup of right translation $\mathcal{T}$ and the linear operator Γ as in (3.4) and (3.23), respectively. Let $\bar{\rho}^\zeta$ and $\tilde{\rho}^\zeta$ be defined by

$$\tilde{\rho}_s^\zeta = \begin{bmatrix} \rho_s^\zeta \\ \bar{\rho}_s^\zeta \end{bmatrix} \in X \times \tilde{X}, \ 0 \le s \le t,$$

$$\left[\bar{\rho}_s^\zeta\right]_\theta = \begin{Bmatrix} \rho_{s+\theta}^\zeta, & s+\theta > 0 \\ -N_{t-s-\theta,t}^{*}, & s+\theta \le 0 \end{Bmatrix}, \text{ a.e. } \theta \in [-\varepsilon, 0], \ 0 \le s \le t.$$

According to the definitions of Sections 3.4.2, $\tilde{\rho}^\zeta$ and $\bar{\rho}^\zeta$ are the tilde and bar functions over ρ^ζ with the initial distribution

$$\lambda_\theta = -N_{t-\theta,t}^{*} l, \text{ a.e. } \theta \in [-\varepsilon, 0].$$

Using the results of Section 3.4.2, the linear regulator problem (9.10)–(9.11) can be written in terms of the process $\tilde{\rho}^\zeta$ in the following form:

$$J_1(\zeta) = \langle \tilde{\rho}_t^\zeta, \tilde{Q}_t \tilde{\rho}_t^\zeta \rangle + \int_0^t \big(\langle \tilde{\rho}_s^\zeta, \tilde{F}_s \tilde{\rho}_s^\zeta \rangle + \langle \zeta_s, G_s \zeta_s \rangle \big) ds, \tag{9.13}$$

$$\tilde{\rho}_s^\zeta = \tilde{\mathcal{R}}_{s,0}\tilde{\rho}_0 + \int_0^s \tilde{\mathcal{R}}_{s,r}\tilde{B}_r \zeta_r \, dr, \ 0 \le s \le t, \tag{9.14}$$

where

$$\tilde{\mathcal{R}} = \mathcal{P}_{\Lambda^*(-BG^{-1}L,\nu)}^{*}(\mathcal{R} \odot \mathcal{T}^{*}) \in \mathcal{E}\big(\Delta_t, \mathcal{L}(X \times \tilde{X})\big), \tag{9.15}$$

$$\tilde{F} = \begin{bmatrix} \check{F} & 0 \\ 0 & 0 \end{bmatrix} \in B_\infty\big(0, t; \mathcal{L}(X \times \tilde{X})\big), \ \tilde{Q}_t = \begin{bmatrix} Q_t & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}), \tag{9.16}$$

$$\tilde{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \in B_\infty\big(0, t; \mathcal{L}(\mathbb{R}^n, X \times \tilde{X})\big). \tag{9.17}$$

Since $\tilde{F}_s \geq 0$, a.e. $s \in [0,t]$, Theorem 5.24 can be applied to the problem (9.13)–(9.14). Accordingly, there exists a unique optimal control in the problem (9.13)–(9.14) (and, respectively, in the problem (9.10)–(9.11)) and it has the form

$$\zeta_s^* = -G_s^{-1}\tilde{B}_s^*\tilde{Q}_s\tilde{\rho}_s^*, \text{ a.e. } s \in [0,t]; \ \tilde{\rho}_s^* = \tilde{\mathcal{K}}_{s,0}\tilde{\rho}_0, \ 0 \leq s \leq t, \tag{9.18}$$

where $\tilde{\rho}^* = \tilde{\rho}^{\zeta^*}$,

$$\tilde{\mathcal{K}} = \mathcal{P}_{-\tilde{B}G^{-1}\tilde{B}^*\tilde{Q}}(\tilde{\mathcal{R}}) \tag{9.19}$$

and $\tilde{Q}$ is a unique solution of the Riccati equation

$$\tilde{Q}_s = \tilde{\mathcal{R}}_{t,s}^*\tilde{Q}_t\tilde{\mathcal{R}}_{t,s} + \int_s^t \tilde{\mathcal{R}}_{r,s}^*(\tilde{F}_r - \tilde{Q}_r\tilde{B}_rG_r^{-1}\tilde{B}_r^*\tilde{Q}_r)\tilde{\mathcal{R}}_{r,s}\,dr, \ 0 \leq s \leq t, \tag{9.20}$$

with $\tilde{Q}_s \geq 0$, $0 \leq s \leq t$. Expressing $\tilde{\rho}_0$ by l, we obtain

$$\tilde{\rho}_0 = \begin{bmatrix} \rho_0 \\ \bar{\rho}_0 \end{bmatrix} \in X \times \tilde{X}, \ \rho_0 = -M_t^*l, \ [\bar{\rho}_0]_\theta = -N_{t-\theta,t}^*l, \text{ a.e. } \theta \in [-\varepsilon, 0].$$

Denote

$$\tilde{M}_t = [M_t \ \ \tilde{N}_t] \in \mathcal{L}(X \times \tilde{X}, U), \ \ \tilde{N}_t h = \int_{-\varepsilon}^0 N_{t-\theta,t}h_\theta\,d\theta, \ h \in \tilde{X}. \tag{9.21}$$

Then

$$\tilde{\rho}_0 = -\tilde{M}_t^*l, \ l \in U. \tag{9.22}$$

Thus, we obtain the following result.

Lemma 9.4. *There exists a unique optimal control in the linear regulator problem* (9.7)–(9.8) *and it has the form*

$$\begin{aligned} \eta_s^* &= G_s^{-1}\tilde{B}_s^*\tilde{Q}_s\tilde{\mathcal{K}}_{s,0}\tilde{M}_t^*l \\ &+ G_s^{-1}L_s \begin{Bmatrix} \tilde{I}\tilde{\mathcal{K}}_{\nu_s,0}\tilde{M}_t^*l, & \nu_s > 0 \\ N_{t-\nu_s,t}^*l, & \nu_s \leq 0 \end{Bmatrix}, \text{ a.e. } s \in [0,t], \end{aligned} \tag{9.23}$$

where

$$\tilde{I} = [I \ \ 0] \in \mathcal{L}(X \times \tilde{X}, X). \tag{9.24}$$

Proof. This follows from (9.9) by use of (9.18) and (9.22). □

Theorem 9.5. *Under the conditions* $(\mathbf{E}_1^s)$–$(\mathbf{E}_3^s)$, *there exists a unique optimal linear feedback filter in the filtering problem* (9.1)–(9.3) *and the best estimate* $\hat{y}_t$ *of* y_t *based on* z_s, $0 \leq s \leq t$, *is equal to* (9.5) *with*

$$\begin{aligned} K_s = K_{t,s} &= \tilde{M}_t\tilde{\mathcal{Y}}_{t,s}\tilde{P}_s\tilde{C}_s^*V_s^{-1} \\ &+ \begin{Bmatrix} \tilde{M}_t\tilde{\mathcal{Y}}_{t,\mu_s}\tilde{I}^*, & t > \mu_s \\ N_{\mu_s,t}, & t \leq \mu_s \end{Bmatrix} R_sV_s^{-1}, \text{ a.e. } s \in [0,t], \end{aligned} \tag{9.25}$$

where

$$\tilde{\mathcal{Y}} = \mathcal{P}_{-\tilde{P}\tilde{C}^* V^{-1}\tilde{C}}(\tilde{\mathcal{U}}),\ \tilde{\mathcal{U}} = \mathcal{P}_{\Lambda(-RV^{-1}C,\,\mu)}(\mathcal{U} \odot \mathcal{T}), \tag{9.26}$$

$\tilde{P}$ is a solution of the Riccati equation

$$\tilde{P}_s = \tilde{\mathcal{U}}_{s,0}\tilde{P}_0\tilde{\mathcal{U}}_{s,0}^* + \int_0^s \tilde{\mathcal{U}}_{s,r}(\tilde{W}_r - \tilde{P}_r\tilde{C}_r^* V_r^{-1}\tilde{C}_r\tilde{P}_r)\tilde{\mathcal{U}}_{s,r}^*\, dr,\ 0 \le s \le T, \tag{9.27}$$

with $\tilde{P}_s \ge 0$, $0 \le s \le T$,

$$\check{W}_s = \begin{cases} W_s - R_{\mu_s^{-1}} V_{\mu_s^{-1}}^{-1} R_{\mu_s^{-1}}^* (\mu^{-1})'_s, & s > \mu_0 \\ W_s, & s \le \mu_0 \end{cases},\ 0 \le s \le T, \tag{9.28}$$

$$\tilde{W} = \begin{bmatrix} \check{W} & 0 \\ 0 & 0 \end{bmatrix} \in B_\infty(\mathbf{T}; \mathcal{L}(X \times \tilde{X})),\ \tilde{P}_0 = \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}), \tag{9.29}$$

$$\tilde{C} = \begin{bmatrix} C & 0 \end{bmatrix} \in B_\infty(\mathbf{T}; \mathcal{L}(X \times \tilde{X}, \mathbb{R}^n)), \tag{9.30}$$

$\tilde{I}$ is defined by (9.24)*, W, V and R are defined by* (9.4) *and $\tilde{M}$ is defined by* (9.21)*.*

Proof. Using Proposition 3.18, one can easily verify that

$$\tilde{F} = D_t(\tilde{W}),\ \tilde{B} = D_t(\tilde{C}) \text{ and } \tilde{Q}_t = \tilde{P}_0,$$

where $\tilde{F}$, $\tilde{Q}_t$ and $\tilde{B}$ are defined by (9.16)–(9.17). By Propositions 3.41(a) and 3.18(a),

$$\tilde{\mathcal{K}} = \mathcal{D}_t(\tilde{\mathcal{Y}}),$$

where $\tilde{\mathcal{K}}$ is defined by (9.19) and (9.15). Using these relations and (9.6), one can show that

$$\tilde{Q} = D_t(\tilde{P}),$$

where $\tilde{Q}$ is a solution of the equation (9.20). Thus, all these relations together with Theorem 9.3 imply the statement. □

9.2.4 About the Riccati Equation (9.27)

The solution $\tilde{P}$ of the Riccati equation (9.27) can be decomposed in the form

$$\tilde{P}_s = \begin{bmatrix} P_s^{00} & \tilde{P}_s^{01} \\ \tilde{P}_s^{01*} & \tilde{P}_s^{11} \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}),\ 0 \le s \le T. \tag{9.31}$$

Our aim in this section is a derivation of the equations for the components P^{00} and $\tilde{P}^{01}$ of $\tilde{P}$ in (9.31).

At first, recall that the mild evolution operator

$$\tilde{\mathcal{U}} = \mathcal{P}_{\Lambda(-RV^{-1}C,\mu)}(\mathcal{U} \odot \mathcal{T})$$

can also be decomposed in the form

$$\tilde{\mathcal{U}}_{s,r} = \begin{bmatrix} \tilde{\mathcal{U}}^{00}_{s,r} & \tilde{\mathcal{U}}^{01}_{s,r} \\ \tilde{\mathcal{U}}^{10}_{s,r} & \tilde{\mathcal{U}}^{11}_{s,r} \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}),\ 0 \le r \le s \le T, \tag{9.32}$$

where, according to the results of Section 3.4.3, $\tilde{\mathcal{U}}^{00}$ is a solution of the equivalent equations

$$\tilde{\mathcal{U}}^{00}_{s,r} = \mathcal{U}_{s,r} - \int_r^{\max(\mu_s^{-1},r)} \tilde{\mathcal{U}}^{00}_{s,\mu_\sigma} R_\sigma V_\sigma^{-1} C_\sigma \mathcal{U}_{\sigma,r}\, d\sigma, \tag{9.33}$$

and

$$\tilde{\mathcal{U}}^{00}_{s,r} = \mathcal{U}_{s,r} - \int_r^{\max(\mu_s^{-1},r)} \mathcal{U}_{s,\mu_\sigma} R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}^{00}_{\sigma,r}\, d\sigma, \tag{9.34}$$

$\tilde{\mathcal{U}}^{01}$ and $\tilde{\mathcal{U}}^{10}$ are defined by

$$\tilde{\mathcal{U}}^{01}_{s,r} = \int_r^s \tilde{\mathcal{U}}^{00}_{s,\sigma} \Gamma \mathcal{T}_{\sigma-r}\, d\sigma, \tag{9.35}$$

and

$$\tilde{\mathcal{U}}^{10}_{s,r} = -\int_{\max(\mu_s^{-1},r)}^s \mathcal{T}^*_{\mu_\sigma - s} \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}^{00}_{\sigma,r}\, d\sigma, \tag{9.36}$$

and $\tilde{\mathcal{U}}^{11}$ is defined by either

$$\tilde{\mathcal{U}}^{11}_{s,r} = \mathcal{T}_{s-r} + \int_r^s \tilde{\mathcal{U}}^{10}_{s,\sigma} \Gamma \mathcal{T}_{\sigma-r}\, d\sigma, \tag{9.37}$$

or

$$\tilde{\mathcal{U}}^{11}_{s,r} = \mathcal{T}_{s-r} - \int_{\max(\mu_s^{-1},r)}^s \mathcal{T}^*_{\mu_\sigma - s} \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}^{01}_{\sigma,r}\, d\sigma. \tag{9.38}$$

Note that in (9.35)–(9.38) the integrals of the integrands containing $\Gamma\mathcal{T}$ and $\mathcal{T}^*\Gamma^*$ must be understood in the senses defined in Section 3.4.1.

Proposition 9.6. *Under the conditions* $(\mathbf{E_1^s})$–$(\mathbf{E_2^s})$, *the pair* $\left(P^{00}, \tilde{P}^{01}\right)$ *satisfies the equations*

$$\begin{aligned} P^{00}_s &= \tilde{\mathcal{U}}^{00}_{s,0} P_0 \tilde{\mathcal{U}}^{00*}_{s,0} + \int_0^s \tilde{\mathcal{U}}^{00}_{s,r} \check{W}_r \tilde{\mathcal{U}}^{00*}_{s,r}\, dr \\ &\quad - \int_0^s (\tilde{\mathcal{U}}^{00}_{s,r} P^{00}_r + \tilde{\mathcal{U}}^{01}_{s,r} \tilde{P}^{01*}_r) C_r^* V_r^{-1} C_r (P^{00}_r \tilde{\mathcal{U}}^{00*}_{s,r} + \tilde{P}^{01}_r \tilde{\mathcal{U}}^{01*}_{s,r})\, dr, \end{aligned} \tag{9.39}$$

$$\begin{aligned} \tilde{P}^{01}_s &= -\int_0^s (\tilde{\mathcal{U}}^{00}_{s,r} P^{00}_r + \tilde{\mathcal{U}}^{01}_{s,r} \tilde{P}^{01*}_r) C_r^* V_r^{-1} C_r \tilde{P}^{01}_r \mathcal{T}^*_{s-r}\, dr \\ &\quad - \int_{\max(\mu_s^{-1},0)}^s (\tilde{\mathcal{U}}^{00}_{s,r} P^{00}_r + \tilde{\mathcal{U}}^{01}_{s,r} \tilde{P}^{01*}_r) C_r^* V_r^{-1} R_r^* \Gamma \mathcal{T}_{\mu_r - s}\, dr, \end{aligned} \tag{9.40}$$

where $0 \le s \le T$.

Proof. From (9.27), one can obtain the equation (9.39). In a similar way, we have

$$
\begin{aligned}
\tilde{P}_s^{01*} =& \tilde{\mathcal{U}}_{s,0}^{10} P_0 \tilde{\mathcal{U}}_{s,0}^{00*} + \int_0^s \tilde{\mathcal{U}}_{s,r}^{10} \check{W}_r \tilde{\mathcal{U}}_{s,r}^{00*}\, dr \\
&- \int_0^s \big(\tilde{\mathcal{U}}_{s,r}^{10} P_r^{00} + \tilde{\mathcal{U}}_{s,r}^{11} \tilde{P}_r^{01*}\big) C_r^* V_r^{-1} C_r \big(P_r^{00} \tilde{\mathcal{U}}_{s,r}^{00*} + \tilde{P}_r^{01} \tilde{\mathcal{U}}_{s,r}^{01*}\big)\, dr. \qquad (9.41)
\end{aligned}
$$

Substitution of (9.36) and (9.38) in (9.41) yields

$$
\begin{aligned}
\tilde{P}_s^{01*} =& -\int_{\max(\mu_s^{-1},0)}^s \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,0}^{00} P_0 \tilde{\mathcal{U}}_{s,0}^{00*}\, d\sigma \\
&- \int_0^s \int_{\max(\mu_s^{-1},r)}^s \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,r}^{00} \check{W}_r \tilde{\mathcal{U}}_{s,r}^{00*}\, d\sigma\, dr \\
&+ \int_0^s \int_{\max(\mu_s^{-1},r)}^s \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,r}^{00} P_r^{00} C_r^* V_r^{-1} \tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{s,r}^* \tilde{I}^*\, d\sigma\, dr \\
&+ \int_0^s \int_{\max(\mu_s^{-1},r)}^s \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,r}^{01} \tilde{P}_r^{01*} C_r^* V_r^{-1} \tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{s,r}^* \tilde{I}^*\, d\sigma\, dr \\
&- \int_0^s \mathcal{T}_{s-r} \tilde{P}_r^{01*} C_r^* V_r^{-1} \tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{s,r}^* \tilde{I}^*\, dr,
\end{aligned}
$$

where for brevity we used the equality

$$
\tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{s,r}^* \tilde{I}^* = C_r \big(P_r^{00} \tilde{\mathcal{U}}_{s,r}^{00*} + \tilde{P}_r^{01} \tilde{\mathcal{U}}_{s,r}^{01*}\big).
$$

The semigroup property and changing the order of integration yield

$$
\begin{aligned}
\tilde{P}_s^{01*} =& -\int_{\max(\mu_s^{-1},0)}^s \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,0}^{00} P_0 \big(\tilde{\mathcal{U}}_{\sigma,0}^{00*} \tilde{\mathcal{U}}_{s,\sigma}^{00*} + \tilde{\mathcal{U}}_{\sigma,0}^{10*} \tilde{\mathcal{U}}_{s,\sigma}^{01*}\big)\, d\sigma \\
&- \int_{\max(\mu_s^{-1},0)}^s \int_0^\sigma \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,r}^{00} \check{W}_r \big(\tilde{\mathcal{U}}_{\sigma,r}^{00*} \tilde{\mathcal{U}}_{s,\sigma}^{00*} + \tilde{\mathcal{U}}_{\sigma,r}^{10*} \tilde{\mathcal{U}}_{s,\sigma}^{01*}\big)\, dr\, d\sigma \\
&+ \int_{\max(\mu_s^{-1},0)}^s \int_0^\sigma \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,r}^{00} P_r^{00} C_r^* V_r^{-1} \tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{\sigma,r}^* \tilde{\mathcal{U}}_{s,\sigma}^* \tilde{I}^*\, dr\, d\sigma \\
&+ \int_{\max(\mu_s^{-1},0)}^s \int_0^\sigma \mathcal{T}_{\mu_\sigma - s}^* \Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \tilde{\mathcal{U}}_{\sigma,r}^{01} \tilde{P}_r^{01*} C_r^* V_r^{-1} \tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{\sigma,r}^* \tilde{\mathcal{U}}_{s,\sigma}^* \tilde{I}^*\, dr\, d\sigma \\
&- \int_0^s \mathcal{T}_{s-r} \tilde{P}_r^{01*} C_r^* V_r^{-1} \tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{s,r}^* \tilde{I}^*\, dr.
\end{aligned}
$$

Since

$$
\tilde{C}_r \tilde{P}_r \tilde{\mathcal{U}}_{\sigma,r}^* \tilde{\mathcal{U}}_{s,\sigma}^* \tilde{I}^* = C_r P_r^{00} \big(\mathcal{U}_{\sigma,r}^{00*} \mathcal{U}_{s,\sigma}^{00*} + \mathcal{U}_{\sigma,r}^{10*} \mathcal{U}_{s,\sigma}^{01*}\big) + C_r \tilde{P}_r^{01} \big(\mathcal{U}_{\sigma,r}^{01*} \mathcal{U}_{s,\sigma}^{00*} + \mathcal{U}_{\sigma,r}^{11*} \mathcal{U}_{s,\sigma}^{01*}\big),
$$

from (9.39) and (9.41), we obtain

$$\begin{aligned}\tilde{P}_s^{01*} &= -\int_{\max(\mu_s^{-1},0)}^{s} \mathcal{T}_{\mu_\sigma - s}^{*}\Gamma^* R_\sigma V_\sigma^{-1} C_\sigma \left(P_\sigma^{00}\tilde{\mathcal{U}}_{s,\sigma}^{00*} + \tilde{P}_\sigma^{01}\tilde{\mathcal{U}}_{s,\sigma}^{01*}\right) d\sigma \\ &\quad - \int_0^s \mathcal{T}_{s-r}\tilde{P}_r^{01*} C_r^* V_r^{-1} C_r \left(P_r^{00}\tilde{\mathcal{U}}_{s,r}^{00*} + \tilde{P}_r^{01}\tilde{\mathcal{U}}_{s,r}^{01*}\right) dr,\end{aligned}$$

which implies (9.40). □

9.2.5 About the Optimal Filter

In this section we will derive more detailed formulae for the best estimate $\hat{y}_t$ than in Theorem 9.5.

Theorem 9.7. *Under the conditions* $(\mathbf{E_1^s})$–$(\mathbf{E_3^s})$*, the best estimate* $\hat{y}_t$ *of* y_t *based on the observations* z_s*,* $0 \le s \le t$*, in the filtering problem* (9.1)–(9.3) *has the form*

$$\hat{y}_t = M_t\hat{x}_t + \tilde{N}_t\tilde{\psi}_t + \int_{\max(\mu_t^{-1},0)}^{t} N_{\mu_s,t} R_s V_s^{-1}\, d\bar{z}_s,\ \ 0 \le t \le T, \tag{9.42}$$

where $\hat{x}$ *and* $\tilde{\psi}$ *are defined by*

$$\begin{aligned}\hat{x}_t &= \int_0^t \mathcal{U}_{t,s}\Gamma\tilde{\psi}_s\, ds + \int_0^t \mathcal{U}_{t,s} P_s^{00} C_s^* V_s^{-1}\, d\bar{z}_s \\ &\qquad + \int_0^{\max(\mu_t^{-1},0)} \mathcal{U}_{t,\mu_s} R_s V_s^{-1}\, d\bar{z}_s,\ \ 0 \le t \le T, \end{aligned}\tag{9.43}$$

$$\tilde{\psi}_t = \int_0^t \mathcal{T}_{t-s}\tilde{P}_s^{01*} C_s^* V_s^{-1}\, d\bar{z}_s,\ \ 0 \le t \le T, \tag{9.44}$$

$\bar{z}$ *is the innovation process defined by*

$$d\bar{z}_s = dz_s - C_s\hat{x}_s ds,\ \ 0 < s \le T,\ \ \bar{z}_0 = 0, \tag{9.45}$$

P^{00} *and* $\tilde{P}^{01}$ *are defined by* (9.39)–(9.40) *and* $\Gamma\tilde{\psi}_s = \left[\tilde{\psi}_s\right]_0$.

Proof. By Theorem 9.5,

$$\begin{aligned}\hat{y}_t &= \int_0^t K_{t,s}\, dz_s \\ &= \tilde{M}_t\left(\int_0^t \tilde{\mathcal{Y}}_{t,s}\tilde{P}_s\tilde{C}_s^* V_s^{-1}\, dz_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{Y}}_{t,\mu_s}\tilde{I}^* R_s V_s^{-1}\, dz_s\right) \\ &\quad + \int_{\max(\mu_t^{-1},0)}^{t} N_{\mu_s,t} R_s V_s^{-1}\, dz_s.\end{aligned}$$

Denote

$$\hat{\tilde{x}}_t = \begin{bmatrix} \hat{x}_t \\ \tilde{\phi}_t \end{bmatrix} = \int_0^t \tilde{\mathcal{Y}}_{t,s} \tilde{P}_s \tilde{C}_s^* V_s^{-1} \, dz_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{Y}}_{t,\mu_s} \tilde{I}^* R_s V_s^{-1} \, dz_s. \tag{9.46}$$

Then

$$\hat{y}_t = M_t \hat{x}_t + \tilde{N}_t \tilde{\phi}_t + \int_{\max(\mu_t^{-1},0)}^t N_{\mu_s,t} R_s V_s^{-1} \, dz_s. \tag{9.47}$$

The equality

$$\tilde{\mathcal{Y}} = \mathcal{P}_{-\tilde{P}\tilde{C}^* V^{-1} \tilde{C}}(\tilde{\mathcal{U}})$$

and (9.46) yield

$$\begin{aligned}
\hat{\tilde{x}}_t &= \int_0^t \tilde{\mathcal{U}}_{t,s} \tilde{P}_s \tilde{C}_s^* V_s^{-1} \, dz_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{U}}_{t,\mu_s} \tilde{I}^* R_s V_s^{-1} \, dz_s \\
&\quad - \int_0^t \int_s^t \tilde{\mathcal{U}}_{t,r} \tilde{P}_r \tilde{C}_r^* V_r^{-1} \tilde{C}_r \tilde{\mathcal{Y}}_{r,s} \tilde{P}_s \tilde{C}_s^* V_s^{-1} \, dr \, dz_s \\
&\quad - \int_0^{\max(\mu_t^{-1},0)} \int_{\mu_s}^t \tilde{\mathcal{U}}_{t,r} \tilde{P}_r \tilde{C}_r^* V_r^{-1} \tilde{C}_r \tilde{\mathcal{Y}}_{r,\mu_s} \tilde{I}^* R_s V_s^{-1} \, dr \, dz_s \\
&= \int_0^t \tilde{\mathcal{U}}_{t,s} \tilde{P}_s \tilde{C}_s^* V_s^{-1} \, dz_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{U}}_{t,\mu_s} \tilde{I}^* R_s V_s^{-1} \, dz_s \\
&\quad - \int_0^t \int_0^r \tilde{\mathcal{U}}_{t,r} \tilde{P}_r \tilde{C}_r^* V_r^{-1} \tilde{C}_r \tilde{\mathcal{Y}}_{r,s} \tilde{P}_s \tilde{C}_s^* V_s^{-1} \, dz_s \, dr \\
&\quad - \int_0^t \int_0^{\max(\mu_r^{-1},0)} \tilde{\mathcal{U}}_{t,r} \tilde{P}_r \tilde{C}_r^* V_r^{-1} \tilde{C}_r \tilde{\mathcal{Y}}_{r,\mu_s} \tilde{I}^* R_s V_s^{-1} \, dz_s \, dr \\
&= \int_0^t \tilde{\mathcal{U}}_{t,s} \tilde{P}_s \tilde{C}_s^* V_s^{-1} \, dz_s - \int_0^t \tilde{\mathcal{U}}_{t,r} \tilde{P}_r \tilde{C}_r^* V_r^{-1} C_r \hat{x}_r \, dr \\
&\quad + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{U}}_{t,\mu_s} \tilde{I}^* R_s V_s^{-1} \, dz_s \\
&= \int_0^t \tilde{\mathcal{U}}_{t,s} \tilde{P}_s \tilde{C}_s^* V_s^{-1} \, d\bar{z}_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{U}}_{t,\mu_s} \tilde{I}^* R_s V_s^{-1} \, dz_s.
\end{aligned}$$

Writing this in componentwise form, we obtain

$$\hat{x}_t = \int_0^t (\tilde{\mathcal{U}}_{t,s}^{00} P_s^{00} + \tilde{\mathcal{U}}_{t,s}^{01} \tilde{P}_s^{01*}) C_s^* V_s^{-1} \, d\bar{z}_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{U}}_{t,\mu_s}^{00} R_s V_s^{-1} \, dz_s, \tag{9.48}$$

$$\tilde{\phi}_t = \int_0^t (\tilde{\mathcal{U}}_{t,s}^{10} P_s^{00} + \tilde{\mathcal{U}}_{t,s}^{11} \tilde{P}_s^{01*}) C_s^* V_s^{-1} \, d\bar{z}_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{U}}_{t,\mu_s}^{10} R_s V_s^{-1} \, dz_s. \tag{9.49}$$

Substituting (9.36) and (9.38) in (9.49) and using (9.48), we obtain

$$
\begin{aligned}
\tilde{\phi}_t = & \int_0^t \mathcal{T}_{t-s}\tilde{P}_s^{01*}C_s^*V_s^{-1}\,d\bar{z}_s \\
& - \int_0^t \int_{\max(\mu_t^{-1},s)}^t \mathcal{T}_{\mu_r-t}^*\Gamma^* R_r V_r^{-1}C_r\big(\tilde{\mathcal{U}}_{r,s}^{00}P_s^{00} + \tilde{\mathcal{U}}_{r,s}^{01}\tilde{P}_s^{01*}\big)C_s^*V_s^{-1}\,dr\,d\bar{z}_s \\
& - \int_0^{\max(\mu_t^{-1},0)} \int_{\max(\mu_t^{-1},\mu_s)}^t \mathcal{T}_{\mu_r-t}^*\Gamma^* R_r V_r^{-1}C_r\tilde{\mathcal{U}}_{r,\mu_s}^{00}R_sV_s^{-1}\,dr\,dz_s \\
= & \int_0^t \mathcal{T}_{t-s}\tilde{P}_s^{01*}C_s^*V_s^{-1}\,d\bar{z}_s \\
& - \int_{\max(\mu_t^{-1},0)}^t \int_0^r \mathcal{T}_{\mu_r-t}^*\Gamma^* R_r V_r^{-1}C_r\big(\tilde{\mathcal{U}}_{r,s}^{00}P_s^{00} + \tilde{\mathcal{U}}_{r,s}^{01}\tilde{P}_s^{01*}\big)C_s^*V_s^{-1}\,d\bar{z}_s\,dr \\
& - \int_{\max(\mu_t^{-1},0)}^t \int_0^{\max(\mu_r^{-1},0)} \mathcal{T}_{\mu_r-t}^*\Gamma^* R_r V_r^{-1}C_r\tilde{\mathcal{U}}_{r,\mu_s}^{00}R_sV_s^{-1}\,dz_s\,dr \\
= & \int_0^t \mathcal{T}_{t-s}\tilde{P}_s^{01*}C_s^*V_s^{-1}\,d\bar{z}_s \\
& - \int_{\max(\mu_t^{-1},0)}^t \mathcal{T}_{\mu_r-t}^*\Gamma^* R_r V_r^{-1}C_r\hat{x}_r\,dr.
\end{aligned}
$$

Denote

$$
\tilde{\psi}_t = \tilde{\phi}_t + \int_{\max(\mu_t^{-1},0)}^t \mathcal{T}_{\mu_r-t}^*\Gamma^* R_r V_r^{-1}C_r\hat{x}_r\,dr,\ 0 \le t \le T. \tag{9.50}
$$

Then, for $\tilde{\psi}$, we obtain

$$
\tilde{\psi}_t = \int_0^t \mathcal{T}_{t-s}\tilde{P}_s^{01*}C_s^*V_s^{-1}\,d\bar{z}_s,
$$

proving (9.44). Also, by (9.35),

$$
\begin{aligned}
\int_0^t \tilde{\mathcal{U}}_{t,s}^{01}\tilde{P}_s^{01*}C_s^*V_s^{-1}\,d\bar{z}_s &= \int_0^t \int_s^t \tilde{\mathcal{U}}_{t,r}^{00}\Gamma\mathcal{T}_{r-s}\tilde{P}_s^{01*}C_s^*V_s^{-1}\,dr\,d\bar{z}_s \\
&= \int_0^t \int_0^r \tilde{\mathcal{U}}_{t,r}^{00}\Gamma\mathcal{T}_{r-s}\tilde{P}_s^{01*}C_s^*V_s^{-1}\,d\bar{z}_s\,dr \\
&= \int_0^t \tilde{\mathcal{U}}_{t,r}^{00}\Gamma\tilde{\psi}_r\,dr.
\end{aligned}
$$

Using this in (9.48), we obtain

$$
\hat{x}_t = \int_0^t \tilde{\mathcal{U}}_{t,s}^{00}\Gamma\tilde{\psi}_s\,ds + \int_0^t \tilde{\mathcal{U}}_{t,s}^{00}P_s^{00}C_s^*V_s^{-1}\,d\bar{z}_s + \int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{U}}_{t,\mu_s}^{00}R_sV_s^{-1}\,dz_s. \tag{9.51}
$$

Thus, from (9.34) and (9.51),

$$\begin{aligned}
\hat{x}_t &= \int_0^t \mathcal{U}_{t,s}\Gamma\tilde{\psi}_s\,ds + \int_0^t \mathcal{U}_{t,s}P_s^{00}C_s^*V_s^{-1}\,d\bar{z}_s \\
&\quad + \int_0^{\max(\mu_t^{-1},0)} \mathcal{U}_{t,\mu_s}R_sV_s^{-1}\,dz_s \\
&\quad - \int_0^t \int_s^{\max(\mu_t^{-1},s)} \mathcal{U}_{t,\mu_r}R_rV_r^{-1}C_r\tilde{\mathcal{U}}_{r,s}^{00}\Gamma\tilde{\psi}_s\,dr\,ds \\
&\quad - \int_0^t \int_s^{\max(\mu_t^{-1},s)} \mathcal{U}_{t,\mu_r}R_rV_r^{-1}C_r\tilde{\mathcal{U}}_{r,s}^{00}P_s^{00}C_s^*V_s^{-1}\,dr\,d\bar{z}_s \\
&\quad - \int_0^{\max(\mu_t^{-1},0)} \int_{\mu_s}^{\max(\mu_t^{-1},\mu_s)} \mathcal{U}_{t,\mu_r}R_rV_r^{-1}C_r\tilde{\mathcal{U}}_{r,\mu_s}^{00}R_sV_s^{-1}\,dr\,dz_s \\
&= \int_0^t \mathcal{U}_{t,s}\Gamma\tilde{\psi}_s\,ds + \int_0^t \mathcal{U}_{t,s}P_s^{00}C_s^*V_s^{-1}\,d\bar{z}_s + \int_0^{\max(\mu_t^{-1},0)} \mathcal{U}_{t,\mu_s}R_sV_s^{-1}\,dz_s \\
&\quad - \int_0^{\max(\mu_t^{-1},0)} \int_0^r \mathcal{U}_{t,\mu_r}R_rV_r^{-1}C_r\tilde{\mathcal{U}}_{r,s}^{00}\Gamma\tilde{\psi}_s\,ds\,dr \\
&\quad - \int_0^{\max(\mu_t^{-1},0)} \int_0^r \mathcal{U}_{t,\mu_r}R_rV_r^{-1}C_r\tilde{\mathcal{U}}_{r,s}^{00}P_s^{00}C_s^*V_s^{-1}\,d\bar{z}_s\,dr \\
&\quad - \int_0^{\max(\mu_t^{-1},0)} \int_0^{\max(\mu_r^{-1},0)} \mathcal{U}_{t,\mu_r}R_rV_r^{-1}C_r\tilde{\mathcal{U}}_{r,\mu_s}^{00}R_sV_s^{-1}\,dz_s\,dr \\
&= \int_0^t \mathcal{U}_{t,s}\Gamma\tilde{\psi}_s\,ds + \int_0^t \mathcal{U}_{t,s}P_s^{00}C_s^*V_s^{-1}\,d\bar{z}_s + \int_0^{\max(\mu_t^{-1},0)} \mathcal{U}_{t,\mu_s}R_sV_s^{-1}\,dz_s \\
&\quad - \int_0^{\max(\mu_t^{-1},0)} \mathcal{U}_{t,\mu_r}R_rV_r^{-1}C_r\hat{x}_r\,dr \\
&= \int_0^t \mathcal{U}_{t,s}\Gamma\tilde{\psi}_s\,ds + \int_0^t \mathcal{U}_{t,s}P_s^{00}C_s^*V_s^{-1}\,d\bar{z}_s + \int_0^{\max(\mu_t^{-1},0)} \mathcal{U}_{t,\mu_s}R_sV_s^{-1}\,d\bar{z}_s.
\end{aligned}$$

Thus, the expression (9.43) for $\hat{x}_t$ is obtained. For the further calculations, recall the following modification of the formula (3.27):

$$\begin{aligned}
&\left[\int_{\max(\mu_t^{-1},0)}^t \mathcal{T}_{\mu_r-t}^*\Gamma^*g_r\,dr\right]_\theta \\
&\quad = \left[\int_{\max(0,\mu_0-t)}^{\mu_t-t} \left(\mu^{-1}\right)'_{t+r}\mathcal{T}_r^*\Gamma^*g_{\mu_{t+r}^{-1}}\,dr\right]_\theta \\
&\quad = \left\{\begin{array}{ll} \left(\mu^{-1}\right)'_{t-\theta}g_{\mu_{t-\theta}^{-1}}, & t-\mu_t \le \theta \le \min(0,t-\mu_0) \\ 0, & \text{otherwise} \end{array}\right\}, \text{a.e. } \theta \in [-\varepsilon,0], \quad (9.52)
\end{aligned}$$

where $g \in L_2(0,T;X)$. Then

$$\begin{aligned}
\tilde{N}_t \int_{\max(\mu_t^{-1},0)}^{t} & \mathcal{T}^*_{\mu_r - t}\Gamma^* R_r V_r^{-1} C_r \hat{x}_r \, dr \\
&= \int_{t-\mu_t}^{\min(0,t-\mu_0)} \left(\mu^{-1}\right)'_{t-\theta} N_{t-\theta,t} R_{\mu^{-1}_{t-\theta}} V^{-1}_{\mu^{-1}_{t-\theta}} C_{\mu^{-1}_{t-\theta}} \hat{x}_{\mu^{-1}_{t-\theta}} \, d\theta \\
&= \int_{\max(\mu_t^{-1},0)}^{t} N_{\mu_s,t} R_s V_s^{-1} C_s \hat{x}_s \, ds.
\end{aligned}$$

By (9.47) and (9.50), this implies

$$\begin{aligned}
\hat{y}_t &= M_t \hat{x}_t + \tilde{N}_t \tilde{\psi}_t + \int_{\max(\mu_t^{-1},0)}^{t} N_{\mu_s,t} R_s V_s^{-1} \left(dz_s - C_s \hat{x}_s ds\right) \\
&= M_t \hat{x}_t + \tilde{N}_t \tilde{\psi}_t + \int_{\max(\mu_t^{-1},0)}^{t} N_{\mu_s,t} R_s V_s^{-1} \, d\bar{z}_s,
\end{aligned}$$

proving (9.42). □

Example 9.8. If $M_t \equiv I$ and $N_{s,t} \equiv 0$ (assuming that $X = U$), then the filtering problem (9.1)–(9.3) is reduced to the filtering problem of finding the best estimate $\hat{x}_t$ of x_t based on the observations z_s, $0 \le s \le t$, where

$$x_t = \mathcal{U}_{t,0} x_0 + \int_0^t \mathcal{U}_{t,s} \Phi_s \, dw_s, \ 0 \le t \le T, \tag{9.53}$$

$$z_t = \int_0^t C_s x_s \, ds + \int_0^t \Psi_s \, dv_{\mu_s}, \ 0 \le t \le T. \tag{9.54}$$

From Theorem 9.7, it follows that under the conditions ($\mathbf{E_1^s}$)–($\mathbf{E_2^s}$) the best estimate $\hat{x}$ has the representation (9.43) where $\tilde{\psi}$ is defined by (9.44) and $\left(P^{00}, \tilde{P}^{01}\right)$ is a solution of (9.39)–(9.40).

Example 9.9 (Navigation of Earth orbiting satellites). Assume that the conditions ($\mathbf{E_1^s}$)–($\mathbf{E_2^s}$) hold so that $\mathcal{U} \in \mathcal{S}(X)$ and A is the infinitesimal generator of $\mathcal{U}$ and let

$$\mu_t = t + \varepsilon, \ 0 \le t \le T.$$

In Section 9.1 it was shown that this function μ describes the shift arising in navigation of Earth orbiting satellites. Then from Theorem 9.7 it follows that the best estimate $\hat{x}_t$ of x_t based on the observations z_s, $0 \le s \le t$, in the filtering problem (9.53)–(9.54) is a mild solution of the equation

$$\begin{aligned}
d\hat{x}_t &= \left(A\hat{x}_t + \Gamma\tilde{\psi}_t\right)dt + P_t^{00} C_t^* \left(\Psi_t \bar{V} \Psi_t^*\right)^{-1} d\bar{z}_t \\
&\quad + \begin{Bmatrix} \Phi_t \bar{R} \bar{V}^{-1} \Psi_{t-\varepsilon}^{-1}, & t > \varepsilon, \\ 0, & t \le \varepsilon, \end{Bmatrix} d\bar{z}_{t-\varepsilon}, \ \hat{x}_0 = 0, \ 0 < t \le T,
\end{aligned}$$

where $\tilde{\psi}$ is a mild solution of the equation

$$d\tilde{\psi}_t = -\frac{d}{d\theta}\tilde{\psi}_t dt + \tilde{P}_t^{01*}C_t^*\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}d\bar{z}_t,\ \tilde{\psi}_0 = 0,\ 0 < t \le T.$$

An application of Proposition 3.26 to the Riccati equation (9.27) yields $\tilde{P}^{01*} \in L_\infty(\mathbf{T}, \mathcal{L}_2(X, \tilde{X}))$. Hence, by Proposition 2.39,

$$\left[\tilde{P}_t^{01*}h\right]_\theta = P_{t,\theta}^{01*}h,\ \text{a.e. } \theta \in [-\varepsilon, 0],\ 0 \le t \le T,\ h \in X,$$

where $P^{01} \in L_\infty(\mathbf{T}, L_2(-\varepsilon, 0; \mathcal{L}_2(X)))$. Also, from (9.44), it follows that

$$\begin{aligned}\left[\tilde{\psi}_t\right]_\theta &= \left[\int_0^t \mathcal{T}_{t-s}\tilde{P}_s^{01*}C_s^*\left(\Psi_s\bar{V}\Psi_s^*\right)^{-1}d\bar{z}_s\right]_\theta \\ &= \int_{\max(0,t-\varepsilon-\theta)}^t P_{s,s-t+\theta}^{01*}C_s^*\left(\Psi_s\bar{V}\Psi_s^*\right)^{-1}d\bar{z}_s.\end{aligned}$$

Hence, if $\left[\tilde{\psi}_t\right]_\theta = \psi_{t,\theta}$, then $\psi_{t,-\varepsilon} = 0$.

Resuming and using (9.4), we obtain the following equations for the best estimate $\hat{x}$:

$$\begin{cases} d\hat{x}_t = (A\hat{x}_t + \psi_{t,0})dt + P_t^{00}C_t^*\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}d\bar{z}_t + \Phi_t\bar{R}\bar{V}^{-1}\Psi_{t-\varepsilon}^{-1}d\bar{z}_{t-\varepsilon}, \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta}dt = P_{t,\theta}^{01*}C_t^*\left(\Psi_t\bar{V}\Psi_t^*\right)^{-1}d\bar{z}_t, \\ \hat{x}_0 = 0,\ \psi_{0,\theta} = \psi_{t,-\varepsilon} = 0,\ \bar{z}_\theta = 0,\ -\varepsilon \le \theta \le 0,\ 0 < t \le T. \end{cases} \tag{9.55}$$

In order to complete the equations (9.55) for the optimal filter, the respective differential equations for the components of the Riccati equation (9.27) must be derived. This will be done in the next section where we will employ another method of study.

9.3 State Noise Delaying Observation Noise: Prediction

The prediction problem under shifted white noises will be discussed for the partially observable linear system (9.53)–(9.54) rather than the system (9.1)–(9.3) that was used in case of the filtering problem from Section 9.2.1. We will assume that the conditions $(\mathbf{E_1^s})$–$(\mathbf{E_2^s})$ hold and use the notation from (9.4). Also, we will assume that $0 \le \tau \le t \le T$.

Introduce the notation from (9.6) and consider the linear regulator problem of minimizing the functional

$$\begin{aligned} J(\eta) = \langle\xi_t^\eta, Q_t\xi_t^\eta\rangle &+ \int_{t-\tau}^t \left(\langle\xi_s^\eta, F_s\xi_s^\eta\rangle + \langle\eta_s, G_s\eta_s\rangle\right)ds \\ &+ 2\int_{t-\tau}^t \left\langle \eta_s, L_s\begin{Bmatrix} \xi_{\nu_s}^\eta, & \nu_s > 0 \\ 0, & \nu_s \le 0 \end{Bmatrix}\right\rangle ds, \end{aligned} \tag{9.56}$$

where

$$\xi_s^\eta = -\mathcal{R}_{s,0}l + \int_{t-\tau}^{\max(s,t-\tau)} \mathcal{R}_{s,r}B_r\eta_r\,dr,\ 0 \le s \le t, \tag{9.57}$$

and η is an admissible control taken from $U_{\text{ad}} = L_2(t-\tau, t; \mathbb{R}^n)$.

Theorem 9.10. *Let $0 \le \tau \le t \le T$. Then under the conditions $(\mathbf{E}_1^s)$–$(\mathbf{E}_2^s)$ and (9.6), the best estimate $\hat{x}_t^\tau$ of x_t based on z_s, $0 \le s \le \tau$, in the prediction problem (9.53)–(9.54) is equal to*

$$\hat{x}_t^\tau = \int_0^\tau K_s\,dz_s,$$

where $K \in B_2(0,\tau; \mathcal{L}(\mathbb{R}^n, X))$, if and only if the function defined by $\eta_s^* = K_{t-s}^* l$, a.e. $s \in [t-\tau, t]$, is an optimal control in the linear regulator problem (9.56)–(9.57) for all $l \in X$.

Proof. This can be proved in a similar way as Theorem 6.3 comparing the respective Wiener–Hopf equations. □

By Theorem 9.10, the linear regulator problem (9.56)–(9.57) is dual to the prediction problem (9.53)–(9.54). To find an optimal control in the linear regulator problem (9.56)–(9.57), write a control $\eta \in L_2(t-\tau, t; \mathbb{R}^n)$ in the form

$$\eta_s = \zeta_s - G_s^{-1}L_s \begin{Bmatrix} \xi_{\nu_s}^\eta, & \nu_s > 0 \\ 0, & \nu_s \le 0 \end{Bmatrix},\ \text{a.e. } s \in [t-\tau, t], \tag{9.58}$$

where $\zeta \in L_2(t-\tau, t; \mathbb{R}^n)$. Substituting (9.58) in (9.56)–(9.57), we obtain that the function $\eta = \eta^*$ is an optimal control in the problem (9.56)–(9.57) if and only if the function $\zeta = \zeta^*$, which is related with $\eta = \eta^*$ as in (9.58), is an optimal control in the linear regulator problem of minimizing the functional

$$J_1(\zeta) = \langle \rho_t^\zeta, Q_t\rho_t^\zeta\rangle + \int_{t-\tau}^t \left(\langle \rho_s^\zeta, \check{F}_s\rho_s^\zeta\rangle + \langle \zeta_s, G_s\zeta_s\rangle\right) ds \tag{9.59}$$

with

$$\begin{aligned} \rho_s^\zeta = {} & -\mathcal{R}_{s,0}l + \int_{t-\tau}^{\max(s,t-\tau)} \mathcal{R}_{s,r}B_r\zeta_r\,dr \\ & - \int_{t-\tau}^{\max(s,t-\tau)} \mathcal{R}_{s,r}B_rG_r^{-1}L_r \begin{Bmatrix} \rho_{\nu_r}^\zeta, & \nu_r > 0 \\ 0, & \nu_r \le 0 \end{Bmatrix} dr,\ 0 \le s \le t, \end{aligned} \tag{9.60}$$

where $\rho^\zeta = \xi^\eta$ if ζ and η are related as in (9.58) and $\check{F}$ is defined by (9.12). Note that, the values of the function $\check{F}$ are nonnegative operators (see Section 9.2.1).

Let $\tilde{X} = L_2(-\varepsilon, 0; X)$ and let $\check{X} = W^{1,2}(-\varepsilon, 0; X)$. Define $\mathcal{T}$ and Γ by (3.4) and (3.23), respectively. Let $\tilde{\rho}^\zeta$ and $\bar{\rho}^\zeta$ be the tilde and bar functions on $[t-\tau, t]$ over ρ^ζ with the initial distribution

$$\lambda_\theta = -\begin{Bmatrix} \mathcal{R}_{t-\tau+\theta,0}l, & \theta \ge \tau - t \\ 0, & \theta < \tau - t \end{Bmatrix},\ -\varepsilon \le \theta \le 0,$$

(see Section 3.4.2), i.e.,

$$\tilde{\rho}_s^\zeta = \begin{bmatrix} \rho_s^\zeta \\ \bar{\rho}_s^\zeta \end{bmatrix} \in X \times \tilde{X},\ t-\tau \le s \le t,\ \text{with } \tilde{\rho}_{t-\tau}^\zeta = \tilde{\rho}_{t-\tau} = -\begin{bmatrix} \mathcal{R}_{t-\tau,0} l \\ \lambda \end{bmatrix},$$

$$[\bar{\rho}_s^\zeta]_\theta = \begin{Bmatrix} \rho_{s+\theta}^\zeta, & s+\theta > 0 \\ 0, & s+\theta \le 0 \end{Bmatrix},\ \text{a.e. } \theta \in [-\varepsilon, 0],\ t-\tau \le s \le t.$$

Using the results of Section 3.4.2, the linear regulator problem (9.59)–(9.60) can be written in terms of the process $\tilde{\rho}^\zeta$ in the following form:

$$J_1(\zeta) = \langle \tilde{\rho}_t^\zeta, \tilde{Q}_t \tilde{\rho}_t^\zeta \rangle + \int_{t-\tau}^t \big(\langle \tilde{\rho}_s^\zeta, \tilde{F}_s \tilde{\rho}_s^\zeta \rangle + \langle \zeta_s, G_s \zeta_s \rangle \big) ds, \tag{9.61}$$

$$\tilde{\rho}_s^\zeta = \tilde{\mathcal{R}}_{s,t-\tau} \tilde{\rho}_{t-\tau} + \int_{t-\tau}^s \tilde{\mathcal{R}}_{s,r} \tilde{B}_r \zeta_r \, dr,\ t-\tau \le s \le t, \tag{9.62}$$

where $\tilde{\mathcal{R}}$, $\tilde{F}$, $\tilde{Q}_t$ and $\tilde{B}$ are defined by (9.15)–(9.17). Theorem 5.24 can be applied to the problem (9.61)–(9.62). Accordingly, there exists a unique optimal control in the problem (9.61)–(9.62) (and, respectively, in the problem (9.59)–(9.60)) and it has the form

$$\begin{cases} \zeta_s^* = -G_s^{-1} \tilde{B}_s^* \tilde{Q}_s \tilde{\rho}_s^*, \ \text{a.e. } s \in [t-\tau, t], \\ \tilde{\rho}_s^* = \tilde{\mathcal{K}}_{s,t-\tau} \tilde{\rho}_{t-\tau},\ t-\tau \le s \le t, \end{cases} \tag{9.63}$$

where $\tilde{\rho}^* = \tilde{\rho}^{\zeta^*}$, $\tilde{\mathcal{K}} = \mathcal{P}_{-\tilde{B}G^{-1}\tilde{B}^*\tilde{Q}}(\tilde{\mathcal{R}})$ and $\tilde{Q}$ is a unique solution of the Riccati equation (9.20). Thus, there exists a unique optimal control in the problem (9.56)–(9.57) too, and it has the form

$$\eta_s^* = -G_s^{-1} \tilde{B}_s^* \tilde{Q}_s \tilde{\mathcal{K}}_{s,t-\tau} \tilde{\rho}_{t-\tau}^* - G_s^{-1} L_s \begin{Bmatrix} \tilde{I} \tilde{\mathcal{K}}_{\nu_s, t-\tau} \tilde{\rho}_{t-\tau}^*, & \nu_s > t-\tau \\ \lambda_{\nu_s - t + \tau}, & \nu_s \le t-\tau \end{Bmatrix}. \tag{9.64}$$

In order to express $\tilde{\rho}_{t-\tau}^*$ in terms of l, define the operator $\mathcal{E}_{t,\tau} \in \mathcal{L}(\tilde{X}, X)$ by

$$\mathcal{E}_{t,\tau} f = \int_{\max(-\varepsilon, \tau - t)}^0 \mathcal{U}_{t,\tau-\theta} f_\theta \, d\theta,\ f \in \tilde{X}. \tag{9.65}$$

One can easily compute that the adjoint $\mathcal{E}_{t,\tau}^* \in \mathcal{L}(X, \tilde{X})$ of $\mathcal{E}_{t,\tau}$ is

$$\begin{aligned} [\mathcal{E}_{t,\tau}^* h]_\theta &= \begin{Bmatrix} \mathcal{U}_{t,\tau-\theta}^* h, & \theta > \tau - t \\ 0, & \theta \le \tau - t \end{Bmatrix} \\ &= \begin{Bmatrix} \mathcal{R}_{t-\tau+\theta,0} h, & \theta > \tau - t \\ 0, & \theta \le \tau - t \end{Bmatrix},\ -\varepsilon \le \theta \le 0,\ h \in X. \end{aligned}$$

Hence, $\lambda = -\mathcal{E}_{t,\tau}^* l$ and $\tilde{\rho}_{t-\tau}^* = -\tilde{\mathcal{E}}_{t,\tau}^* l$ where

$$\tilde{\mathcal{E}}_{t,\tau}^* = \begin{bmatrix} \mathcal{R}_{t-\tau,0} \\ \mathcal{E}_{t,\tau}^* \end{bmatrix} \in \mathcal{L}(X, X \times \tilde{X}).$$

Consequently, we obtain that there exists a unique optimal control in the linear regulator problem (9.56)–(9.57) and it has the form

$$\eta_s^* = G_s^{-1}\tilde{B}_s^*\tilde{Q}_s\tilde{\mathcal{K}}_{s,t-\tau}\tilde{\mathcal{E}}_{t,\tau}^* l + G_s^{-1}L_s \left\{ \begin{array}{ll} \tilde{I}\tilde{\mathcal{K}}_{\nu_s,t-\tau}\tilde{\mathcal{E}}_{t,\tau}^* l, & \nu_s > t-\tau \\ \mathcal{R}_{\nu_s,0}l, & 0 \le \nu_s \le t-\tau \\ 0, & \nu_s < 0 \end{array} \right\}, \text{ a.e. } s \in [t-\tau, t], \tag{9.66}$$

where $\tilde{I}$ is defined by (9.24). Thus, we can state the following theorem.

Theorem 9.11. *Let* $0 \le \tau \le t \le T$. *Then under the conditions* ($\mathbf{E_1^s}$)–($\mathbf{E_2^s}$) *there exist a unique optimal linear feedback predictor in the prediction problem* (9.53)–(9.54) *and the best estimate* $\hat{x}_t^\tau$ *of* x_t *based on* z_s, $0 \le s \le \tau$, *is equal to*

$$\hat{x}_t^\tau = \mathcal{U}_{t,\tau}\hat{x}_\tau + \mathcal{E}_{t,\tau}\tilde{\psi}_\tau + \int_{\max(\mu_\tau^{-1},0)}^{\min(\mu_t^{-1},\tau)} \left\{ \begin{array}{ll} \mathcal{U}_{t,\mu_s}, & t \ge \mu_0 \\ 0, & t < \mu_0 \end{array} \right\} R_s V_s^{-1}\, d\bar{z}_s, \tag{9.67}$$

where $\hat{x}$, $\tilde{\psi}$ *and* $\bar{z}$ *are defined by* (9.43)–(9.45).

Proof. Let $\tilde{\mathcal{Y}}$, $\tilde{\mathcal{U}}$, $\tilde{P}$, $\tilde{C}$ and $\tilde{I}$ be defined by (9.26), (9.27), (9.30) and (9.24). Then using (9.66) and Theorem 9.10, in a similar way as in the proof of Theorem 9.5, one can obtain

$$\hat{x}_t^\tau = \tilde{\mathcal{E}}_{t,\tau}\left(\int_0^\tau \tilde{\mathcal{Y}}_{\tau,s}\tilde{P}_s\tilde{C}_s^* V_s^{-1}\, dz_s + \int_0^{\max(\mu_\tau^{-1},0)} \tilde{\mathcal{Y}}_{\tau,\mu_s}\tilde{I}^* R_s V_s^{-1}\, dz_s \right) + \int_{\max(\mu_\tau^{-1},0)}^{\min(\mu_t^{-1},\tau)} \left\{ \begin{array}{ll} \mathcal{U}_{t,\mu_s}, & t \ge \mu_0 \\ 0, & t < \mu_0 \end{array} \right\} R_s V_s^{-1}\, dz_s.$$

Hence, using (9.46) and (9.50), we obtain

$$\begin{aligned} \hat{x}_t^\tau &= \mathcal{U}_{t,\tau}\hat{x}_\tau + \mathcal{E}_{t,\tau}\tilde{\phi}_\tau + \int_{\max(\mu_\tau^{-1},0)}^{\min(\mu_t^{-1},\tau)} \left\{ \begin{array}{ll} \mathcal{U}_{t,\mu_s}, & t \ge \mu_0 \\ 0, & t < \mu_0 \end{array} \right\} R_s V_s^{-1}\, dz_s \\ &= \mathcal{U}_{t,\tau}\hat{x}_\tau + \mathcal{E}_{t,\tau}\tilde{\psi}_\tau - \mathcal{E}_{t,\tau}\int_{\max(\mu_\tau^{-1},0)}^{\tau} \mathcal{T}_{\mu_r-\tau}^*\Gamma^* R_r V_r^{-1} C_r \hat{x}_r\, dr \\ &\quad + \int_{\max(\mu_\tau^{-1},0)}^{\min(\mu_t^{-1},\tau)} \left\{ \begin{array}{ll} \mathcal{U}_{t,\mu_s}, & t \ge \mu_0 \\ 0, & t < \mu_0 \end{array} \right\} R_s V_s^{-1}\, dz_s. \end{aligned} \tag{9.68}$$

In a similar way as in the proof of Theorem 9.7, one can show that

$$\mathcal{E}_{t,\tau}\int_{\max(\mu_\tau^{-1},0)}^{\tau} \mathcal{T}_{\mu_r-\tau}^*\Gamma^* R_r V_r^{-1} C_r \hat{x}_r\, dr = \int_{\max(\mu_\tau^{-1},0)}^{\min(\mu_t^{-1},\tau)} \left\{ \begin{array}{ll} \mathcal{U}_{t,\mu_s}, & t \ge \mu_0 \\ 0, & t < \mu_0 \end{array} \right\} R_s V_s^{-1} C_s \hat{x}_s\, ds.$$

Using this equality in (9.68), we obtain (9.67). □

9.4 State Noise Delaying Observation Noise: Smoothing

Similar to the prediction problem from Section 9.2.2, the smoothing problem under shifted white noises will be discussed for the system (9.53)–(9.54). We will assume that the conditions $(\mathbf{E_1^s})$–$(\mathbf{E_2^s})$ hold and use the notation from (9.4). Also, we will assume that $0 \le t \le \tau \le T$.

Introduce the notation

$$\begin{cases} \mathcal{R} = \mathcal{D}_\tau(\mathcal{U}), \ B = D_\tau(C), \ F = D_\tau(W), \\ G = D_\tau(V), \ L = D_\tau(R), \ Q_\tau = P_0, \ \nu = D_\tau(\mu), \end{cases} \tag{9.69}$$

which is very similar to the notation from (9.6) and consider the linear regulator problem of minimizing the functional

$$\begin{aligned} J(\eta) = \langle \xi_\tau^\eta, Q_\tau \xi_\tau^\eta \rangle + \int_0^\tau (\langle \xi_s^\eta, F_s \xi_s^\eta \rangle + \langle \eta_s, G_s \eta_s \rangle)\, ds \\ + 2 \int_0^\tau \left(\left\langle \eta_s, L_s \begin{Bmatrix} \xi_{\nu_s}^\eta, & \nu_s > 0, \\ 0, & \nu_s \le 0 \end{Bmatrix} \right\rangle \right) ds, \end{aligned} \tag{9.70}$$

where

$$\xi_s^\eta = - \begin{Bmatrix} \mathcal{R}_{s,\tau-t} l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix} + \int_0^s \mathcal{R}_{s,r} B_r \eta_r \, dr, \ 0 \le s \le \tau, \tag{9.71}$$

and η is an admissible control taken from $U_{\text{ad}} = L_2(0, \tau; \mathbb{R}^n)$.

Theorem 9.12. *Let $0 \le t \le \tau \le T$. Then under the conditions $(\mathbf{E_1^s})$–$(\mathbf{E_2^s})$ and (9.69), the best estimate $\hat{x}_t^\tau$ of x_t based on z_s, $0 \le s \le \tau$, in the smoothing problem (9.53)–(9.54) is equal to*

$$\hat{x}_t^\tau = \int_0^\tau K_s \, dz_s,$$

where $K \in B_2(0, \tau; \mathcal{L}(\mathbb{R}^n, X))$, if and only if the function defined by $\eta_s^* = K_{\tau-s}^* l$, a.e. $s \in [0, \tau]$, is an optimal control in the linear regulator problem (9.70)–(9.71) for all $l \in X$.

Proof. This can be proved in a similar way as Theorem 6.3 comparing the respective Wiener–Hopf equations. □

Thus, the smoothing problem (9.53)–(9.54) and the linear regulator problem (9.70)–(9.71) are dual. To find an optimal control in the linear regulator problem (9.70)–(9.71), write a control $\eta \in L_2(0, \tau; \mathbb{R}^n)$ in the form

$$\eta_s = \zeta_s - G_s^{-1} L_s \begin{Bmatrix} \xi_{\nu_s}^\eta, & \nu_s > 0 \\ 0, & \nu_s \le 0 \end{Bmatrix}, \text{ a.e. } s \in [0, \tau], \tag{9.72}$$

where $\zeta \in L_2(0, \tau; \mathbb{R}^n)$. Substituting (9.72) in (9.70)–(9.71), we obtain that the function $\eta = \eta^*$ is an optimal control in the problem (9.70)–(9.71) if and only if the function $\zeta = \zeta^*$, which is related with $\eta = \eta^*$ as in (9.72), is an optimal control in the linear regulator problem of minimizing the functional

$$J_1(\zeta) = \langle \rho_\tau^\zeta, Q_\tau \rho_\tau^\zeta \rangle + \int_0^\tau \left(\langle \rho_s^\zeta, \check{F}_s \rho_s^\zeta \rangle + \langle \zeta_s, G_s \zeta_s \rangle \right) ds \tag{9.73}$$

with

$$\rho_s^\zeta = -\begin{Bmatrix} \mathcal{R}_{s,\tau-t} l, & s \geq \tau - t \\ 0, & s < \tau - t \end{Bmatrix} + \int_0^s \mathcal{R}_{s,r} B_r \zeta_r \, dr$$
$$- \int_0^s \mathcal{R}_{s,r} B_r G_r^{-1} L_r \begin{Bmatrix} \rho_{\nu_r}^\zeta, & \nu_r > 0 \\ 0, & \nu_r \leq 0 \end{Bmatrix} dr, \ 0 \leq s \leq \tau, \tag{9.74}$$

where $\rho^\zeta = \xi^\eta$ if ζ and η are related as in (9.72) and $\check{F}$ is defined by

$$\check{F}_s = \begin{Bmatrix} F_s - L_{\nu_s^{-1}}^* G_{\nu_s^{-1}}^{-1} L_{\nu_s^{-1}} \left(\nu^{-1}\right)_s', & s < \nu_t \\ F_s, & s \geq \nu_t \end{Bmatrix}, \ \text{a.e. } s \in [0, \tau].$$

Note that the values of the function $\check{F}$ are nonnegative operators (see Section 9.2.3).

Now let $\tilde{\mathcal{U}} = \mathcal{P}_{\Lambda(-RV^{-1}C,\mu)}(\mathcal{U} \odot \mathcal{T})$ be decomposed in the form of (9.32) with the components $\tilde{\mathcal{U}}^{00}$, $\tilde{\mathcal{U}}^{01}$, $\tilde{\mathcal{U}}^{10}$ and $\tilde{\mathcal{U}}^{11}$ as defined by (9.33)–(9.38). Define $\tilde{\mathcal{R}} = \mathcal{D}_\tau(\tilde{\mathcal{U}})$ and decompose it in the form

$$\tilde{\mathcal{R}}_{s,r} = \begin{bmatrix} \tilde{\mathcal{R}}_{s,r}^{00} & \tilde{\mathcal{R}}_{s,r}^{01} \\ \tilde{\mathcal{R}}_{s,r}^{10} & \tilde{\mathcal{R}}_{s,r}^{11} \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}), \ 0 \leq r \leq s \leq \tau.$$

Consider

$$\tilde{\rho}_s^\zeta = \begin{bmatrix} \rho_{1,s}^\zeta \\ \rho_{2,s}^\zeta \end{bmatrix} = -\begin{Bmatrix} \tilde{\mathcal{R}}_{s,\tau-t} \tilde{I}^* l, & s \geq \tau - t \\ 0, & s < \tau - t \end{Bmatrix} + \int_0^s \tilde{\mathcal{R}}_{s,r} \tilde{B}_r \zeta_r \, dr, \ 0 \leq s \leq \tau, \tag{9.75}$$

where $\tilde{I}$ is defined by (9.24) and

$$\tilde{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \in B_\infty\left(0, \tau; \mathcal{L}(\mathbb{R}^n, X \times \tilde{X})\right).$$

Lemma 9.13. *Under the above conditions, the upper component ρ_1^ζ of $\tilde{\rho}^\zeta$, defined by* (9.75), *is equal to the state process ρ^ζ, defined by* (9.74).

Proof. From $\tilde{\mathcal{R}} = \mathcal{D}_\tau(\tilde{\mathcal{U}})$ and from (9.34),

$$\mathcal{R}_{s,r} = \tilde{\mathcal{R}}_{s,r}^{00} + \int_{\min(\nu_r^{-1}, s)}^s \tilde{\mathcal{R}}_{s,\sigma}^{00} B_\sigma G_\sigma^{-1} L_\sigma \mathcal{R}_{\nu_\sigma, r} \, d\sigma, \ 0 \leq r \leq s \leq \tau.$$

Using this equality and (9.74), we obtain

$$
\begin{aligned}
\rho_s^\zeta - \rho_{1,s}^\zeta &= \rho_s^\zeta + \begin{Bmatrix} \tilde{\mathcal{R}}^{00}_{s,\tau-t} l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix} + \int_0^s \tilde{\mathcal{R}}^{00}_{s,r} B_r \zeta_r \, dr \\
&= -\int_0^s \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \begin{Bmatrix} \rho^\zeta_{\nu_\sigma}, & \nu_\sigma > 0 \\ 0, & \nu_\sigma \le 0 \end{Bmatrix} d\sigma \\
&\quad - \int_{\min(\nu_{\tau-t}^{-1}, s)}^s \begin{Bmatrix} \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \mathcal{R}_{\nu_\sigma,\tau-t} l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix} d\sigma \\
&\quad + \int_0^s \int_{\min(\nu_r^{-1}, s)}^s \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \mathcal{R}_{\nu_\sigma, r} B_r \zeta_r \, d\sigma \, dr \\
&\quad - \int_0^s \int_{\min(\nu_r^{-1}, s)}^s \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \mathcal{R}_{\nu_\sigma, r} B_r G_r^{-1} L_r \begin{Bmatrix} \rho^\zeta_{\nu_r}, & \nu_r > 0 \\ 0, & \nu_r \le 0 \end{Bmatrix} d\sigma \, dr \\
&= -\int_{\min(\nu_0^{-1}, s)}^s \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \rho^\zeta_{\nu_\sigma} \, d\sigma \\
&\quad - \int_{\min(\nu_0^{-1}, s)}^s \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \begin{Bmatrix} \mathcal{R}_{\nu_\sigma,\tau-t} l, & \nu_\sigma \ge \tau - t \\ 0, & \nu_\sigma < \tau - t \end{Bmatrix} d\sigma \\
&\quad + \int_{\min(\nu_0^{-1}, s)}^s \int_0^{\nu_\sigma} \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \mathcal{R}_{\nu_\sigma, r} B_r \zeta_r \, dr \, d\sigma \\
&\quad - \int_{\min(\nu_0^{-1}, s)}^s \int_0^{\nu_\sigma} \tilde{\mathcal{R}}^{00}_{s,\sigma} B_\sigma G_\sigma^{-1} L_\sigma \mathcal{R}_{\nu_\sigma, r} B_r G_r^{-1} L_r \rho^\zeta_{\nu_r} \, dr \, d\sigma = 0.
\end{aligned}
$$

Thus, $\rho_s^\zeta = \rho_{1,s}^\zeta$, $0 \le s \le \tau$. □

Remark 9.14. In fact $\tilde{\rho}^\zeta$ and $\bar{\rho}^\zeta = \rho_2^\zeta$ from (9.75) are the tilde and bar functions on $[0,\tau]$ over ρ^ζ with the initial distribution $\lambda = 0$ (see Section 3.4.2). In the rest of this section this fact will not be used.

By Lemma 9.13, the functional (9.73) can be rewritten in the form

$$
J_1(\zeta) = \langle \tilde{\rho}_\tau^\zeta, \tilde{Q}_\tau \tilde{\rho}_\tau^\zeta \rangle + \int_0^\tau \left(\langle \tilde{\rho}_s^\zeta, \tilde{F}_s \tilde{\rho}_s^\zeta \rangle + \langle \zeta_s, G_s \zeta_s \rangle \right) ds, \tag{9.76}
$$

where $\tilde{\rho}^\zeta$ is the state process defined by (9.75), ζ is a control in $L_2(0,\tau;\mathbb{R}^n)$ and

$$
\tilde{F} = \begin{bmatrix} \check{F} & 0 \\ 0 & 0 \end{bmatrix} \in B_\infty(0,\tau;\mathcal{L}(X \times \tilde{X})), \ \tilde{Q}_\tau = \begin{bmatrix} Q_\tau & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{L}(X \times \tilde{X}).
$$

Let

$$
\tilde{\beta}_s = \begin{Bmatrix} \tilde{\mathcal{R}}_{s,\tau-t} \tilde{I}^* l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix}, \ \tilde{\alpha}_s^\zeta = \tilde{\rho}_s^\zeta + \tilde{\beta}_s, \ 0 \le s \le \tau. \tag{9.77}
$$

Then the linear regulator problem (9.75)–(9.76) can be reduced to the problem of minimizing

$$J_1(\zeta) = \langle \tilde{\alpha}^\zeta_\tau - \tilde{\beta}_\tau, \tilde{Q}_\tau(\tilde{\alpha}^\zeta_\tau - \tilde{\beta}_\tau)\rangle + \int_0^\tau \left(\langle \tilde{\alpha}^\zeta_s - \tilde{\beta}_s, \tilde{F}_s(\tilde{\alpha}^\zeta_s - \tilde{\beta}_s)\rangle + \langle \zeta_s, G_s\zeta_s\rangle\right) ds, \tag{9.78}$$

where

$$\tilde{\alpha}^\zeta_s = \int_0^s \tilde{\mathcal{R}}_{s,r}\tilde{B}_r\zeta_r\, dr,\ 0 \le s \le \tau. \tag{9.79}$$

Applying Theorem 5.24 to the problem (9.78)–(9.79), we obtain that there exists a unique optimal control ζ^* in the problem (9.78)–(9.79) (and (9.75)–(9.76) as well) and it has the form:

$$\zeta^*_s = -G_s^{-1}\tilde{B}^*_s(\tilde{Q}_s\tilde{\alpha}^*_s + \tilde{\gamma}_s),\ 0 \le s \le \tau, \tag{9.80}$$

$$\tilde{\alpha}^*_s = \int_0^s \tilde{\mathcal{K}}_{s,r}\tilde{B}_r G_r^{-1}\tilde{B}^*_r\tilde{\gamma}_r\, dr,\ 0 \le s \le \tau, \tag{9.81}$$

$$\tilde{\gamma}_s = -\tilde{\mathcal{K}}^*_{\tau,s}\tilde{Q}_\tau\tilde{\beta}_\tau - \int_s^\tau \tilde{\mathcal{K}}^*_{r,s}\tilde{F}_r\tilde{\beta}_r\, dr,\ 0 \le s \le \tau, \tag{9.82}$$

where $\tilde{\alpha}^* = \tilde{\alpha}^{\zeta^*}$, $\tilde{\mathcal{K}} = \mathcal{P}_{-\tilde{B}G^{-1}\tilde{B}^*\tilde{Q}}(\tilde{\mathcal{R}})$ and $\tilde{Q}$ is a unique solution of the Riccati equation

$$\tilde{Q}_s = \tilde{\mathcal{R}}^*_{\tau,s}\tilde{Q}_\tau\tilde{\mathcal{R}}_{\tau,s} + \int_s^\tau \tilde{\mathcal{R}}^*_{r,s}(\tilde{F}_r - \tilde{Q}_r\tilde{B}_r G_r^{-1}\tilde{B}^*_r\tilde{Q}_r)\tilde{\mathcal{R}}_{r,s}\, dr,\ 0 \le s \le \tau. \tag{9.83}$$

Lemma 9.15. *Under the above notation,*

$$\tilde{\gamma}_s = -\begin{Bmatrix} \tilde{Q}_s\tilde{\mathcal{R}}_{s,\tau-t}\tilde{I}^*l, & s \ge \tau - t \\ \tilde{\mathcal{K}}^*_{\tau-t,s}\tilde{Q}_{\tau-t}\tilde{I}^*l, & s < \tau - t \end{Bmatrix},\ 0 \le s \le \tau.$$

Proof. By Proposition 3.23, the equation (9.83) is equivalent to

$$\tilde{Q}_s = \tilde{\mathcal{K}}^*_{\tau,s}\tilde{Q}_\tau\tilde{\mathcal{R}}_{\tau,s} + \int_s^\tau \tilde{\mathcal{K}}^*_{r,s}\tilde{F}_r\tilde{\mathcal{R}}_{r,s}\, dr,\ 0 \le s \le \tau.$$

Therefore, if $\tau - t \le s \le \tau$, then by (9.82) and (9.77), we have

$$\begin{aligned}\tilde{\gamma}_s &= -\tilde{\mathcal{K}}^*_{\tau,s}\tilde{Q}_\tau\tilde{\mathcal{R}}_{\tau,\tau-t}\tilde{I}^*l - \int_s^\tau \tilde{\mathcal{K}}^*_{r,s}\tilde{F}_r\tilde{\mathcal{R}}_{r,\tau-t}\tilde{I}^*l\, dr \\ &= -\left(\tilde{\mathcal{K}}^*_{\tau,s}\tilde{Q}_\tau\tilde{\mathcal{R}}_{\tau,s} + \int_s^\tau \tilde{\mathcal{K}}^*_{r,s}\tilde{F}_r\tilde{\mathcal{R}}_{r,s}\, dr\right)\tilde{\mathcal{R}}_{s,\tau-t}\tilde{I}^*l \\ &= -\tilde{Q}_s\tilde{\mathcal{R}}_{s,\tau-t}\tilde{I}^*l.\end{aligned}$$

If $0 \le s < \tau - t$, then in a similar way, we have

$$\begin{aligned}\tilde{\gamma}_s &= -\tilde{\mathcal{K}}^*_{\tau,s}\tilde{Q}_\tau\tilde{\mathcal{R}}_{\tau,\tau-t}l - \int_{\tau-t}^{\tau} \tilde{\mathcal{K}}^*_{r,s}\tilde{F}_r\tilde{\mathcal{R}}_{r,\tau-t}\tilde{I}^*l\,dr \\ &= -\tilde{\mathcal{K}}^*_{\tau-t,s}\left(\tilde{\mathcal{K}}^*_{\tau,\tau-t}\tilde{Q}_\tau\tilde{\mathcal{R}}_{\tau,\tau-t} + \int_{\tau-t}^{\tau} \tilde{\mathcal{K}}^*_{r,\tau-t}\tilde{F}_r\tilde{\mathcal{R}}_{r,\tau-t}\,dr\right)\tilde{I}^*l \\ &= -\tilde{\mathcal{K}}^*_{\tau-t,s}\tilde{Q}_{\tau-t}\tilde{I}^*l.\end{aligned}$$

Thus, the proof is completed. □

Lemma 9.16. *Under the above notation,*

$$\tilde{\alpha}^*_s = \int_0^{\min(s,\tau-t)} \tilde{\mathcal{K}}_{s,r}\tilde{B}_rG_r^{-1}\tilde{B}^*_r\tilde{\mathcal{K}}^*_{\tau-t,r}\tilde{Q}_{\tau-t}\tilde{I}^*l\,dr + \begin{Bmatrix} (\tilde{\mathcal{R}}_{s,\tau-t} - \tilde{\mathcal{K}}_{s,\tau-t})\tilde{I}^*l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix}, \quad 0 \le s \le \tau.$$

Proof. By (9.81) and by Lemma 9.15, we have

$$\tilde{\alpha}^*_s = \int_0^{\min(s,\tau-t)} \tilde{\mathcal{K}}_{s,r}\tilde{B}_rG_r^{-1}\tilde{B}^*_r\tilde{\mathcal{K}}^*_{\tau-t,r}\tilde{Q}_{\tau-t}\tilde{I}^*l\,dr + \int_{\min(s,\tau-t)}^{s} \tilde{\mathcal{K}}_{s,r}\tilde{B}_rG_r^{-1}\tilde{B}^*_r\tilde{Q}_r\tilde{\mathcal{R}}_{r,\tau-t}\tilde{I}^*l\,dr.$$

Since $\tilde{\mathcal{K}} = \mathcal{P}_{-\tilde{B}G^{-1}\tilde{B}^*\tilde{Q}}(\tilde{\mathcal{R}})$, the last equality implies the statement. □

From Lemmas 9.15 and 9.16,

$$\tilde{Q}_s\tilde{\alpha}^*_s + \tilde{\gamma}_s = \int_0^{\min(s,\tau-t)} \tilde{Q}_s\tilde{\mathcal{K}}_{s,r}\tilde{B}_rG_r^{-1}\tilde{B}^*_r\tilde{\mathcal{K}}^*_{\tau-t,r}\tilde{Q}_{\tau-t}\tilde{I}^*l\,dr - \begin{Bmatrix} \tilde{Q}_s\tilde{\mathcal{K}}_{s,\tau-t}\tilde{I}^*l, & s \ge \tau - t \\ \tilde{\mathcal{K}}_{\tau-t,s}\tilde{Q}_{\tau-t}\tilde{I}^*l, & s < \tau - t \end{Bmatrix}, \quad 0 \le s \le \tau.$$

Also, from (9.77) and Lemma 9.16,

$$\tilde{\rho}^{\zeta^*}_s = \tilde{\alpha}^*_s - \tilde{\beta}_s = \int_0^{\min(s,\tau-t)} \tilde{\mathcal{K}}_{s,r}\tilde{B}_rG_r^{-1}\tilde{B}^*_r\tilde{\mathcal{K}}^*_{\tau-t,r}\tilde{Q}_{\tau-t}\tilde{I}^*l\,dr - \begin{Bmatrix} \tilde{\mathcal{K}}_{s,\tau-t}\tilde{I}^*l, & s \ge \tau - t \\ 0, & s < \tau - t \end{Bmatrix}, \quad 0 \le s \le \tau.$$

Using the last two equalities and (9.80) in (9.72), we obtain that there exists a unique optimal control in the linear regulator problem (9.70)–(9.71) and it has the form

$$
\begin{aligned}
\eta_s^* = &- G_s^{-1}\tilde{B}_s^*\tilde{Q}_s \int_0^{\min(s,\tau-t)} \tilde{\mathcal{K}}_{s,r}\tilde{B}_r G_r^{-1}\tilde{B}_r^*\tilde{\mathcal{K}}_{\tau-t,r}^*\tilde{Q}_{\tau-t}\tilde{I}^* l\, dr \\
&- G_s^{-1}L_s\tilde{I}\int_{\min(\nu_s,0)}^{\min(\nu_s,\tau-t)} \tilde{\mathcal{K}}_{\nu_s,r}\tilde{B}_r G_r^{-1}\tilde{B}_r^*\tilde{\mathcal{K}}_{\tau-t,r}^*\tilde{Q}_{\tau-t}\tilde{I}^* l\, dr \\
&+ G_s^{-1}\tilde{B}_s^*\left\{\begin{matrix} \tilde{Q}_s\tilde{\mathcal{K}}_{s,\tau-t}\tilde{I}^* l, & s \ge \tau - t \\ \tilde{\mathcal{K}}_{\tau-t,s}^*\tilde{Q}_{\tau-t}\tilde{I}^* l, & s < \tau - t \end{matrix}\right\} \\
&+ G_s^{-1}L_s\tilde{I}\left\{\begin{matrix} \tilde{\mathcal{K}}_{\nu_s,\tau-t}\tilde{I}^* l, & \nu_s \ge \tau - t \\ 0, & \nu_s < \tau - t \end{matrix}\right\}, \ 0 \le s \le \tau. \qquad (9.84)
\end{aligned}
$$

This leads to the following.

Theorem 9.17. *Let $0 \le t \le \tau \le T$. Then under the conditions $(\mathbf{E_1^s})$–$(\mathbf{E_2^s})$, there exists a unique optimal linear feedback smoother in the smoothing problem* (9.53)–(9.54) *and the best estimate $\hat{x}_t^\tau$ of x_t based on z_s, $0 \le s \le \tau$, is equal to*

$$\hat{x}_t^\tau = \hat{x}_t + \tilde{I}\tilde{P}_t\int_t^\tau \tilde{\mathcal{Y}}_{s,t}^*\tilde{C}_s^* V_s^{-1}\, d\bar{z}_s, \qquad (9.85)$$

where $\hat{x}_t$ and $\bar{z}_t$ are defined by (9.43) *and* (9.45), *$\tilde{P}$ is a solution of the Riccati equation* (9.27) *and $\tilde{C}$, $\tilde{\mathcal{Y}}$ and $\tilde{I}$ are defined by* (9.30), (9.26) *and* (9.24), *respectively.*

Proof. Using the relations in (9.69), one can show that $\tilde{\mathcal{Y}} = \mathcal{D}_\tau(\tilde{\mathcal{K}})$, $\tilde{P} = D_\tau(\tilde{Q})$ and $\tilde{C} = D_\tau(\tilde{B})$, where $\tilde{\mathcal{K}}$, $\tilde{Q}$ and $\tilde{B}$ are the operator-valued functions from the right-hand side of (9.84). Hence, from Theorem 9.12 and from the formula (9.84), we have

$$
\begin{aligned}
\hat{x}_t^\tau = &\tilde{I}\int_0^t \tilde{\mathcal{Y}}_{t,s}\tilde{P}_s\tilde{C}_s^* V_s^{-1}\, dz_s + \tilde{I}\int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{Y}}_{t,\mu_s}\tilde{I}^* R_s V_s^{-1}\, dz_s \\
&+ \tilde{I}\int_t^\tau \tilde{P}_t\tilde{\mathcal{Y}}_{s,t}^*\tilde{C}_s^* V_s^{-1}\, dz_s \\
&- \tilde{I}\int_0^\tau\int_{\max(t,s)}^\tau \tilde{P}_t\tilde{\mathcal{Y}}_{r,t}^*\tilde{C}_r^* V_r^{-1}\tilde{C}_r\tilde{\mathcal{Y}}_{r,s}\tilde{P}_s\tilde{C}_s^* V_s^{-1}\, dr\, dz_s \\
&- \tilde{I}\int_0^\tau\int_{\max(\mu_s,t)}^{\max(\mu_s,\tau)} \tilde{P}_t\tilde{\mathcal{Y}}_{r,t}^*\tilde{C}_r^* V_r^{-1}\tilde{C}_r\tilde{\mathcal{Y}}_{r,\mu_s}\tilde{I}^* R_s V_s^{-1}\, dr\, dz_s.
\end{aligned}
$$

Letting $t = \tau$ in the last equality, we observe that the optimal estimate $\hat{x}_t$, defined by (9.43), can be represented in the form

$$\hat{x}_t = \hat{x}_t^t = \tilde{I}\int_0^t \tilde{\mathcal{Y}}_{t,s}\tilde{P}_s\tilde{C}_s^* V_s^{-1}\, dz_s + \tilde{I}\int_0^{\max(\mu_t^{-1},0)} \tilde{\mathcal{Y}}_{t,\mu_s}\tilde{I}^* R_s V_s^{-1}\, dz_s, \ 0 \le t \le T.$$

Hence,

$$
\begin{aligned}
\hat{x}_t^\tau &= \hat{x}_t + \tilde{I}\int_t^\tau \tilde{P}_t \tilde{\mathcal{Y}}_{s,t}^* \tilde{C}_s^* V_s^{-1}\, dz_s \\
&\quad - \tilde{I}\int_t^\tau \tilde{P}_t \tilde{\mathcal{Y}}_{r,t}^* \tilde{C}_r^* V_r^{-1} C_r \tilde{I}\Big(\int_0^r \tilde{\mathcal{Y}}_{r,s}\tilde{P}_s \tilde{C}_s^* V_s^{-1}\, dz_s \\
&\qquad + \int_0^{\max(\mu_r^{-1},0)} \tilde{\mathcal{Y}}_{r,\mu_s} \tilde{I}^* R_s V_s^{-1}\, dz_s\Big)\, dr \\
&= \hat{x}_t + \tilde{I}\int_t^\tau \tilde{P}_t \tilde{\mathcal{Y}}_{s,t}^* \tilde{C}_s^* V_s^{-1}\, dz_s - \tilde{I}\int_t^\tau \tilde{P}_t \tilde{\mathcal{Y}}_{r,t}^* \tilde{C}_r^* V_r^{-1} C_r \hat{x}_r\, dr \\
&= \hat{x}_t + \tilde{I}\int_t^\tau \tilde{P}_t \tilde{\mathcal{Y}}_{s,t}^* \tilde{C}_s^* V_s^{-1}\, d\bar{z}_s
\end{aligned}
$$

proving the theorem. □

Remark 9.18. The form of the optimal predictor under shifted white noises from Theorem 9.11 differs from that under correlated white noises from Theorem 6.13. This is because the observations on $[0,\tau]$ are related with the signal process x_t on $\tau < t \le \tau + \varepsilon$ in case of the signal noise delaying the observation noise while they are independent in case of correlated white noises. But the optimal smoothers from Theorems 6.18 and 9.17 are very similar to each other.

9.5 State Noise Delaying Observation Noise: Stochastic Regulator Problem

Consider the problem (5.1)–(5.4) in which the state-observation system (5.1)–(5.2) and the functional (5.4) are defined in the form

$$x_t^u = \mathcal{U}_{t,0}x_0 + \int_0^t \mathcal{U}_{t,s}B_s u_s\, ds + \int_0^t \mathcal{U}_{t,s}\Phi_s\, dw_s,\ 0 \le t \le T, \tag{9.86}$$

$$z_t^u = \int_0^t C_s x_s^u\, ds + \int_0^t \Psi_s\, dv_{\mu_s},\ 0 \le t \le T, \tag{9.87}$$

$$J(u) = \mathbf{E}\Bigg(\langle x_T^u, Q_T x_T^u\rangle + \int_0^T \Bigg\langle \begin{bmatrix} x_t^u \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix}\begin{bmatrix} x_t^u \\ u_t \end{bmatrix}\Bigg\rangle\, dt\Bigg), \tag{9.88}$$

and a control u is taken from the set of admissible controls U_{ad} as defined by (5.3) in Section 5.1.2. This problem will be called the linear stochastic regulator problem (9.86)–(9.88).

In this section we assume that the following conditions hold:

$(\mathbf{R}_1^{\mathbf{s}})$ $\mathcal{U} \in \mathcal{E}(\Delta_T, \mathcal{L}(X))$, $B \in B_\infty(\mathbf{T}, \mathcal{L}(U,X))$, $C \in B_\infty(\mathbf{T}, \mathcal{L}(X, \mathbb{R}^n))$;

($\mathbf{R_2^s}$) $\Phi \in B_\infty(\mathbf{T}, \mathcal{L}(H, X))$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, $\begin{bmatrix} w \\ v \end{bmatrix}$ is an $H \times \mathbb{R}^n$-valued Wiener process on $[0, T+\varepsilon]$ with $\operatorname{cov} v_T > 0$, $0 < \varepsilon < T$, $\mu \in W^{1,\infty}(\mathbf{T}, \mathbb{R})$ is a function satisfying $t \le \mu_t \le t + \varepsilon$ for $0 \le t \le T$, $\mu_s < \mu_t$ for $0 \le s < t \le T$ and $\mu_t' \ge c$ for a.e. $t \in \mathbf{T}$ and for some $c > 0$, x_0 is a an X-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0 and (w, v) are independent;

($\mathbf{R_3^s}$) $Q_T \in \mathcal{L}(X)$, $Q_T \ge 0$, $F \in B_\infty(\mathbf{T}, \mathcal{L}(X))$, $G, G^{-1} \in B_\infty(\mathbf{T}, \mathcal{L}(U))$, $L \in B_\infty(\mathbf{T}, \mathcal{L}(X, U))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \ge 0$ for a.e. $t \in \mathbf{T}$.

Note that ($\mathbf{R_1^s}$) is the same as ($\mathbf{E_1^s}$) completed with the condition about B, and ($\mathbf{R_2^s}$) and ($\mathbf{R_3^s}$) are the same as ($\mathbf{E_2^s}$) and ($\mathbf{C_3}$), respectively.

In this section we will use the operator-valued functions W, R and V defined by (9.4).

Theorem 9.19. *Under the conditions* ($\mathbf{R_1^s}$)–($\mathbf{R_3^s}$), *there exists a unique optimal stochastic regulator in the problem* (9.86)–(9.88) *and the respective optimal control has the form*

$$\begin{aligned} u_t^* = &- G_t^{-1}(B_t^* Q_t + L_t)\hat{x}_t^* - G_t^{-1} B_t^* \int_t^{\min(T, t+\varepsilon)} \mathcal{Y}_{s,t}^* Q_s \big[\tilde{\psi}_t\big]_{t-s} \, ds \\ &- G_t^{-1} B_t^* \int_{\max\left(0, \mu_t^{-1}\right)}^{\min\left(t, \mu_T^{-1}\right)} \mathcal{Y}_{\mu_s, t}^* Q_{\mu_s} R_s V_s^{-1} \, d\bar{z}_s, \ \textit{a.e. } t \in \mathbf{T}, \end{aligned} \tag{9.89}$$

where

$$\begin{aligned} \hat{x}_t^* = &\int_0^t \mathcal{U}_{t,s}\big(\Gamma \tilde{\psi}_s + B_s u_s^*\big) \, ds + \int_0^t \mathcal{U}_{t,s} P_s^{00} C_s^* V_s^{-1} \, d\bar{z}_s \\ &+ \int_0^{\max\left(0, \mu_t^{-1}\right)} \mathcal{U}_{t, \mu_s} R_s V_s^{-1} \, d\bar{z}_s, \ 0 \le t \le T, \end{aligned} \tag{9.90}$$

$$\tilde{\psi}_t = \int_0^t \mathcal{T}_{t-s} \tilde{P}_s^{01*} C_s^* V_s^{-1} \, d\bar{z}_s, \ 0 \le t \le T, \tag{9.91}$$

$\mathcal{Y} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{U})$, *$Q$ is a solution of the Riccati equation* (3.9), *the pair* $\big(P^{00}, \tilde{P}^{01}\big)$ *is a solution of the system of equations* (9.39)–(9.40) *and $\bar{z}$ is the innovation process defined by*

$$d\bar{z}_t = dz_t^* - C_t \hat{x}_t^* \, dt, \ 0 < t \le T, \ \bar{z}_0 = 0, \tag{9.92}$$

with $z^* = z^{u^*}$.

Proof. Let $M_t = B_t^* Q_t + L_t$ and $N_{s,t} = B_t^* \mathcal{Y}_{s,t}^* Q_s$ in Theorem 9.7. Then from Theorems 5.28 and 9.7, it is easily seen that there exists a unique optimal stochastic regulator in the problem (9.86)–(9.88) and, by Theorem 5.16, the respective optimal control has the form

$$u_t^* = - G_t^{-1}(B_t^* Q_t + L_t)\mathbf{E}_t^* x_t^* - G_t^{-1} B_t^* \mathbf{E}_t^* \int_t^T \mathcal{Y}_{s,t}^* Q_s \Phi_s \, dw_s, \ \text{a.e. } t \in \mathbf{T}, \tag{9.93}$$

where $x_t^* = x_t^{u^*}$ and $\mathbf{E}_t^* = \mathbf{E}(\cdot\,|z_s^*; 0 \le s \le t)$. Since $u^* \in U_{\mathrm{ad}}$, by Proposition 5.5, $\mathbf{E}_t^* = \mathbf{E}_t^0 = \mathbf{E}\big(\cdot\,|z_s^0; 0 \le s \le t\big)$, where z^0 is the observation process corresponding to zero control $u = 0$. Similar to the proof of Proposition 6.19, one can show that in the definition (9.92) of the innovation process $\bar{z}$, the processes z^* and x^* can be replaced by z^u and x^u, respectively, for any $u \in U_{\mathrm{ad}}$. Hence, (9.92) and (9.45) define the same process $\bar{z}$. Thus, using (9.42)–(9.44) in (9.93), we obtain the formulae (9.89)–(9.91) for the optimal control u^* in the problem (9.86)–(9.88). □

Example 9.20 (Navigation of Earth orbiting satellites). Consider a linear regulator version of the filtering problem from Example 9.9. For this assume that the conditions $(\mathbf{R}_1^s)$–$(\mathbf{R}_3^s)$ hold so that $\mathcal{U} \in \mathcal{S}(X)$ and A is the infinitesimal generator of $\mathcal{U}$ and let $\mu_t = t + \varepsilon$, $0 \le t \le T$. Then from Theorem 9.19 and Example 9.9, it follows that the optimal control u^* in the problem (9.86)–(9.88) has the form

$$\begin{aligned} u_t^* = &- G_t^{-1}(B_t^* Q_t + L_t)\hat{x}_t^* - G_t^{-1} B_t^* \int_t^{\min(T,t+\varepsilon)} \mathcal{Y}_{s,t}^* Q_s \psi_{t,t-s}\, ds \\ &- G_t^{-1} B_t^* \int_{\max(\varepsilon,t)}^{\min(T,t+\varepsilon)} \mathcal{Y}_{s,t}^* Q_s \Phi_s \bar{R} \bar{V}^{-1} \Psi_{s-\varepsilon}^{-1}\, d\bar{z}_{s-\varepsilon},\ 0 \le t \le T, \end{aligned}$$

where $\hat{x}^*$ satisfies (in the mild sense)

$$\begin{cases} d\hat{x}_t^* = (A\hat{x}_t^* + \psi_{t,0} + B_t u_t^*)dt + P_t^{00} C_t^* \big(\Psi_t \bar{V} \Psi_t^*\big)^{-1} d\bar{z}_t + \Phi_t \bar{R} \bar{V}^{-1} \Psi_{t-\varepsilon}^{-1} d\bar{z}_{t-\varepsilon}, \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta} dt = P_{t,\theta}^{01*} C_t^* \big(\Psi_t \bar{V} \Psi_t^*\big)^{-1} d\bar{z}_t, \\ \hat{x}_0^* = 0,\ \psi_{0,\theta} = \psi_{t,-\varepsilon} = 0, \bar{z}_\theta = 0, a.e.\ \theta \in [-\varepsilon, 0],\ 0 < t \le T, \end{cases}$$

Q is a solution of the equation (3.9) and P^{01} is as defined in Example 9.9.

9.6 Concluding Remarks

After hard calculations in this section we obtained the formulae (9.89)–(9.91) for the optimal control and for the optimal filter in the linear stochastic regulator and filtering problems determined by (9.86)–(9.88). The application of these formulae to navigation of Earth orbiting satellites was considered in Example 9.20. A disadvantage of the method applied in this section is that it requires a routine of calculations. More malleable and promising method of study will be presented in the next section.

Finalizing this section, note that the formulae (9.89)–(9.91) as well as (9.67) can be written in more convenient form. To demonstrate, let us use the informal equality

$$d\bar{z}_t = \bar{z}_t' dt.$$

Introduce a new function φ by

$$\tilde{\varphi}_t = \tilde{\psi}_t + \int_{\max(\mu_t^{-1},0)}^t \mathcal{T}_{\mu_r - t}^* \Gamma^* R_r V_r^{-1} \bar{z}_r'\, dr,\ 0 \le t \le T,$$

where $\tilde{\psi}$ is defined by (9.91). Similar to (9.52), we can write

$$\begin{aligned}
&\left[\int_{\max(\mu_t^{-1},0)}^{t} \mathcal{T}^*_{\mu_r - t}\Gamma^* R_r V_r^{-1} \bar{z}'_r \, dr\right]_\theta \\
&= \left[\int_{\max(0,\mu_0 - t)}^{\mu_t - t} \left(\mu^{-1}\right)'_{t+r} \mathcal{T}^*_r \Gamma^* R_{\mu^{-1}_{t+r}} V^{-1}_{\mu^{-1}_{t+r}} \bar{z}'_{\mu^{-1}_{t+r}} \, dr\right]_\theta \\
&= \left\{ \begin{array}{ll} \left(\mu^{-1}\right)'_{t-\theta} R_{\mu^{-1}_{t-\theta}} V^{-1}_{\mu^{-1}_{t-\theta}} \bar{z}'_{\mu^{-1}_{t-\theta}}, & t - \mu_t \le \theta \le \min(0, t - \mu_0) \\ 0, & \text{otherwise} \end{array} \right\}.
\end{aligned}$$

Using this equality one can derive

$$\begin{aligned}
\int_t^{\min(T,t+\varepsilon)} \mathcal{Y}^*_{s,t} Q_s [\tilde{\varphi}_t]_{t-s} \, ds = \int_t^{\min(T,t+\varepsilon)} \mathcal{Y}^*_{s,t} Q_s [\tilde{\psi}_t]_{t-s} \, ds \\
+ \int_{\max(0,\mu_t^{-1})}^{\min(t,\mu_T^{-1})} \mathcal{Y}^*_{\mu_s,t} Q_{\mu_s} R_s V_s^{-1} \, d\bar{z}_s, \ \text{a.e. } t \in \mathbf{T}, \\
\int_0^t \mathcal{U}_{t,s} \Gamma \tilde{\varphi}_s \, ds = \int_0^t \mathcal{U}_{t,s} \Gamma \tilde{\psi}_s \, ds + \int_0^{\max(0,\mu_t^{-1})} \mathcal{U}_{t,\mu_s} R_s V_s^{-1} \, d\bar{z}_s.
\end{aligned}$$

Hence, the equations (9.89)–(9.91) can be written as

$$\begin{aligned}
u_t^* &= - G_t^{-1}(B_t^* Q_t + L_t)\hat{x}_t^* - G_t^{-1} B_t^* \int_t^{\min(T,t+\varepsilon)} \mathcal{Y}^*_{s,t} Q_s [\tilde{\varphi}_t]_{t-s} \, ds, \ \text{a.e. } t \in \mathbf{T}, \\
\hat{x}_t^* &= \int_0^t \mathcal{U}_{t,s}\left(\Gamma\tilde{\varphi}_s + B_s u_s^*\right) ds + \int_0^t \mathcal{U}_{t,s} P_s^{00} C_s^* V_s^{-1} \, d\bar{z}_s, \ 0 \le t \le T, \\
\tilde{\varphi}_t &= \int_0^t \mathcal{T}_{t-s} \tilde{P}_s^{01*} C_s^* V_s^{-1} \, d\bar{z}_s + \int_{\max(\mu_t^{-1},0)}^{t} \mathcal{T}^*_{\mu_s - t} \Gamma^* R_s V_s^{-1} \, d\bar{z}_s, \ 0 \le t \le T.
\end{aligned}$$

Also, one can derive

$$\mathcal{E}_{t,\tau}\tilde{\varphi}_\tau = \mathcal{E}_{t,\tau}\tilde{\psi}_\tau + \int_{\max(\mu_\tau^{-1},0)}^{\min(\mu_t^{-1},\tau)} \left\{ \begin{array}{ll} \mathcal{U}_{t,\mu_s}, & t \ge \mu_0 \\ 0, & t < \mu_0 \end{array} \right\} R_s V_s^{-1} \, d\bar{z}_s,$$

implying a more simple equation

$$\hat{x}_t^\tau = \mathcal{U}_{t,\tau}\hat{x}_\tau + \mathcal{E}_{t,\tau}\tilde{\varphi}_\tau, \ 0 \le \tau \le t \le T,$$

for the optimal predictor than in (9.67).

Chapter 10

Control and Estimation under Shifted White Noises (Revised)

In this chapter we apply a method, based on the convergence of wide band noise processes to white noise, to derive differential equations for the optimal controls and the optimal filters in the linear stochastic regulator and estimation problems under shifted white noises. This chapter is a logical continuation of the previous one but it can be read independently. The reader is recommended to read Section 9.1 for preliminary discussion of shifted white noises.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $T > 0$ and $\mathbf{T} = [0, T]$.

10.1 Preliminaries

In Chapter 9 the linear stochastic regulator and estimation problems under shifted white noises with the state noise delaying the observation noise were studied via the duality principle and the extended separation principle. Some formulae in the integral form were obtained for the respective best estimates and the optimal control. The method of study used in Chapter 9 does not allow further developments in the theory and getting more delicate results. Therefore, in this chapter we change the method of study.

In this chapter we will assume that $0 < \varepsilon < T$ and consider the function $\lambda \in W^{1,\infty}(\mathbf{T}, \mathbb{R})$ satisfying

$$t - \varepsilon \le \lambda_t \le t,\ 0 \le t \le T,\ \text{and}\ \lambda_s < \lambda_t,\ 0 \le s < t \le T. \tag{10.1}$$

Since λ is increasing and continuous, its inverse λ^{-1} exists on $[\lambda_0, \lambda_T]$ and it is continuous.

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

We will consider two different partially observable systems under shifted noises. For this, we will consider a Wiener process w on the interval $[-\varepsilon, T]$. In the first system the noises of the state (signal) and the observations will be formed by the random processes

$$w_{\lambda_t} \text{ and } w_t, \ 0 \le t \le T,$$

respectively. Thus, the noise process of the state (signal) will be a pointwise delay of the noise process acting on the observations. This is similar to the system studied in Chapter 9 where the state (signal) and observation noises were formed by the Wiener processes w_t and v_{μ_t}, respectively, with the function μ satisfying the conditions in $(\mathbf{E}_2^{\mathrm{s}})$ or in $(\mathbf{R}_2^{\mathrm{s}})$. In this chapter we prefer to consider a left translation of w instead of the right translation of v that was used in Chapter 9. One can observe that both these translations lead to the same case when the state (signal) noise is a delay of the observation noise.

In the second system the noises of the state and the observations will be formed by the random processes

$$w_t \text{ and } (v_t, w_{\lambda_t}), \ 0 \le t \le T,$$

respectively, where v is a nondegenerate Wiener process independent of w. Thus, the second component of the noise process of the observations will be a pointwise delay of the noise process acting on the signal. Consequently, both the state (signal) noise delaying and anticipating the observation noise will be covered.

Three particular cases of the function λ, which are symmetric to the respective cases of the function μ mentioned in Chapter 9, are as follows:

(a) $\lambda_t = t$ (identity);

(b) $\lambda_t = t - \varepsilon$ with $\varepsilon > 0$ (left translation);

(c) $\lambda_t = ct$ with $0 < c < 1$ (rotation).

These cases will form our examples to demonstrate the theory. Note that in accordance to the discussion given in Section 9.1, the cases (b) and (c) form shifts arising in navigation of spacecraft.

Our main aim in this chapter is to derive the formulae for the optimal filters and optimal controls in differential form. For this, we will try two methods. At first, in Section 10.2 we discuss the reduction method, successfully used in Chapters 7 and 8, and show that this method meets a lot of difficulties as applied to shifted noises. Then in succeeding sections of this chapter we create a new promising method that is based on the convergence of wide band noise processes to white noise. We use the control and filtering results from Chapter 8 tending the wide band noise processes in these problems to white noises in a certain way.

For simplicity the underlying state, control and observation spaces will be considered as finite dimensional Euclidean spaces. But, the results can be extended to infinite dimensional spaces as well.

10.2 Shifted White Noises and Boundary Noises

Let us investigate if the reduction method, successfully used in Chapters 7 and 8, could be useful in this chapter. For this consider the first order partial differential equation

$$\begin{cases} (\partial/\partial t + \partial/\partial\theta)\varphi_{t,\theta} = 0, \\ \varphi_{0,\theta} = 0, \ -\lambda_0^{-1} \le \theta \le 0, \\ \varphi_{t,t-\lambda_t^{-1}} dt = \lambda'_{\lambda_t^{-1}} \Phi_{\lambda_t^{-1}} dw_t, \ 0 \le t \le \lambda_T, \\ \varphi_{t,t-T} = 0, \ \lambda_T < t \le T. \end{cases} \tag{10.2}$$

This is a kind of stochastic partial differential equation disturbed by white noise along the boundary curve

$$\theta = t - \lambda_t^{-1}, \ 0 \le t \le \lambda_T.$$

Similar to the terminology used for the systems with boundary control and boundary observations, we will refer to the equation (10.2) as a stochastic equation with boundary noise.

Obviously, the solution of the equation (10.2) has the form $\varphi_{t,\theta} = f_{t-\theta}$ for some random process f. Hence, the zero initial condition in (10.2) implies $\varphi_{t,0} = \varphi_{0,-t} = 0$ for $0 \le t \le \lambda_0^{-1}$. Furthermore, substituting $s = \lambda_t^{-1}$ in the boundary condition, we obtain

$$\varphi_{\lambda_s,\lambda_s-s}\lambda'_s ds = \lambda'_s \Phi_s dw_{\lambda_s}, \ \lambda_0^{-1} \le s \le T.$$

Since $\varphi_{s,0} = f_s = \varphi_{\lambda_s,\lambda_s-s}$,

$$\varphi_{s,0} ds = \Phi_s dw_{\lambda_s}, \ \lambda_0^{-1} \le s \le T.$$

Thus, the differential $\Phi_s dw_{\lambda_s}$ that forms a shifted noise can be replaced by $\varphi_{s,0} ds$ where φ is the solution of the equation (10.2). Joining the equation (10.2) to the state (signal) system and enlarging the state (signal) space, we can thus reduce the partially observable system under shifted white noises to the system with correlated or independent white noises.

A disadvantage of this reduction is that the equation (10.2) contains a boundary noise. Respectively, the diffusion coefficient of the equation (10.2) is a function with unbounded operator values. Consequently, the control and filtering results from Chapter 6 can not be directly used for the reduced system. This means that the reduction method used in Chapters 7 and 8 is useless in this chapter. On the other hand, the limited nature of the duality principle as applied to problems with shifted noises was discussed in Chapter 9. Therefore, we need a more handle method for such problems. In the rest of this chapter we develop such a method that is based on approximations of white noises by wide band noises. This method is especially useful to derive optimal controls and optimal filters in differential form. It will be applied to the partially observable linear systems with the state (signal) noise delaying or anticipating the observation noise.

10.3 Convergence of Wide Band Noise Processes

10.3.1 Approximation of White Noises

Let $0 < \varepsilon < T$ and denote $\sigma = 2\varepsilon$ and $\varepsilon_n = 2^{-n}\varepsilon$. Consider the sequence of the deterministic real-valued functions $f_{n,\theta}$, $-\sigma \le \theta \le 0$:

$$f_{n,\theta} = \begin{Bmatrix} 0, & -\sigma \le \theta \le \varepsilon_n \\ 4\varepsilon_n^{-2}\theta + 4\varepsilon_n^{-1}, & -\varepsilon_n < \theta \le -\varepsilon_{n+1} \\ -4\varepsilon_n^{-2}\theta, & -\varepsilon_{n+1} < \theta \le 0 \end{Bmatrix}, \quad n = 1, 2, \ldots. \tag{10.3}$$

The function f_n has the properties: it vanishes on $[-\sigma, -\varepsilon_n]$ and at $\theta = 0$, but on the interval $(-\varepsilon_n, 0)$ it has a single peak $2\varepsilon_n^{-1}$ that corresponds to $\theta = -\varepsilon_{n+1}$. Moreover, f_n linearly increases on $(-\varepsilon_n, -\varepsilon_{n+1})$ and linearly decreases on $(-\varepsilon_{n+1}, 0)$ so that the triangle formed by the graph of this function and the θ-axis has a unit area, i.e.,

$$\int_{-\sigma}^{0} f_{n,\theta}\, d\theta = \int_{-\varepsilon_n}^{0} f_{n,\theta}\, d\theta = 1.$$

Since $f_{n,0} = 0$ and the length of the intervals $(-\varepsilon_n, 0)$ goes to 0 as $n \to \infty$, the pointwise limit of the functions f_n is the zero function on $[-\sigma, 0]$. But a uniform limit of the functions f_n does not exist since the interchange of the limit (as $n \to \infty$) and the integral does not hold for this sequence of functions:

$$\lim_{n\to\infty} \int_{-\sigma}^{0} f_{n,\theta}\, d\theta = 1 \ne 0 = \int_{-\sigma}^{0} \lim_{n\to\infty} f_{n,\theta}\, d\theta.$$

One can also calculate

$$\int_{-\sigma}^{0} f_{n,\theta}^2\, d\theta = \frac{4}{3\varepsilon_n}, \quad n = 1, 2, \ldots,$$

that will be used later for technical purposes.

Now let $\Phi \in C(\mathbf{T}, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k))$ and define the sequence of relaxing functions

$$\Phi_{t,\theta}^n = \Phi_t f_{n,\theta}, \; -\sigma \le \theta \le 0, \; 0 \le t \le T, \; n = 1, 2, \ldots. \tag{10.4}$$

Construct the sequence of wide band noise processes φ^n in the form

$$\varphi_t^n = \int_{\max(0, t-\sigma)}^{t} \Phi_{t,\theta-t}^n\, dw_\theta, \; 0 \le t \le T, \; n = 1, 2, \ldots, \tag{10.5}$$

where w is an $\mathbb{R}^n$-valued Wiener process.

Theorem 10.1. *Let $\Phi \in C(\mathbf{T}, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k))$ and let w be an $\mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with the unit covariance matrix. Then for every $N \in C(\mathbf{T}, \mathcal{L}(\mathbb{R}^k))$, the limit*

$$\lim_{n\to\infty} \int_0^t N_s \varphi_s^n\, ds = \int_0^t N_s \Phi_s\, dw_s$$

holds uniformly for $t \in \mathbf{T}$ *in the mean square sense, where* φ^n *is defined by* (10.3)–(10.5).

Proof. The convergence holds trivially for $t = 0$. Let $0 < t \leq T$. For sufficiently large n, we have $t > \varepsilon_n$. For these values of n,

$$\begin{aligned}
&\mathbf{E}\left\| \int_0^t N_s \varphi_s^n \, ds - \int_0^t N_s \Phi_s \, dw_s \right\|^2 \\
&\quad = \mathbf{E}\left\| \int_0^t \int_{\max(0, s-\varepsilon_n)}^s N_s \Phi_s f_{n,\theta-s} \, dw_\theta \, ds - \int_0^t N_s \Phi_s \, dw_s \right\|^2 \\
&\quad = \mathbf{E}\left\| \int_0^t \int_\theta^{\min(\theta+\varepsilon_n, t)} N_s \Phi_s f_{n,\theta-s} \, ds \, dw_\theta - \int_0^t N_\theta \Phi_\theta \, dw_\theta \right\|^2 \\
&\quad = \int_0^t \left\| \int_\theta^{\min(\theta+\varepsilon_n, t)} N_s \Phi_s f_{n,\theta-s} \, ds - N_\theta \Phi_\theta \right\|^2 d\theta \\
&\quad \leq \int_0^{t-\varepsilon_n} \left\| \int_\theta^{\theta+\varepsilon_n} N_s \Phi_s f_{n,\theta-s} \, ds - N_\theta \Phi_\theta \right\|^2 d\theta \\
&\qquad + 2 \int_{t-\varepsilon_n}^t \left\| \int_\theta^t N_s \Phi_s f_{n,\theta-s} \, ds \right\|^2 d\theta + 2 \int_{t-\varepsilon_n}^t \|N_\theta \Phi_\theta\|^2 d\theta. \qquad (10.6)
\end{aligned}$$

The first term in the right-hand side of (10.6) can be estimated as

$$\begin{aligned}
&\int_0^{t-\varepsilon_n} \left\| \int_\theta^{\theta+\varepsilon_n} N_s \Phi_s f_{n,\theta-s} \, ds - N_\theta \Phi_\theta \right\|^2 d\theta \\
&\quad = \int_0^{t-\varepsilon_n} \left\| \int_\theta^{\theta+\varepsilon_n} (N_s \Phi_s - N_\theta \Phi_\theta) f_{n,\theta-s} \, ds \right\|^2 d\theta \\
&\quad \leq \int_0^{t-\varepsilon_n} \left(\int_\theta^{\theta+\varepsilon_n} \|N_s \Phi_s - N_\theta \Phi_\theta\|^2 \, ds \cdot \int_\theta^{\theta+\varepsilon_n} f_{n,\theta-s}^2 \, ds \right) d\theta \\
&\quad \leq \frac{4}{3\varepsilon_n} \int_0^{t-\varepsilon_n} \int_\theta^{\theta+\varepsilon_n} \|N_s \Phi_s - N_\theta \Phi_\theta\|^2 \, ds \, d\theta \\
&\quad \leq \frac{4T}{3} \max_{\theta \in [0, t-\varepsilon_n]} \max_{s \in [0, \varepsilon_n]} \|N_{\theta+s} \Phi_{\theta+s} - N_\theta \Phi_\theta\|^2.
\end{aligned}$$

Since $N\Phi$ is continuous on $\mathbf{T}$, it is uniformly continuous on $\mathbf{T}$. Hence,

$$\max_{\theta \in [0, t-\varepsilon_n]} \max_{s \in [0, \varepsilon_n]} \|N_{\theta+s} \Phi_{\theta+s} - N_\theta \Phi_\theta\|^2 \to 0, \ n \to \infty, \ \text{uniformly in } t \in \mathbf{T},$$

proving that the uniform (for $t \in \mathbf{T}$) limit of the first term in the right-hand side of (10.6) is 0. To estimate the second term in the right-hand side of (10.6) let

$$K = \max_{t \in \mathbf{T}} \|N_t \Phi_t\| < \infty.$$

Clearly, $0 \le K < \infty$ since $N\Phi$ is continuous on the compact set $\mathbf{T}$. Then

$$\begin{aligned} 2\int_{t-\varepsilon_n}^{t} &\left\| \int_{\theta}^{t} N_s \Phi_s f_{n,\theta-s}\, ds \right\|^2 d\theta \\ &\le 2\int_{t-\varepsilon_n}^{t} \left(\int_{t-\varepsilon_n}^{t} \|N_s\Phi_s\|^2\, ds \cdot \int_{\theta}^{\theta+\varepsilon_n} f_{n,\theta-s}^2\, ds \right) d\theta \\ &\le \frac{8K^2\varepsilon_n^2}{3\varepsilon_n} \to 0,\ n \to \infty,\ \text{uniformly in}\ \ t \in \mathbf{T}. \end{aligned}$$

And, finally, for the last term in the right-hand side of (10.6), we have

$$2\int_{t-\varepsilon_n}^{t} \|N_\theta \Phi_\theta\|^2 d\theta \le 2K^2\varepsilon_n \to 0,\ n \to \infty,\ \text{uniformly in}\ \ t \in \mathbf{T}.$$

Thus, the limit holds uniformly in $t \in \mathbf{T}$. □

Theorem 10.1 demonstrates that a white noise process can be approximated by wide noise processes in a certain way. It has the following deterministic analog as well.

Corollary 10.2. *Let $\Phi \in C\big(-\sigma, 0; \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)\big)$. Then the following limit holds:*

$$\lim_{n\to\infty} \int_{-\sigma}^{0} \Phi_\theta f_{n,\theta}\, d\theta = \Phi_0.$$

Proof. This can be proved in the same manner as Theorem 10.1. □

10.3.2 Approximation of Shifted White Noises

Now let $\lambda \in W^{1,\infty}(\mathbf{T}, \mathbb{R})$ satisfy the conditions in (10.1). Substituting the relaxing functions from (10.4) by

$$\Phi_{t,\theta}^n = \left\{ \begin{array}{ll} \Phi_t f_{n,t+\theta-\lambda_t}\lambda_t', & \lambda_t - \varepsilon_n \le t+\theta \le \lambda_t \\ 0, & \text{otherwise} \end{array} \right\},\ 0 \le t \le T,\ -\sigma \le \theta \le 0, \quad (10.7)$$

we obtain the following modification of Theorem 10.1:

Theorem 10.3. *Let $\lambda \in W^{1,\infty}(\mathbf{T}, \mathbb{R})$ satisfy the conditions in* (10.1), *let $\Phi \in C\big(\mathbf{T}, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^k)\big)$ and let w be an $\mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with the unit covariance matrix. Then for every $N \in C\big(\mathbf{T}, \mathcal{L}(\mathbb{R}^k)\big)$, the limit*

$$\lim_{n\to\infty} \int_0^t N_s \varphi_s^n\, ds = \int_{\min\left(t,\lambda_0^{-1}\right)}^{t} N_s \Phi_s\, dw_{\lambda_s}$$

holds uniformly for $t \in \mathbf{T}$ in the mean square sense, where φ^n is defined by (10.3), (10.5) *and* (10.7).

Proof. This can be done in a similar way as the proof of Theorem 10.1. Assume that $0 \le t \le \lambda_0^{-1}$. Then

$$\int_0^t N_s \varphi_s^n \, ds = 0$$

and, hence, the limit holds trivially. Let $\lambda_0^{-1} < t \le T$. For sufficiently large n, we have $\varepsilon_n < \lambda_t$. For these values of n,

$$\begin{aligned}
&\mathbf{E}\left\| \int_0^t N_s \varphi_s^n \, ds - \int_{\lambda_0^{-1}}^t N_s \Phi_s \, dw_{\lambda_s} \right\|^2 \\
&\quad = \mathbf{E}\left\| \int_0^t \int_{\max(0,\lambda_s-\varepsilon_n)}^{\lambda_s} N_s \Phi_s f_{n,\theta-\lambda_s} \lambda_s' \, dw_\theta \, ds - \int_{\lambda_0^{-1}}^t N_s \Phi_s \, dw_{\lambda_s} \right\|^2 \\
&\quad = \mathbf{E}\left\| \int_0^{\lambda_t} \int_{\lambda_\theta^{-1}}^{\min\left(\lambda_{\theta+\varepsilon_n}^{-1},t\right)} N_s \Phi_s f_{n,\theta-\lambda_s} \lambda_s' \, ds \, dw_\theta - \int_0^{\lambda_t} N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \, dw_\theta \right\|^2 \\
&\quad = \int_0^{\lambda_t} \left\| \int_{\lambda_\theta^{-1}}^{\min\left(\lambda_{\theta+\varepsilon_n}^{-1},t\right)} N_s \Phi_s f_{n,\theta-\lambda_s} \lambda_s' \, ds - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 d\theta \\
&\quad \le \int_0^{\lambda_t-\varepsilon_n} \left\| \int_{\lambda_\theta^{-1}}^{\lambda_{\theta+\varepsilon_n}^{-1}} N_s \Phi_s f_{n,\theta-\lambda_s} \lambda_s' \, ds - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 d\theta \\
&\qquad + 2\int_{\lambda_t-\varepsilon_n}^{\lambda_t} \left\| \int_{\lambda_\theta^{-1}}^t N_s \Phi_s f_{n,\theta-\lambda_s} \lambda_s' \, ds \right\|^2 d\theta + 2\int_{\lambda_t-\varepsilon_n}^{\lambda_t} \left\| N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 d\theta. \qquad (10.8)
\end{aligned}$$

The first term in the right-hand side of (10.8) can be estimated as

$$\begin{aligned}
&\int_0^{\lambda_t-\varepsilon_n} \left\| \int_{\lambda_\theta^{-1}}^{\lambda_{\theta+\varepsilon_n}^{-1}} N_s \Phi_s f_{n,\theta-\lambda_s} \lambda_s' \, ds - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 d\theta \\
&\quad = \int_0^{\lambda_t-\varepsilon_n} \left\| \int_\theta^{\theta+\varepsilon_n} N_{\lambda_s^{-1}} \Phi_{\lambda_s^{-1}} f_{n,\theta-s} \, ds - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 d\theta \\
&\quad = \int_0^{\lambda_t-\varepsilon_n} \left\| \int_\theta^{\theta+\varepsilon_n} \left(N_{\lambda_s^{-1}} \Phi_{\lambda_s^{-1}} - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}}\right) f_{n,\theta-s} \, ds \right\|^2 d\theta \\
&\quad \le \int_0^{\lambda_t-\varepsilon_n} \left(\int_\theta^{\theta+\varepsilon_n} \left\| N_{\lambda_s^{-1}} \Phi_{\lambda_s^{-1}} - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 ds \cdot \int_\theta^{\theta+\varepsilon_n} f_{n,\theta-s}^2 \, ds \right) d\theta \\
&\quad \le \frac{4}{3\varepsilon_n} \int_0^{\lambda_t-\varepsilon_n} \int_\theta^{\theta+\varepsilon_n} \left\| N_{\lambda_s^{-1}} \Phi_{\lambda_s^{-1}} - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 ds \, d\theta \\
&\quad \le \frac{4T}{3} \max_{\theta\in[0,\lambda_{t-\varepsilon_n}]} \max_{s\in[0,\varepsilon_n]} \left\| N_{\lambda_{\theta+s}^{-1}} \Phi_{\lambda_{\theta+s}^{-1}} - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2.
\end{aligned}$$

Since the composition of the functions $N\Phi$ and λ^{-1} is continuous on the closed interval $[0, \lambda_T]$, it is uniformly continuous on $[0, \lambda_T]$. Hence,

$$\max_{\theta\in[0,\lambda_{t-\varepsilon_n}]} \max_{s\in[0,\varepsilon_n]} \left\| N_{\lambda_{\theta+s}^{-1}} \Phi_{\lambda_{\theta+s}^{-1}} - N_{\lambda_\theta^{-1}} \Phi_{\lambda_\theta^{-1}} \right\|^2 \to 0, \; n \to \infty, \text{ uniformly in } t \in \left[\lambda_0^{-1}, T\right],$$

proving that the uniform limit of the first term in the right-hand side of (10.8) is equal to 0. To estimate the second term in the right-hand side of (10.8) let

$$K = \max_{t\in[\lambda_0^{-1},T]} \|N_t\Phi_t\|.$$

Then

$$\begin{aligned}
&2\int_{\lambda_t-\varepsilon_n}^{\lambda_t} \left\| \int_{\lambda_\theta^{-1}}^{t} N_s\Phi_s f_{n,\theta-\lambda_s}\lambda_s'\, ds \right\|^2 d\theta \\
&\quad = 2\int_{\lambda_t-\varepsilon_n}^{\lambda_t} \left\| \int_{\theta}^{\lambda_t} N_{\lambda_s^{-1}}\Phi_{\lambda_s^{-1}} f_{n,\theta-s}\, ds \right\|^2 d\theta \\
&\quad \le 2\int_{\lambda_t-\varepsilon_n}^{\lambda_t} \left(\int_{\lambda_t-\varepsilon_n}^{\lambda_t} \left\|N_{\lambda_s^{-1}}\Phi_{\lambda_s^{-1}}\right\|^2 ds \cdot \int_{\theta}^{\theta+\varepsilon_n} f_{n,\theta-s}^2\, ds \right) d\theta \\
&\quad \le \frac{8K^2\varepsilon_n^2}{3\varepsilon_n} \to 0,\ n\to\infty,\ \text{uniformly in}\ \ t\in\left[\lambda_0^{-1},T\right].
\end{aligned}$$

And, finally, for the last term in the right-hand side of (10.8), we have

$$2\int_{\lambda_t-\varepsilon_n}^{\lambda_t} \|N_{\lambda_\theta^{-1}}\Phi_{\lambda_\theta^{-1}}\|^2 d\theta \le 2K^2\varepsilon_n \to 0,\ n\to\infty,\ \text{uniformly in}\ \ t\in\left[\lambda_0^{-1},T\right].$$

Thus, the limit holds uniformly for $t\in\mathbf{T}$. □

Theorem 10.3 demonstrates how to approximate a shifted white noise process by wide band noise processes.

10.4 State Noise Delaying Observation Noise

10.4.1 Setting of the Problem

Consider the problem (9.86)–(9.88) from Section 9.5 which we write in the following differential form:

$$dx_t = (Ax_t + B_t u_t)dt + \Phi_t dw_{\lambda_t},\ x_0 \text{ is given },\ 0 < t \le T, \tag{10.9}$$

$$dz_t = C_t x_t dt + \Psi_t dw_t,\ z_0 = 0,\ 0 < t \le T, \tag{10.10}$$

$$J(u) = \mathbf{E}\left(\langle x_T, Q_T x_T\rangle + \int_0^T \left\langle \begin{bmatrix} x_t \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} \right\rangle dt\right). \tag{10.11}$$

In this section $A\in\mathcal{L}(\mathbb{R}^k)$, $B\in L_\infty(\mathbf{T},\mathcal{L}(\mathbb{R}^m,\mathbb{R}^k))$, $C\in L_\infty(\mathbf{T},\mathcal{L}(\mathbb{R}^k,\mathbb{R}^n))$, $\Phi\in W^{1,2}(\mathbf{T},\mathcal{L}(\mathbb{R}^n,\mathbb{R}^k))$, $\Psi,\Psi^{-1}\in L_\infty(\mathbf{T},\mathcal{L}(\mathbb{R}^n))$, w is an $\mathbb{R}^n$-valued Wiener process on $[-\varepsilon,T]$ with the covariance matrix I, $0<\varepsilon<T$, $\lambda\in W^{1,\infty}(\mathbf{T},\mathbb{R})$ is a function satisfying the conditions in (10.1), x_0 is an $\mathbb{R}^k$-valued Gaussian random

variable with $\mathbf{E}x_0 = 0$, x_0 and w are independent, $Q_T \in \mathcal{L}(\mathbb{R}^k)$, $Q_T \geq 0$, $F \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^k))$, $G, G^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^m))$, $L \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^k, \mathbb{R}^m))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \geq 0$ for a.e. $t \in \mathbf{T}$.

These conditions are the same as the conditions $(\mathbf{R}_1^s)$–$(\mathbf{R}_3^s)$ under $X = \mathbb{R}^k$, $U = \mathbb{R}^m$, $H = \mathbb{R}^n$, fortified with the smoothness condition on Φ. For simplicity, we let $v = w$. Note that we take A as independent of the time variable, since we will refer to the results from Section 8.3.5. In general the results presented below are valid if, say, $A \in C(\mathbf{T}, \mathcal{L}(\mathbb{R}^k))$. We will denote

$$P_0 = \text{cov} x_0, \ \sigma = 2\varepsilon \text{ and } \varepsilon_n = 2^{-n}\varepsilon.$$

10.4.2 Approximating Problems

One can observe that in the problem (10.9)–(10.11) the noise processes of the state and the observations are independent of the interval $[0, \lambda_0^{-1}]$. Starting at time moment λ_0^{-1} the state noise is a pointwise delay of the observation noise. Hence, we can keep the state noise up to time moment λ_0^{-1} and then approximate it by wide band noise processes in accordance with Theorem 10.3. Then we obtain the following sequence of state-observation systems:

$$dx_t = (Ax_t + B_t u_t + \varphi_t^n)dt + \Phi_t \chi_{[0,\lambda_0^{-1})}(t) dw_{\lambda_t}, \ x_0 \text{ is given}, \ 0 < t \leq T, \quad (10.12)$$

$$dz_t = C_t x_t dt + \Psi_t dw_t, \ z_0 = 0, \ 0 < t \leq T, \quad (10.13)$$

where φ^n is defined by (10.3), (10.5) and (10.7).

Note that while w_t, $-\varepsilon \leq t \leq T$ is assumed to be a Wiener process, w_{λ_t}, $0 \leq t \leq T$, may not be a Wiener process. In two particular cases when $\lambda_t = t - \varepsilon$ and $\lambda_t = ct$ with $0 < c < 1$ it can be considered as a Wiener process. Indeed, in the first case, $w_{t-\varepsilon}$ is clearly a Wiener process on $[0, T]$. In the second case, by Proposition 4.16, the random process

$$w_t^1 = \frac{1}{\sqrt{c}}(w_{ct} - w_0), \ 0 \leq t \leq T,$$

is a Wiener process with the covariance matrix I. So, we can replace the term $\Phi_t \chi_{[0,\lambda_0^{-1})}(t) dw_{ct}$ in (10.12) by $\sqrt{c}\Phi_t \chi_{[0,\lambda_0^{-1})}(t) dw_t^1$.

In the general case of function λ satisfying the conditions in (10.1), following from

$$\begin{aligned}
\text{cov}\left(\int_0^t \Phi_r \, dw_{\lambda_r}, \int_0^s \Phi_r \, dw_{\lambda_r}\right) &= \text{cov}\left(\int_{\lambda_0}^{\lambda_t} \Phi_{\lambda_r^{-1}} \, dw_r, \int_{\lambda_0}^{\lambda_s} \Phi_{\lambda_r^{-1}} \, dw_r\right) \\
&= \int_{\lambda_0}^{\min(\lambda_t,\lambda_s)} \Phi_{\lambda_r^{-1}} \Phi_{\lambda_r^{-1}}^* \, dr \\
&= \int_0^{\min(t,s)} \left(\sqrt{\lambda_r'}\Phi_r\right)\left(\sqrt{\lambda_r'}\Phi_r\right)^* dr,
\end{aligned}$$

we can replace the term $\Phi_t\chi_{[0,\lambda_0^{-1})}(t)dw_{\lambda_t}$ in (10.12) by $\sqrt{\lambda_t'}\Phi_t\chi_{[0,\lambda_0^{-1})}(t)dw_t^1$ assuming that w^1 is an $\mathbb{R}^n$-valued Wiener process on $[0,T]$ that is independent of (x_0,w) and has the covariance matrix I.

This enables us to write the following formulae from Example 8.25 for the optimal control u^n and for the best estimate $\hat{x}^n$ in the linear stochastic regulator and filtering problems determined by (10.11)–(10.13):

$$u_t^n = -G_t^{-1}\big(B_t^*Q_t^{00} + L_t\big)\hat{x}_t^n - G_t^{-1}B_t^*\int_{\max(-\sigma,t-T)}^0 \mathcal{Y}_{t-\theta,t}^*Q_{t-\theta}^{00}\psi_{t,\theta}^n\,d\theta,\ \text{a.e. } t\in\mathbf{T}, \tag{10.14}$$

$$\begin{cases} d\hat{x}_t^n = (A\hat{x}_t^n + \psi_{t,0}^n + B_tu_t^n)dt + P_{n,t}^{00}C_t^*(\Psi_t\Psi_t^*)^{-1} \\ \quad\times(dz_t^n - C_t\hat{x}_t^n dt),\ \hat{x}_0^n = 0,\ 0<t\le T, \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta}^n dt = \big(P_{n,t,\theta}^{01*}C_t^* + \bar{\Phi}_{t,\theta}^n\Psi_t^*\big)(\Psi_t\Psi_t^*)^{-1} \\ \quad\times(dz_t^n - C_t\hat{x}_t^n dt),\ \psi_{0,\theta}^n = \psi_{t,-\sigma}^n = 0,\ -\sigma\le\theta\le 0,\ 0<t\le T, \end{cases} \tag{10.15}$$

where $\big(P_n^{00}, P_n^{01}, P_n^{11}\big)$ is a solution of the system of equations

$$\frac{d}{dt}P_{n,t}^{00} - P_{n,t}^{00}A^* - AP_{n,t}^{00} - P_{n,t,0}^{01} - P_{n,t,0}^{01*} - \Phi_t\Phi_t^*\chi_{[0,\lambda_0^{-1})}(t)\lambda_t' + P_{n,t}^{00}C_t^*(\Psi_t\Psi_t^*)^{-1}C_tP_{n,t}^{00} = 0,\ P_{n,0}^{00} = P_0,\ 0<t\le T, \tag{10.16}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial\theta}\right)P_{n,t,\theta}^{01} - AP_{n,t,\theta}^{01} - P_{n,t,0,\theta}^{11} + P_{n,t}^{00}C_t^*(\Psi_t\Psi_t^*)^{-1} \times\big(C_tP_{n,t,\theta}^{01} + \Psi_t\bar{\Phi}_{t,\theta}^{n*}\big) = 0,\ P_{n,0,\theta}^{01} = P_{n,t,-\sigma}^{01} = 0,\ -\sigma\le\theta\le 0,\ 0<t\le T, \tag{10.17}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial\theta} + \frac{\partial}{\partial\tau}\right)P_{n,t,\theta,\tau}^{11} - \bar{\Phi}_{t,\theta}^n\bar{\Phi}_{t,\tau}^{n*} + \big(P_{n,t,\theta}^{01*}C_t^* + \bar{\Phi}_{t,\theta}^n\Psi_t^*\big)(\Psi_t\Psi_t^*)^{-1} \times\big(C_tP_{n,t,\tau}^{01} + \Psi_t\bar{\Phi}_{t,\tau}^{n*}\big) = 0,\ P_{n,0,\theta,\tau}^{11} = P_{n,t,-\sigma,\tau}^{11} = P_{n,t,\theta,-\sigma}^{11} = 0,\ -\sigma\le\theta\le 0,\ -\sigma\le\tau\le 0,\ 0<t\le T, \tag{10.18}$$

Q^{00} is a solution of

$$\frac{d}{dt}Q_t^{00} + Q_t^{00}A + A^*Q_t^{00} + F_t - \big(Q_t^{00}B_t + L_t^*\big)G_t^{-1}\big(B_t^*Q_t^{00} + L_t\big) = 0,\ Q_T^{00} = Q_T,\ 0\le t<T, \tag{10.19}$$

Q^{01} and Q^{11} are defined by

$$Q_{t,\theta}^{01} = \begin{Bmatrix} \mathcal{Y}_{t-\theta,t}^*Q_{t-\theta}^{00}, & \theta\ge t-T \\ 0, & \theta<t-T \end{Bmatrix},\ -\sigma\le\theta\le 0,\ 0\le t\le T, \tag{10.20}$$

$$Q_{t,\theta,\tau}^{11} = \begin{Bmatrix} Q_{t-\theta,\tau-\theta}^{01}, & \theta\ge\tau \\ Q_{t-\tau,\theta-\tau}^{01*}, & \theta<\tau \end{Bmatrix} - \int_t^{\min(t-\theta,t-\tau)} Q_{s,\theta+s-t}^{01*}B_sG_s^{-1}B_s^*Q_{s,\tau+s-t}^{01}\,ds,\ -\sigma\le\theta\le 0,\ -\sigma\le\tau\le 0,\ 0\le t\le T, \tag{10.21}$$

$\mathcal{Y}$ is the fundamental matrix of $A - B_t G_t^{-1}(B_t^* Q_t^{00} + L_t)$ and $\bar{\Phi}^n$ is defined by

$$\bar{\Phi}^n_{t,\theta} = \begin{cases} \Phi_{t-\theta} f_{n,t-\lambda_{t-\theta}} \lambda'_{t-\theta}, & \min\left(\lambda_t^{-1}, T\right) \le t-\theta \le \min\left(\lambda_{t+\varepsilon_n}^{-1}, T\right), \\ 0, & \text{otherwise}, \end{cases} \tag{10.22}$$

for $0 \le t \le T$. The covariance of the error of estimation and the minimum of the functional are given by

$$\operatorname{cov}(x_t^n - \hat{x}_t^n) = P_{n,t}^{00}, \; 0 \le t \le T, \tag{10.23}$$

and

$$\begin{aligned} J(u^*) = \operatorname{tr}\bigl(Q_T^{00} P_{n,T}^{00}\bigr) &+ \operatorname{tr} \int_0^T \bigl(F_t P_{n,t}^{00} + \Psi_t^{-1} C_t P_{n,t}^{00} Q_t^{00} P_{n,t}^{00} C_t^* \Psi_t^{-1*}\bigr) dt \\ &+ \operatorname{tr} \int_0^T \int_{-\sigma}^0 \Psi_t^{-1} C_t P_{n,t}^{00} Q_{t,\theta}^{01} \bigl(P_{n,t,\theta}^{01*} C_t^* + \bar{\Phi}^n_{t,\theta} \Psi_t^*\bigr) \Psi_t^{-1*} \, d\theta \, dt \\ &+ \operatorname{tr} \int_0^T \int_{-\sigma}^0 \Psi_t^{-1} \bigl(C_t P_{n,t,\theta}^{01} + \Psi_t \bar{\Phi}^{n*}_{t,\theta}\bigr) Q_{t,\theta}^{01*} P_{n,t}^{00} C_t^* \Psi_t^{-1*} \, d\theta \, dt \\ &+ \operatorname{tr} \int_0^T \int_{-\sigma}^0 \int_{-\sigma}^0 \Psi_t^{-1} \bigl(C_t P_{n,t,\theta}^{01} + \Psi_t \bar{\Phi}^{n*}_{t,\theta}\bigr) Q_{t,\theta,\tau}^{11} \\ &\qquad \times \bigl(P_{n,t,\tau}^{01*} C_t^* + \bar{\Phi}^n_{t,\tau} \Psi_t^*\bigr) \Psi_t^{-1*} \, d\tau \, d\theta \, dt. \end{aligned} \tag{10.24}$$

Note that since the underlying spaces are finite dimensional, the solutions of the equations in (10.15)–(10.19) are in the ordinary sense and the integrals in (10.21) and (10.24) are uniform Bochner integrals.

10.4.3 Optimal Control and Optimal Filter

The results presented in this section are not yet mathematically strictly proved, but they are evident. The starting point in this section is as follows. Let u^* be the optimal control, x^* and z^* be the state and observation processes, respectively, corresponding to the optimal control and $\hat{x}_t^*$ be the best estimate of x_t^* based on z_s^*, $0 \le s \le t$, in the linear stochastic regulator and filtering problems determined by (10.9)–(10.11). From the results of Chapter 9, it follows that the processes u^*, x^*, z^* and $\hat{x}^*$ exist and they are unique. Denote by u^n, x^n, z^n and $\hat{x}^n$ the respective processes in the linear stochastic regulator and filtering problems determined by (10.11)–(10.13). From the results of Chapter 8, they exist and are unique as well. The formulae for them are given by (10.14)–(10.22) and (10.3). Assume that

$$\hat{x}_t^* = \int_0^t K_{t,s}^* \, dz_s^* \text{ and } \hat{x}_t^n = \int_0^t K_{t,s}^n \, dz_s^n, \; 0 \le t \le T.$$

We assert that if $K_{t,\cdot}^n \to K_{t,\cdot}^*$ as $n \to \infty$ in the mean square sense, then

$$u_t^n \to u_t^*, \; x_t^n \to x_t^*, \; z_t^n \to z_t^* \text{ and } \hat{x}_t^n \to \hat{x}_t^* \text{ as } n \to \infty.$$

in the mean square sense. Indeed, at first one can show that, if the systems (10.9)–(10.10) and (10.12)–(10.13) are considered under the zero control $u = 0$, then the convergence of the signal, observation and best estimate processes in the above mentioned sense, is assured. Then write the optimal controls in terms of differentials of the respective observation processes corresponding to the zero control (see Lemma 5.25) and show the convergence of the respective kernel functions. In this step the convergence properties of Volterra integral equations must be used. As far as the convergence of optimal controls is proved, one can easily go on to prove the convergence of the sequences of the other related processes.

Thus, we conclude that if we move n to ∞ in the formulae (10.14)–(10.21), then in the limit we obtain the equations for the optimal control and for the optimal filter in the linear stochastic regulator and filtering problems determined by (10.9)–(10.11). For this, note that by Corollary 10.2,

$$\begin{aligned}\int_{-\sigma}^{0} \bar{\Phi}^n_{t,\theta}\, d\theta &= \int_{t-\lambda^{-1}_{t+\varepsilon_n}}^{t-\lambda_t^{-1}} \Phi_{t-\theta} f_{n,t-\lambda_{t-\theta}} \lambda'_{t-\theta}\, d\theta \\ &= \int_{-\varepsilon_n}^{0} \Phi_{\lambda^{-1}_{t-\theta}} f_{n,\theta}\, d\theta \;\; \to \Phi_{\lambda_t^{-1}} \text{ as } n \to \infty \text{ if } \; 0 < t \le \lambda_T,\end{aligned}$$

and

$$\int_{-\sigma}^{0} \bar{\Phi}^n_{t,\theta}\, d\theta = 0 \;\text{ if }\; \lambda_T < t \le T.$$

So,

$$\int_{-\sigma}^{0} \bar{\Phi}^n_{t,\theta}\, d\theta \to \left\{ \begin{matrix} \Phi_{\lambda_t^{-1}}, & 0 < t \le \lambda_T \\ 0, & \lambda_T < t \le T \end{matrix} \right\} \text{ as } n \to \infty. \tag{10.25}$$

Now denote by P^{00}, P^{01} and P^{11} the limits of P^{00}_n, P^{01}_n and P^{11}_n assuming that they exist. Then the equation (10.16) in the limit produces

$$\begin{aligned}&\frac{d}{dt} P^{00}_t - P^{00}_t A^* - A P^{00}_t - P^{01}_{t,0} - P^{01*}_{t,0} - \Phi_t \Phi^*_t \chi_{[0,\lambda_0^{-1})}(t) \lambda'_t \\ &\qquad + P^{00}_t C^*_t (\Psi_t \Psi^*_t)^{-1} C_t P^{00}_t = 0, \; P^{00}_0 = P_0, \; 0 < t \le T.\end{aligned} \tag{10.26}$$

To study the limits in equations (10.17) and (10.18), define the function

$$\mu_t = \begin{cases} t - \lambda_t^{-1}, & 0 < t \le \lambda_T, \\ t - T, & \lambda_T < t \le T, \end{cases} \tag{10.27}$$

and let

$$\begin{aligned} S_1 &= \{(t,\theta) : -\sigma \le \theta < \mu_t, \; 0 < t \le T\}, \\ S_2 &= \{(t,\theta,\tau) : (t,\theta) \in S_1 \text{ or } (t,\tau) \in S_1\}. \end{aligned}$$

Since $\bar{\Phi}^n_{t,\theta} \to 0$ as $n \to \infty$ for $(t,\theta) \in S_1$, in the limit the equations (10.17) and (10.18) produce homogenous equations on S_1 and S_2, respectively, with the zero initial and boundary conditions. Hence, $P^{01}_{t,\theta} = 0$ on S_1 and $P^{11}_{t,\theta,\tau} = 0$ on S_2.

Furthermore, taking limits as $n \to \infty$ in (10.17)–(10.18) and using (10.25), at the value $\theta = \tau = \mu_t$, we obtain

$$P^{11}_{t,\mu_t,\mu_t} = 0 \text{ and } P^{01}_{t,\mu_t} = \begin{cases} -P^{00}_t C^*_t \Psi_t^{-1*} \Phi^*_{\lambda_t^{-1}}, & 0 < t \le \lambda_T, \\ 0, & \lambda_T < t \le T. \end{cases}$$

If $\theta = \mu_t$ and $\mu_t < \tau \le 0$, then the limit in (10.18) produces

$$\left[\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau}\right) P^{11}_{t,\theta,\tau}\right]_{\theta=\mu_t} + \begin{Bmatrix} \Phi_{\lambda_t^{-1}} \Psi_t^{-1} C_t P^{01}_{t,\tau}, & 0 < t \le \lambda_T \\ 0, & \lambda_T < t \le T \end{Bmatrix} = 0.$$

Similarly, if $\tau = \mu_t$ and $\mu_t < \theta \le 0$, then

$$\left[\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta}\right) P^{11}_{t,\theta,\tau}\right]_{\tau=\mu_t} + \begin{Bmatrix} P^{01*}_{t,\theta} C^*_t \Psi_t^{-1*} \Phi^*_{\lambda_t^{-1}}, & 0 < t \le \lambda_T \\ 0, & \lambda_T < t \le T \end{Bmatrix} = 0.$$

For $\mu_t < \theta \le 0$, the limit in (10.17) produces

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta}\right) P^{01}_{t,\theta} - A P^{01}_{t,\theta} - P^{11}_{t,0,\theta} + P^{00}_t C^*_t (\Psi_t \Psi^*_t)^{-1} C_t P^{01}_{t,\theta} = 0,$$

and for $\mu_t < \theta \le 0$ and $\mu_t < \tau \le 0$, the limit in (10.18) produces

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta} + \frac{\partial}{\partial \tau}\right) P^{11}_{t,\theta,\tau} + P^{01*}_{t,\theta} C^*_t (\Psi_t \Psi^*_t)^{-1} C_t P^{01}_{t,\tau} = 0.$$

Combining these conditions, we obtain for P^{01}:

$$\begin{cases} (\partial/\partial t + \partial/\partial \theta) P^{01}_{t,\theta} - A P^{01}_{t,\theta} - P^{11}_{t,0,\theta} + P^{00}_t C^*_t (\Psi_t \Psi^*_t)^{-1} C_t P^{01}_{t,\theta} = 0, \\ \qquad \mu_t < \theta \le 0, \ 0 < t \le T; \\ P^{01}_{0,\theta} = 0, \ -\lambda_0^{-1} \le \theta \le 0; \\ P^{01}_{t,t-\lambda_t^{-1}} = -P^{00}_t C^*_t \Psi_t^{-1*} \Phi^*_{\lambda_t^{-1}}, \ 0 < t \le \lambda_T; \\ P^{01}_{t,t-T} = 0, \ \lambda_T < t \le T. \end{cases} \tag{10.28}$$

Also combining the above mentioned conditions for P^{11} and using the symmetry in P^{11}, we obtain

$$\begin{cases} (\partial/\partial t + \partial/\partial \theta + \partial/\partial \tau) P^{11}_{t,\theta,\tau} + P^{01*}_{t,\theta} C^*_t (\Psi_t \Psi^*_t)^{-1} C_t P^{01}_{t,\tau} = 0, \\ \qquad \mu_t < \theta \le 0, \ \mu_t < \tau \le 0, \ 0 < t \le T; \\ P^{11}_{0,\theta,\tau} = 0, \ -\lambda_0^{-1} \le \theta \le 0, \ -\lambda_0^{-1} \le \tau \le 0; \\ \left[(\partial/\partial t + \partial/\partial \theta) P^{11}_{t,\theta,\tau}\right]_{\tau=t-\lambda_t^{-1}} + P^{01*}_{t,\theta} C^*_t \Psi_t^{-1*} \Phi^*_{\lambda_t^{-1}} = 0, \\ \qquad t - \lambda_t^{-1} < \theta \le 0, \ 0 < t \le \lambda_T; \\ \left[(\partial/\partial t + \partial/\partial \theta) P^{11}_{t,\theta,\tau}\right]_{\tau=t-T} = 0, \ t - T < \theta \le 0, \ \lambda_T < t \le T; \\ P^{11}_{t,\mu_t,\mu_t} = 0, \ 0 < t \le T. \end{cases} \tag{10.29}$$

A similar argument, as applied to (10.15), yields

$$\begin{cases} d\hat{x}_t^* = (A\hat{x}_t^* + \psi_{t,0} + B_t u_t^*)dt + P_t^{00} C_t^* (\Psi_t \Psi_t^*)^{-1} \\ \qquad \times (dz_t^* - C_t \hat{x}_t^* dt), \; \hat{x}_0^* = 0, \; 0 < t \leq T; \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta} dt = P_{t,\theta}^{01*} C_t^* (\Psi_t \Psi_t^*)^{-1} (dz_t^* - C_t \hat{x}_t^* dt), \\ \qquad \mu_t < \theta \leq 0, \; 0 < t \leq T; \\ \psi_{0,\theta} = 0, \; -\lambda_0^{-1} \leq \theta \leq 0; \\ \psi_{t,t-\lambda_t^{-1}} dt = \Phi_{\lambda_t^{-1}} \Psi_t^{-1} (dz_t^* - C_t \hat{x}_t^* dt), \; 0 < t \leq \lambda_T; \\ \psi_{t,t-T} = 0, \; \lambda_T < t \leq T. \end{cases} \tag{10.30}$$

The second equation in (10.30) is significant because it is a stochastic partial differential equation with the boundary condition that is again determined by a stochastic differential equation. Also, in the limit, (10.14) yields

$$u_t^* = -G_t^{-1}(B_t^* Q_t^{00} + L_t)\hat{x}_t^* \\ - G_t^{-1} B_t^* \int_{\max(t-\lambda_t^{-1}, t-T)}^{0} \mathcal{Y}_{t-\theta,t}^* Q_{t-\theta}^{00} \psi_{t,\theta} \, d\theta, \text{ a.e. } t \in \mathbf{T}. \tag{10.31}$$

The equations (10.26)–(10.31) together with (10.19)–(10.21) represent the optimal control and the optimal filter in the linear stochastic regulator and filtering problems determined by (10.9)–(10.11). We add two more formulae for the covariance of the error and for the minimum of the functional which easily follow from (10.23)–(10.24):

$$\operatorname{cov}(x_t - \hat{x}_t) = P_t^{00}, \; 0 \leq t \leq T, \tag{10.32}$$

and

$$\begin{aligned} J(u^*) &= \operatorname{tr}(Q_T^{00} P_T^{00}) + \operatorname{tr} \int_0^T (F_t P_t^{00} + \Psi_t^{-1} C_t P_t^{00} Q_t^{00} P_t^{00} C_t^* \Psi_t^{-1*}) dt \\ &+ \operatorname{tr} \int_0^T \int_{\max(t-\lambda_t^{-1}, t-T)}^0 \Psi_t^{-1} C_t (P_t^{00} Q_{t,\theta}^{01} P_{t,\theta}^{01*} + P_{t,\theta}^{01} Q_{t,\theta}^{01*} P_t^{00}) C_t^* \Psi_t^{-1*} d\theta \, dt \\ &+ \operatorname{tr} \int_0^T \int_{\max(t-\lambda_t^{-1}, t-T)}^0 \int_{\max(t-\lambda_t^{-1}, t-T)}^0 \Psi_t^{-1} C_t P_{t,\theta}^{01} Q_{t,\theta,\tau}^{11} P_{t,\tau}^{01*} C_t^* \Psi_t^{-1*} d\tau \, d\theta \, dt \\ &+ \operatorname{tr} \int_0^{\lambda_T} (\Psi_t^{-1} C_t P_t^{00} Q_{t,t-\lambda_t^{-1}}^{01} \Phi_{\lambda_t^{-1}} + \Phi_{\lambda_t^{-1}}^* Q_{t,t-\lambda_t^{-1}}^{01*} P_t^{00} C_t^* \Psi_t^{-1*}) \, dt \\ &+ \operatorname{tr} \int_0^{\lambda_T} \int_{t-\lambda_t^{-1}}^0 \Psi_t^{-1} C_t P_{t,\theta}^{01} Q_{t,\theta,t-\lambda_t^{-1}}^{11} \Phi_{\lambda_t^{-1}} \, d\theta \, dt \\ &+ \operatorname{tr} \int_0^{\lambda_T} \int_{t-\lambda_t^{-1}}^0 \Phi_{\lambda_t^{-1}}^* Q_{t,t-\lambda_t^{-1},\theta}^{11} P_{t,\theta}^{01*} C_t^* \Psi_t^{-1*} \, d\theta \, dt \\ &+ \operatorname{tr} \int_0^{\lambda_T} \Phi_{\lambda_t^{-1}}^* Q_{t,t-\lambda_t^{-1},t-\lambda_t^{-1}}^{11} \Phi_{\lambda_t^{-1}} \, dt. \end{aligned} \tag{10.33}$$

Finally, note that by the zero initial condition in (10.28), we have $P^{01}_{t,0} = 0$ for $0 \le t \le \lambda_0^{-1}$. Hence, (10.26) becomes an ordinary Riccati equation on $[0, \lambda_0^{-1}]$ that is natural since on $[0, \lambda_0^{-1}]$ the noises of the state and the observations are independent.

10.4.4 Application to Space Navigation and Guidance

Linear stochastic regulator and filtering theories have significant applications in space navigation and guidance. In fact, up to now these applications use only the results concerning independent and correlated white noise disturbances. In Section 9.1 it was illustrated that the noise processes arising in space navigation and guidance can be more adequately modelled by shifted white noises. The function $\lambda_t = t - \varepsilon$ with $\varepsilon > 0$ in the state-observation system (10.9)–(10.10) describes the shift arising in navigation of Earth orbiting satellites since they have nearly constant distance from the Earth. But the case $\lambda_t = ct$ with $0 < c < 1$ describes the shift arising in navigation of space probes flying away from the Earth, since their distance from the Earth increases with nearly constant rate of change. Note that in this case it is reasonable to take $\varepsilon = (1-c)T$ that is the maximum of $t - \lambda_t$ on $[0, T]$. One can deduce that for navigation of space probes flying toward the Earth the function λ must be taken as $\lambda_t = -at + b$, where a and b are positive constants, that is the combination of the previous two cases.

Example 10.4 (No shift). Let $\lambda_t = t$ in the system (10.9)–(10.10). How does this effect the equations (10.26)–(10.33) and (10.19)–(10.21)? The equation (10.28) reduces to

$$P^{01}_{t,0} = -P^{00}_t C^*_t \Psi_t^{-1*} \Phi^*_t, \ 0 < t \le T.$$

Substituting this in (10.26), we obtain the familiar Riccati equation

$$\frac{d}{dt} P^{00}_t - P^{00}_t A^* - A P^{00}_t - \Phi_t \Phi^*_t + \left(P^{00}_t C^*_t + \Phi_t \Psi^*_t\right) \times (\Psi_t \Psi^*_t)^{-1} \left(C_t P^{00}_t + \Psi_t \Phi^*_t\right) = 0, \ 0 < t \le T, \ P_0 = \operatorname{cov} x_0. \tag{10.34}$$

From (10.30), we obtain

$$\psi_{t,0} dt = \Phi_t \Psi_t^{-1} (dz^*_t - C_t \hat{x}^*_t dt), \ 0 < t \le T,$$

and, consequently,

$$d\hat{x}^*_t = (A\hat{x}^*_t + B_t u^*_t) dt + \left(P^{00}_t C^*_t + \Phi_t \Psi^*_t\right) \times (\Psi_t \Psi^*_t)^{-1} (dz^*_t - C_t \hat{x}^*_t dt), \ \hat{x}_0 = 0, \ 0 < t \le T. \tag{10.35}$$

Also, (10.31) produces

$$u^*_t = -G_t^{-1} \left(B^*_t Q^{00}_t + L_t\right) \hat{x}^*_t, \ \text{a.e. } t \in \mathbf{T}, \tag{10.36}$$

where Q^{00} satisfies (10.19). For the covariance of the error process we have the equality in (10.32) without any modification. For the minimum of the functional, note that $Q^{11}_{t,0,0} = Q^{01}_{t,0} = Q^{01*}_{t,0} = Q^{00}_t$ by (10.20)–(10.21). Hence, from (10.33),

$$J(u^*) = \operatorname{tr}\big(Q^{00}_T P^{00}_T\big) + \operatorname{tr}\int_0^T F_t P^{00}_t \, dt$$
$$+ \operatorname{tr}\int_0^T \Psi_t^{-1}\big(C_t P^{00}_t + \Psi_t \Phi_t^*\big) Q^{00}_t \big(P^{00}_t C_t^* + \Phi_t \Psi_t^*\big)\Psi_t^{-1*}\, dt. \qquad (10.37)$$

The equations (10.34)–(10.37) together with (10.19) and (10.32) are well-known equations describing the optimal control and the optimal filter for the correlated white noise disturbances of the state-observation system (cf. Example 6.23).

Example 10.5 (Navigation of Earth orbiting satellites). Now let $\lambda_t = t - \varepsilon$ in the system (10.9)–(10.10). Then $\lambda_t^{-1} = t + \varepsilon$. Consequently, the equations (10.26), (10.28) and (10.29) produce

$$\frac{d}{dt} P^{00}_t - P^{00}_t A^* - A P^{00}_t - P^{01}_{t,0} - P^{01*}_{t,0} - \Phi_t \Phi_t^* \chi_{[0,\varepsilon)}(t)$$
$$+ P^{00}_t C_t^* (\Psi_t \Psi_t^*)^{-1} C_t P^{00}_t = 0, \; P^{00}_0 = P_0, \; 0 < t \le T, \qquad (10.38)$$

$$\begin{cases} (\partial/\partial t + \partial/\partial\theta) P^{01}_{t,\theta} - A P^{01}_{t,\theta} - P^{11}_{t,0,\theta} + P^{00}_t C_t^* (\Psi_t \Psi_t^*)^{-1} C_t P^{01}_{t,\theta} = 0, \\ \qquad \max(-\varepsilon, t-T) < \theta \le 0, \; 0 < t \le T; \\ P^{01}_{0,\theta} = 0, \; -\varepsilon \le \theta \le 0; \\ P^{01}_{t,-\varepsilon} = -P^{00}_t C_t^* \Psi_t^{-1*} \Phi^*_{t+\varepsilon}, \; 0 < t \le T - \varepsilon; \\ P^{01}_{t,t-T} = 0, \; T - \varepsilon < t \le T, \end{cases} \qquad (10.39)$$

$$\begin{cases} (\partial/\partial t + \partial/\partial\theta + \partial/\partial\tau) P^{11}_{t,\theta,\tau} + P^{01*}_{t,\theta} C_t^* (\Psi_t \Psi_t^*)^{-1} C_t P^{01}_{t,\tau} = 0, \\ \qquad \max(-\varepsilon, t-T) < \theta \le 0, \; \max(-\varepsilon, t-T) < \tau \le 0, \; 0 < t \le T; \\ P^{11}_{0,\theta,\tau} = 0, \; -\varepsilon \le \theta \le 0, \; -\varepsilon \le \tau \le 0; \\ (\partial/\partial t + \partial/\partial\theta) P^{11}_{t,\theta,-\varepsilon} + P^{01*}_{t,\theta} C_t^* \Psi_t^{-1*} \Phi^*_{t+\varepsilon} = 0, \\ \qquad -\varepsilon < \theta \le 0, \; 0 < t \le T - \varepsilon; \\ \big[(\partial/\partial t + \partial/\partial\theta) P^{11}_{t,\theta,\tau}\big]_{\tau = t-T} = 0, \; t - T < \theta \le 0, \; T - \varepsilon < t \le T; \\ P^{11}_{t,-\varepsilon,-\varepsilon} = 0, \; 0 < t \le T - \varepsilon; \\ P^{11}_{t,t-T,t-T} = 0, \; T - \varepsilon < t \le T. \end{cases} \qquad (10.40)$$

Similarly, (10.30) yields

$$\begin{cases} d\hat{x}^*_t = (A\hat{x}^*_t + \psi_{t,0} + B_t u^*_t)dt + P^{00}_t C_t^* (\Psi_t \Psi_t^*)^{-1} \\ \qquad \times (dz^*_t - C_t \hat{x}^*_t dt), \; \hat{x}^*_0 = 0, \; 0 < t \le T; \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta} dt = P^{01*}_{t,\theta} C_t^* (\Psi_t \Psi_t^*)^{-1} (dz^*_t - C_t \hat{x}^*_t dt), \\ \qquad \max(-\varepsilon, t-T) < \theta \le 0, \; 0 < t \le T; \\ \psi_{0,\theta} = 0, \; -\varepsilon \le \theta \le 0; \\ \psi_{t,-\varepsilon} dt = \Phi_{t+\varepsilon} \Psi_t^{-1} (dz^*_t - C_t \hat{x}^*_t dt), \; 0 < t \le T - \varepsilon; \\ \psi_{t,t-T} = 0, \; T - \varepsilon < t \le T. \end{cases} \qquad (10.41)$$

Note that by substitution the equations in (10.41) can be reduced to the respective equations in Example 9.20. Furthermore, from (10.31), for the optimal control we have

$$u_t^* = -G_t^{-1}(B_t^* Q_t^{00} + L_t)\hat{x}_t^* - G_t^{-1} B_t^* \int_{\max(-\varepsilon, t-T)}^{0} \mathcal{Y}_{t-\theta,t}^* Q_{t-\theta}^{00} \psi_{t,\theta}\, d\theta, \text{ a.e. } t \in \mathbf{T}. \tag{10.42}$$

The formula (10.32) remains valid for the covariance of the error process and, from (10.33), we obtain a formula for the minimum of the functional:

$$\begin{aligned} J(u^*) &= \operatorname{tr}(Q_T^{00} P_T^{00}) + \operatorname{tr} \int_0^T (F_t P_t^{00} + \Psi_t^{-1} C_t P_t^{00} Q_t^{00} P_t^{00} C_t^* \Psi_t^{-1*})\, dt \\ &+ \operatorname{tr} \int_0^T \int_{\max(-\varepsilon,t-T)}^0 \Psi_t^{-1} C_t (P_t^{00} Q_{t,\theta}^{01} P_{t,\theta}^{01*} + P_{t,\theta}^{01} Q_{t,\theta}^{01*} P_t^{00}) C_t^* \Psi_t^{-1*}\, d\theta\, dt \\ &+ \operatorname{tr} \int_0^T \int_{\max(-\varepsilon,t-T)}^0 \int_{\max(-\varepsilon,t-T)}^0 \Psi_t^{-1} C_t P_{t,\theta}^{01} Q_{t,\theta,\tau}^{11} P_{t,\tau}^{01*} C_t^* \Psi_t^{-1*}\, d\tau\, d\theta\, dt \\ &+ \operatorname{tr} \int_0^{T-\varepsilon} (\Psi_t^{-1} C_t P_t^{00} Q_{t,-\varepsilon}^{01} \Phi_{t-\varepsilon} + \Phi_{t-\varepsilon}^* Q_{t,-\varepsilon}^{01*} P_t^{00} C_t^* \Psi_t^{-1*})\, dt \\ &+ \operatorname{tr} \int_0^{T-\varepsilon} \int_{-\varepsilon}^0 \Psi_t^{-1} C_t P_{t,\theta}^{01} Q_{t,\theta,-\varepsilon}^{11} \Phi_{t+\varepsilon}\, d\theta\, dt \\ &+ \operatorname{tr} \int_0^{T-\varepsilon} \int_{-\varepsilon}^0 \Phi_{t+\varepsilon}^* Q_{t,-\varepsilon,\theta}^{11} P_{t,\theta}^{01*} C_t^* \Psi_t^{-1*}\, d\theta\, dt \\ &+ \operatorname{tr} \int_0^{T-\varepsilon} \Phi_{t+\varepsilon}^* Q_{t,-\varepsilon,-\varepsilon}^{11} \Phi_{t+\varepsilon}\, dt. \end{aligned} \tag{10.43}$$

The equations (10.38)–(10.43) together with (10.19)–(10.21) and (10.32) represent the complete set of formulae for the optimal control and the optimal filter in the linear stochastic regulator and filtering problems determined by (10.9)–(10.11) under $\lambda_t = t - \varepsilon$. Note that in Example 9.20 we could derive only the equations (10.41) and (10.42) (in a slightly different form).

Example 10.6 (Navigation of space probes). Let $\lambda_t = ct$ in the system (10.9)–(10.10) assuming $0 < c < 1$. Then $\lambda_t^{-1} = c^{-1}t$. Consequently, the equations (10.26), (10.28) and (10.29) produce

$$\frac{d}{dt} P_t^{00} - P_t^{00} A^* - A P_t^{00} - P_{t,0}^{01} - P_{t,0}^{01*} + P_t^{00} C_t^* (\Psi_t \Psi_t^*)^{-1} C_t P_t^{00} = 0, \; P_0^{00} = P_0, \; 0 < t \le T, \tag{10.44}$$

$$\begin{cases} (\partial/\partial t + \partial/\partial\theta) P_{t,\theta}^{01} - A P_{t,\theta}^{01} - P_{t,0,\theta}^{11} + P_t^{00} C_t^* (\Psi_t \Psi_t^*)^{-1} C_t P_{t,\theta}^{01} = 0, \\ \qquad \max(t - c^{-1}t, t - T) < \theta \le 0, \; 0 < t \le T; \\ P_{0,0}^{01} = 0; \; P_{t,t-c^{-1}t}^{01} = -P_t^{00} C_t^* \Psi_t^{-1*} \Phi_{c^{-1}t}^*, \; 0 < t \le cT; \\ P_{t,t-T}^{01} = 0, \; cT < t \le T, \end{cases} \tag{10.45}$$

$$
\left\{\begin{array}{l}
(\partial/\partial t + \partial/\partial\theta + \partial/\partial\tau)P^{11}_{t,\theta,\tau} + P^{01*}_{t,\theta}C^*_t(\Psi_t\Psi^*_t)^{-1}C_tP^{01}_{t,\tau} = 0, \\
\qquad \max(t - c^{-1}t, t - T) < \theta \le 0,\ \max(t - c^{-1}t, t - T) < \tau \le 0, \\
\qquad 0 < t \le T; \\
\left[(\partial/\partial t + \partial/\partial\theta)P^{11}_{t,\theta,\tau}\right]_{\tau = t - c^{-1}t} + P^{01*}_{t,\theta}C^*_t\Psi^{-1*}_t\Phi^*_{c^{-1}t} = 0, \\
\qquad t - c^{-1}t < \theta \le 0,\ 0 < t \le cT; \\
\left[(\partial/\partial t + \partial/\partial\theta)P^{11}_{t,\theta,\tau}\right]_{\tau = t - T} = 0,\ t - T < \theta \le 0,\ cT < t \le T; \\
P^{11}_{t,t-c^{-1}t,t-c^{-1}t} = 0,\ 0 \le t \le cT;\ P^{11}_{t,t-T,t-T} = 0,\ cT < t \le T.
\end{array}\right. \tag{10.46}
$$

Similarly, (10.30) and (10.31) yield

$$
\left\{\begin{array}{l}
d\hat{x}^*_t = (A\hat{x}^*_t + \psi_{t,0} + B_tu^*_t)dt + P^{00}_tC^*_t(\Psi_t\Psi^*_t)^{-1} \\
\qquad \times(dz^*_t - C_t\hat{x}^*_tdt),\ \hat{x}^*_0 = 0,\ 0 < t \le T; \\
(\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta}dt = P^{01*}_{t,\theta}C^*_t(\Psi_t\Psi^*_t)^{-1}(dz^*_t - C_t\hat{x}^*_tdt), \\
\qquad \max(t - c^{-1}t, t - T) < \theta \le 0,\ 0 < t \le T; \\
\psi_{0,0} = 0;\ \psi_{t,t-c^{-1}t}dt = \Phi_{c^{-1}t}\Psi^{-1}_t(dz^*_t - C_t\hat{x}^*_tdt),\ 0 < t \le cT; \\
\psi_{t,t-T} = 0,\ cT < t \le T,
\end{array}\right. \tag{10.47}
$$

$$
\begin{aligned}
u^*_t &= -G^{-1}_t(B^*_tQ^{00}_t + L_t)\hat{x}^*_t \\
&\quad - G^{-1}_tB^*_t\int_{\max(t-c^{-1}t,t-T)}^0 \mathcal{Y}^*_{t-\theta,t}Q^{00}_{t-\theta}\psi_{t,\theta}\,d\theta,\ \text{a.e. } t \in \mathbf{T}.
\end{aligned} \tag{10.48}
$$

The formula (10.32) remains valid for the covariance of the error process and, from (10.33), we obtain a formula for the minimum of the functional:

$$
\begin{aligned}
J(u^*) &= \operatorname{tr}(Q^{00}_TP^{00}_T) + \operatorname{tr}\int_0^T (F_tP^{00}_t + \Psi^{-1}_tC_tP^{00}_tQ^{00}_tP^{00}_tC^*_t\Psi^{-1*}_t)dt \\
&\quad + \operatorname{tr}\int_0^T\int_{\max(t-c^{-1}t,t-T)}^0 \Psi^{-1}_tC_t(P^{00}_tQ^{01}_{t,\theta}P^{01*}_{t,\theta} + P^{01}_{t,\theta}Q^{01*}_{t,\theta}P^{00}_t)C^*_t\Psi^{-1*}_t d\theta\,dt \\
&\quad + \operatorname{tr}\int_0^T\int_{\max(t-c^{-1}t,t-T)}^0\int_{\max(t-c^{-1}t,t-T)}^0 \Psi^{-1}_tC_tP^{01}_{t,\theta}Q^{11}_{t,\theta,\tau}P^{01*}_{t,\tau}C^*_t\Psi^{-1*}_t d\tau\,d\theta\,dt \\
&\quad + \operatorname{tr}\int_0^{cT} \left(\Psi^{-1}_tC_tP^{00}_tQ^{01}_{t,t-c^{-1}t}\Phi_{c^{-1}t} + \Phi^*_{c^{-1}t}Q^{01*}_{t,t-c^{-1}t}P^{00}_tC^*_t\Psi^{-1*}_t\right)dt \\
&\quad + \operatorname{tr}\int_0^{cT}\int_{t-c^{-1}t}^0 \Psi^{-1}_tC_tP^{01}_{t,\theta}Q^{11}_{t,\theta,t-c^{-1}t}\Phi_{c^{-1}t}\,d\theta\,dt \\
&\quad + \operatorname{tr}\int_0^{cT}\int_{t-c^{-1}t}^0 \Phi^*_{c^{-1}t}Q^{11}_{t,t-c^{-1}t,\theta}P^{01*}_{t,\theta}C^*_t\Psi^{-1*}_t\,d\theta\,dt \\
&\quad + \operatorname{tr}\int_0^{cT} \Phi^*_{c^{-1}t}Q^{11}_{t,t-c^{-1}t,t-c^{-1}t}\Phi_{c^{-1}t}\,dt.
\end{aligned} \tag{10.49}
$$

The equations (10.44)–(10.49) together with (10.19)–(10.21) and (10.32) represent the complete set of formulae for the optimal control and the optimal filter in the linear stochastic regulator and filtering problems determined by (10.9)–(10.11) under $\lambda_t = ct$ with $0 < c < 1$.

10.5 State Noise Anticipating Observation Noise

10.5.1 Setting of the Problem

Consider the problem (10.9)–(10.11) in which the state noise anticipates the observation noise:

$$dx_t = (Ax_t + B_t u_t)dt + \Upsilon_t dw_t,\ x_0 \text{ is given },\ 0 < t \le T, \tag{10.50}$$
$$dz_t = C_t x_t dt + \Phi_t \chi_{(\lambda_0^{-1},T]}(t) dw_{\lambda_t} + \Psi_t dv_t,\ z_0 = 0,\ 0 < t \le T, \tag{10.51}$$
$$J(u) = \mathbf{E}\Big(\langle x_T, Q_T x_T\rangle + \int_0^T \Big\langle \begin{bmatrix} x_t \\ u_t \end{bmatrix}, \begin{bmatrix} F_t & L_t^* \\ L_t & G_t \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix} \Big\rangle dt\Big). \tag{10.52}$$

In this section $A \in \mathcal{L}(\mathbb{R}^k)$, $B \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^k))$, $C \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^k, \mathbb{R}^n))$, $\Phi \in W^{1,2}(\mathbf{T}, \mathcal{L}(\mathbb{R}^l, \mathbb{R}^n))$, $\Upsilon \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^l, \mathbb{R}^k))$, $\Psi, \Psi^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^n))$, w is an $\mathbb{R}^l$-valued Wiener process on $\mathbf{T}$ with the covariance matrix I, $0 < \varepsilon < T$, $\lambda \in W^{1,\infty}(\mathbf{T}, \mathbb{R})$ is a function satisfying the conditions in (10.1), v is an $\mathbb{R}^n$-valued Wiener process on $\mathbf{T}$ with the covariance matrix I, x_0 is an $\mathbb{R}^k$-valued Gaussian random variable with $\mathbf{E}x_0 = 0$, x_0, w and v are mutually independent, $Q_T \in \mathcal{L}(\mathbb{R}^k)$, $Q_T \ge 0$, $F \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^k))$, $G, G^{-1} \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^m))$, $L \in L_\infty(\mathbf{T}, \mathcal{L}(\mathbb{R}^k, \mathbb{R}^m))$, $G_t > 0$ and $F_t - L_t^* G_t^{-1} L_t \ge 0$ for a.e. $t \in \mathbf{T}$. We will denote

$$P_0 = \operatorname{cov} x_0,\ \sigma = 2\varepsilon \text{ and } \varepsilon_n = 2^{-n}\varepsilon.$$

The role of the function $\chi_{(\lambda_0^{-1},T]}(t)$ in (10.51) is as follows: write

$$\Phi_t dw_{\lambda_t} = \Phi_t \chi_{(0,\lambda_0^{-1}]}(t) dw_{\lambda_t} + \Phi_t \chi_{(\lambda_0^{-1},T]}(t) dw_{\lambda_t},$$

assuming that w is a Wiener process on $[-\varepsilon, T]$. Here the first term in the right-hand side is independent of the state noise in (10.50). Hence, in the setting of the problem we can assume that $\Phi_t \chi_{(0,\lambda_0^{-1}]}(t) dw_{\lambda_t}$ is contained in $\Psi_t dw_t$.

10.5.2 Approximating Problems

Similar to Section 10.4.2, we will approximate part of the observation noise, that is a pointwise delay of the state noise, by wide band noise processes in accordance with Theorem 10.3. Then we obtain the following sequence of state-observation systems:

$$dx_t = (Ax_t + B_t u_t)dt + \Upsilon_t dw_t,\ x_0 \text{ is given},\ 0 < t \le T, \tag{10.53}$$
$$dz_t = (C_t x_t + \varphi_t^n)dt + \Psi_t dw_t,\ z_0 = 0,\ 0 < t \le T, \tag{10.54}$$

where φ^n is defined by (10.3), (10.5) and (10.7).

We can write the following formulae from Example 8.25 for the optimal control u^n and for the best estimate $\hat{x}^n$ in the linear stochastic regulator and filtering problems determined by (10.52)–(10.54):

$$u_t^n = -G_t^{-1}(B_t^* Q_t^{00} + L_t)\hat{x}_t^n,\ \text{a.e. } t \in \mathbf{T}, \tag{10.55}$$

$$\begin{cases} d\hat{x}_t^n = (A\hat{x}_t^n + B_t u_t^n)dt + \big(P_{n,t}^{00} C_t^* + P_{n,t,0}^{02}\big)(\Psi_t \Psi_t^*)^{-1} \\ \quad \times (dz_t^n - C_t \hat{x}_t^n dt - \psi_{t,0}^n dt), \ \hat{x}_0^n = 0, \ 0 < t \le T, \\ (\partial/\partial t + \partial/\partial \theta)\psi_{t,\theta}^n dt = \big(P_{n,t,\theta}^{02*} C_t^* + P_{n,t,\theta,0}^{22}\big)(\Psi_t \Psi_t^*)^{-1} \\ \quad \times (dz_t^n - C_t \hat{x}_t^n dt - \psi_{t,0}^n dt), \ \psi_{0,\theta}^n = \psi_{t,-\sigma}^n = 0, \ -\sigma \le \theta \le 0, \ 0 < t \le T, \end{cases} \tag{10.56}$$

where $\big(P_n^{00}, P_n^{02}, P_n^{22}\big)$ is a solution of the system of equations

$$\frac{d}{dt} P_{n,t}^{00} - P_{n,t}^{00} A^* - A P_{n,t}^{00} - \Upsilon_t \Upsilon_t^* + \big(P_{n,t}^{00} C_t^* + P_{n,t,0}^{02}\big) \\ \times (\Psi_t \Psi_t^*)^{-1} \big(C_t P_{n,t}^{00} + P_{n,t,0}^{02*}\big) = 0, \ P_{n,0}^{00} = P_0, \ 0 < t \le T, \tag{10.57}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta}\right) P_{n,t,\theta}^{02} - A P_{n,t,\theta}^{02} - \Upsilon_t \bar{\Phi}_{t,\theta}^{n*} + \big(P_{n,t}^{00} C_t^* + P_{n,t,0}^{02}\big) \\ \times (\Psi_t \Psi_t^*)^{-1} \big(C_t P_{n,t,\theta}^{02} + P_{n,t,\theta,0}^{22*}\big) = 0, \ P_{n,0,\theta}^{02} = P_{n,t,-\sigma}^{02} = 0, \\ -\sigma \le \theta \le 0, \ 0 < t \le T, \tag{10.58}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \theta} + \frac{\partial}{\partial \tau}\right) P_{n,t,\theta,\tau}^{22} - \bar{\Phi}_{t,\theta}^n \bar{\Phi}_{t,\tau}^{n*} + \big(P_{n,t,\theta}^{02*} C_t^* + P_{n,t,\theta,0}^{22}\big)(\Psi_t \Psi_t^*)^{-1} \\ \times \big(C_t P_{n,t,\tau}^{02} + P_{n,t,\tau,0}^{22*}\big) = 0, \ P_{n,0,\theta,\tau}^{22} = P_{n,t,-\sigma,\tau}^{11} = P_{n,t,\theta,-\sigma}^{22} = 0, \\ -\sigma \le \theta \le 0, \ -\sigma \le \tau \le 0, \ 0 < t \le T, \tag{10.59}$$

Q^{00} is a solution of

$$\frac{d}{dt} Q_t^{00} + Q_t^{00} A + A^* Q_t^{00} + F_t \\ - \big(Q_t^{00} B_t + L_t^*\big) G_t^{-1} \big(B_t^* Q_t^{00} + L_t\big) = 0, \ Q_T^{00} = Q_T, \ 0 \le t < T, \tag{10.60}$$

and $\bar{\Phi}^n$ is defined by (10.22). The covariance of the error of estimation and the minimum of the functional are given by

$$\operatorname{cov}(x_t^n - \hat{x}_t^n) = P_t^{00}, \ 0 \le t \le T, \tag{10.61}$$

and

$$J(u^*) = \operatorname{tr}\big(Q_T^{00} P_{n,T}^{00}\big) + \operatorname{tr} \int_0^T F_t P_{n,t}^{00} \, dt \\ + \operatorname{tr} \int_0^T \Psi_t^{-1} \big(C_t P_{n,t}^{00} + P_{n,t,0}^{02*}\big) Q_t^{00} \big(P_{n,t}^{00} C_t^* + P_{n,t,\theta}^{02}\big) \Psi_t^{-1*} \, dt. \tag{10.62}$$

10.5.3 Optimal Control and Optimal Filter

The arguments used in Section 10.4.3 are applicable to the linear stochastic regulator and filtering problems determined by (10.50)–(10.52). Hence, moving n to ∞ in the formulae (10.55)–(10.62), we will derive the formulae for the optimal control and for the optimal filter in the linear stochastic regulator and filtering problems for (10.50)–(10.52). For this, we will use the limit in (10.25).

Denote by P^{00}, P^{02} and P^{22} the limits of P_n^{00}, P_n^{02} and P_n^{22} assuming that they exist. Then the equation (10.57) in the limit produces

$$\frac{d}{dt}P_t^{00} - P_t^{00}A^* - AP_t^{00} - \Upsilon_t\Upsilon_t^* + \big(P_t^{00}C_t^* + P_{t,0}^{02}\big) \\ \times (\Psi_t\Psi_t^*)^{-1}\big(C_tP_t^{00} + P_{t,0}^{02*}\big) = 0,\ P_0^{00} = P_0,\ 0 < t \le T. \tag{10.63}$$

Let μ be defined by (10.27). Similar to the derivation of the equations (10.28)–(10.30), from (10.58)–(10.59) and (10.56), by using (10.25), we obtain the equation

$$\begin{cases} (\partial/\partial t + \partial/\partial\theta)P_{t,\theta}^{02} - AP_{t,\theta}^{02} + \big(P_t^{00}C_t^* + P_{t,0}^{02}\big)(\Psi_t\Psi_t^*)^{-1} \\ \qquad \times\big(C_tP_{t,\theta}^{02} + P_{t,\theta,0}^{22*}\big) = 0,\ \mu_t < \theta \le 0,\ 0 < t \le T; \\ P_{0,\theta}^{02} = 0,\ -\lambda_0^{-1} \le \theta \le 0; \\ P_{t,t-\lambda_t^{-1}}^{02} = \Upsilon_t\Phi_{\lambda_t^{-1}}^*,\ 0 < t \le \lambda_T; \\ P_{t,t-T}^{02} = 0,\ \lambda_T < t \le T, \end{cases} \tag{10.64}$$

for P^{02},

$$\begin{cases} (\partial/\partial t + \partial/\partial\theta + \partial/\partial\tau)P_{t,\theta,\tau}^{22} + \big(P_{t,\theta}^{02*}C_t^* + P_{t,\theta,0}^{22}\big)(\Psi_t\Psi_t^*)^{-1} \\ \qquad \times\big(C_tP_{t,\tau}^{02} + P_{t,\tau,0}^{22*}\big) = 0,\ \mu_t < \theta \le 0,\ \mu_t < \tau \le 0,\ 0 < t \le T; \\ P_{0,\theta,\tau}^{22} = 0,\ -\lambda_0^{-1} \le \theta \le 0,\ -\lambda_0^{-1} \le \tau \le 0; \\ \big[(\partial/\partial t + \partial/\partial\theta)P_{t,\theta,\tau}^{22}\big]_{\tau=\mu_t} = 0,\ \mu_t < \theta \le 0,\ 0 < t \le T; \\ P_{t,t-\lambda_t^{-1},t-\lambda_t^{-1}}^{22} = \Phi_{\lambda_t^{-1}}\Phi_{\lambda_t^{-1}}^*,\ 0 < t \le \lambda_T; \\ P_{t,t-T,t-T}^{22} = 0,\ \lambda_T < t \le T, \end{cases} \tag{10.65}$$

for P^{22} and the system

$$\begin{cases} d\hat{x}_t^* = (A\hat{x}_t^* + B_tu_t^*)dt + \big(P_t^{00}C_t^* + P_{t,0}^{02}\big)(\Psi_t\Psi_t^*)^{-1} \\ \qquad \times(dz_t^* - C_t\hat{x}_t^*dt - \psi_{t,0}dt),\ \hat{x}_0^* = 0,\ 0 < t \le T; \\ (\partial/\partial t + \partial/\partial\theta)\psi_{t,\theta}dt = \big(P_{t,\theta}^{02*}C_t^* + P_{t,\theta,0}^{22}\big)(\Psi_t\Psi_t^*)^{-1} \\ \qquad \times(dz_t^* - C_t\hat{x}_t^*dt - \psi_{t,0}dt),\ \mu_t < \theta \le 0,\ 0 < t \le T; \\ \psi_{0,\theta} = 0,\ -\lambda_0^{-1} < \theta \le 0; \\ \psi_{t,\mu_t} = 0,\ 0 < t \le T. \end{cases} \tag{10.66}$$

for the best estimate. Also, in the limit, (10.55) yields

$$u_t^* = -G_t^{-1}\big(B_t^*Q_t^{00} + L_t\big)\hat{x}_t^*,\ \text{a.e, } t \in \mathbf{T}. \tag{10.67}$$

The equations (10.63)–(10.67) together with (10.60) and (10.27) represent the optimal control and the optimal filter in the linear stochastic regulator and filtering problems determined by (10.50)–(10.52). We add two more formulae for the covariance of the error and for the minimum of the functional which easily follow from (10.61)–(10.62):

$$\operatorname{cov}(x_t - \hat{x}_t) = P_t^{00},\ 0 \le t \le T, \tag{10.68}$$

and

$$J(u^*) = \operatorname{tr}\bigl(Q_T^{00} P_T^{00}\bigr) + \operatorname{tr} \int_0^T F_t P_t^{00}\, dt$$
$$+ \int_0^T \Psi_t^{-1} \bigl(C_t P_t^{00} + P_{t,0}^{02*}\bigr) Q_t^{00} \bigl(P_t^{00} C_t^* + P_{t,0}^{02}\bigr) \Psi_t^{-1*}\, dt. \qquad (10.69)$$

Chapter 11

Duality

In this chapter the distinction between the classical and extended forms of the separation principle is explained by use of duality between the control and estimation problems.

Convention. In this chapter it is always assumed that $(\Omega, \mathcal{F}, \mathbf{P})$ is a complete probability space, $X, Z \in \mathcal{H}$, $T > 0$, $\mathbf{T} = [0, T]$ is a finite time interval and $\Delta_T = \{(t,s) : 0 \le s \le t \le T\}$.

11.1 Classical Separation Principle and Duality

In Sections 6.2–6.4 and 9.2–9.4 the duality principle was used to reduce estimation problems to linear regulator problems. In this section we try to substitute a linear regulator problem by a linear stochastic optimal control problem in this duality.

Under the conditions $(\mathbf{E_1^w})$–$(\mathbf{E_2^w})$, consider the filtering problem (6.1)–(6.2). By Theorem 6.4, the best estimate in this problem has the linear feedback form (6.16) with $\mathcal{Y}$ and P as defined in Theorem 6.4.

Introduce the functions W, V, R and the operator P_0 from (6.3) and let

$(\mathbf{D_1})$ $\mathcal{R} = \mathcal{D}_T(\mathcal{U})$, $B = D_T(C)$, $F = D_T(W)$, $G = D_T(V)$, $L = D_T(R)$, $Q_T = P_0$;

$(\mathbf{D_2})$ $b \in L_2(\mathbf{T} \times \Omega, X)$, $c \in L_2(\mathbf{T} \times \Omega, Z)$, $\mathbf{E}b_t = 0$ for a.e. $t \in \mathbf{T}$, $D \in B_\infty(\mathbf{T}, \mathcal{L}(X, Z))$, $m \in M_2(\mathbf{T}, X)$, $n \in M_2(\mathbf{T}, Z)$.

Under $(\mathbf{D_1})$ and $(\mathbf{D_2})$, consider the partially observable linear quadratic optimal control problem of minimizing the functional

$$J(u) = \mathbf{E}\left(\langle x_T^u, Q_T x_T^u \rangle + \int_0^T \left(\langle x_t^u, F_t x_t^u \rangle + 2 \langle u_t, L_t x_t^u \rangle + \langle u_t, G_t u_t \rangle \right) dt \right), \quad (11.1)$$

where

$$x_t^u = \int_0^t \mathcal{R}_{t,s}(B_s u_s + b_s)\, ds + \int_0^t \mathcal{R}_{t,s}\, dm_s,\ 0 \le t \le T, \tag{11.2}$$

$$z_t^u = \int_0^t (D_s x_s^u + c_s)\, ds + n_t,\ 0 \le t \le T, \tag{11.3}$$

and

$$u \in U_{\text{ad}} = \big\{u \in L_2(\mathbf{T}, L_2(\Omega, \mathbb{R}^n)) : \\ u_t \in L_2\big(\Omega, \sigma\big(z^{u,t}\big) \cap \sigma\big(z^{0,t}\big), \mathbf{P}, \mathbb{R}^n\big),\ \text{for a.e. } t \in \mathbf{T}\big\}. \tag{11.4}$$

Assume

($\mathbf{D_3}$) $\sigma(b^t, c^t, n_s, m_s;\ 0 \le s \le t)$ and $\sigma(b^{t+}, m_s - m_t;\ t < s \le T)$ are independent for all $0 < t \le T$;

where the symbols b^t and b^{t+} are introduced in Section 5.1.1. Then by the classical separation principle (see Theorem 5.17), if there exists an optimal control u^* in the problem (11.1)–(11.4), then it has the form

$$u_t^* = -G_t^{-1}(B_t^* Q_t + L_t)\mathbf{E}_t^* x_t^*,\ \text{a.e. } t \in \mathbf{T}, \tag{11.5}$$

where $x^* = x^{u^*}$, $\mathbf{E}_t^* = \mathbf{E}\big(\cdot\,\big|z_s^{u^*};\ 0 \le s \le t\big)$ and Q is a solution of the Riccati equation

$$Q_t = \mathcal{R}_{T,t}^* Q_T \mathcal{R}_{T,t} + \int_t^T \mathcal{R}_{s,t}^*\big(F_s \\ - (Q_s B_s + L_s^*)G_s^{-1}(B_s^* Q_s + L_s)\big)\mathcal{R}_{s,t}\, ds,\ 0 \le t \le T. \tag{11.6}$$

Let $\mathcal{K} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{R})$ and let

$$\hat{u}_t = u_t^* + G_t^{-1}(B_t^* Q_t + L_t)\int_0^t \mathcal{K}_{t,s} B_s G_s^{-1} \\ \times\ (B_s^* Q_s + L_s)\,(\mathbf{E}_t^* x_s^* - \mathbf{E}_s^* x_s^*)\, ds,\ \text{a.e. } t \in \mathbf{T}. \tag{11.7}$$

To find another expression for $\hat{u}$, substitute (11.5) in (11.2). Then using Proposition 4.28, we obtain

$$x_t^* = -\int_0^t \mathcal{R}_{t,s} B_s G_s^{-1}(B_s^* Q_s + L_s)\mathbf{E}_s^* x_s^*\, ds + \int_0^t \mathcal{R}_{t,s}(b_s\, ds + dm_s) \\ = \int_0^t \mathcal{K}_{t,s} B_s G_s^{-1}(B_s^* Q_s + L_s)\,(x_s^* - \mathbf{E}_s^* x_s^*)\, ds + \int_0^t \mathcal{K}_{t,s}(b_s\, ds + dm_s),$$

and hence,

$$\begin{aligned}\mathbf{E}_t^* x_t^* - \mathbf{E}_t^* \int_0^t \mathcal{K}_{t,s}(b_s\,ds + dm_s) \\ = \int_0^t \mathcal{K}_{t,s} B_s G_s^{-1}(B_s^* Q_s + L_s)\,(\mathbf{E}_t^* x_s^* - \mathbf{E}_s^* x_s^*)\,ds.\end{aligned} \tag{11.8}$$

Thus, the substitution of (11.5) and (11.8) in (11.7) yields

$$\hat{u}_t = -\,\mathbf{E}_t^* \int_0^t G_t^{-1}(B_t^* Q_t + L_t)\mathcal{K}_{t,s}(b_s\,ds + dm_s), \text{ a.e. } t \in \mathbf{T}. \tag{11.9}$$

Using $(\mathbf{D_1})$, in a similar way as in the proof of Theorem 6.4, one can show that $Q = D_T(P)$ and $\mathcal{K} = \mathcal{D}_T(\mathcal{Y})$, where P and $\mathcal{Y}$ are defined in Theorem 6.4. Therefore, comparing (6.16) and (11.9), we conclude that if K is defined by (6.20), then

$$\hat{x}_t = \int_0^t K_{t,s}\,dz_s \quad\text{and}\quad \hat{u}_t = -\mathbf{E}_t^* \int_0^t K_{T-s,T-t}^*(b_s\,ds + dm_s). \tag{11.10}$$

Thus, we can state the following result.

Theorem 11.1. *Under the conditions* $(\mathbf{E_1^w})$–$(\mathbf{E_2^w})$ *and* $(\mathbf{D_1})$–$(\mathbf{D_3})$*, the filtering problem* (6.1)–(6.2) *and the partially observable linear quadratic optimal control problem* (11.1)–(11.4) *are dual in the sense that if the operator-valued functions used in setting of these problems are related as in* $(\mathbf{D_1})$*, then the best estimate* $\hat{x}$ *and the process* $\hat{u}$*, defined by* (11.7)*, are related in the same form.*

Proof. This follows from (11.10). □

One may think that the above duality is artificial and refer to the process $\hat{u}$ which is used in this duality instead of the optimal control u^*. The result of the next section shows that Theorem 11.1 should be considered more seriously than the initial impression.

11.2 Extended Separation Principle and Duality

In this section Theorem 11.1 will be generalized to the case of the extended separation principle.

Under the conditions $(\mathbf{E_1^w})$–$(\mathbf{E_2^w})$ and the notation in (6.3), consider the smoothing problem (6.1)–(6.2). By Theorem 6.18, the best estimate $\hat{x}_t^T$ in this problem has the form

$$\hat{x}_t^T = \int_0^t \mathcal{Y}_{t,s}(P_s C_s^* + R_s)V_s^{-1}\,dz_s + \int_t^T P_t \mathcal{Y}_{s,t}^* C_s^* V_s^{-1}\,d\bar{z}_s,\ 0 \le t \le T, \tag{11.11}$$

where $\mathcal{Y}$, P and $\bar{z}$ are defined in Theorem 6.18.

Let $(\mathbf{D_1})$ and $(\mathbf{D_2})$ hold and consider the partially observable linear quadratic optimal control problem (11.1)–(11.4). Since we are not restricted by the condition $(\mathbf{D_3})$, the extended separation principle (see Theorem 5.16) must be applied to the problem (11.1)–(11.4). Consequently, we obtain that if there exists an optimal control u^* in this problem, then it has the form

$$\begin{aligned} u_t^* =& -G_t^{-1}(B_t^* Q_t + L_t)\mathbf{E}_t^* x_t^* \\ & - G_t^{-1} B_t^* \mathbf{E}_t^* \int_t^T \mathcal{K}_{s,t}^* Q_s(b_s\, ds + dm_s), \text{ a.e. } t \in \mathbf{T}, \end{aligned} \tag{11.12}$$

where $x^* = x^{u^*}$, $\mathbf{E}_t^* = \mathbf{E}(\cdot\,|z_s^{u^*};\, 0 \le s \le t)$, Q is the solution of the Riccati equation (11.6) and $\mathcal{K} = \mathcal{P}_{-BG^{-1}(B^*Q+L)}(\mathcal{R})$.

Let

$$\bar{m}_t = m_t + \int_0^t b_s\, ds + \int_0^t B_s \tilde{u}_s\, ds,\ 0 \le t \le T, \tag{11.13}$$

$$\tilde{u}_t = -G_t^{-1} B_t^* \mathbf{E}_t^* \int_t^T \mathcal{K}_{s,t}^* Q_s(b_s\, ds + dm_s), \text{ a.e. } t \in \mathbf{T}, \tag{11.14}$$

and define the process $\hat{u}$ by (11.7). To find another expression for $\hat{u}$, substitute (11.12) in (11.2). Then by (11.13)–(11.14) and by Proposition 4.28, we obtain

$$\begin{aligned} x_t^* =& -\int_0^t \mathcal{R}_{t,s} B_s G_s^{-1}(B_s^* Q_s + L_s)\mathbf{E}_s^* x_s^*\, ds + \int_0^t \mathcal{R}_{t,s}(b_s\, ds + dm_s) \\ & - \int_0^t \mathcal{R}_{t,s} B_s G_s^{-1} B_s^* \mathbf{E}_s^* \int_s^T \mathcal{K}_{r,s}^* Q_r(b_r\, dr + dm_r)\, ds \\ =& -\int_0^t \mathcal{R}_{t,s} B_s G_s^{-1}(B_s^* Q_s + L_s)\mathbf{E}_s^* x_s^*\, ds + \int_0^t \mathcal{R}_{t,s}\, d\bar{m}_s \\ =& \int_0^t \mathcal{K}_{t,s} B_s G_s^{-1}(B_s^* Q_s + L_s)\,(x_s^* - \mathbf{E}_s^* x_s^*)\, ds + \int_0^t \mathcal{K}_{t,s}\, d\bar{m}_s, \end{aligned}$$

and hence,

$$\mathbf{E}_t^* x_t^* - \mathbf{E}_t^* \int_0^t \mathcal{K}_{t,s}\, d\bar{m}_s = \int_0^t \mathcal{K}_{t,s} B_s G_s^{-1}(B_s^* Q_s + L_s)\,(\mathbf{E}_t^* x_s^* - \mathbf{E}_s^* x_s)\, ds. \tag{11.15}$$

Substituting (11.12) and (11.15) in (11.7), we obtain

$$\begin{aligned} \hat{u}_t =& -\mathbf{E}_t^* \int_0^t G_t^{-1}(B_t^* Q_t + L_t)\mathcal{K}_{t,s}\, d\bar{m}_s \\ & - \mathbf{E}_t^* \int_t^T G_t^{-1} B_t^* \mathcal{K}_{s,t}^* Q_s(b_s\, ds + dm_s), \text{ a.e. } t \in \mathbf{T}. \end{aligned}$$

By $(\mathbf{D_1})$, we have $Q = D_T(P)$ and $\mathcal{K} = \mathcal{D}_T(\mathcal{Y})$, where P and $\mathcal{Y}$ are as defined in Theorem 6.18. Therefore, if K is defined by (6.20) and

$$G_{t,s} = P_t \mathcal{Y}_{s,t}^* C_s^* (\Psi_s V \Psi_s^*)^{-1}, \ 0 \le s \le t \le T, \tag{11.16}$$

then

$$\hat{x}_t^T = \int_0^t K_{t,s} \, dz_s + \int_t^T G_{t,s} \, d\bar{z}_s, \tag{11.17}$$

$$\hat{u}_t = -\mathbf{E}_t^* \int_0^t K_{T-s,T-t}^* \, d\bar{m}_s - \mathbf{E}_t^* \int_t^T G_{T-s,T-t}^* \, dm_s. \tag{11.18}$$

Theorem 11.2. *Under the conditions* $(\mathbf{E_1^w})$–$(\mathbf{E_2^w})$ *and* $(\mathbf{D_1})$–$(\mathbf{D_2})$*, the smoothing problem* (6.1)–(6.2) *and the partially observable linear quadratic optimal control problem* (11.1)–(11.4) *are dual in the sense that if the operator-valued functions used in setting of these problems are related as in* $(\mathbf{D_1})$*, then the best estimate process* $\hat{x}^T$ *in the smoothing problem* (6.1)–(6.2) *and the process* $\hat{u}$*, defined by* (11.7)*, are related in the same form.*

Proof. This follows from (11.17) and (11.18). □

Summarizing Theorems 11.1 and 11.2, one can observe that the distinction between the classical and extended forms of the separation principle is the same as the distinction between optimal filter and optimal smoother for the system (6.1)–(6.2).

11.3 Innovation Process for Control Actions

By the duality stated in Theorem 11.2, the innovation process $\bar{z}$ for the smoothing problem (6.1)–(6.2) corresponds to the process $\bar{m}$ for the control problem (11.1)–(11.4). Therefore, the process $\bar{m}$, defined by (11.13)–(11.14), will be called a *dual analogue of the innovation process for control actions*. Innovation processes play a significant role in studying estimation problems; especially they are helpful in derivation of nonlinear filtering equations. Hence, we can expect the same from their dual analogue for control actions. In particular, if the process $\tilde{u}$ is given beforehand, then substituting $u = \eta + \tilde{u}$, the state-observation system (11.2)–(11.3) can be reduced to

$$x_t^\eta = x_t^u = \int_0^t \mathcal{R}_{t,s} B_s \eta_s \, ds + \int_0^t \mathcal{R}_{t,s} \, d\bar{m}_s, \ 0 \le t \le T,$$

$$z_t^\eta = z_t^u = \int_0^t (D_s x_s^\eta + c_s) \, ds + n_t, \ 0 \le t \le T.$$

If we write the functional (11.1) in terms of the new control action η, then the optimal control η^* in the reduced control problem is

$$\eta_t^* = u_t^* - \tilde{u}_t = -G^{-1}(B_t^* Q_t + L_t) \mathbf{E}_t^* x_t^*, \text{ a.e. } t \in \mathbf{T}.$$

Thus, for the reduced control problem, which is set in terms of the new noise process $\bar{m}$, the classical separation principle holds instead of the extended one that holds for the original control problem.

Example 11.3. To illustrate the ideas mentioned in this section, consider the linear stochastic regulator problem (8.68)–(8.70) under the conditions of Example 8.25 assuming $x_0 = 0$. The optimal control u^* in this problem was found as a function of three random processes $\hat{x}^*$, ψ^1 and ψ^2. Here $\hat{x}^*$ is the basic best estimate process, but ψ^1 and ψ^2 are two associated random processes. From the equations in (8.92) it is clear that the process ψ^2 plays a role in forming the innovation process $\bar{z}$ for estimation problems:

$$d\bar{z}_t = dz_t^* + \big(C_t\hat{x}_t^* + \psi_{t,0}^2\big)dt, \ \ \bar{z}_0 = 0.$$

The role of the process ψ^1 becomes clear due to the dual analogue of the innovation process for control actions. The process ψ^1 forms the process $\tilde{u}$ and, hence, the process $\bar{m}$ defined by (11.13)–(11.14) for the linear stochastic regulator problem from Example 8.25.

It is useful to mention the following: the reduction method used in Chapter 8 is in fact an application of the dual analogue of the innovation process for control actions. Exactly this was the reason for obtaining the formulae for the optimal control in Chapter 8 without referring to the extended separation principle.

The arguments mentioned in this example are applicable to the linear stochastic regulator problem from Chapter 7 and they can be used to analyze the linear stochastic regulator problem from Chapter 9.

Chapter 12

Controllability

In this chapter the concepts of controllability for the deterministic and stochastic systems are discussed.

12.1 Preliminaries

12.1.1 Definitions

Let $T > 0$ and consider a deterministic or stochastic control system on the time interval $[0, T]$. Let x_T^u be its (random or not) state value at time T corresponding to the control u taken from the set of admissible controls U_{ad}. If the control system under consideration is stochastic, then by $\mathcal{F}_T^u$ we denote the smallest σ-algebra generated by the observations on the time interval $[0, T]$ corresponding to the control u. If the considered control system is deterministic, then simply $\mathcal{F}_T^u = \{\varnothing, \Omega\}$. Suppose that X is the state space. Introduce the set

$$D(T) = \{x_T^u : u \in U_{\mathrm{ad}}\}. \tag{12.1}$$

Definition 12.1. Given $T > 0$, a deterministic control system will be called

(a) D_T^c*-controllable* if $D(T) = X$;

(b) D_T^a*-controllable* if $\overline{D(T)} = X$.

It is clear that the D_T^c-controllability is the well-known complete controllability and the D_T^a-controllability is the approximate controllability for the time T for deterministic control systems. Originally, the D_T^c-controllability was introduced in Kalman [61] as a concept for finite dimensional deterministic control systems. The natural extension of this concept to infinite dimensional control systems is too strong for many of them. Therefore, the D_T^a-controllability was introduced as a weakened version of the D_T^c-controllability.

A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*

The natural extension of the complete and approximate controllability concepts to stochastic control systems is meaningless since now a terminal value is a random variable. Therefore, there is a need for further weakening of these concepts in order to extend them to stochastic control systems. The two different interconnections of controllability and randomness define the two principally different methods of extending the controllability concepts to stochastic systems. In the first method the state space in the definition of controllability concepts is replaced by a suitable space of random variables, say, the space of square integrable random variables. Thus, attaining random variables, even those with large entropy, is necessary to be controllable in this sense. In this chapter we will follow the second method that is more practical: it assumes attaining only those random variables that have small entropy, excluding the needless random variables with large entropy.

Given $T > 0$, $0 \le \varepsilon < \infty$ and $0 \le p \le 1$, introduce the sets

$$\begin{aligned} S(T,\varepsilon,p) = \big\{h \in X : \exists u \in U_{\mathrm{ad}} \text{ such that} \\ \mathbf{P}\big(\|\mathbf{E}(x_T^u|\mathcal{F}_T^u) - h\|^2 > \varepsilon\big) \le 1-p\big\}, \end{aligned} \tag{12.2}$$

$$\begin{aligned} C(T,\varepsilon,p) = \big\{h \in X : \exists u \in U_{\mathrm{ad}} \text{ such that } h = \mathbf{E}x_T^u \text{ and} \\ \mathbf{P}\big(\|\mathbf{E}(x_T^u|\mathcal{F}_T^u) - h\|^2 > \varepsilon\big) \le 1-p\big\}. \end{aligned} \tag{12.3}$$

The following definitions will be used as a step in discussing the main concepts of controllability for stochastic systems. A stochastic control system will be called

(a) $S^c_{T,\varepsilon,p}$-*controllable* if $S(T,\varepsilon,p) = X$;

(b) $S^a_{T,\varepsilon,p}$-*controllable* if $\overline{S(T,\varepsilon,p)} = X$;

(c) $C^c_{T,\varepsilon,p}$-*controllable* if $C(T,\varepsilon,p) = X$;

(d) $C^a_{T,\varepsilon,p}$-*controllable* if $\overline{C(T,\varepsilon,p)} = X$;

(e) $S^0_{T,\varepsilon,p}$-*controllable* if $0 \in S(T,\varepsilon,p)$.

Geometrically, the $S^c_{T,\varepsilon,p}$-controllability ($S^a_{T,\varepsilon,p}$-controllability) can be interpreted as follows. If a control system with the initial state x_0 is $S^c_{T,\varepsilon,p}$-controllable ($S^a_{T,\varepsilon,p}$-controllable), then with probability not less than p it can pass from x_0 for the time T into the $\sqrt{\varepsilon}$-neighborhood of an arbitrary point in the state space (in a set that is dense in the state space). The interpretation of the $C^c_{T,\varepsilon,p}$- and $C^a_{T,\varepsilon,p}$-controllability differs from the same of the $S^c_{T,\varepsilon,p}$- and $S^a_{T,\varepsilon,p}$-controllability since among the controls, with the help of which the $\sqrt{\varepsilon}$-neighborhood of any point h is achieved, there exists one with the property that the expectation of the state at the time T, corresponding to this control, coincides with h. One can easily observe that a $C^c_{T,\varepsilon,p}$-controllable ($C^a_{T,\varepsilon,p}$-controllable) control system is $S^c_{T,\varepsilon,p}$-controllable ($S^a_{T,\varepsilon,p}$-controllable), but the converse is not true.

The smaller ε is and the larger p is for a control system, the more controllable it is, i.e., it is possible to hit into a smaller neighborhood with a higher probability. One can observe that for all $T > 0$, all control systems are $S^c_{T,\varepsilon,p}$-, $S^a_{T,\varepsilon,p}$-, $C^c_{T,\varepsilon,p}$- and $C^a_{T,\varepsilon,p}$-controllable with $\varepsilon \geq 0$ and $p = 0$ or $\varepsilon = \infty$ and $0 \leq p \leq 1$ if we admit ∞ as a value for ε. At the same time it is clear that a D^c_T-controllable (D^a_T-controllable) deterministic system is $S^c_{T,0,1}$- and $C^c_{T,0,1}$-controllable ($S^a_{T,0,1}$- and $C^a_{T,0,1}$-controllable) with parameters $\varepsilon = 0$ and $p = 1$, since for deterministic systems, $D(T) = S(T, 0, 1) = C(T, 0, 1)$. Also, each kind of controllability, introduced above, with a smaller ε and a greater p implies the same kind of controllability with a greater ε and a smaller p.

Summarizing, we can give the following easy necessary and sufficient conditions for the D^c_T- and D^a_T-controllability.

Proposition 12.2. *Given $T > 0$, for a deterministic control system the following three conditions are equivalent:*

(a) *the D^c_T-controllability;*

(b) *the $S^c_{T,\varepsilon,p}$-controllability for all $\varepsilon \geq 0$ and for all $0 \leq p \leq 1$;*

(c) *the $C^c_{T,\varepsilon,p}$-controllability for all $\varepsilon \geq 0$ and for all $0 \leq p \leq 1$.*

Proposition 12.3. *Given $T > 0$, for a deterministic control system the following three conditions are equivalent:*

(a) *the D^a_T-controllability;*

(b) *the $S^a_{T,\varepsilon,p}$-controllability for all $\varepsilon \geq 0$ and for all $0 \leq p \leq 1$;*

(c) *the $C^a_{T,\varepsilon,p}$-controllability for all $\varepsilon \geq 0$ and for all $0 \leq p \leq 1$.*

Excepting the limit values $\varepsilon = 0$ and $p = 1$ from the above mentioned necessary and sufficient conditions of the complete and approximate controllability, one can obtain the weakened versions of these concepts. For a moment call a given stochastic system

(a) *S^c_T-controllable* if it is $S^c_{T,\varepsilon,p}$-controllable for all $\varepsilon > 0$ and for all $0 \leq p < 1$;

(b) *S^a_T-controllable* if it is $S^a_{T,\varepsilon,p}$-controllable for all $\varepsilon > 0$ and for all $0 \leq p < 1$;

(c) *C^c_T-controllable* if it is $C^c_{T,\varepsilon,p}$-controllable for all $\varepsilon > 0$ and for all $0 \leq p < 1$;

(d) *C^a_T-controllable* if it is $C^a_{T,\varepsilon,p}$-controllable for all $\varepsilon > 0$ and for all $0 \leq p < 1$.

These concepts of controllability can be easily described in terms of the sets

$$S(T) = \bigcap_{\varepsilon,p} S(T, \varepsilon, p) \text{ and } C(T) = \bigcap_{\varepsilon,p} C(T, \varepsilon, p), \tag{12.4}$$

where the intersections are taken over all $\varepsilon > 0$ and all $0 \leq p < 1$.

Proposition 12.4. *For a given stochastic system and for $T > 0$, let $S(T)$ and $C(T)$ be defined by* (12.4) *and let X be the state space. Then this stochastic system is*

(a) *S_T^c-controllable if and only if $S(T) = X$;*

(b) *S_T^a-controllable if and only if $\overline{S(T)} = X$;*

(c) *C_T^c-controllable if and only if $C(T) = X$;*

(d) *C_T^a-controllable if and only if $\overline{C(T)} = X$.*

In the next result we establish that S_T^c-controllability and S_T^a-controllability are equivalent concepts for every stochastic system.

Proposition 12.5. *For a given stochastic system and for a given $T > 0$, let $S(T)$ be the set defined by* (12.4) *and let X be the respective state space. Then the following statements hold.*

(a) *$S(T)$ is a closed set in X.*

(b) *The stochastic system is S_T^c-controllable if and only if it is S_T^a-controllable.*

Proof. Part (b) easily follows from part (a) and from Propositions 12.4(a) and 12.4(b). Hence, it suffices to prove only part (a). Let $h \in \overline{S(T)}$. Fix arbitrary $\varepsilon > 0$ and $0 \le p < 1$. Then from

$$h \in \overline{S(T)} \subset \overline{S(T, \varepsilon/4, p)},$$

it follows that there exists $h_0 \in S(T, \varepsilon/4, p)$ such that $\|h_0 - h\|^2 \le \varepsilon/4$. At the same time, $h_0 \in S(T, \varepsilon/4, p)$ implies that there exists $u \in U_{\text{ad}}$ with

$$\mathbf{P}\{\|\mathbf{E}(x_T^u|\mathcal{F}_T^u) - h_0\|^2 > \varepsilon/4\} \le 1 - p.$$

Hence, for this $u \in U_{\text{ad}}$, we have

$$\begin{aligned}
\mathbf{P}\{\|\mathbf{E}(x_T^u|\mathcal{F}_T^u) - h\|^2 > \varepsilon\} &\le \mathbf{P}\{\|\mathbf{E}(x_T^u|\mathcal{F}_T^u) - h_0\| + \|h_0 - h\| > \sqrt{\varepsilon}\} \\
&\le \mathbf{P}\{\|\mathbf{E}(x_T^u|\mathcal{F}_T^u) - h_0\| + \sqrt{\varepsilon}/2 > \sqrt{\varepsilon}\} \\
&= \mathbf{P}\{\|\mathbf{E}(x_T^u|\mathcal{F}_T^u) - h_0\|^2 > \varepsilon/4\} \\
&\le 1 - p.
\end{aligned}$$

This implies $h \in S(T, \varepsilon, p)$. Since $\varepsilon > 0$ and $0 \le p < 1$ are arbitrary, we obtain $h \in S(T)$, proving that $\overline{S(T)} \subset S(T)$, i.e., $S(T)$ is closed. □

Also, it will be shown that for every partially observable linear stationary control system with an additive Gaussian white noise disturbance, the C_T^a-controllability is equivalent to the S_T^c-controllability and the S_T^a-controllability (see Proposition 12.32). So, we can define two basic and one additional concepts of controllability for stochastic systems.

Definition 12.6. For a given stochastic system and for $T > 0$, let $S(T)$ and $C(T)$ be the sets defined by (12.4) and let X be the respective state space. Then this stochastic control system will be called

(a) S_T*-controllable* if $S(T) = X$;

(b) C_T*-controllable* if $C(T) = X$;

(c) S_T^0*-controllable* if $0 \in S(T)$.

Geometrically, the S_T-controllability can be interpreted as follows: an S_T-controllable control system can attain for the time T an arbitrarily small neighborhood of each point in the state space with a probability that is arbitrarily near to 1. The C_T-controllability is the S_T-controllability fortified with some uniformity. The S_T^0-controllability is useful in discussing S_T- and C_T-controllability.

Finally, notice that the abbreviations D, S, C, c and a in the previously introduced controllability concepts mean deterministic, stochastic, combined, complete and approximate, respectively.

12.1.2 Description of the System

We will examine the S_T- and C_T-controllability of the partially observable linear system

$$\begin{cases} dx_t^u = (Ax_t^u + Bu_t + f_t)\,dt + d\varphi_t, \ 0 < t \le T, \ x_0^u = x_0, \\ d\xi_t^u = Cx_t^u\,dt + d\psi_t, \ 0 < t \le T, \ \xi_0^u = 0, \end{cases} \tag{12.5}$$

where x, u and ξ are the state, control and observation processes. Under the set U_{ad} of admissible controls we consider the set of all controls u in the linear feedback form

$$u_t = \bar{u}_t + \int_0^t K_{t,s}\,d\xi_s^u, \ \text{a.e. } t \in [0,T], \tag{12.6}$$

with $\bar{u} \in L_2(0,T;U)$ and $K \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$.

Throughout this chapter it is assumed that

(**C**) $X, U \in \mathcal{H}$, $T > 0$, A is the infinitesimal generator of $\mathcal{U} \in \mathcal{S}(X)$, $B \in \mathcal{L}(U,X)$, $C \in \mathcal{L}(X, \mathbb{R}^n)$, $f \in L_2(0,T;X)$, x_0 is an X-valued Gaussian random variable, $\begin{bmatrix} \varphi \\ \psi \end{bmatrix}$ is an $X \times \mathbb{R}^n$-valued Wiener process on $[0,T]$, the covariance operator of ψ is I, x_0 and (φ, ψ) are independent.

Also, we use the notation

$$P_0 = \text{cov}x_0 \ \text{ and } \ \begin{bmatrix} \Phi & R \\ R^* & I \end{bmatrix} = T^{-1}\text{cov}\begin{bmatrix} \varphi_T \\ \psi_T \end{bmatrix}. \tag{12.7}$$

Note that for the Wiener processes under consideration we use the symbols φ and ψ, reserving their traditional symbols w and v for the control actions as defined below.

One can associate two systems with the system (12.5). The first of them is the deterministic system

$$\frac{d}{dt}y_t^v = Ay_t^v + Bv_t + f_t, \ 0 < t \le T, \ y_0^v = y_0 = \mathbf{E}x_0, \tag{12.8}$$

with the admissible controls v taken from $V_{\mathrm{ad}} = L_2(0,T;U)$. The second one is the partially observable stochastic system

$$\begin{cases} dz_t^w = (Az_t^w + Bw_t)\,dt + d\varphi_t, \ 0 < t \le T, \ z_0^w = z_0 = x_0 - \mathbf{E}x_0, \\ d\eta_t^w = Cz_t^w\,dt + d\psi_t, \ 0 < t \le T, \ \eta_0^w = 0, \end{cases} \tag{12.9}$$

where w is a control from the set of admissible controls W_{ad} consisting of all controls in the form

$$w_t = \int_0^t K_{t,s}\,d\eta_s^w, \ \text{a.e. } t \in [0,T], \tag{12.10}$$

where $K \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$.

12.2 Controllability: Deterministic Systems

In this section the D_T^c- and D_T^a-controllability of the deterministic control system (12.8) will be discussed.

12.2.1 CCC, ACC and Rank Condition

With the deterministic control system (12.8), one can associate the operator-valued function

$$\mathcal{Q}_T = \int_0^T \mathcal{U}_s BB^* \mathcal{U}_s^*\,ds, \ T \ge 0, \tag{12.11}$$

which is called a *controllability operator*.

Theorem 12.7. *The control system* (12.8) *on* V_{ad} *is*

(a) D_T^c*-controllable if and only if* $\mathcal{Q}_T > 0$;

(b) D_T^a*-controllable if and only if* $B^*\mathcal{U}_t^* x = 0$ *for all* $0 \le t \le T$ *implies* $x = 0$.

Proof. This theorem is proved in various books, for example in Curtain and Pritchard [40]. □

The condition in Theorem 12.7(a) is called the *complete controllability condition* (CCC) and in Theorem 12.7(b) the *approximate controllability condition* (ACC). These conditions are demonstrated in the following example.

Example 12.8. Let $X = U = l_2$ (see Example 1.10 for this space), let $A = 0$ and let

$$B = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1/2 & 0 & \cdots \\ 0 & 0 & 1/3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

By Example 3.3, the semigroup generated by $A = 0$ is $\mathcal{U}_t \equiv I$. Consider the standard basis

$$e_1 = (1, 0, 0, \ldots),\ e_2 = (0, 1, 0, \ldots),\ e_3 = (0, 0, 1, \ldots),\ \ldots$$

in l_2. Since

$$\sum_{n=1}^{\infty} \langle Be_n, Be_n \rangle = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

B is a Hilbert–Schmidt operator on l_2 ($B \in \mathcal{L}_2(l_2)$) and, therefore, $B \in \mathcal{L}(l_2)$. Obviously, $B = B^*$. Therefore, $B^* \mathcal{U}_t^* x = 0$ implies $Bx = 0$ and, hence, $x = 0$. So, by Theorem 12.7(b), for each $T > 0$, the control system (12.8) with the operators A and B as defined above is D_T^a-controllable. But by Theorem 12.7(a), there is no $T > 0$ for which it is D_T^c-controllable since

$$\left\| B^2 e_n \right\|_{l_2} = \frac{1}{n^2} \to 0 \text{ as } n \to \infty,$$

and, therefore,

$$\mathcal{Q}_T = \int_0^T \mathcal{U}_s B B^* \mathcal{U}_s^* \, ds = TB^2$$

is not a coercive operator.

If the state space X and the control space U are finite dimensional Euclidean spaces, say $X = \mathbb{R}^n$ and $U = \mathbb{R}^m$ for $n, m \in \mathbb{N}$, then the operators A and B in the control system (12.8) are simply $(n \times n)$- and $(n \times m)$-matrices. Therefore, we can form the $(n \times nm)$-matrix

$$\left[B, AB, \ldots, A^{n-1} B \right] \tag{12.12}$$

consisting of the columns of the matrices $B, AB, \ldots, A^{n-1}B$.

Theorem 12.9. *Assume $X = \mathbb{R}^n$ and $U = \mathbb{R}^m$ for $n, m \in \mathbb{N}$. Then the following statements are equivalent:*

(a) *the control system* (12.8) *on V_{ad} is D_T^c-controllable;*

(b) *the control system* (12.8) *on V_{ad} is D_T^a-controllable;*

(c) *the rank of the matrix* (12.12) *is equal to n.*

Proof. This theorem is proved in various books, for example in Curtain and Pritchard [40]. □

The condition in part (c) of Theorem 12.9 is called the *Kalman rank condition.*

12.2.2 Resolvent Conditions

In this section the necessary and sufficient conditions in terms of convergence of operators will be obtained for the control system (12.8) on V_{ad} to be D_T^c- and D_T^a-controllable.

Consider the controllability operator $\mathcal{Q}_T$ of the control system (12.8) defined by (12.11). For $T \geq 0$, the operator $\mathcal{Q}_T$ is nonnegative ($\mathcal{Q}_T \geq 0$) and, hence, $R(\lambda, -\mathcal{Q}_T) = (\lambda I + \mathcal{Q}_T)^{-1}$ is a well-defined bounded linear operator for all $\lambda > 0$ and for all $T \geq 0$. If $\mathcal{Q}_T > 0$, then $R(\lambda, -\mathcal{Q}_T)$ is defined for $\lambda = 0$ as well. The operator $R(\lambda, -\mathcal{Q}_T)$ is called the *resolvent* of $-\mathcal{Q}_T$. This resolvent will be used to represent the optimal control in the linear regulator problem of minimizing the functional

$$J(v) = \|y_T^v - h\|^2 + \lambda \int_0^T \|v_t\|^2 \, dt, \tag{12.13}$$

where y^v is defined by (12.8), v is a control taken from $L_2(0,T;U)$, $h \in X$ and $\lambda > 0$ are parameters.

Lemma 12.10. *Given $h \in X$ and $\lambda > 0$, there exists a unique optimal control $v^\lambda \in L_2(0,T;U)$ at which the functional* (12.13) *takes its minimum value on $L_2(0,T;U)$. Furthermore,*

$$v_t^\lambda = -B^*\mathcal{U}_{T-t}^* R(\lambda, -\mathcal{Q}_T)(\mathcal{U}_T y_0 - h + g), \text{ a.e. } t \in [0,T], \tag{12.14}$$

and

$$y_T^{v^\lambda} - h = \lambda R(\lambda, -\mathcal{Q}_T)(\mathcal{U}_T y_0 - h + g), \tag{12.15}$$

where $R(\lambda, -\mathcal{Q}_T)$ is the resolvent of $-\mathcal{Q}_T$ and

$$g = \int_0^T \mathcal{U}_{T-t} f_t \, dt.$$

Proof. By Theorem 5.24, there exists a unique optimal control v^λ in the above mentioned linear regulator problem. Computing the variation of the functional (12.13), one can easily obtain

$$v_t^\lambda = -\lambda^{-1} B^*\mathcal{U}_{T-t}^*\left(y_T^{v^\lambda} - h\right), \text{ a.e. } t \in [0,T]. \tag{12.16}$$

Substituting this in (12.8) and using (12.11), we obtain

$$\begin{aligned} y_T^{v^\lambda} &= \mathcal{U}_T y_0 + \int_0^T \mathcal{U}_{T-t}(Bv_t^\lambda + f_t)\,dt \\ &= \mathcal{U}_T y_0 + g - \lambda^{-1} \int_0^T \mathcal{U}_{T-t} BB^* \mathcal{U}_{T-t}^* \left(y_T^{v^\lambda} - h\right) dt \\ &= \mathcal{U}_T y_0 + g - \lambda^{-1} \mathcal{Q}_T \left(y_T^{v^\lambda} - h\right). \end{aligned}$$

Hence,

$$\lambda y_T^{v^\lambda} = \lambda(\mathcal{U}_T y_0 + g) - \mathcal{Q}_T\left(y_T^{v^\lambda} - h\right),$$

which implies

$$(\lambda I + \mathcal{Q}_T) y_T^{v^\lambda} = \lambda(\mathcal{U}_T y_0 + g) + \mathcal{Q}_T h$$

and, consequently,

$$\begin{aligned} y_T^{v^\lambda} &= \lambda(\lambda I + \mathcal{Q}_T)^{-1}(\mathcal{U}_T y_0 + g) + (\lambda I + \mathcal{Q}_T)^{-1}(\lambda I + \mathcal{Q}_T - \lambda I)h \\ &= \lambda R(\lambda, -\mathcal{Q}_T)(\mathcal{U}_T y_0 + g - h) + h. \end{aligned}$$

Thus, (12.15) holds. Substituting (12.15) in (12.16), we obtain (12.14). □

Theorem 12.11. *The following statements are equivalent:*

(a) *the control system* (12.8) *on* V_{ad} *is* D_T^c*-controllable;*

(b) $\mathcal{Q}_T > 0$;

(c) $R(\lambda, -\mathcal{Q}_T)$ *converges uniformly as* $\lambda \to 0$;

(d) $R(\lambda, -\mathcal{Q}_T)$ *converges strongly as* $\lambda \to 0$;

(e) $R(\lambda, -\mathcal{Q}_T)$ *converges weakly as* $\lambda \to 0$;

(f) $\lambda R(\lambda, -\mathcal{Q}_T)$ *converges uniformly to the zero operator as* $\lambda \to 0$;

Proof. The equivalence (a) ⇔ (b) is well-known and it is already stated in Theorem 12.7(a). For the implication (b) ⇒ (c), let $\mathcal{Q}_T > 0$. Then for all $x \in X$ and for all $\lambda \geq 0$,

$$\langle x, (\lambda I + \mathcal{Q}_T)x\rangle \geq (\lambda + k)\|x\|^2,$$

where $k > 0$ is a constant. Therefore, for all $\lambda \geq 0$,

$$\|R(\lambda, -\mathcal{Q}_T)\| = \left\|(\lambda I + \mathcal{Q}_T)^{-1}\right\| \leq \frac{1}{\lambda + k} \leq \frac{1}{k}.$$

We obtain that $\|R(\lambda, -\mathcal{Q}_T)\|$ is bounded with respect to $\lambda \geq 0$. This implies

$$\begin{aligned}\left\|R(\lambda, -\mathcal{Q}_T) - \mathcal{Q}_T^{-1}\right\| &= \left\|(\lambda I + \mathcal{Q}_T)^{-1} - \mathcal{Q}_T^{-1}\right\| \\ &= \left\|\mathcal{Q}_T^{-1}(\mathcal{Q}_T - \lambda I - \mathcal{Q}_T)(\lambda I + \mathcal{Q}_T)^{-1}\right\| \\ &\leq \lambda\|\mathcal{Q}_T^{-1}\|\,\|(\lambda I + \mathcal{Q}_T)^{-1}\| \\ &\leq \lambda k^{-2}.\end{aligned}$$

So, $R(\lambda, -\mathcal{Q}_T)$ converges uniformly to $\mathcal{Q}_T^{-1}$ as $\lambda \to 0$. The implications (c) $\Rightarrow$ (d) $\Rightarrow$ (e) are obvious. The implication (e) $\Rightarrow$ (f) follows from the boundedness of a weakly convergent sequence of operators (see Proposition 1.32(a)). For the implication (f) $\Rightarrow$ (b), suppose

$$\lambda\|R(\lambda, -\mathcal{Q}_T)\| = \lambda\left\|(\lambda I + \mathcal{Q}_T)^{-1}\right\| \to 0, \ \lambda \to 0.$$

Then $\lambda^{1/2}\left\|(\lambda I + \mathcal{Q}_T)^{-1/2}\right\| \to 0$ as $\lambda \to 0$. For sufficiently small $\lambda_0 > 0$, we can write

$$\lambda_0^{1/2}\left\|(\lambda_0 I + \mathcal{Q}_T)^{-1/2}\right\| \leq \frac{1}{\sqrt{2}}.$$

So, for all $x \in X$, we have

$$\begin{aligned}\|x\|^2 &= \left\|\left(\lambda_0^{1/2}(\lambda_0 I + \mathcal{Q}_T)^{-1/2}\right)\left(\lambda_0^{-1/2}(\lambda_0 I + \mathcal{Q}_T)^{1/2}\right)x\right\|^2 \\ &\leq \frac{1}{2}\left\|\lambda_0^{-1/2}(\lambda_0 I + \mathcal{Q}_T)^{1/2}x\right\|^2 \\ &= \frac{1}{2}\left\langle \lambda_0^{-1}(\lambda_0 I + \mathcal{Q}_T)x, x\right\rangle,\end{aligned}$$

which implies

$$\left\langle \lambda_0^{-1}(\lambda_0 I + \mathcal{Q}_T)x, x\right\rangle \geq 2\|x\|^2$$

and, consequently,

$$\langle \mathcal{Q}_T x, x\rangle \geq \lambda_0\|x\|^2.$$

Thus, $\mathcal{Q}_T > 0$. □

Theorem 12.12. *The following statements are equivalent:*

(a) *the control system* (12.8) *on* V_{ad} *is* D_T^a*-controllable;*

(b) *if* $B^*\mathcal{U}_t^*x = 0$ *for all* $0 \leq t \leq T$, *then* $x = 0$;

(c) $\lambda R(\lambda, -\mathcal{Q}_T)$ *converges strongly to the zero operator as* $\lambda \to 0$;

(d) $\lambda R(\lambda, -\mathcal{Q}_T)$ *converges weakly to the zero operator as* $\lambda \to 0$.

Proof. The equivalence (a) $\Leftrightarrow$ (b) is well-known and it is already stated in Theorem 12.7(b). For the implication (c) $\Rightarrow$ (a), let $\lambda R(\lambda, -\mathcal{Q}_T)$ be strongly convergent to the zero operator as $\lambda \to 0$. Consider an arbitrary $h \in X$ and the functional (12.13) with this h. By Lemma 12.10, there is a control $v^\lambda \in L_2(0,T;U)$ such that (12.15) holds. Hence, selecting λ sufficiently small, we can make $y_T^{v^\lambda}$ to be close to h, proving that the control system (12.8) on V_{ad} is D_T^a-controllable. For the implication (a) $\Rightarrow$ (c), let the control system (12.8) on V_{ad} be D_T^a-controllable. Then for arbitrary $h \in X$, there exists a sequence $\{\bar{v}^n\}$ in $L_2(0,T;U)$ such that $\left\|y_T^{\bar{v}^n} - h\right\| \to 0$ as $n \to \infty$. We have

$$\left\|y_T^{v^\lambda} - h\right\|^2 \le \left\|y_T^{v^\lambda} - h\right\|^2 + \lambda \int_0^T \left\|v_t^\lambda\right\|^2 dt$$
$$\le \left\|y_T^{\bar{v}^n} - h\right\|^2 + \lambda \int_0^T \left\|\bar{v}_t^n\right\|^2 dt,$$

where v^λ is the control at which the functional (12.13) takes on its minimum value. If $\varepsilon > 0$ is given, then we can make

$$\left\|y_T^{\bar{v}^n} - h\right\| < \frac{\varepsilon}{\sqrt{2}}$$

for some sufficiently large n and then we can select $\delta > 0$ to be sufficiently small so that for all $0 < \lambda < \delta$,

$$\lambda \int_0^T \left\|\bar{v}_t^n\right\|^2 dt < \frac{\varepsilon^2}{2}.$$

Thus, $\left\|y_T^{v^\lambda} - h\right\| < \varepsilon$ for all $0 < \lambda < \delta$, i.e., $y_T^{v^\lambda}$ converges to h as $\lambda \to 0$. By (12.15) and by the arbitrariness of h, this implies the strong convergence of $\lambda R(\lambda, -\mathcal{Q}_T)$ to the zero operator as $\lambda \to 0$. Finally, the equivalence (c) $\Leftrightarrow$ (d) is a consequence of $\lambda R(\lambda, -\mathcal{Q}_T) \ge 0$. □

The conditions (f) in Theorem 12.11 and (c) in Theorem 12.12 clearly distinguish the D_T^c-controllability and the D_T^a-controllability of the control system (12.8) showing that the distinction between them is in a kind of convergence of the operator $\lambda R(\lambda, -\mathcal{Q}_T)$ to the zero operator as $\lambda \to 0$. We call these conditions the *resolvent conditions* for the control system (12.8) to be D_T^c- and D_T^a-controllable, respectively.

12.2.3 Applications of Resolvent Conditions

An application of the resolvent conditions to a concrete control system requires a computation of the respective resolvent and then a verification of the respective convergence. These are illustrated below in the examples of controlled one-dimensional heat and wave equations.

Example 12.13. Consider the controlled one-dimensional heat equation from Example 3.5:

$$\frac{\partial}{\partial t} y_{t,\theta} = \frac{\partial^2}{\partial \theta^2} y_{t,\theta} + v_{t,\theta}, \ 0 \le \theta \le 1, \ 0 < t \le T, \tag{12.17}$$

with the initial and boundary conditions

$$\begin{cases} y_{0,\theta} = f_\theta, \ 0 \le \theta \le 1, \\ y_{t,0} = y_{t,1} = 0, \ 0 \le t \le T. \end{cases} \tag{12.18}$$

Let $X = U = L_2(0,1;\mathbb{R})$ and let $f \in X$. In the system (12.17)–(12.18), the second order differential operator $d^2/d\theta^2$ stands for the operator A with the domain

$$D(A) = \{h \in W^{2,2}(0,1;\mathbb{R}) : h_0 = h_1 = 0\}$$

and it generates the strongly continuous semigroup $\mathcal{U}$ defined by

$$[\mathcal{U}_t h]_\theta = \sum_{n=1}^{\infty} 2e^{-n^2\pi^2 t} \sin(n\pi\theta) \int_0^1 h_\alpha \sin(n\pi\alpha)\, d\alpha, \ 0 \le \theta \le 1, \ t \ge 0, \ h \in X.$$

If v is considered as a control action taken from the set of admissible controls $V_{\text{ad}} = L_2(0,T;L_2(0,1;\mathbb{R}))$, then it is easily seen that $B = B^* = I$ and, since $\mathcal{U}_t$ is self-adjoint (see Example 3.5),

$$\mathcal{Q}_T = \int_0^T \mathcal{U}_s B B^* \mathcal{U}_s^*\, ds = \int_0^T \mathcal{U}_{2s}\, ds.$$

Therefore, for $h \in X$,

$$\begin{aligned} [\mathcal{Q}_T h]_\theta &= \left[\int_0^T \mathcal{U}_{2s} h\, ds \right]_\theta \\ &= \sum_{n=1}^{\infty} \int_0^T 2e^{-2n^2\pi^2 s} \sin(n\pi\theta) \int_0^1 h_\alpha \sin(n\pi\alpha)\, d\alpha\, ds \\ &= \sum_{n=1}^{\infty} \frac{1 - e^{-2n^2\pi^2 T}}{n^2\pi^2} \sin(n\pi\theta) \int_0^1 h_\alpha \sin(n\pi\alpha)\, d\alpha. \end{aligned}$$

The half-range Fourier sine expansion of $h \in X$ is

$$h_\theta = \sum_{n=1}^{\infty} 2\sin(n\pi\theta) \int_0^1 h_\alpha \sin(n\pi\alpha)\, d\alpha, \ 0 \le \theta \le 1.$$

Using this, we obtain

$$[(\lambda I + \mathcal{Q}_T)h]_\theta = \sum_{n=1}^{\infty} \frac{2n^2\pi^2\lambda + 1 - e^{-2n^2\pi^2 T}}{n^2\pi^2} \sin(n\pi\theta) \int_0^1 h_\alpha \sin(n\pi\alpha)\, d\alpha.$$

Let $(\lambda I + \mathcal{Q}_T)h = g$. If we use the half-range Fourier sine expansion of $g \in X$, then

$$\sum_{n=1}^{\infty} 2\sin(n\pi\theta)\int_0^1 g_\alpha \sin(n\pi\alpha)\,d\alpha$$
$$= \sum_{n=1}^{\infty} \frac{2n^2\pi^2\lambda + 1 - e^{-2n^2\pi^2 T}}{n^2\pi^2}\sin(n\pi\theta)\int_0^1 h_\alpha \sin(n\pi\alpha)\,d\alpha,$$

which for all $n \in \mathbb{N}$ implies

$$\int_0^1 h_\alpha \sin(n\pi\alpha)\,d\alpha = \frac{2n^2\pi^2}{2n^2\pi^2\lambda + 1 - e^{-2n^2\pi^2 T}}\int_0^1 g_\alpha \sin(n\pi\alpha)\,d\alpha.$$

Therefore,

$$\begin{aligned} h_\theta &= \left[(\lambda I + \mathcal{Q}_T)^{-1} g\right]_\theta = [R(\lambda, -\mathcal{Q}_T)g]_\theta \\ &= \sum_{n=1}^{\infty} \frac{4n^2\pi^2}{2n^2\pi^2\lambda + 1 - e^{-2n^2\pi^2 T}}\sin(n\pi\theta)\int_0^1 g_\alpha \sin(n\pi\alpha)\,d\alpha. \end{aligned}$$

If $g_\alpha \equiv 1$, then by Parseval identity,

$$\begin{aligned} \|R(\lambda, -\mathcal{Q}_T)g\|_X^2 &= \frac{1}{2}\sum_{n=1}^{\infty} \frac{(4n^2\pi^2)^2}{(2n^2\pi^2\lambda + 1 - e^{-2n^2\pi^2 T})^2}\left(\int_0^1 \sin(n\pi\alpha)\,d\alpha\right)^2 \\ &= \sum_{n=1}^{\infty} \frac{8n^2\pi^2(1-(-1)^n)^2}{(2n^2\pi^2\lambda + 1 - e^{-2n^2\pi^2 T})^2} \\ &\geq \sum_{n=1}^{\infty} \frac{8n^2\pi^2(1-(-1)^n)^2}{(2n^2\pi^2\lambda + 1)^2} = \sum_{n=1,3,5,\ldots} \frac{32n^2\pi^2}{(2n^2\pi^2\lambda + 1)^2}. \end{aligned}$$

One can verify that the inequality

$$\frac{n}{2n^2\pi^2\lambda + 1} > \frac{n+1}{2(n+1)^2\pi^2\lambda + 1}$$

holds whenever n is an integer that is greater than the number $1/\sqrt{2\lambda}\pi$. Let N_λ be the smallest odd integer that is greater than $1/\sqrt{2\lambda}\pi$. Then the sequence

$$\left\{n^2\pi^2/(2n^2\pi^2\lambda + 1)^2\right\}_{n=1,2,\ldots}$$

is decreasing for $n \geq N_\lambda$. The following limits are obvious:

$$N_\lambda \to \infty \text{ and } \lambda N_\lambda^2 \to \frac{1}{2\pi^2} \text{ as } \lambda \to 0.$$

Using these, for $g_\alpha \equiv 1$, we obtain

$$\|R(\lambda, -\mathcal{Q}_T)g\|_X^2 \geq \sum_{n=N_\lambda}^{\infty} \frac{16n^2\pi^2}{(2n^2\pi^2\lambda+1)^2} \geq \int_{N_\lambda}^{\infty} \frac{16\pi^2 t^2}{(2\pi^2\lambda t^2+1)^2}\,dt$$
$$\geq \int_{N_\lambda}^{\infty} \frac{4\pi^2 t}{(2\pi^2\lambda t^2+1)^2}\,dt = \frac{1}{\lambda(2\pi^2\lambda N_\lambda^2+1)} \to \infty$$

as $\lambda \to 0$. So, by $(a) \Leftrightarrow (d)$ in Theorem 12.11, there is no $T > 0$ for which the system (12.17)–(12.18) is D_T^c-controllable. At the same time, for all $g \in X$,

$$\|\lambda R(\lambda, -\mathcal{Q}_T)g\|_X^2 = \sum_{n=1}^{\infty} \frac{8n^4\pi^4\lambda^2}{(2n^2\pi^2\lambda+1-e^{-2n^2\pi^2T})^2}\left(\int_0^1 g_\alpha \sin(n\pi\alpha)\,d\alpha\right)^2 \to 0$$

as $\lambda \to 0$ and, hence, by $(a) \Leftrightarrow (c)$ in Theorem 12.12, for each $T > 0$, the system (12.17)–(12.18) is D_T^a-controllable.

Example 12.14. Consider the controlled wave equation from Example 3.6:

$$\frac{\partial^2}{\partial t^2}u_{t,\theta} = \frac{\partial^2}{\partial \theta^2}u_{t,\theta} + b_\theta v_t,\ 0 \leq \theta \leq 1,\ 0 < t \leq T, \tag{12.19}$$

with the initial and boundary conditions

$$\begin{cases} u_{0,\theta} = f_\theta,\ (\partial/\partial t)u_{t,\theta}\big|_{t=0} = g_\theta,\ 0 \leq \theta \leq 1, \\ u_{t,0} = u_{t,1} = 0,\ 0 \leq t \leq T. \end{cases} \tag{12.20}$$

We assume that f, g and b are functions in $L_2(0,1;\mathbb{R})$ and with these functions we associate the respective sequences $\{\check{f}_n\}$, $\{\check{g}_n\}$ and $\{\check{b}_n\}$ of Fourier coefficients in the half-range Fourier sine expansions

$$f_\theta = \sum_{n=1}^{\infty} \check{f}_n\sqrt{2}\sin(n\pi\theta),\ g_\theta = \sum_{n=1}^{\infty} \check{g}_n\sqrt{2}\sin(n\pi\theta),\ b_\theta = \sum_{n=1}^{\infty} \check{b}_n\sqrt{2}\sin(n\pi\theta)$$

and suppose that

$$\sum_{n=1}^{\infty} n^2\check{f}_n^2 < \infty.$$

Let X be the Hilbert space introduced in Example 3.6 and let $U = \mathbb{R}$. Then the set of admissible controls v is $V_{\mathrm{ad}} = L_2(0,T;\mathbb{R})$. For the operator

$$A = \begin{bmatrix} 0 & I \\ d^2/d\theta^2 & 0 \end{bmatrix},$$

where I is the identity operator on $W^{2,2}(0,1;\mathbb{R})$ and $d^2/d\theta^2$ has the domain

$$D(d^2/d\theta^2) = \{\xi \in W^{2,2}(0,1;\mathbb{R}) : \xi_0 = \xi_1 = 0\},$$

and for $B \in \mathcal{L}(\mathbb{R}, X)$ defined by

$$[Bv]_\theta = \begin{bmatrix} 0 \\ b_\theta v \end{bmatrix}, \quad 0 \le \theta \le 1, \; v \in \mathbb{R},$$

the system (12.19)–(12.20) can be formulated in the abstract form

$$\frac{d}{dt} y_t = A y_t + B v_t, \; t > 0, \tag{12.21}$$

where

$$[y_t]_\theta = \begin{bmatrix} u_{t,\theta} \\ (\partial/\partial t) u_{t,\theta} \end{bmatrix}, \quad 0 \le \theta \le 1, \; t > 0; \; y_0 = \begin{bmatrix} f \\ g \end{bmatrix}.$$

By Example 3.6, the operator A generates a strongly continuous semigroup $\mathcal{U}$ as defined by

$$[\mathcal{U}_t h]_\theta = \sum_{n=1}^{\infty} \begin{bmatrix} \cos(n\pi t) & (n\pi)^{-1}\sin(n\pi t) \\ -n\pi \sin(n\pi t) & \cos(n\pi t) \end{bmatrix} \begin{bmatrix} \check{\xi}_n \\ \check{\eta}_n \end{bmatrix} \sqrt{2}\sin(n\pi\theta), \; 0 \le \theta \le 1, \; t \ge 0,$$

where

$$h = \begin{bmatrix} \xi \\ \eta \end{bmatrix} \in X$$

and $\check{\xi}_n$ and $\check{\eta}_n$ are Fourier coefficients of ξ and η, respectively. It was mentioned in Example 3.6 that the natural extension of $\mathcal{U}$ to $\mathbb{R}$ is a group. Therefore, the controllability operator $\mathcal{Q}_T$ of the system (12.21) is

$$\mathcal{Q}_T h = \int_0^T \mathcal{U}_t B B^* \mathcal{U}_t^* h \, dt = \int_0^T \mathcal{U}_t B B^* \mathcal{U}_{-t} h \, dt, \; h \in X.$$

We have

$$[\mathcal{U}_{-t} h]_\theta = \sum_{n=1}^{\infty} \begin{bmatrix} \check{\xi}_n \cos(n\pi t) - \check{\eta}_n (n\pi)^{-1} \sin(n\pi t) \\ \check{\xi}_n n\pi \sin(n\pi t) + \check{\eta}_n \cos(n\pi t) \end{bmatrix} \sqrt{2}\sin(n\pi\theta).$$

One can calculate that

$$B^* h = \sum_{n=1}^{\infty} \check{b}_n \check{\eta}_n, \; h \in X.$$

Hence,

$$B^* \mathcal{U}_{-t} h = \sum_{n=1}^{\infty} \check{b}_n \big(\check{\xi}_n n\pi \sin(n\pi t) + \check{\eta}_n \cos(n\pi t) \big)$$

and, consequently,

$$[\mathcal{U}_t B B^* \mathcal{U}_{-t} h]_\theta = \sum_{n=1}^{\infty} \begin{bmatrix} \check{b}_n (n\pi)^{-1} \sin(n\pi t) \\ \check{b}_n \cos(n\pi t) \end{bmatrix} \sqrt{2}\sin(n\pi\theta)$$
$$\times \sum_{k=1}^{\infty} \check{b}_k \big(\check{\xi}_k k\pi \sin(k\pi t) + \check{\eta}_k \cos(k\pi t) \big).$$

Thus, for $T = 2$,

$$[\mathcal{Q}_2 h]_\theta = \int_0^2 [\mathcal{U}_t BB^*\mathcal{U}_{-t}h]_\theta \, dt = \sum_{n=1}^{\infty} \begin{bmatrix} \breve{b}_n^2 \breve{\xi}_n \\ \breve{b}_n^2 \breve{\eta}_n \end{bmatrix} \sqrt{2}\sin(n\pi\theta).$$

We obtain that

$$[(\lambda I + \mathcal{Q}_2)h]_\theta = \sum_{n=1}^{\infty} (\lambda + \breve{b}_n^2) \begin{bmatrix} \breve{\xi}_n \\ \breve{\eta}_n \end{bmatrix} \sqrt{2}\sin(n\pi\theta),$$

which implies

$$[R(\lambda, -\mathcal{Q}_2)h]_\theta = \left[(\lambda I + \mathcal{Q}_2)^{-1}h\right]_\theta = \sum_{n=1}^{\infty} \frac{1}{\lambda + \breve{b}_n^2} \begin{bmatrix} \breve{\xi}_n \\ \breve{\eta}_n \end{bmatrix} \sqrt{2}\sin(n\pi\theta).$$

Finally, for all $h \in X$,

$$\|\lambda R(\lambda, -\mathcal{Q}_2)h\|^2 = \sum_{n=1}^{\infty} \frac{\lambda^2}{\left(\lambda + \breve{b}_n^2\right)^2} \left(n^2\pi^2\breve{\xi}_n^2 + \breve{\eta}_n^2\right) \to 0$$

as $\lambda \to 0$ if $\breve{b}_n \neq 0$ for all $n \in \mathbb{N}$. Thus, by $(a) \Leftrightarrow (c)$ in Theorem 12.12, we obtain the following sufficient condition for the D_T^a-controllability of the system (12.19)–(12.20) which agrees with Theorem 2.10 (p. 219) in Zabczyk [95]: if $T \geq 2$ and b is so that

$$\forall n \in \mathbb{N}, \quad \int_0^1 b_\theta \sin(n\pi\theta)\, d\theta \neq 0,$$

then the system (12.19)–(12.20) is D_T^a-controllable.

12.3 Controllability: Stochastic Systems

To study the S_T- and C_T-controllability of the control system (12.5), we will use the results about the D_T^c- and D_T^a-controllability of the control system (12.8) from the previous section and the results about the S_T^0-controllability of the control system (12.9) given below.

12.3.1 S_T^0-Controllability

Consider the Riccati equations

$$\frac{d}{dt}Q_t + Q_t A + A^* Q_t - \lambda^{-1} Q_t BB^* Q_t = 0, \qquad 0 \leq t < T, \; Q_T = I, \; \lambda > 0, \tag{12.22}$$

$$\frac{d}{dt}P_t - AP_t - P_t A^* - \Phi + (P_t C^* + R)(CP_t + R^*) = 0, \qquad 0 < t \leq T, \; P_0 = \operatorname{cov} z_0. \tag{12.23}$$

Recall that by Theorems 3.27 and 3.28, there exist unique solutions (in the scalar product sense) Q^λ and P of these equations, respectively, and $Q_t^\lambda \geq 0$ and $P_t \geq 0$ for all $0 \leq t \leq T$. Moreover, by Theorem 3.29, the solution Q^λ of equation (12.22) has the explicit form

$$Q_t^\lambda = \lambda \mathcal{U}_{T-t}^* R(\lambda, -\mathcal{Q}_{T-t}) \mathcal{U}_{T-t}, \ 0 \leq t \leq T, \ \lambda > 0. \tag{12.24}$$

Lemma 12.15. *There exists the finite limit*

$$a_T = \lim_{\lambda \to 0} \int_0^T \operatorname{tr}\big((CP_s + R^*)Q_s^\lambda (P_s C^* + R)\big)\, ds, \tag{12.25}$$

where Q^λ and P are the solutions of the equations (12.22) *and* (12.23).

Proof. Consider the family of the stochastic optimal control problems on W_{ad} with the state-observation system (12.9) and the functional

$$J^\lambda(w) = \mathbf{E}\left(\|z_T^w\|^2 + \lambda \int_0^T \|w_t\|^2\, dt \right), \ \lambda > 0, \tag{12.26}$$

to be minimized. By Theorem 6.20, the functional J^λ takes its minimum value at some control $w^\lambda \in W_{\text{ad}}$ and, by Proposition 6.22,

$$J^\lambda(w^\lambda) = \operatorname{tr} P_T + \int_0^T \operatorname{tr}\big((CP_s + R^*)Q_s^\lambda (P_s C^* + R)\big)\, ds.$$

Therefore, to prove the lemma it suffices to show that the sequence $\{J^\lambda(w^\lambda)\}$ has a finite limit. Let $0 < \nu < \lambda$. Then

$$\begin{aligned} J^\nu(w^\nu) &= \mathbf{E}\left(\left\|z_T^{w^\nu}\right\|^2 + \nu \int_0^T \left\|w_t^\nu\right\|^2 dt \right) \\ &\leq \mathbf{E}\left(\left\|z_T^{w^\lambda}\right\|^2 + \nu \int_0^T \left\|w_t^\lambda\right\|^2 dt \right) \\ &\leq \mathbf{E}\left(\left\|z_T^{w^\lambda}\right\|^2 + \lambda \int_0^T \left\|w_t^\lambda\right\|^2 dt \right) = J^\lambda(w^\lambda). \end{aligned}$$

We conclude that $J^\lambda(w^\lambda)$ is a nonnegative and nondecreasing function of $\lambda > 0$. Hence, there exists a finite limit of $J^\lambda(w^\lambda)$ as $\lambda \to 0$ proving the lemma. □

Lemma 12.16. *Let a_T be defined by* (12.25) *and let Q^λ and P be the solutions of the equations* (12.22) *and* (12.23), *respectively. Then*

$$\inf_{W_{\text{ad}}} \mathbf{E}\|\mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})\|^2 = a_T. \tag{12.27}$$

Proof. We will compare the functional (12.26) and

$$\tilde{J}^\lambda(w) = \mathbf{E}\Big(\|\mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})\|^2 + \lambda \int_0^T \|w_t\|^2 \, dt \Big),$$

where $w \in W_{\text{ad}}$ and z^w is the state of the system (12.9). By Proposition 6.19, P_T is the covariance of the error $z_T^w - \mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})$ independently of $w \in W_{\text{ad}}$. Hence,

$$\text{tr} P_T = \mathbf{E}\|z_T^w - \mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})\|^2 = \mathbf{E}\|z_T^w\|^2 - \mathbf{E}\|\mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})\|^2,$$

and, consequently,

$$\tilde{J}^\lambda(w^\lambda) = J^\lambda(w^\lambda) - \text{tr} P_T = \int_0^T \text{tr}\big((CP_s + R^*)Q_s^\lambda(P_s C^* + R)\big) \, ds.$$

If we denote by $\{\tilde{w}^n\}$ any minimizing sequence of the functional

$$J_0(w) = \mathbf{E}\|\mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})\|^2,$$

then

$$\inf_{W_{\text{ad}}} \mathbf{E}\|\mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})\|^2 \le \tilde{J}^\lambda(w^\lambda) \le \mathbf{E}\Big(\Big\|\mathbf{E}\Big(z_T^{\tilde{w}^n} \Big| \mathcal{F}_T^{\tilde{w}^n,\eta}\Big)\Big\|^2 + \lambda \int_0^T \|\tilde{w}_t^n\|^2 \, dt \Big).$$

Consequently, taking the limit as $\lambda \to 0$ and $n \to \infty$, we obtain the statement of the lemma. □

Theorem 12.17. *Given $\varepsilon > 0$ and $0 \le p < 1$, the control system* (12.9) *on W_{ad} is $S_{T,\varepsilon,p}^0$-controllable if*

$$a_T < \varepsilon(1 - p), \tag{12.28}$$

where a_T is defined by (12.25).

Proof. By Lemma 12.16, we have

$$\inf_{W_{\text{ad}}} \mathbf{E}\|\mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta})\|^2 = a_T < \varepsilon(1 - p).$$

Therefore, there exists $w^0 \in W_{\text{ad}}$ such that

$$\mathbf{E}\Big\|\mathbf{E}\Big(z_T^{w^0} \Big| \mathcal{F}_T^{w^0,\eta}\Big)\Big\|^2 < \varepsilon(1 - p).$$

Using Chebyshev's inequality, we obtain

$$\mathbf{P}\Big(\Big\|\mathbf{E}\Big(z_T^{w^0} \Big| \mathcal{F}_T^{w^0,\eta}\Big)\Big\|^2 > \varepsilon \Big) \le \frac{1}{\varepsilon} \mathbf{E}\Big\|\mathbf{E}\Big(z_T^{w^0} \Big| \mathcal{F}_T^{w^0,\eta}\Big)\Big\|^2 < (1 - p).$$

Thus, the control system (12.9) on W_{ad} is $S_{T,\varepsilon,p}^0$-controllable. □

It should be noted that the condition (12.28) being a sufficient condition for $S^0_{T,\varepsilon,p}$-controllability is not necessary in general. In view of this we present the following arguments. For a given control system, define the functions

$$\alpha_p = \inf \Upsilon_p,\ \Upsilon_p = \{\varepsilon : \text{the system is } S^0_{T,\varepsilon,p}\text{-controllable}\}, \tag{12.29}$$

$$\beta_\varepsilon = \sup \Pi_\varepsilon,\ \Pi_\varepsilon = \{p : \text{the system is } S^0_{T,\varepsilon,p}\text{-controllable}\}. \tag{12.30}$$

Obviously, α and β are nondecreasing functions with $\alpha_0 = 0$ and $\lim_{\varepsilon\to\infty} \beta_\varepsilon = 1$. It follows from the definitions that the necessary and sufficient condition for the system to be $S^0_{T,\varepsilon,p}$-controllable is

$$\begin{cases} \alpha_p < \varepsilon & \text{if } \inf \Upsilon_p \text{ is not achieved,} \\ \alpha_p \le \varepsilon & \text{if } \inf \Upsilon_p \text{ is achieved,} \end{cases} \tag{12.31}$$

which can also be written in the following equivalent form:

$$\begin{cases} \beta_\varepsilon > p & \text{if } \sup \Pi_\varepsilon \text{ is not achieved,} \\ \beta_\varepsilon \ge p & \text{if } \sup \Pi_\varepsilon \text{ is achieved.} \end{cases} \tag{12.32}$$

Using (12.28), define the functions

$$\tilde{\alpha}_p = \begin{cases} a_T(1-p)^{-1}, & 0 \le p < 1, \\ \infty, & p = 1, \end{cases} \qquad \tilde{\beta}_\varepsilon = \begin{cases} 1 - a_T\varepsilon^{-1}, & a_T < \varepsilon < \infty, \\ 0, & 0 \le \varepsilon \le a_T. \end{cases}$$

By (12.29), (12.30) and Theorem 12.17, it follows that

$$\alpha_p \le \tilde{\alpha}_p,\ 0 \le p \le 1, \quad \text{and} \quad \beta_\varepsilon \ge \tilde{\beta}_\varepsilon,\ 0 \le \varepsilon < \infty,$$

i.e., in the case of the control system (12.9) the functions $\tilde{\alpha}$ and $\tilde{\beta}$, defined with the help of (12.28), give only approximations of the functions α and β and may not be equal to them. In case $\alpha_p < \tilde{\alpha}_p$ or $\beta_\varepsilon > \tilde{\beta}_\varepsilon$ the inequality in (12.28) cannot be a necessary condition for $S^0_{T,\varepsilon,p}$-controllability of the system (12.9).

Theorem 12.18. *The control system* (12.9) *on* W_{ad} *is* S^0_T*-controllable if* $a_T = 0$.

Proof. By Theorem 12.17, $a_T = 0$ implies that the control system (12.9) is $S^0_{T,\varepsilon,p}$-controllable for all ε and for all p satisfying $\varepsilon(1-p) > 0$. This condition includes all pairs (ε, p) with $\varepsilon > 0$ and $0 \le p < 1$. So, the system (12.9) is $S^0_{T,\varepsilon,p}$-controllable for all $\varepsilon > 0$ and for all $0 \le p < 1$. □

Theorem 12.19. *The control system* (12.9) *on* W_{ad} *is* S^0_T*-controllable if the system* (12.8) *is* D^a_t*-controllable for each* $0 < t \le T$.

Proof. From (a) ⇒ (c) in Theorem 12.12, we obtain that $\lambda R(\lambda, -\mathcal{Q}_{T-t})$ strongly converges to the zero operator as $\lambda \to 0$ for all $0 \le t < T$. Hence, by (12.24), Q^λ_t strongly converges to the zero operator as $\lambda \to 0$ for all $0 \le t < T$, where Q^λ is the solution of the equation (12.22). Furthermore, substituting

$$h = \lambda^{1/2}(\lambda I + \mathcal{Q}_{T-t})^{-1/2}x$$

in

$$\langle \lambda^{-1}(\lambda I + \mathcal{Q}_{T-t})h, h\rangle \geq \langle h, h\rangle,$$

we obtain

$$\langle \lambda(\lambda I + \mathcal{Q}_{T-t})^{-1}x, x\rangle \leq \|x\|^2.$$

So,

$$\lambda R(\lambda, -\mathcal{Q}_{T-t}) \leq I$$

and, by (12.24),

$$Q_t^\lambda \leq \mathcal{U}_{T-t}^* \mathcal{U}_{T-t}$$

for all $\lambda > 0$ and for all $0 \leq t \leq T$. Hence, we can change the places of the limit, the integral and the trace in (12.25) to obtain $a_T = 0$. Thus by Theorem 12.18, we obtain the S_T^0-controllability of the control system (12.9). □

Theorem 12.20. *The control system* (12.9) *on* W_{ad} *is* S_T^0*-controllable for each* $T > 0$ *if the control system* (12.8) *on* V_{ad} *is* D_T^a*-controllable for each* $T > 0$.

Proof. This is a direct consequence from Theorem 12.19. □

12.3.2 C_T-Controllability

In this section the C_T-controllability of the control system (12.5) on U_{ad} will be studied. We will use the results about the D_T^c-controllability of the control system (12.8) on V_{ad} from Section 12.2.2 and about the S_T^0-controllability of the control system (12.9) on W_{ad} from Section 12.3.1.

Lemma 12.21. $U_{\text{ad}} = V_{\text{ad}} + W_{\text{ad}}$, *where* $+$ *is the sign of the sum of sets.*

Proof. Let $u \in U_{\text{ad}}$ be of the form (12.5) with $K \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$ and $\bar{u} \in L_2(0,T;U)$. Then $\mathbf{E}u = \bar{u} \in V_{\text{ad}}$ and, if $w = u - \bar{u}$, then

$$\begin{aligned} w_t &= \int_0^t K_{t,s} C(x_s^u - \mathbf{E}x_s^u)\, ds + \int_0^t K_{t,s}\, d\varphi_s \\ &= \int_0^t K_{t,s} C z_s^w\, ds + \int_0^t K_{t,s}\, d\varphi_s = \int_0^t K_{t,s}\, d\eta_s^w. \end{aligned}$$

Thus, $w = u - \bar{u} \in W_{\text{ad}}$ and, consequently, $u \in V_{\text{ad}} + W_{\text{ad}}$. On the other hand, if $v \in V_{\text{ad}}$ and $w \in W_{\text{ad}}$ where w has the form of (12.10) with $K \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$, then

$$\begin{aligned} u_t &= v_t + \int_0^t K_{t,s} C z_s^w\, ds + \int_0^t K_{t,s}\, d\varphi_s \\ &= v_t - \int_0^t K_{t,s} C y_s^v\, ds + \int_0^t K_{t,s} C x_s^u\, ds + \int_0^t K_{t,s}\, d\varphi_s. \end{aligned}$$

Denote

$$\bar{u}_t = v_t - \int_0^t K_{t,s} C y_s^v \, ds. \tag{12.33}$$

Then u has the form of (12.6) with $\bar{u}$ as in (12.33), i.e., $u \in U_{\mathrm{ad}}$. Thus, $U_{\mathrm{ad}} = V_{\mathrm{ad}} + W_{\mathrm{ad}}$. □

Lemma 12.22. *If $u = v + w$ where $v \in V_{\mathrm{ad}}$ and $w \in W_{\mathrm{ad}}$, then the σ-algebras $\mathcal{F}_T^{u,\xi}$ and $\mathcal{F}_T^{w,\eta}$, generated by ξ_s^u, $0 \le s \le T$, and η_s^w, $0 \le s \le T$, respectively, are equal.*

Proof. It is easy to show that

$$\xi_t^u = \eta_t^w + C \int_0^t y_s^v \, ds, \quad 0 \le t \le T. \tag{12.34}$$

Since the second term in the right-hand side of (12.34) is nonrandom, we conclude that $\mathcal{F}_T^{u,\xi}$ and $\mathcal{F}_T^{w,\eta}$ are equal. □

Lemma 12.23. *Given $\varepsilon > 0$ and $0 \le p < 1$, the control system* (12.5) *on U_{ad} is $C_{T,\varepsilon,p}^c$-controllable if and only if the control system* (12.8) *on V_{ad} is D_T^c-controllable and the control system* (12.9) *on W_{ad} is $S_{T,\varepsilon,p}^0$-controllable.*

Proof. Let $C(T, \varepsilon, p)$ be the set (12.3) corresponding to the control system (12.5). Similarly, let $D(T)$ be the set (12.1) corresponding to the control system (12.8). Assume that the control system (12.5) is $C_{T,\varepsilon,p}^c$-controllable. Then from the inclusion $C(T, \varepsilon, p) \subset D(T)$, it follows that the control system (12.8) is D_T^c-controllable. Let $h \in C(T, \varepsilon, p)$. Then there exists $u \in U_{\mathrm{ad}}$ such that $h = \mathbf{E} x_T^u$ and

$$\mathbf{P}\left\{ \left\| \mathbf{E}\left(x_T^u | \mathcal{F}_T^{u,\xi}\right) - h \right\|^2 > \varepsilon \right\} \le 1 - p.$$

Consider $w = u - \mathbf{E}u \in W_{\mathrm{ad}}$. By Lemma 12.22, $\mathcal{F}_T^{u,\xi} = \mathcal{F}_T^{w,\eta}$. Therefore,

$$\mathbf{P}\left\{ \| \mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta}) \|^2 > \varepsilon \right\} = \mathbf{P}\left\{ \left\| \mathbf{E}\left(x_T^u | \mathcal{F}_T^{u,\xi}\right) - \mathbf{E} x_T^u \right\|^2 > \varepsilon \right\} \le 1 - p,$$

i.e., the control system (12.9) is $S_{T,\varepsilon,p}^0$-controllable. So, the necessity is proved. To prove the sufficiency, let $h \in D(T)$. Then there exists $v \in V_{\mathrm{ad}}$ such that $h = y_T^v$. Also, from the $S_{T,\varepsilon,p}^0$-controllability of the control system (12.9), we conclude that there exists $w \in W_{\mathrm{ad}}$ with

$$\mathbf{P}\left\{ \| \mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta}) \|^2 > \varepsilon \right\} \le 1 - p.$$

Consider $u = v + w$. By Lemma 12.21, $u \in U_{\mathrm{ad}} = V_{\mathrm{ad}} + W_{\mathrm{ad}}$. Moreover,

$$\mathbf{P}\left\{ \left\| \mathbf{E}\left(x_T^u | \mathcal{F}_T^{u,\xi}\right) - h \right\|^2 > \varepsilon \right\} = \mathbf{P}\left\{ \| \mathbf{E}(z_T^w | \mathcal{F}_T^{w,\eta}) \|^2 > \varepsilon \right\} \le 1 - p,$$

i.e., $h \in C(T, \varepsilon, p)$. Therefore, $D(T) \subset C(T, \varepsilon, p)$. Since $D(T) = X$, we obtain $C(T, \varepsilon, p) = X$. Thus, the control system (12.5) is $C_{T,\varepsilon,p}^c$-controllable. □

Theorem 12.24. *The control system* (12.5) *on* $U_{\rm ad}$ *is* C_T*-controllable if and only if the control system* (12.8) *on* $V_{\rm ad}$ *is* D_T^c*-controllable and the control system* (12.9) *on* $W_{\rm ad}$ *is* S_T^0*-controllable.*

Proof. This is a direct consequence from Lemma 12.23. □

Theorem 12.25. *The control system* (12.5) *on* $U_{\rm ad}$ *is* C_T*-controllable for each* $T > 0$ *if and only if the control system* (12.8) *on* $V_{\rm ad}$ *is* D_T^c*-controllable for each* $T > 0$.

Proof. The necessity follows from Theorem 12.24. For sufficiency, note that by Theorem 12.20, the D_T^c-controllability of the control system (12.8) for each $T > 0$ implies the S_T^0-controllability of the control system (12.9) for each $T > 0$. Thus, by Theorem 12.24, the control system (12.5) is C_T-controllable for each $T > 0$. □

Example 12.26. Consider the control system (12.5) with $X = \mathbb{R}^n$ and $U = \mathbb{R}^m$. Then the operators A and B are $(n \times n)$- and $(n \times m)$-matrices. By Theorem 12.9, the deterministic part of this system is D_T^c-controllable for each $T > 0$ if the rank condition holds, i.e., if the rank of the matrix (12.12) is n. Hence, by Theorem 12.25, this rank condition implies the C_T-controllability of this system for each $T > 0$.

Example 12.27. Consider the control system (12.5) with the operators A and B as defined in Example 12.13. It was shown in Example 12.13 that there is no $T > 0$ such that the deterministic part of this system is D_T^c-controllable. Thus, by Theorem 12.24, we conclude that there is no $T > 0$ such that this system is C_T-controllable.

12.3.3 S_T-Controllability

In this section the S_T-controllability of the control system (12.5) on $U_{\rm ad}$ will be studied. At first we present the results about the C_T^a-controllability which are similar to those of the C_T-controllability from Section 12.3.2.

Lemma 12.28. *Given* $\varepsilon > 0$ *and* $0 \le p < 1$, *the control system* (12.5) *on* $U_{\rm ad}$ *is* $C_{T,\varepsilon,p}^a$*-controllable if and only if the control system* (12.8) *on* $V_{\rm ad}$ *is* D_T^a*-controllable and the control system* (12.9) *on* $W_{\rm ad}$ *is* $S_{T,\varepsilon,p}^0$*-controllable.*

Proof. This can be proved in a similar way as Lemma 12.23. □

Theorem 12.29. *The control system* (12.5) *on* $U_{\rm ad}$ *is* C_T^a*-controllable if and only if the control system* (12.8) *on* $V_{\rm ad}$ *is* D_T^a*-controllable and the control system* (12.9) *on* $W_{\rm ad}$ *is* S_T^0*-controllable.*

Proof. This is a direct consequence from Lemma 12.28. □

It turns out that Theorem 12.29 is true if the C_T^a-controllability in it is replaced by the S_T-controllability. To prove this result, we will use the following fact.

Lemma 12.30. *U_{ad} is a convex set.*

Proof. By Lemma 5.25, if the control $w \in W_{\text{ad}}$ is of the form (12.10) with $K \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$, then there exists $M \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$ such that

$$w_t = \int_0^t M_{t,s}\, d\eta_s^0, \;\; 0 \le t \le T,$$

and vice versa, where η^0 is the observation process of the system (12.9) corresponding to the zero control. Therefore, if $u^1, u^2 \in U_{\text{ad}}$, then by Lemma 12.21,

$$u_t^i = v_t^i + \int_0^t M_{t,s}^i\, d\eta_s^0, \;\; 0 \le t \le T, \;\; i = 1, 2,$$

for some $v^1, v^2 \in L_2(0, T; U)$ and $M^1, M^2 \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$. Let $\alpha_1 > 0$ and $\alpha_2 > 0$ be such that $\alpha_1 + \alpha_2 = 1$. Then for $v = \alpha_1 v^1 + \alpha_2 v^2$ and for $M = \alpha_1 M^1 + \alpha_2 M^2$, we have

$$u_t = \alpha_1 u_t^1 + \alpha_2 u_t^2 = v_t + \int_0^t M_{t,s}\, d\eta_s^0, \;\; 0 \le t \le T,$$

with $v \in L_2(0, T; U)$ and with $M \in B_2(\Delta_T, \mathcal{L}(\mathbb{R}^n, U))$. Thus, $u \in U_{\text{ad}}$. □

Theorem 12.31. *The control system* (12.5) *on U_{ad} is S_T-controllable if and only if the control system* (12.8) *on V_{ad} is D_T^a-controllable and the control system* (12.9) *on W_{ad} is S_T^0-controllable.*

Proof. If the control system (12.8) is D_T^a-controllable and the control system (12.9) is S_T^0-controllable, then by Theorem 12.29, the control system (12.5) is C_T^a-controllable which implies its S_T-controllability since $C(T, \varepsilon, p) \subset S(T, \varepsilon, p)$. Sufficiency is proved. For the necessity, let the control system (12.5) be S_T-controllable. Take an arbitrary $h \in X$ and consider the sequences $\{\varepsilon_n\}$ and $\{p_n\}$ with

$$\varepsilon_n > 0, \;\; 0 \le p_n < 1 \text{ and } \varepsilon_n \to 0, \;\; p_n \to 1 \text{ as } n \to \infty.$$

From $S_{T,\varepsilon_n,p_n}^c$-controllability of the system (12.5), we obtain the existence of the sequence $\{u^n\}$ in U_{ad} such that

$$\mathbf{P}\left\{\left\|\mathbf{E}\left(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi}\right) - h\right\|^2 > \varepsilon_n\right\} \le 1 - p_n.$$

The obtained inequality implies the convergence in probability of the sequence of random variables $\mathbf{E}\left(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi}\right)$ to h as $n \to \infty$. Indeed, for $\varepsilon > 0$, we can find a number N such that $0 < \varepsilon_n < \varepsilon^2$ for all $n > N$. Therefore, for $n > N$,

$$\begin{aligned} \mathbf{P}\left\{\left\|\mathbf{E}\left(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi}\right) - h\right\| > \varepsilon\right\} &\le \mathbf{P}\left\{\left\|\mathbf{E}\left(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi}\right) - h\right\|^2 > \varepsilon_n\right\} \\ &\le 1 - p_n \to 0, \quad n \to \infty. \end{aligned}$$

Hence, $\mathbf{E}(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi})$ converges to h in probability. Since for all n, $\mathbf{E}(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi})$ is a Gaussian random variable, by Theorem 4.7, its characteristic function has the form

$$\chi_n(x) = \exp\left(i\langle m_n, x\rangle - \frac{1}{2}\langle \Lambda_n x, x\rangle \right), \quad x \in X,$$

where

$$m_n = \mathbf{E}\big(\mathbf{E}(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi})\big) = \mathbf{E}x_T^{u^n}$$

and

$$\Lambda_n = \mathrm{cov}\mathbf{E}(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi})$$

and i is the imaginary unit. Also, the vector $h \in X$ is considered as a degenerate Gaussian random variable with the characteristic function

$$\chi(x) = \exp\left(i\langle h, x\rangle\right), \quad x \in X.$$

By Theorem 4.1. the convergence of $\mathbf{E}(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi})$ to h in probability implies

$$\chi_n(x) \to \chi(x) \text{ for all } x \in X.$$

The last convergence is possible when for all $x \in X$,

$$\langle m_n, x\rangle = \langle \mathbf{E}x_T^{u^n}, x\rangle \to \langle h, x\rangle \text{ and } \langle \Lambda_n x, x\rangle \to 0 \text{ as } n \to \infty. \tag{12.35}$$

The first convergence in (12.35) means the convergence of $\mathbf{E}x_T^{u^n}$ to h in the weak topology of the Hilbert space X. By Theorem 1.29(a), we can construct the sequence

$$h_n = \sum_{k=1}^{n} c_k^n \mathbf{E}x_T^{u^k}, \quad c_k^n \ge 0, \quad \sum_{k=1}^{n} c_k^n = 1, \quad k = 1, 2, \dots, n, \quad n = 1, 2, \dots,$$

of convex combinations of $\mathbf{E}x_T^{u^n}$ such that h_n converges to h in the strong topology of X. Denote

$$\tilde{u}^n = \sum_{k=1}^{n} c_k^n u^k, \ n = 1, 2, \dots .$$

By Lemma 12.30, $\tilde{u}^n \in U_{\mathrm{ad}}$ for all n. Moreover, in view of the affineness of the system (12.5), $h_n = \mathbf{E}x_T^{\tilde{u}^n}$. In terms of the system (12.8) this means that for the sequence of controls $\tilde{v}^n = \mathbf{E}\tilde{u}^n$ in V_{ad}, the sequence of vectors $h_n = \mathbf{E}x_T^{\tilde{u}^n} = y_T^{\tilde{v}^n}$ converges to h in the strong topology of X. Since h is an arbitrary point of X, we conclude that the set $D(T)$ defined by (12.1) for the control system (12.8) is dense in X, i.e., the control system (12.8) is D_T^a-controllable. Now consider the second convergence in (12.35). Let $\{e_k\}$ be a basis in X. We can select a subsequence $\{n_m^1\}$ of $\{n\}$ so that the sequence $\{\langle \Lambda_{n_m^1} e_1, e_1\rangle\}$ decreases and goes to 0. Then we can select a subsequence $\{n_m^2\}$ of $\{n_m^1\}$ so that the sequence $\{\langle \Lambda_{n_m^2} e_2, e_2\rangle\}$ decreases and goes to 0. Continuing this procedure for all e_k and taking the diagonal sequence

$\{n_m^m\}$, we obtain that for all e_k, the sequence $\{\langle \Lambda_{n_m^m} e_k, e_k \rangle\}$ decreases and goes to 0. Thus, in

$$\lim_{m\to\infty} \mathrm{tr}\Lambda_{n_m^m} = \lim_{m\to\infty} \sum_{k=1}^{\dim X} \langle \Lambda_{n_m^m} e_k, e_k \rangle \tag{12.36}$$

the series is such that for all m and for all k,

$$\langle \Lambda_{n_m^m} e_k, e_k \rangle \le \langle \Lambda_{n_1^1} e_k, e_k \rangle.$$

So, we can interchange the places of the limit and the sum in (12.36) and obtain that

$$\lim_{m\to\infty} \mathrm{tr}\Lambda_{n_m^m} = 0.$$

Hence, without loss of generality, assume that $\lim_{n\to\infty} \mathrm{tr}\Lambda_n = 0$. By Lemma 12.21 and Lemma 12.22, if $w^n = u^n - \mathbf{E}u^n$, then $w^n \in W_{\mathrm{ad}}$ and

$$\Lambda_n = \mathrm{cov}\mathbf{E}\big(x_T^{u^n} | \mathcal{F}_T^{u^n,\xi}\big) = \mathrm{cov}\mathbf{E}\big(z_T^{w^n} | \mathcal{F}_T^{w^n,\eta}\big).$$

Therefore,

$$\lim_{n\to\infty} \mathbf{E}\big\|\mathbf{E}\big(z_T^{w^n} | \mathcal{F}_T^{w^n,\eta}\big)\big\|^2 = \lim_{n\to\infty} \mathrm{tr}\Lambda_n = 0.$$

By Lemma 12.16, this implies $a_T = 0$. Finally from Theorem 12.18, we obtain that the control system (12.9) is S_T^0-controllable. □

Proposition 12.32. *The control system* (12.5) *on* U_{ad} *is* S_T*-controllable if and only if it is* C_T^a*-controllable.*

Proof. This follows from Theorems 12.29 and 12.31. □

Theorem 12.33. *The control system* (12.5) *on* U_{ad} *is* S_T*-controllable for each* $T > 0$ *if and only if the control system* (12.8) *on* V_{ad} *is* D_T^a*-controllable for each* $T > 0$.

Proof. The necessity follows from Theorem 12.31. For sufficiency, note that by Theorem 12.20, the D_T^a-controllability of the control system (12.8) for each $T > 0$ implies the S_T^0-controllability of the control system (12.9) for each $T > 0$. Thus, by Theorem 12.31, the control system (12.5) is S_T-controllable for each $T > 0$. □

Example 12.34. Consider the control system (12.5) with $A = 0$ and with the operator B as defined in Example 12.8. As it was shown in Example 12.8, the deterministic part of this system is D_T^a-controllable for each $T > 0$. Hence, by Theorem 12.33, this system is S_T-controllable for each $T > 0$.

Example 12.35. Consider the control system (12.5) with the operators A and B as defined in Example 12.13. It was shown in Example 12.13 that the deterministic part of this system is D_T^a-controllable for each $T > 0$. Hence, by Theorem 12.33, this system is S_T-controllable for each $T > 0$.

Example 12.36. Consider the control system (12.5) with the operators A and B as defined in Example 12.14. It was shown in Example 12.14 that the deterministic part of this system is D_T^a-controllable for each $T \geq 2$ if some additional condition holds. However, Theorem 12.33 does not guarantee the S_T-controllability of this system for any $T > 0$.

Comments

Chapter 1. The reader can use any textbook on functional analysis to study this chapter in more detail.

Chapter 2. The recommended books are Hille and Phillips [54], Dunford and Schwartz [45], Warga [89], Yosida [94], Kato [63], Balakrishnan [4], etc.

Chapter 3. Theory of semigroups is presented in a number of books including Balakrishnan [4], Bensoussan *et al.* [31, 32], Curtain and Zwart [41] etc. The concept of mild evolution operator was introduced by Curtain and Pritchard [39].

Chapter 4. Section 4.1. Random variables and processes in Hilbert spaces are considered in Curtain and Pritchard [40], Metivier [77] and Rozovskii [84]. Recommended book on Gaussian systems is Shiryaev [86]. **Section 4.2.** Physical Brownian motion was first observed in 1827 by the botanist Brown [34]. Many distinguished scientists such as Einstein, Smolukhovskii and Bachelier studied this phenomenon. Wiener [90] initiated the approach of looking at a physical Brownian motion as a path of a specific random process. We recommend the books of Davis [43], Hida [53], Krylov [67] and Gihman and Skorohod [50] to study this section in greater detail. **Section 4.3.** An early stochastic integral was considered in [83] for nonrandom functions and it is called a Wiener integral. For nonanticipative random functions it was defined by Ito [56] and it is called an Ito integral. Afterwards the Ito integral was an object of intensive study and generalizations. There are a number of sources on stochastic integration, for example, Liptser and Shiryaev [70, 72], Gihman and Skorohod [50, 51], Kallianpur [59], Elliot [46] etc. We follow Metivier [77] with some supplements from Rozovskii [84]. The set $\tilde{\Lambda}(0,T;X,Z)$ is introduced in [77]. To make it a Hilbert space, we consider its quotient set. The next step in developing stochastic integration is a Skorohod integral [87] defined for anticipative random functions that is intensively used to construct a general theory of stochastic calculus. **Section 4.4.** An intensive study of stochastic differential equations became possible since the Ito integral was discovered [57]. There are a number of sources on stochastic differential equations, for example, Liptser and Shiryaev [70, 72], Gihman and Skorohod [50, 51], Ikeda and Watanabe [58] etc. In infinite dimensional spaces this subject was studied in Rozovskii [84], Da Prato and Zabczyk [42] for nonlinear and in Curtain and Pritchard [40] for linear cases.

Section 4.5. We recommend the books Curtain and Pritchard [40] and Shiryaev [86]. **Section 4.6.** White noise is a stationary random process having constant spectral density and it does not exist in an ordinary sense. Therefore, theory of white noise is rather advanced and uses generalized random processes (see Hida [53]). Mathematically, a white noise driven system is simply an Ito stochastic differential equation. Colored noise is introduced in Bucy and Joseph [35]. The importance of wide band noise driven systems is mentioned in Fleming and Rishel [48]. There are different approaches to handle wide band noises. In Kushner [69] an approach based on approximations is developed. We follow another approach that is based on a certain integral representation. This was first suggested in Bashirov [9]. The theorem about existence of infinitely many different integral representations under given autocovariance function is proved in Bashirov and Uğural [26, 27].

Chapter 5. Section 5.1. We use the ideas from Bensoussan and Viot [33] and Curtain and Ichikawa [38] to set a linear quadratic optimal control problem for partially observable systems. **Section 5.2.** There are two basic approaches to optimal control problems. One of them concerns necessary conditions of optimality and it is called Pontryagin's maximum principle (see Pontryagin *et al.* [82]). The first result concerning the maximum principle for stochastic systems was obtained by Kushner [68] in the case of noncontrolled diffusion. Afterwards this result was extended to different stochastic systems, even for those with controlled diffusion, but it was mentioned in Arkin and Saksonov [3] that the stochastic maximum principle of Pontryagin's form is not true for general controlled diffusion. The general stochastic maximum principle, that fundamentally differs from Pontryagin's for deterministic systems covering both controlled drift and controlled diffusion, was obtained independently by Mahmudov [73] and Peng [81]. Later, Cadenillas and Karatzas [36] extended this result to systems with random coefficients, Elliot and Kohlmann [47] studied the subject employing stochastic flows and Mahmudov and Bashirov [76] proved this result for constrained stochastic control problems. The discussion of the general stochastic maximum principle is given in Yong and Zhou [93]. The other approach to optimal control problems, giving sufficient conditions of optimality, is Bellman's dynamic programming (see Bellman [28]). For stochastic systems this approach is discussed in Krylov [66], Fleming and Rishel [48], Fleming and Soner [49] etc. The recent book of Yong and Zhou [93] establishes relations between stochastic maximum principle and dynamic programming and many other issues concerning stochastic optimal control. For linear quadratic optimal control problems under partial observations both these approaches, maximum principle and dynamic programming, lead to the same result called the separation principle. In the continuous time case the separation principle was stated and studied in Wonham [92]. This result in Hilbert spaces was considered in a number of works, for example, Bensoussan and Viot [33], Curtain and Ichikawa [38] etc. The extended form of the separation principle was first mentioned in Bashirov [5] and afterwards it was studied in Bashirov [7, 14] which we follow in this section. **Section 5.3.** This section is written on the basis of Bashirov [8, 14]. **Section 5.4.**

The idea of minimizing sequence considered in this section belongs to Bensoussan and Viot [33]. **Section 5.5.** In infinite dimensional spaces the linear regulator problem is studied in a number of works, for example, Curtain and Pritchard [39]. **Section 5.6.** Generally, the results on existence of optimal control are based on weak convergence and weak compactness. In the linear quadratic case it is possible to reduce the existence of optimal control to a certain linear filtering problem when the observations are incomplete.

Chapter 6. The first estimation problems were independently studied by Kolmogorov [64] and Wiener [91] on the basis of the spectral expansion of stationary random processes. A significant stage in developing of estimation theory was the famous works by Kalman [61] and Kalman and Bucy [62]. For complete discussion of estimation problems, see Liptser and Shiryaev [70, 71], Kallianpur [59], Elliot [46] etc. In infinite dimensional spaces linear estimation problems are discussed in Curtain and Pritchard [40].

Chapter 7. Colored noise in estimation problems was used in Bucy and Joseph [35]. In general, in Arato [2] it is proved that in the one-dimensional case any stationary random process with the rational spectral density can be shown as a solution of linear stochastic differential equations with white noise disturbance, i.e., it is a colored noise.

Chapter 8. Estimation and control of wide band noise driven systems is actual in engineering. The results of this chapter are obtained in the works of the author and his colleagues. In the early papers [9, 11, 19] the duality principle was used to investigate the estimation problems for wide band noise driven systems. In the recent papers [13, 26, 27] a more flexible method based on a reduction was developed. The basic differential equations in (8.60)–(8.65) and in (8.92) are derived in [15]. Applications to space engineering and geophysics are discussed in [10, 18].

Chapter 9. This chapter is written on the basis of the early papers [6, 7, 11, 25] of the author and his colleagues, completed with some new progress in the theory not published previously. The reduction method faces difficulties when it is applied to the control and estimation problems under shifted noises. Therefore, the duality principle and the extended separation principle are used as methods of study in this chapter.

Chapter 10. The results of this chapter are most recent and they are reflected in [17] only. The method of approximations discussed in this chapter is promising but the control and estimation results are not yet proved precisely. The basic differential equations in (10.26),(10.28)–(10.30) and (10.63)–(10.66) as well as the other formulae for the optimal controls and optimal filters are derived intuitively and all of them can be considered as conjectures. The discussion of possible applications of control and filtering under shifted white noises in space navigation and guidance, given in this and previous chapters, is an illustrative but significant argument toward engineering.

Chapter 11. There is a remarkable relation between the control and estimation problems. This relation was discovered by Kalman [60] between the linear regulator and linear filtering problems and stated as the principle of duality. In Allahverdiev and Bashirov [1] this duality is extended to linear stochastic optimal control and estimation problems. Later this subject was developed in Bashirov [7] where the dual analogue of the innovation process was introduced.

Chapter 12. Theory of controllability originates from the famous work of Kalman [61] where the concept of complete controllability was defined. Later it was clear that the natural extension of this concept to infinite dimensional systems is too strong for many of them. Therefore, the approximate controllability was defined as a weakened version of the complete controllability in the early works of Triggiani, Fattorini, Russel etc. The significant achievements in controllability theory for deterministic linear systems are the Kalman rank condition, the complete controllability condition and the approximate controllability condition and they are well discussed in a number of books, for example, Balakrishnan [4], Curtain and Pritchard [40], Curtain and Zwart [41], Bensoussan *et al.* [32], Zabczyk [95] etc. Afterwards the resolvent conditions for complete and approximate controllability were discovered in Bashirov and Mahmudov [23]. Both the concepts of complete and approximate controllability lose sense for stochastic systems since now a terminal value is a random variable. The two different interconnections of controllability and randomness define the two principally different methods of extending the controllability concepts to stochastic systems. In the first method the state space in the definitions of controllability concepts is replaced by a suitable space of random variables, for example, the space of square integrable random variables. Thus, attaining random variables, even those with large entropy, is necessary to be controllable in this sense. This direction is employed by Mahmudov [74, 75]. The second method is more practical: it assumes attaining only those random variables which have small entropy excluding the needless random variables with large entropy. In this chapter we follow the second method. The concepts of C- and S-controllability for stochastic systems are studied in the works of the author and his colleagues [20, 21, 22, 23, 12, 24]. In the early papers [20, 21] these concepts are not yet formulated but significant progress has been made. The papers [22, 23] are basic for the C- and S-controllability and their conditions.

Bibliography

[1] J. E. Allahverdiev and A. E. Bashirov, *On the duality between control and interpolation problems.* Dokl. Akad. Nauk Azerb. SSR, **40** (1984), 3–6 (in Russian).

[2] M. Arato, *Linear Stochastic Systems with Constant Coefficients. A Statistical Approach.* Springer-Verlag, 1982.

[3] V. I. Arkin and M. T. Saksonov, *Necessary conditions of optimality for control problems with stochastic differential equations.* Dokl. Akad. Nauk SSSR, **244** (1979), 11–15 (in Russian).

[4] A. V. Balakrishnan, *Applied Functional Analysis.* Springer-Verlag, 1976.

[5] A. E. Bashirov, *A separation theorem for a stochastic problem of optimal control in Hilbert space.* Izv. Akad. Nauk Azerb. SSR, Ser. Fiz.-Tekhn.-Mat. Nauk, **1** (1980), 137–141 (in Russian).

[6] A. E. Bashirov, *A problem of filtering with correlated noises.* Izv. Akad. Nauk Azerb. SSR, Ser. Fiz.-Tekhn.-Mat. Nauk, **5** (1984), 97–102 (in Russian).

[7] A. E. Bashirov, *Optimal control of partially observed systems with arbitrary dependent noises: linear quadratic case.* Stochastics, **17** (1986), 163–205.

[8] A. E. Bashirov, *A partially observable linear quadratic game with dependent random noise.* Matematicheskaya Kibernetika i Prikladnaya Matematika, No. 7, Elm, Baku, 1986, 20–37 (in Russian).

[9] A. E. Bashirov, *On linear filtering under dependent wide band noises.* Stochastics, **23** (1988), 413–437.

[10] A. E. Bashirov, *Control and filtering for wide band noise driven linear systems.* Journal on Guidance, Control and Dynamics, **16** (1993), 983–985.

[11] A. E. Bashirov, *Linear filtering under dependent white and wide band noises.* Proc. Steklov Mathematical Institute, **202** (1993), 11–24 (in Russian); translation to English is published by AMS, Issue 4 (1994), 7–17.

[12] A. E. Bashirov, *On weakening of the controllability concepts.* Proc. 35th IEEE Conference on Decision and Control, Kobe, Japan, December 11–13, 1996, **1**, 640–645.

[13] A. E. Bashirov, *On linear systems disturbed by wide band noise.* Proc. 14th International Conference on Mathematical Theory of Networks and Systems, Perpignan, France, June 19–23, 2000, 7 p.

[14] A. E. Bashirov, *Stochastic control and game under arbitrarily dependent noises.* Proc. 8th Silivri-Gazimagusa Workshop on Stochastic Analysis and Related Topics, Gazimagusa, North Cyprus, September 18–27, 2000, 26 p.

[15] A. E. Bashirov, *Control and filtering of linear sytems driven by wide band noise.* Proc. 1st IFAC Symposium on Systems Structure and Control, Prague, Czech Republic, August 29–31, 2001, 6 p.

[16] A. E. Bashirov, *On error of estimation and minimum of cost for wide band noise driven systems.* 2002 MTNS Problem Book "Open Problems on the Mathematical Theory of Systems", Notre-Dame, USA, August 12–16, 2002, 24–26.

[17] A. E. Bashirov, *Approximation of white noise by wide band noise and its application to control and filtering under shifted white noises.* Submitted for publication in IEEE Trans. Automatic Control.

[18] A. E. Bashirov, L. V. Eppelbaum and L. R. Mishne, *Improving Eötvös corrections by wide band noise Kalman filtering.* Geophysical J. International, **108** (1992), 193–197.

[19] A. E. Bashirov, H. Etikan and N. Şemi, *Filtering, smoothing and prediction for wide band noise driven linear systems.* J. Franklin Institute, **334B** (1997), 667–683.

[20] A. E. Bashirov and R. R. Hajiyev, *The notion of controllability for partially observable stochastic systems, 1.* Izv. Akad. Nauk Azerb. SSR, Ser. Fiz.-Tekhn.-Mat. Nauk, **4** (1983), 109–114 (in Russian).

[21] A. E. Bashirov and R. R. Hajiyev, *The notion of controllability for partially observable stochastic systems, 2.* Izv. Akad. Nauk Azerb. SSR, Ser. Fiz.-Tekhn.-Mat. Nauk, **5** (1984), 99–103 (in Russian).

[22] A. E. Bashirov and K. R. Kerimov, *On controllability conception for stochastic systems.* SIAM J. Control and Optimization, **35** (1997), 384–398.

[23] A. E. Bashirov and N. I. Mahmudov, *On concepts of controllability for deterministic and stochastic systems.* SIAM J. Control and Optimization, **37** (1999), 1808–1821.

[24] A. E. Bashirov and N. I. Mahmudov, *Some new results in theory of controllability.* Proc. 7th IEEE Mediterranean Conference on Control and Automation, Haifa, Israel, June 28–30, 1999, 323–343.

[25] A. E. Bashirov and L. R. Mishne, *On linear filtering under dependent white noises.* Stochastics and Stochastics Reports, **35** (1991), 1–23.

[26] A. E. Bashirov and S. Uğural, *Analyzing wide band noise with application to control and filtering.* IEEE Trans. Automatic Control, **47** (2002), 323–327.

[27] A. E. Bashirov and S. Uğural, *Representation of systems disturbed by wide band noise.* Appl. Math. Letters, **15** (2002), 607–613.

[28] R. Bellman, *Dynamic Programming.* Princeton University Press, 1957.

[29] A. Bensoussan, *Filtrage Optimal des Systemes Lineaires.* Dunod, 1971.

[30] A. Bensoussan, *Stochastic Control of Partially Observable Systems.* Cambridge University Press, 1992.

[31] A. Bensoussan, G. Da Prato, M. C. Delfour and S. K. Mitter, *Representation and Control of Infinite Dimensional Systems, 1.* Birkäuser, 1992.

[32] A. Bensoussan, G. Da Prato, M. C. Delfour and S. K. Mitter, *Representation and Control of Infinite Dimensional Systems, 2.* Birkhäuser, 1993.

[33] A. Bensoussan and M. Viot, *Optimal control of stochastic linear distributed parameter systems.* SIAM J. Control, **13** (1975), 904–926.

[34] R. Brown, *A brief account of microscopical observations made in the months of June, July and August, 1927, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies.* Philos. Mag. Ann. of Philos. New Ser. **4** (1928), 161–178.

[35] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance.* Interscience, 1968.

[36] A. Cadenillas and I. Karatzas, *The stochastic maximum principle for linear convex optimal control with random coefficients.* SIAM J. Control and Optimization, **33** (1995), 590–624.

[37] R. R. Curtain, *Estimation theory for abstract evolution equations excited by general white noise processes.* SIAM J. Control and Optimization, **14** (1976), 1124–1150.

[38] R. F. Curtain and A. Ichikawa, *The separation principle for stochastic evolution equations.* SIAM J. Control and Optimization, **15** (1977), 367–383.

[39] R. F. Curtain and A. J. Pritchard, *The infinite dimensional Riccati equation for systems defined by evolution operators.* SIAM J. Control and Optimization, **14** (1976), 951–983.

[40] R. F. Curtain and A. J. Pritchard, *Infinite Dimensional Linear Systems Theory.* Lecture Notes on Control and Inform. Sci., **8**, Springer-Verlag, 1978.

[41] R. F. Curtain and H. J. Zwart, *An Introduction to Infinite Dimensional Linear Systems Theory.* Springer-Verlag, 1995.

[42] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, 1992.

[43] M. H. A. Davis, *Linear Estimation and Stochastic Control.* Chapman and Hall, 1977.

[44] M. C. Delfour and S. K. Mitter, *Controllability, observability and optimal feedback control of affine hereditary differential systems.* SIAM J. Control, **10** (1972), 298–328.

[45] N. Dunford and J. T. Schwartz, *Linear operators, Part I: General Theory.* Interscience, 1967.

[46] R. J. Elliot, *Stochastic Calculus and Applications.* Springer-Verlag, 1982.

[47] R. J. Elliot and M. Kohlmann, *The second order minimum principle and adjoint processes.* Stochastics and Stochastics Reports, **46** (1994), 25–39.

[48] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control.* Springer-Verlag, 1975.

[49] W. H. Fleming and H. M. Soner, *Contrlled Markov Processes and Viscosity Solutions.* Springer-Verlag, 1993.

[50] I. I. Gihman and A. V. Skorohod, *Introduction to the Theory of Random Processes.* Saunders, 1969.

[51] I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations.* Springer-Verlag, 1972.

[52] M. S. Grewal and A. P. Andrews, *Kalman Filtering. Theory and Applications.* Prentice Hall, 1993.

[53] T. Hida, *Brownian Motion.* Springer-Verlag, 1980.

[54] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups.* Amer. Math. Soc. Coll. Publ., **31**, Providence, 1957.

[55] A. Ichikawa, *Quadratic control of evolution equations with delays in control.* SIAM J. Control and Optimization, **20** (1982), 645–668.

[56] K. Ito, *Stochastic integral.* Proc. Imperial Acad., Tokyo, **20** (1944), 519–524.

[57] K. Ito, *On stochastic differential equations.* Mem. Amer. Math. Soc., **4** (1951), 645–668.

[58] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes.* North Holland, 1979.

[59] G. Kallianpur, *Stochastic Filtering Theory.* Springer-Verlag, 1980.

[60] R. E. Kalman, *Contributions to the theory of optimal control.* Bull. Soc. Math. Mexicana, **5** (1960), 102–119.

[61] R. E. Kalman, *A new approach to linear filtering and prediction problems.* Trans. ASME, Ser. D, J. Basic Engineering, **82** (1960), 35–45.

[62] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory.* Trans. ASME, Ser. D, J. Basic Engineering, **83** (1961), 95–108.

[63] T. Kato, *Perturbation Theory of Linear Operators.* Springer-Verlag, 1980.

[64] A. N. Kolmogorov, *Interpolation and extrapolation of stationary random sequences.* Izv. Akad. Nauk SSSR, Ser. Mat., **5** (1941), 3–14 (in Russian).

[65] A. N. Kolmogorov, *Foundations of the Theory of Probability.* Chelsea, 1956.

[66] N. V. Krylov, *Controlled Diffusion Processes.* Springer-Verlag, 1080.

[67] N. V. Krylov, *Introduction to the Theory of Diffusion Processes.* Amer. Math. Soc., Providence, 1995.

[68] H. J. Kushner, *Necessary conditions for continuous parameter stochastic optimization problems.* SIAM J. Control, **10** (1972), 550–565.

[69] H. J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems.* Birkhäuser, 1990.

[70] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes 1: General Theory.* Springer-Verlag, 1977.

[71] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes 2: Applications.* Springer-Verlag, 1978.

[72] R. S. Liptser and A. N. Shiryaev, *Theory of Martingales.* Kluwer, 1989.

[73] N. I. Mahmudov, *General necessary conditions of optimality for stochastic systems with controllable diffusion.* Proc. Workshop "Statistics and Control of Random Processes, Preila, 1987", Nauka, Moscow, 1989, 135–138 (in Russian).

[74] N. I. Mahmudov, *Controllability of linear stochastic systems.* IEEE Trans. Automatic Control, **46** (2001), 724–732.

[75] N. I. Mahmudov, *Controllability of linear stochastic systems in Hilbert spaces.* J. Math. Anal. Appl., **259** (2001), 64–82.

[76] N. I. Mahmudov and A. E. Bashirov, *First order and second order necessary conditions of optimality for stochastic systems.* Proc. Steklov Mathematical Institute Seminar "Statistics and Control of Stochastic Processes, The Liptser Festschrift", Editors: Yu. M. Kabanov, B. L. Rozovskii and A. N. Shiryaev, World Scientific, 1997, 283–295.

[77] M. Metivier, *Semimartingales.* Walter de Gruyter, 1983.

[78] P. A. Meyer, *Probability and Potential.* Blaisdell Pablishing Company, 1966.

[79] A. Papoulis, *The Fourier Integral and Its Applications.* McGraw-Hill, 1987.

[80] K. R. Parthasarathy, *Probability Measures on Metric Spaces.* Academic Press, 1967.

[81] S. Peng, *A general stochastic maximum principle for optimal control problems.* SIAM J. Control and Optimization, **28** (1990), 966–979.

[82] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkreledze and E. F. Mischenko, *The Mathematical Theory of Optimal Processes.* Interscience, 1962.

[83] R. K. A. Paley, N. Wiener and A. Zygmund, *Note on random functions.* Math. Z., **37** (1933), 647–668.

[84] B. L. Rozovskii, *Stochastic Evolution Systems.* Kluwer, 1990.

[85] W. Rudin, *Principles of Mathematical Analysis.* McGraw-Hill, 1976.

[86] A. N. Shiryaev, *Probability.* Springer-Verlag, 1984.

[87] A. V. Skorohod, *On a generalization of a stochastic integral.* Theory of Probability Appl., **20** (1975), 219–233.

[88] E. G. F. Thomas, *Vector-valued integration with applications to the operator-valued H^∞ space.* IMA J. Mathematical Control and Information, **14** (1997), 109–136.

[89] J. Warga, *Optimal Control of Differential and Functional Equations.* Academic Press, 1972.

[90] N. Wiener, *Differential space.* J. Math. and Phys., **2** (1923), 131–174.

[91] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications.* Wiley, 1950.

[92] W. M. Wonham, *On the separation theorem of stochastic control.* SIAM J. Control, **6** (1968), 312–326.

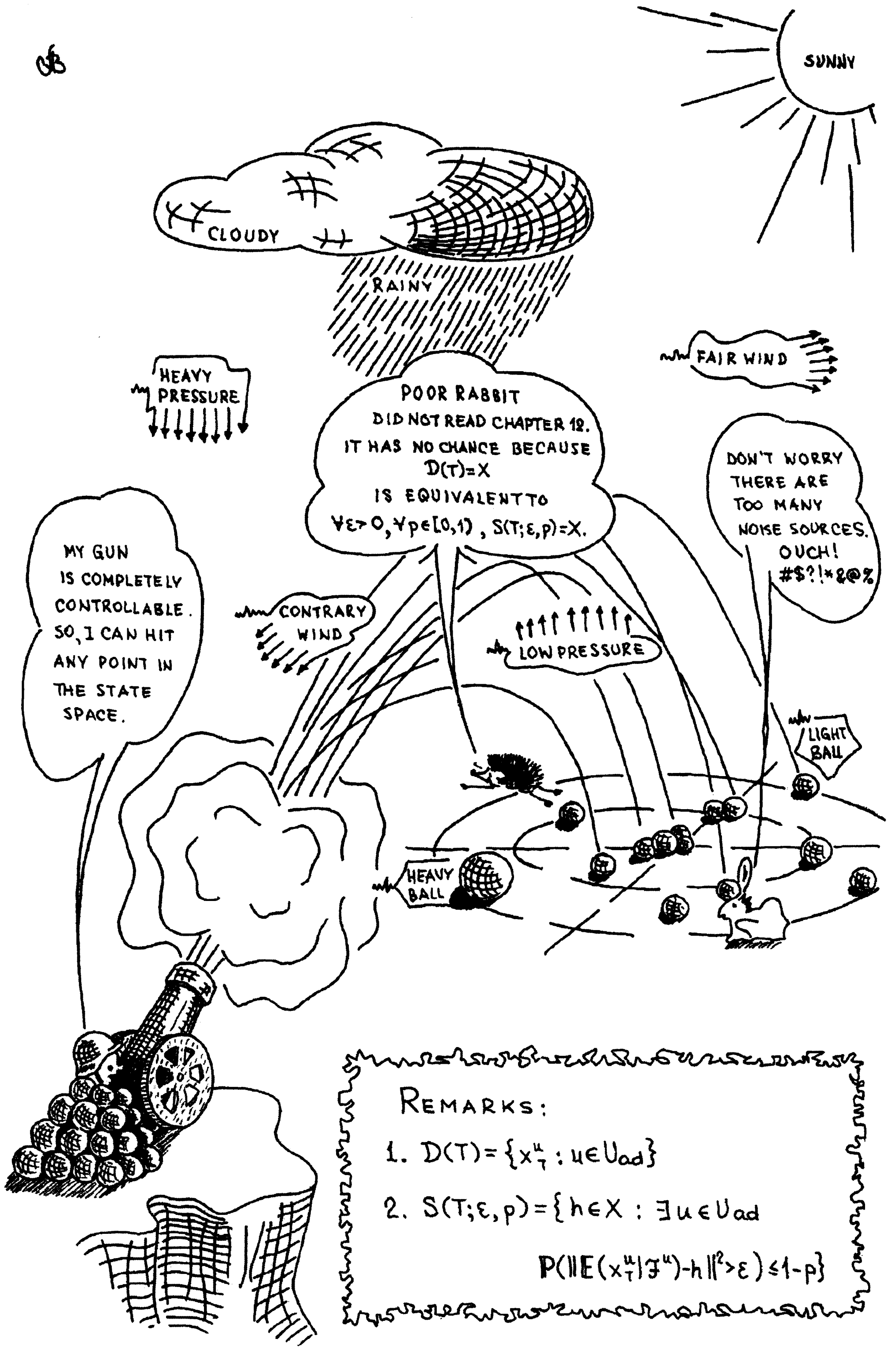
[93] J. Yong and X. Y. Zhou, *Stochastic Controls. Hamiltonian Systems and HJB Equations.* Springer-Verlag, 1999.

[94] K. Yosida, *Functional Analysis.* Springer-Verlag, 1980.

[95] J. Zabczyk, *Mathematical Control Theory: An Introduction.* Birkhäuser, 1992.

Index of Notation

Index



REMARKS:

1. $D(T)=\{x^u_T : u\in U_{ad}\}$

2. $S(T;\varepsilon,p)=\{h\in X : \exists u\in U_{ad}$

$\mathbb{P}(\|E(x^u_T|\mathcal{F}^u)-h\|^2>\varepsilon)\le 1-p\}$