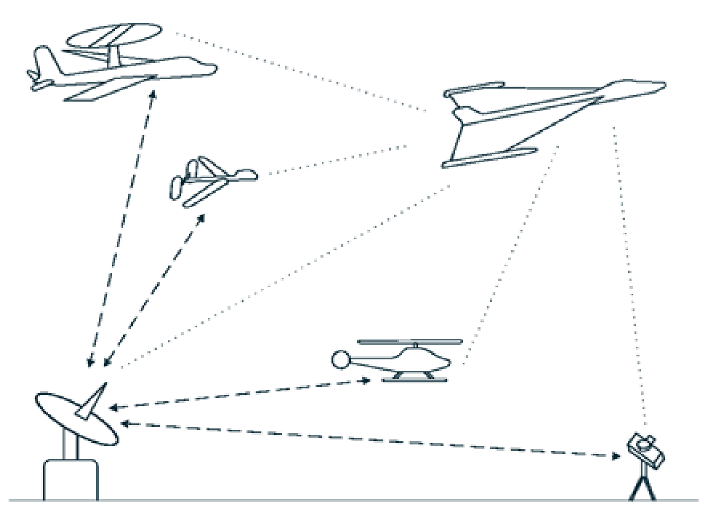
CONTROL
ENGINEERING

Alexey S. Matveev
Andrey V. Savkin

Estimation and Control over Communication Networks



Birkhäuser

Control Engineering

Alexey S. Matveev
Andrey V. Savkin

Estimation and Control over Communication Networks

Birkhäuser
Boston • Basel • Berlin

Alexey S. Matveev
Department of Mathematics
and Mechanics
Saint Petersburg University
198504 Petrodovoretz
Saint Petersburg
Russia

Andrey V. Savkin
School of Electrical Engineering
and Telecommunications
University of New South Wales
NSW2052 Sydney
Australia

ISBN: 978-0-8176-4494-9 e-ISBN: 978-0-8176-4607-3
DOI: 10.1007/978-0-8176-4067-3

Library of Congress Control Number: 2008929989

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

www.birkhauser.com

Preface

Rapid advances in communication technology have opened up the possibility of large-scale control systems in which the control task is distributed among several processors and the communication among the processors, sensors, and actuators is via communication channels. Such control systems may be distributed over large distances and may use large numbers of actuators and sensors. The possibility of such networked control systems motivates the development of a new chapter of control theory in which control and communication issues are integrated, and all the limitations of communication channels are taken into account. There is an emerging literature on this topic; however, at present there is no systematic theory of estimation and control over communication networks. This book is concerned with the development of such a theory.

This book is primarily a research monograph that presents, in a unified manner, some recent research on control and estimation over communication channels. It is essentially self-contained and is intended both for researchers and advanced postgraduate students working in the areas of control engineering, communications, information theory, signal processing or applied mathematics with an interest in the emerging field of networked control systems. The reader is assumed to be competent in the basic mathematical techniques of modern control theory.

By restricting ourselves to several selected problems of estimation and control over communication networks, we are able to present and prove a number of results concerning optimality, stability, and robustness that are of practical significance for networked control system design. In particular, various problems of Kalman filtering, stabilization, and optimal control over communication channels are considered and solved. The results establish fundamental links among mathematical control theory, Shannon information theory, and entropy theory of dynamical systems. We hope that the reader finds this work both useful and interesting and is inspired to explore further the diverse and challenging area of networked control systems. This book is one of the first research monographs on estimation and control over communication networks.

The material presented in this book derives from a period of fruitful research collaboration between the authors on the area of networked control systems beginning

in 1999 and is still ongoing. Some of the material contained herein has appeared as isolated results in journal papers and conference proceedings. This work presents this material in an integrated and coherent manner and presents many new results. Much of the material arose from joint work with students and colleagues, and the authors wish to acknowledge the major contributions made by Veerachai Malyavej, Ian Petersen, Rob Evans, Teddy Cheng, Efstratios Skafidas, and Valery Ugrinovskii. Our thanks for the help with some figures in the book go to Teddy Cheng and Veerachai Malyavej. We are also grateful to our colleagues Girish Nair, Daniel Liberzon, Victor Solo, Tamer Başar, David Clements, and Andrey Barabanov who have provided useful comments and suggestions.

The authors wish to acknowledge the support they have received throughout the preparation of this work from the School of Electrical Engineering and Telecommunications at the University of New South Wales, Sydney, and the Faculty of Mathematics and Mechanics at the Saint Petersburg University. The authors are also extremely grateful for the financial support they have received from the Australian Research Council, the Russian Foundation for Basic Research (grant 06-08-01386), and the Research Council of the President of the Russian Federation (grant 2387.2008.1).

Furthermore, the first author is grateful for the enormous support he has received from his wife Elena and daughter Julia. Also, the second author is indebted to the endless love and support he has received from his wife Natalia and children Mikhail and Katerina.

Alexey S. Matveev — Saint Petersburg, Russia
Andrey V. Savkin — Sydney, Australia
March 2008

Contents

1

Introduction

1.1 Control Systems and Communication Networks

Control and communications have traditionally been different areas with little overlap. Until the 1990s it was common to decouple the communication issues from consideration of state estimation or control problems. In particular, in the classic control and state estimation theory, the standard assumption is that all data transmission required by the algorithm can be performed with infinite precision in value. In such an approach, control and communication components are treated as totally independent. This considerably simplifies the analysis and design of the overall system and mostly works well for engineering systems with large communication bandwidth. However, in some recently emerging applications, situations are encountered where observation and control signals are transmitted via a communication channel with a limited capacity. For instance, this issue may arise with the transmission of control signals when a large number of mobile units needs to be controlled remotely by a single decision maker. Since the radio spectrum is limited, communication constraints are a real concern. In [199], the design of large-scale control systems for platoons of underwater vehicles highlights the need for control strategies that address reduced communications, since communication bandwidth is severely limited underwater. Other recent emerging applications are micro-electromechanical systems and mobile telephony.

On the other hand, for complex networked sensor systems containing a very large number of low-power sensors, the amount of data collected by the sensors is too large to be transmitted in full via the existing communication channel. In these problems, classic control and state estimation theory cannot be applied since the controller/state estimator only observes the transmitted sequence of finite-valued symbols. So it is natural to ask how much transmission capacity is needed to achieve a certain control goal or a specified state estimation accuracy. The problem becomes even more challenging when the system contains multiple sensors and actuators transmitting and receiving data over a shared communication network. In such systems, each module is effectively allocated only a small portion of the network total communication capacity.

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_1,

Another shortcoming of the classic control and estimation theory is the assumption that data transmission and information processing required by the control/estimation algorithm can be performed instantaneously. However, in complex real-world networked control systems, data arrival times are often delayed, irregular, time-varying, and not precisely known, and data may arrive out of order. Moreover, data transferred via a communication network may be corrupted or even lost due to noise in the communication medium, congestion of the communication network, or protocol malfunctions. The problem of missing data may also arise from temporary sensor failures. Examples arise in planetary rovers, arrays of microactuators, and power control in mobile communications. Other examples are offered by complex dynamic processes like advanced aircraft, spacecraft, and manufacturing processes, where time division multiplexed computer networks are employed for exchange of information between spatially distributed plant components.

On the other hand, for many complex control systems, it can be desirable to distribute the control task among several processors, rather than using a single central processor. If these processors are not triggered by a common clock pulse, and their computation, sampling, and hold activities are not synchronized, we call them asynchronous controllers. In addition, these processors need not operate with the same sampling rate, and so-called multirate sampling in control systems has been of interest since the 1950s (see, e.g., [54, 80, 230]). The sampling rates of the controllers are typically assumed to be precisely known and integrally proportional, and sampling is synchronized to make the sampling process periodic, with a period equal to an integral multiple of the largest sampling period. However, in many practical situations, the sampling times are irregular and not precisely known. This occurs, for example, when a large-scale computer controller is time-shared by several plants so that control signals are sent out to each plant at random times. It should be pointed out that the multitask allocation for large multiprocessor computers is a very complex and practically nondeterministic process. In fact, the problem of uncertain and irregular sampling times often faces engineers when they use multiprocessor computer systems and communication networks for operation and control of complex physical processes. In all these applications, communication issues are of real concern.

Another rapidly emerging area is cooperative control of multiagent networked systems, especially formations of autonomous unmanned vehicles; see, e.g., [9, 51, 76, 159, 160, 169]. The key challenge in this area is the problem of cooperation between a group of agents performing a shared task using interagent communication. The system is decentralized, and decisions are made by each agent using limited information about other agents and the environment. Applications include mobile robots, unmanned aerial vehicles (UAVs), automated highway systems, sensor networks for spatially distributed sensing, and microsatellite clusters. In all these applications, the interplay between communication network properties and vehicle dynamics is crucial. This class of problems represents a difficult and exciting challenge in control engineering and is expected to be one of the most important areas of control theory in the near future.

A slightly different approach was proposed in the signal processing community where problems of parameter estimation in sensor networks with limited communication capacity were studied (see, e.g., [6, 7] for a survey).

These new engineering applications have attracted considerable research interest in the last decade; however, the interplay between control and communication is a fundamental topic, and its origins go back much earlier than that. For example, in 1948 Wiener introduced the term *cybernetics* and defined it as control and communication in the animal and the machine [218]. Furthermore, ideas on importance of the information-based approach to control can be found in the work of many researchers over several decades.

All these engineering applications and fundamental questions motivate development of a new chapter of mathematical electrical engineering in which control and communication issues are combined, and all the limitations of the communication channels are taken into account. The emerging area of networked control systems lies at the crossroads of control, information, communication, and dynamical system theory. The importance of this area is quickly increasing due to the growing use of communication networks and very large numbers of sensors in modern control systems. There is now an emerging literature on this topic, (see, e.g., [15, 27, 39, 42, 48, 58, 64, 74, 128, 133, 135]) describing a number of models, algorithms, and stability criteria. However, currently there is no systematic theory of estimation and control over communication networks. This book is concerned with the development of such a new theory that utilizes communications, control, information, and dynamical systems theory and is motivated by and applied to advanced networking scenarios.

The literature in the field of control over communication networks is vast, and we have limited ourselves to references that we found most useful or that contain material supplementing the text. The coverage of the literature in this book is by no means complete. We apologize in advance to the many authors whose contributions have not been mentioned. Also, an excellent overview of the literature in the field can be found in [142].

In conclusion, the area of networked control systems is a fascinating new discipline bridging control engineering, communications, information theory, signal processing, and dynamical system theory. The study of networked control systems represents a difficult and exciting challenge in control engineering. We hope that this monograph will help in some small way to meet this challenge.

1.2 Overview of the Book

In this section, we briefly describe the results presented in the book.

1.2.1 Estimation and Control over Limited Capacity Deterministic Channels

Chapter 2 provides basic results on connections between problems of estimation and control over limited capacity communication channels and the entropy theory of dynamical systems originated in the work of Kolmogorov [82, 83]. The paper [141] imported the concept of topological entropy into the area of control over communication channels. In Chap. 2, we use the so-called "metric definition" of topological entropy and derive several important properties of it. In particular, we present a simple proof of the well-known result asserting that the topological entropy of a discrete-time linear system is given by

$$H(A) = \sum_{i=1,2,\ldots,n} \log_2(\max\{1, |\lambda_i|\}), \tag{1.2.1}$$

where A is the matrix of the linear system and $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is the set of eigenvalues of the matrix A. In this chapter, we give a necessary and sufficient condition for observability over a limited capacity channel in terms of an inequality between the channel capacity and the topological entropy of the open-loop plant. Furthermore, we show that the similar inequality

$$H(A) < R \tag{1.2.2}$$

is a necessary and sufficient condition for stabilizability of a linear plant via a digital channel. Here $H(A)$ is defined by (1.2.1) and R is the capacity of the deterministic digital channel. It should be pointed out that similar results were first proved in the work of Nair and Evans [137, 138]. Furthermore, we prove that under the same inequality (1.2.2) between the channel capacity and the topological entropy of the plant, the cost in the problem of linear-quadratic (LQ) optimal control via a digital channel can be brought as close as desired to the cost in the classic LQ optimal control problem.

Chapter 3 extends the stabilization result of Chap. 3 to the much more general case of linear plants with multiple sensors and multiple digital communication channels. Moreover, it is not assumed that the channels are perfect; i.e., time-varying delays and data losses are possible.

In Chap. 4, we consider problems of detectability and output feedback stabilizability via limited capacity communication channels for a class of nonlinear systems, with nonlinearities satisfying a globally Lipschitz condition. We derive sufficient conditions for stabilizability and detectability and present a constructive procedure for the design of state estimators and stabilizing output feedback controllers. Finally, we present an illustrative example in which a stabilizing output feedback controller is designed for a robotic flexible joint with video measurement transmitted to the controller location via a wireless limited capacity communication channel.

Chapter 5 addresses the problem of robust state estimation over limited capacity communication channels. Robustness is one key requirement for any control system. That is, the requirement that the control system will maintain an adequate level of performance in the face of significant plant uncertainty. Such plant uncertainties may be due to variation in the plant parameters and to the effects on nonlinearities and unmodeled dynamics that have not been included in the plant model. In fact, the requirement for robustness is one of the main reasons for using feedback in control system design. In this chapter, we consider a plant modeled by an uncertain system with uncertainties satisfying so-called integral quadratic constraint. This uncertainty description was first introduced in the work of Yakubovich on absolute stability (see, e.g., [222]). A robust coder–decoder–estimator is designed for such uncertain plants.

Chapter 13 studies the problem of linear-quadratic Gaussian (LQG) optimal control over a limited capacity communication channel. This problem is considered for a discrete-time linear plant and a finite time interval. We derive an optimal coding–decoding–control strategy for this problem. One consequence of the main result of this chapter is that an analog of the separation principle from linear stochastic control does not hold for problems of optimal Gaussian control via limited capacity channels.

1.2.2 An Analog of Shannon Information Theory: Estimation and Control over Noisy Discrete Channels

In Chaps. 6–8, we present several results that can be viewed as an analog of Shannon information theory for networked control systems. We consider problems of stabilization and state estimation of unstable linear discrete-time plants via stationary, memoryless, noisy discrete channels, which are common in classic information theory.

The main result of Chap. 6 is that stabilizability (detectability) with probability 1 of a linear unstable plant without plant disturbances is "almost" equivalent to the inequality

$$H(A) < \mathfrak{c}, \tag{1.2.3}$$

where $\mathfrak{c}$ is the Shannon ordinary capacity of the channel and $H(A)$ is the topological entropy of the open-loop plant defined by (1.2.1).

In Chap. 7, we address similar stabilization and state detection problems; however, it is assumed that the plant is affected by disturbances. We prove that an "almost" necessary and sufficient condition for existence of a coder–decoder pair such that solutions of the closed-loop system are bounded with probability 1 is the inequality

$$H(A) < \mathfrak{c}_0, \tag{1.2.4}$$

where $\mathfrak{c}_0$ is the zero error capacity of the channel. The Shannon ordinary capacity $\mathfrak{c}$ of the channel is the least upper bound of rates at which information can be transmitted with as small a probability of error as desired, whereas the zero error capacity $\mathfrak{c}_0$ is the least upper bound of rates at which it is possible to transmit information with zero probability of error. The concept of the zero error capacity was also introduced by Shannon in 1956 [189]. Unlike the Shannon ordinary capacity, the zero error capacity may depend on whether the communication feedback is available. The general formula for $\mathfrak{c}$ is well known, whereas the general formula for $\mathfrak{c}_0$ is still missing.

The results of these two chapters have significant shortcomings. The results of Chap. 6 do not guarantee any robustness subject to disturbances. On the other hand, the results of Chap. 7 are quite conservative. Indeed, usually, $\mathfrak{c}_0$ is significantly less than $\mathfrak{c}$. Moreover, $\mathfrak{c}_0 = 0$ for many channels. Also, despite 50 years of research in information theory started by Shannon, there is no general formula for $\mathfrak{c}_0$.

To overcome these shortcomings, in Chap. 8, we introduce the concept of stabilizability in probability. This kind of stabilizability means that one can find a coder–decoder pair such that the closed-loop system satisfies the following condition: For any probability $0 < p < 1$, a constant $b_p > 0$ exists such that:

$$\mathbf{P}\left[\|x(t)\| \leq b_p\right] \geq p \qquad \forall t = 1, 2, \dots. \tag{1.2.5}$$

The main result of Chap. 8 is that stabilizability in probability is almost equivalent to the inequality (1.2.3).

Combining the results of Chaps. 6 and 8, it can be shown that if the inequality (1.2.4) holds, then the constants b_p in (1.2.5) can be taken so that

$$\sup_{p \to 1} b_p < \infty.$$

On the other hand, if $\mathfrak{c}_0 < H(A) < \mathfrak{c}$, then

$$\sup_{p \to 1} b_p = \infty.$$

Similar results were derived in Chaps. 7 and 8 for state estimation problems.

It should be pointed out that the procedures for the design of controllers and state estimators proposed in Chaps. 6–8 are quite constructive. Furthermore, it is very important that all these coder–decoder pairs require uniformly bounded over infinite time memory and computational power.

1.2.3 Decentralized Stabilization via Limited Capacity Communication Networks

The advanced networking scenario is considered in Chap. 9. In this chapter, we study linear plants with multiple sensors and actuators. The sensors and actuators are connected via a complex communication networks with a very general topology. The network contains a large number of spatially distributed nodes that receive and transmit data. Each node is equipped with a CPU. For some nodes, coding and decoding algorithms are fixed, for other nodes, they need to be designed. Moreover, data may arrive with delays, be lost, or become corrupted. The goal is to stabilize a linear plant via such a network. We give a necessary and sufficient condition for stabilizability. This condition is given in terms of the so-called rate (capacity) domain of the communication network. Our results show that the problem of networked stabilization is reduced to the very hard long-standing problem of information theory: calculating the capacity domain of communication networks.

1.2.4 H^∞ State Estimation via Communication Channels

In Chap. 10, a different approach to state estimation via communication channels is presented. In this new problem statement, the channel transmits a continuous-time vector signal. The limited capacity of the channel means that the dimension of the signal to be transmitted is smaller than the dimension of the plant measured output. Our goal is to design a coder at the transmitting end of the channel and a decoder–estimator at the receiving end so that the state estimate produced by the coder–decoder pair satisfies a standard requirement from H^∞ filtering theory. It should be pointed out that the state estimator designed in Chap. 10 is linear and time-invariant.

1.2.5 Kalman Filtering and Optimal Control via Asynchronous Channels with Irregular Delays

In Chaps. 11 and 12, discrete-time linear plants with Gaussian disturbances are considered. The system under consideration has several sensors and measurements are transmitted to the estimator or controller via parallel communication channels with independent delays. Unlike Chaps. 2–9 where the communication channels transmit symbols from finite alphabets, in these chapters we assume that transmissions are performed with infinite precision in value; i.e., the channels transmit discrete-time sequences of real numbers or vectors. However, data may be lost or arrive out of order.

In Chap. 11, we assume that the probability distributions of the channels delays are known. Under this assumption, we derive an analog of the Kalman filter and solve the LQG optimal control problem. In Chap. 12, we consider a more complicated situation where the control loop is not perfect and control signals arrive to several actuators via asynchronous communication channels. Based on the results of the previous chapter, we give a solution of the optimal control problem for such systems. It should be pointed out that unlike Chaps. 2–9, all the optimal state estimators and controllers of Chaps. 11 and 12 are linear.

1.2.6 Kalman Filtering with Switched Sensors

In Chaps. 14 and 15 we consider plants with multiple sensors communicating to the state estimator via a set of independent channels. The bandwidth limitation constraint is modeled in such a manner that the state estimator can communicate with only one sensor at any time. So the state estimation problem is reduced to finding a suitable sensor scheduling algorithm. In Chap. 14 we consider the system with asynchronous communication channels between the sensors and the state estimator. As in Chaps. 11 and 12, sensor data arrive with irregular delays and may be lost. Using the results of Chap. 11, we derive an optimal sensor scheduling rule. The construction of the optimal state estimator is based on solving the Riccati equations and a dynamic programming equation.

Chapter 15 considers the sensor switching problem for uncertain plants, with uncertainties satisfying integral quadratic constraints. Such uncertain system models were studied in Chap. 5. Furthermore, we use robust state estimation results from Chap. 5. As in Chap. 14, our sensor switching algorithm requires solving a set of Riccati equations and a dynamic programming equation. Because solving a dynamic programming equation is a computationally expansive procedure, in both Chaps. 14

and 15, we propose suboptimal state estimators that are designed using ideas of so-called model predictive control. Such state estimators require much less computational power and are more implementable in real time.

1.2.7 Some Other Remarks

The chapters of this book can be divided into two groups. In Chaps. 2–9, and 13, we consider communication channels that transmit a finite number of bits; in other words, elements form a finite set. On the other hand, in Chaps. 10–12, 14, and 15, transmissions are performed with infinite precision in value; i.e., the communication channels under consideration transmit real numbers or vectors.

It should be pointed out that the state estimation and control systems designed in Chaps. 2–9, and 13–15 can be naturally viewed as so-called hybrid dynamical systems; see, e.g., [104, 173, 174, 177, 185, 211]. The term "hybrid dynamical system" has many meanings, the most common of which is a dynamic system that involves the interaction of discrete and continuous dynamics. Such dynamic systems typically contain variables that take values from a continuous set (usually the set of real numbers) and symbolic variables that take values from a finite set (e.g., the set of symbols $\{q_1, q_2, \ldots, q_k\}$). A model of this type can be used to describe accurately a wide range of real-time industrial processes and their associated supervisory control and monitoring systems. In Chaps. 2–9, and 13–15, the plant state variables are continuous, whereas data transmitted via digital finite capacity channels can be naturally modeled as symbolic variables.

Discrete-time plants are under consideration in Chaps. 2 and 3, 6–9, and 11–14. Chapters 4 and 5, 10 and 15 study continuous-time plants.

Stochastic models are addressed in Chaps. 6–8 and 10–14, whereas all other chapters consider deterministic models.

The design procedures of Chaps. 10–12 result in linear state estimators and controllers. The state estimators and controllers in all other chapters are highly nonlinear.

Finally, plants with parametric uncertainties are studied in Chaps. 2, 5, and 15.

1.3 Frequently Used Notations

$:=$	is defined (set) to be
$\wedge, \&$	and
$\vee$	or
$\Rightarrow$	implies
$\Leftrightarrow$	is equivalent to
$\longleftrightarrow$	corresponds to, is associated with
$\equiv$	is identical to
$\square$	the end of proof
$\{e_1, e_2, \ldots, e_n\}$	the set composed by the elements $e_1, e_2, \ldots, e_n$
$\{e \in E : \mathfrak{P}(e) \text{ holds}\}$	the set of all elements $e \in E$ with the property $\mathfrak{P}(e)$

$E = \{e\}$	this means that the elements of the set E are denoted by e
$\emptyset$	the empty set
$\|S\|$	the size (cardinality) of the set S
ds	the counting measure $S' \subset S \mapsto \|S'\|$ if the set $S = \{s\}$ is finite, and the Lebesgue measure if $S = \mathbb{R}^k$
$f(\cdot)$	the dot $\cdot$ in brackets underscores that the preceding symbol is used to denote a function
$\circledast$	a ”void” symbol; it is used to symbolize that the contents of something are null; for example, the memory is empty, no message is received (which may be interpreted as receiving a message with the null content)
✠	the ”alarm” symbol; it is used to mark ”undesirable” events
$\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}$	the sets of real, complex, and integer numbers, respectively
sgn x	the sign of the real number x
$f(t+0)$	the limit of the function $f(\cdot)$ at the point t from the right $f(t+0) := \lim_{\epsilon>0,\ \epsilon\to 0} f(t+\epsilon)$
$f(t-0)$	the limit of the function $f(\cdot)$ at the point t from the left $f(t-0) := \lim_{\epsilon>0,\ \epsilon\to 0} f(t-\epsilon)$

	Real interval with end points $t_0, t_1 \in \mathbb{R}$	Integer interval with end points $t_0, t_1 \in \mathbb{Z}$
a)	$[t_0, t_1] := \{t \in \mathbb{R} : t_0 \le t \le t_1\}$	$[t_0 : t_1] := \{t \in \mathbb{Z} : t_0 \le t \le t_1\}$
b)	$[t_0, t_1) := \{t \in \mathbb{R} : t_0 \le t < t_1\}$	$[t_0 : t_1) := \{t \in \mathbb{Z} : t_0 \le t < t_1\}$
c)	$(t_0, t_1] := \{t \in \mathbb{R} : t_0 < t \le t_1\}$	$(t_0 : t_1] := \{t \in \mathbb{Z} : t_0 < t \le t_1\}$
d)	$(t_0, t_1) := \{t \in \mathbb{R} : t_0 < t < t_1\}$	$(t_0 : t_1) := \{t \in \mathbb{Z} : t_0 < t < t_1\}$
t_1 may equal $+\infty$ in the cases b) and d). t_0 may equal $-\infty$ in the cases c) and d).		

$\overline{\lim}_{i\to\infty} x_i = \limsup_{i\to\infty} x_i$ the upper limit of the real-valued sequence:

$$\overline{\lim}_{i\to\infty} x_i = \limsup_{i\to\infty} x_i := \lim_{k\to\infty} \sup_{i\ge k} x_i$$

$\underline{\lim}_{i\to\infty} x_i = \liminf_{i\to\infty} x_i$ the lower limit of the real-valued sequence:

$$\underline{\lim}_{i\to\infty} x_i = \liminf_{i\to\infty} x_i := \lim_{k\to\infty} \inf_{i\ge k} x_i$$

$\imath$	the imaginary unit
$\mathrm{Re}z, \mathrm{Im}z$	the real and imaginary parts of the complex number z

$\lceil x \rceil$	$:= \min_{k \in \mathbb{Z}: k \geq x} k$: the integer ceiling of the real number x
$\lfloor x \rfloor$	$:= \max_{k \in \mathbb{Z}: k \leq x} k$: the integer floor of the real number x
$\log_a$	the logarithm base $a \in (0, 1) \cup (1, \infty)$
$\log_a 0$	$:= -\infty$
$\log_a \infty$	$:= +\infty$
ln	the natural logarithm
$0 \cdot (\pm\infty)$	$:= 0$
$x_i \uparrow \infty$	this means that the real-valued sequence $x_1, x_2, \ldots$ increases $x_i < x_{i+1}$ and $x_i \to \infty$ as $i \to \infty$
$x(\cdot)\vert_{t_0}^{t_1}$	the restriction of the function $x(t)$ of t on $[t_0 : t_1]$ if t is the integer variable and on $[t_0, t_1]$ if t is the real variable
B_z^r	the open ball centered at z with the radius r
$\boldsymbol{V}_k(S)$	the volume (Lebesgue measure) of the set $S \subset \mathbb{R}^k$; the index may be dropped if k is clear from the context
$\overline{S}$	the closure of the set S
int$S$	the interior of the set S
$\dim L$	the dimension of a linear space L
$\dim(x)$	the dimension of the vector x
$^{\mathrm{T}}$	the transpose
col $(D_1, \ldots, D_p)$	$:= \left(D_1^{\mathrm{T}}, \ldots, D_p^{\mathrm{T}}\right)^{\mathrm{T}}$, where D_i is a $q_i \times r$ matrix
$\\|\cdot\\|_p$	the norm in $\mathbb{R}^n$ given by

$$\|x\|_p := \begin{cases} \left(\sum_{i=1}^n |x_i^p|\right)^{\frac{1}{p}} & \text{if } p \in [1, \infty) \\ \max_{i=1,\ldots,n} |x_i| & \text{if } p = \infty \end{cases}$$
for $\quad x = \mathbf{col}\,(x_1, \ldots, x_n)$

$\\|\cdot\\|$	the standard Eucledian norm $\\|\cdot\\| = \\|\cdot\\|_2$
L_p	the space of $p \in [1, \infty)$ power integrable vector-functions $f(\cdot)$:

$$\|f(\cdot)\|_p := \left(\int \|f(t)\|^p \, dt\right)^{\frac{1}{p}} < \infty$$

$\oplus$	the direct sum of linear subspaces; the direct sum of the empty group of subspaces is defined to be $\{0\}$
$\langle \cdot, \cdot \rangle$	the standard inner product in the Eucledian space $\mathbb{R}^n$

$$\langle x, y \rangle := \sum_{i=1}^n x_i y_i \quad \text{for} \quad \begin{array}{l} x = \mathbf{col}\,(x_1, \ldots, x_n) \\ y = \mathbf{col}\,(y_1, \ldots, y_n) \end{array}$$

$\mathrm{Lin}S$	the linear hull of a subset S of a linear space
I_m	the unit $m \times m$ matrix; the index may be dropped if m is clear from the context
$0_{m\times n}$	the zero $m \times n$ matrix
$\ker A$	the kernel of the matrix (operator) A: $\ker A := \{x : Ax = 0\}$
$\mathrm{Im}A$	the image of the matrix (operator) A: $\mathrm{Im}A := \{y : \exists x, y = Ax\}$
$A\|_L$	the restriction of the operator (matrix) A on the linear subspace L
tr A	the trace (the sum of the diagonal elements) of the square matrix A
$\det A$	the determinant of the square matrix (operator) A
$\sigma(A)$	the spectrum of A
$\sigma^+(A)$	$:= \{\lambda \in \sigma(A) : \|\lambda\| \geq 1\}$: the unstable part of the spectrum
$\sigma^-(A)$	$:= \sigma(A) \setminus \sigma^+(A)$: the stable part of the spectrum
$M_\sigma(A)$	the invariant subspace of A related to the spectrum set $\sigma \subset \sigma(A)$
$M_{\text{st}}(A)$	$:= M_{\sigma^-(A)}$: the invariant subspace related to the stable part of the spectrum
$M_{\text{unst}}(A)$	$:= M_{\sigma^+(A)}$ the invariant subspace related to the unstable part of the spectrum
$\mathbf{diag}(A_1, \ldots, A_k)$	the diagonal block-matrix with the square matrices A_i along the diagonal and zero blocks outside the diagonal
$\\|A\\|$	the norm of the matrix (operator) A: $\\|A\\| := \sup_{x\neq 0} \frac{\\|Ax\\|_*}{\\|x\\|_*} = \sup_{\\|x\\|_*=1} \\|Ax\\|_*$ where $\\|\cdot\\|_*$ is a given vector norm
$\sum_{i=m}^n Q_i := 0$	whenever $m > n$, where Q_i are elements of a common linear space, and 0 is the zero of this space
$\prod_{i=m}^n A_i := I_s$	whenever $m > n$, where A_i are $s \times s$-matrices
$\deg \varphi(\cdot)$	the degree of the polynomial $\varphi(\cdot)$
$\lesssim$	the inequality up to a polynomial factor: $f(t) \lesssim g(t) \Leftrightarrow \left\{ \begin{array}{l} \text{a polynomial } \varphi(\cdot) \text{ exists} \\ \text{such that } f(t) \leq \varphi(t)g(t)\ \forall t \end{array} \right\}$
$\eqsim$	the equality up to a polynomial factor: $f(r) \eqsim g(r) \Leftrightarrow f(r) \lesssim g(r)\ \\&\ g(r) \lesssim f(r)$

$\boldsymbol{P}$	the propability
$\boldsymbol{E}$	the mathematical expectation
$\boldsymbol{P}(f)$	$:= \boldsymbol{P}(F = f)$ for a random variable $F \in \mathfrak{F}$ and an element $f \in \mathfrak{F}$
$\left.\begin{array}{l}\boldsymbol{P}(\cdot\|F = f)\\ = \boldsymbol{P}(\cdot\|f)\end{array}\right\}$	the conditional probability given that $F = f$
$\boldsymbol{P}(E\|f) := 0$	whenever $\boldsymbol{P}(f) = 0$
$\boldsymbol{P}_G(dg)$	the probability distribution of the random variable G
$\left.\begin{array}{l}\boldsymbol{P}_G(dg\|F = f)\\ = \boldsymbol{P}_G(dg\|f)\end{array}\right\}$	the probability distribution of the random variable G given that $F = f$
$p_V(\cdot)$	the probability density of a random vector $V \in \mathbb{R}^s$
$\left.\begin{array}{l}p_V(\cdot\|F = f)\\ = p_V(\cdot\|f)\end{array}\right\}$	the probability density of a random vector $V \in \mathbb{R}^s$ given that $F = f$
I_E	the indicator of the random event E: $I_E = 1$ if E holds, and $I_E = 0$ otherwise
a.s.	”almost surely, with probability 1”
$h(V)$	the differential entropy of the random vector V

$$h(V) := -\int p_V(v) \log_2 p_V(v)\, dv$$

In conclusion, we note that the capital script letters will be mostly used to denote deterministic functions. The *measurable space* is a pair $[\mathfrak{V}, \Sigma]$, where $\mathfrak{V}$ is a set and Σ is a σ-algebra of subsets $V \subset \mathfrak{V}$.

2

Topological Entropy, Observability, Robustness, Stabilizability, and Optimal Control

2.1 Introduction

In this chapter, we study connections among observability, stabilizability, and optimal control via digital channels on the one hand, and topological entropy of the open-loop system on the other hand. The concept of entropy of dynamic systems was originated in the work of Kolmogorov [82, 83] and was inspired by the Shannon's pioneering paper [188]. Kolmogorov's work started a whole new research direction in which entropy appears as a numerical invariant of a class of deterministic dynamic systems (see also [162]). Later, Adler and his co-authors introduced topological entropy of dynamic systems [2], which is a modification of Kolmogorov's metric entropy. The paper [140] imported the concept of topological entropy into the theory of networked control systems. The concept of feedback topological entropy was introduced, and the condition of a local stabilizability of nonlinear systems via a limited capacity channel was given. In this chapter, we extend the concept of topological entropy to the case of uncertain dynamic systems with noncompact state space. Unlike [140], we use a less common "metric"definition of topological entropy introduced by Bowen (see, e.g., [26]). The "metric definition" is, in our opinion, more suitable to the theory of networked control systems. The main results of the chapter are necessary and sufficient conditions of robust observability, stabilizability, and solvability of the optimal control problem that are given in terms of inequalities between the communication channel data rate and the topological entropy of the open-loop system. The main results of the chapter were originally published in [171]. Notice that the results on stabilizability of linear plants via limited capacity communication channels were proved by Nair and Evans (see, e.g., [137, 138]).

The remainder of the chapter is organized as follows. Section 2.2 introduces the concept of observability of a nonlinear uncertain system via a digital communication channel. The definition of topological entropy and several conditions for observability in terms of topological entropy are given in Sect. 2.3. In Sect. 2.4, we calculate the topological entropy for some important classes of linear systems. Section 2.5 addresses the problem of stabilization of linear systems. The problem of linear-quadratic (LQ) optimal control via a limited capacity digital channel is solved

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_2,

in Sect. 2.6. Finally, Section 2.7 presents the proofs of some results from Sects. 2.4, 2.5, and 2.6.

2.2 Observability via Communication Channels

In this section, we consider a nonlinear, uncertain, discrete-time dynamic system of the form:

$$x(t+1) = F(x(t), \omega(t)), \quad x(1) \in \mathfrak{X}_1, \quad x(t) \in \mathfrak{X}, \tag{2.2.1}$$

where $t = 1, 2, 3, \ldots$, $x(t) \in \mathbb{R}^n$ is the state; $\omega(t) \in \Omega$ is the uncertainty input; $\mathfrak{X} \subset \mathbb{R}^n$ is a given set; $\mathfrak{X}_1 \subset \mathfrak{X}$ is a given nonempty compact set; and $\Omega \subset \mathbb{R}^m$ is a given set. Notice that we do not assume that the function $F(\cdot, \cdot)$ is continuous.

In our observability problem, a sensor measures the state $x(t)$ and is connected to the controller that is at the remote location. Moreover, the only way of communicating information from the sensor to that remote location is via a digital communication channel that carries one discrete-valued symbol $h(jT)$ at time jT, selected from a coding alphabet $\mathfrak{H}$ of size l. Here $T \geq 1$ is a given integer period, and $j = 1, 2, 3, \ldots$.

This restricted number l of codewords $h(jT)$ is determined by the transmission data rate of the channel. For example, if μ is the number of bits that our channel can transmit, then $l = 2^\mu$ is the number of admissible codewords. We assume that the channel is a perfect noiseless channel and that there is no time delay. Let $R \geq 0$ be a given constant. We consider the class $\mathfrak{C}_R$ of such channels with any period T satisfying the following transmission data rate constraint:

$$\frac{\log_2 l}{T} \leq R. \tag{2.2.2}$$

The rate $R = 0$ corresponds to the case when the channel does not transmit data at all.

We consider the problem of estimation of the state $x(t)$ via a digital communication channel with a bit-rate constraint. Our state estimator consists of two components. The first component is developed at the measurement location by taking the measured state $x(\cdot)$ and coding to the codeword $h(jT)$. This component will be called a "coder." Then the codeword $h(jT)$ is transmitted via a limited capacity communication channel to the second component, which is called a "decoder." The second component developed at the remote location takes the codeword $h(jT)$ and produces the estimated states $\hat{x}((j-1)T+1), \ldots, \hat{x}(jT-1), \hat{x}(jT)$. This situation is illustrated in Fig. 2.1 (where $y \equiv x$ now).

The coder and the decoder are of the following forms, respectively:

$$h(jT) = \mathcal{F}_j\left(x(\cdot)|_1^{jT}\right); \tag{2.2.3}$$

$$\begin{pmatrix} \hat{x}((j-1)T+1) \\ \vdots \\ \hat{x}(jT-1) \\ \hat{x}(jT) \end{pmatrix} = \mathcal{G}_j\left[h(T), h(2T), ..., h((j-1)T), h(jT)\right]. \tag{2.2.4}$$

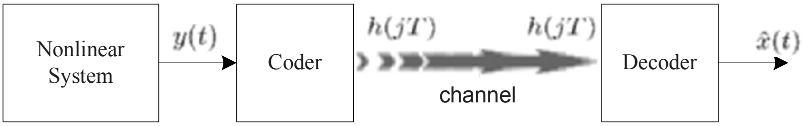


Fig. 2.1. State estimation via digital communication channel

Here $j = 1, 2, 3, \ldots$.

We recall that for a vector $x = \mathbf{col}\begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}$ from $\mathbb{R}^n$,

$$\|x\|_\infty := \max_{j=1,\ldots,n} |x_j|. \tag{2.2.5}$$

Furthermore, $\|\cdot\|$ denotes the standard Euclidean vector norm:

$$\|x\| := \sqrt{\sum_{j=1}^{n} x_j^2}.$$

Definition 2.2.1. *The system* (2.2.1) *is said to be observable in the communication channel class* $\mathfrak{C}_R$ *if for any* $\epsilon > 0$*, a period* $T \geq 1$ *and a coder–decoder pair* (2.2.3), (2.2.4) *with a coding alphabet of size* l *satisfying the constraint* (2.2.2) *exist such that*

$$\|x(t) - \hat{x}(t)\|_\infty < \epsilon \quad \forall t = 1, 2, 3, \ldots \tag{2.2.6}$$

for any solution of (2.2.1).

2.3 Topological Entropy and Observability of Uncertain Systems

In this section, we introduce the concept of topological entropy for the system (2.2.1). In general, we follow the scheme of [154]; however, unlike [154], we consider uncertain dynamic systems.

Notation 2.3.1. *For any* $k \geq 1$*, let* $\mathfrak{X}_k := \{x(1), x(2), \ldots, x(k)\}$ *be the set of solutions of* (2.2.1) *with uncertainty inputs from* Ω.

Definition 2.3.2. *Consider the system* (2.2.1)*. For* $k \geq 1$ *and* $\epsilon > 0$ *we call a finite set* $Q \subset \mathfrak{X}_k$ *an* $(k, \epsilon)-$*spanning set if for any* $x_a(\cdot) \in \mathfrak{X}_k$*, an element* $x_b(\cdot) \in Q$ *exists such that* $\|x_a(t) - x_b(t)\|_\infty < \epsilon$ *for all* $t = 1, 2, \ldots, k$*. If at least one finite* $(k, \epsilon)-$*spanning set exists, then* $q(k, \epsilon)$ *denotes the least cardinality of any* $(k, \epsilon)-$*spanning set. If a finite* $(k, \epsilon)-$*spanning set does not exist, then* $q(k, \epsilon) := \infty$.

Now we are in a position to give a definition of topological entropy for the uncertain dynamic system (2.2.1).

Definition 2.3.3. *The quantity*

$$H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega) := \lim_{\epsilon\to 0}\limsup_{k\to\infty}\frac{1}{k}\log_2(q(k,\epsilon)) \tag{2.3.1}$$

is called the topological entropy of the uncertain system (2.2.1)*.*

Remark 2.3.4. Notice that the topological entropy may be equal to infinity. In the case of a system without uncertainty with continuous $F(\cdot,\cdot)$ and compact $\mathfrak{X}$, the topological entropy is always finite [154].

Remark 2.3.5. We use Bowen's "metric" definition of topological entropy that is different from the more common "topological" definition (see, e.g., p. 20 of [154]). In the case of a continuous system without uncertainty, both definitions are equivalent [154].

Now we are in a position to present the main result of this section.

Theorem 2.3.6. *Consider the system* (2.2.1)*, and assume that $\mathfrak{X} = \mathfrak{X}_1$ (hence, $\mathfrak{X}$ is compact). Let $R \geq 0$ be a given constant. Then the following two statements hold:*

(i) If $R < H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega)$, then the system (2.2.1) *is not observable in the communication channel class $\mathfrak{C}_R$;*
(ii) If $R > H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega)$, then the system (2.2.1) *is observable in the communication channel class $\mathfrak{C}_R$.*

In order to prove Theorem 2.3.6, we will need the following definition and lemma.

Definition 2.3.7. *Consider the system* (2.2.1)*. For $k \geq 1$ and $\epsilon > 0$, we call a finite set $S \subset \mathfrak{X}_k$ an $(k,\epsilon)-$separated set if for distinct points $x_a(\cdot), x_b(\cdot) \in S$, we have that $\|x_a(t) - x_b(t)\|_\infty \geq \epsilon$ for some $t = 1, 2, \ldots, k$. Let $s(k,\epsilon)$ denote the least upper bound of the cardinality of all $(k,\epsilon)-$separated sets. Notice that $s(k,\epsilon)$ may be equal to infinity.*

Lemma 2.3.8. *For any system* (2.2.1)*,*

$$\lim_{\epsilon\to 0}\limsup_{k\to\infty}\frac{1}{k}\log_2(s(k,\epsilon)) = H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega). \tag{2.3.2}$$

Proof of Lemma 2.3.8. We first observe that $s(k,\epsilon) \geq q(k,\epsilon)$. Indeed, if $s(k,\epsilon) = \infty$, then this inequality always holds. If $s(k,\epsilon) < \infty$, then a finite $(k,\epsilon)-$separated set S of maximal cardinality exists and any such set must also be an $(k,\epsilon)-$spanning set. Furthermore, we prove that $s(k,2\epsilon) \leq q(k,\epsilon)$. Indeed, if $q(k,\epsilon) = \infty$, then this inequality obviously holds. If $q(k,\epsilon) < \infty$, then a finite $(k,\epsilon)-$spanning set Q of cardinality $q(k,\epsilon)$ exists. Let S be any $(k,2\epsilon)-$separated finite set and s be its cardinality. We take s open balls of radius ϵ centered at the points of this $(k,2\epsilon)-$separated set S. Then all these open balls do not intersect with each other. On the other hand, each of these balls must contain an element of the $(k,\epsilon)-$spanning set Q. Since the balls do not intersect, we have $s \leq q(k,\epsilon)$. It means we have proved that $q(k,\epsilon)$ is no less than the cardinality of any $(k,2\epsilon)-$separated set. Therefore,$s(k,2\epsilon) \leq q(k,\epsilon)$. We have proved that

$$s(k,2\epsilon) \leq q(k,\epsilon) \leq s(k,\epsilon).$$

This obviously implies that

$$\lim_{\epsilon\to 0} \limsup_{k\to\infty} \frac{1}{k} \log_2(s(k,\epsilon)) = \lim_{\epsilon\to 0} \limsup_{k\to\infty} \frac{1}{k} \log_2(q(k,\epsilon)).$$

Now the statement of the lemma immediately follows from the definition of the topological entropy (2.3.1). □

Remark 2.3.9. Notice that $q(k,\epsilon)$ and $s(k,\epsilon)$ increase with decreasing ϵ. Therefore, the corresponding limits $\lim_{\epsilon\to 0}$ in (2.3.1) and (2.3.2) may be replaced by $\sup_{\epsilon>0}$.

Proof of Theorem 2.3.6. Statement (i). We prove this statement by contradiction. Indeed, assume that the system is observable in the communication channel class $\mathfrak{C}_R$ with $R < H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega)$. Let α be any number such that $R < \alpha < H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega)$. Then, it follows from Lemma 2.3.8 that a constant $\epsilon > 0$ exists such that

$$\limsup_{k\to\infty} \frac{1}{k} \log_2(s(k,2\epsilon)) > \alpha. \tag{2.3.3}$$

Consider a coder–decoder pair (2.2.3), (2.2.4) such that the condition (2.2.6) holds, and let $T > 0$ be its period. The inequality (2.3.3) implies that an integer $k > 0$ and an $(k,2\epsilon)-$separated set S of cardinality N exist such that

$$\frac{\log_2 N}{k} > \alpha \tag{2.3.4}$$

and

$$\frac{k}{k+T} > \frac{R}{\alpha}. \tag{2.3.5}$$

Notice that inequality (2.3.5) is satisfied for all large enough k. Let $j > 0$ be the integer such that

$$(j-1)T \leq k < jT. \tag{2.3.6}$$

Furthermore, let $\tilde{S}$ be any set of solutions on the time interval $t = 1,2,\ldots,jT$ coinciding with S for $t = 1,2,\ldots,k$. Then $\tilde{S}$ is obviously an $(k,2\epsilon)-$separated set of cardinality N. We now prove that

$$\frac{\log_2 N}{jT} > R. \tag{2.3.7}$$

Indeed, from (2.3.4)–(2.3.6), we obtain

$$\frac{\log_2 N}{jT} = \frac{\log_2 N}{k} \frac{k}{jT} > \alpha \frac{k}{jT} > \alpha \frac{k}{k+T} > R.$$

Furthermore, let $\hat{S}$ be the set of all sequences $\hat{x}(1), \hat{x}(2), \ldots, \hat{x}(jT)$ produced by (2.2.3), (2.2.4). Then, the cardinality of $\hat{S}$ does not exceed l^j. Since $\frac{\log_2 N}{jT} > R$, condition (2.2.2) implies that $l^j < N$. Because condition (2.2.6) must be satisfied for any solution of (2.2.1) with some $\hat{x}(1), \hat{x}(2), \ldots, \hat{x}(jT) \in \hat{S}$, this implies that two elements $x_a(\cdot), x_b(\cdot)$ of $\tilde{S}$ and an element $\hat{x}(\cdot)$ of $\hat{S}$ exist such that condition (2.2.6) holds with $x(\cdot) = x_a(\cdot)$ and $x(\cdot) = x_b(\cdot)$. This implies that $\|x_a(t) - x_b(t)\|_\infty < 2\epsilon$ for all $t = 1, 2, \ldots, jT$. However, the latter inequality contradicts to our assumption that the set $\tilde{S}$ is $(jT, 2\epsilon)-$separated. This completes the proof of this part of the theorem.

Statement (ii). If the inequality $R > H(F(\cdot,\cdot), \mathfrak{X}_1, \mathfrak{X}, \Omega)$ holds, then for any $\epsilon > 0$, an integer $k > 1$ and an $(k, \epsilon)-$spanning set Q of cardinality N exist such that $\frac{\log_2 N}{k} \leq R$. Now introduce a coder–decoder pair of the form (2.2.3), (2.2.4) with $T = k$ and $l = N$ as follows. Because Q is an $(k, \epsilon)-$spanning set, for any solution $x(\cdot)$ of (2.2.1), an element $x_a^{(1)}(\cdot)$ of Q exists such that $\|x_a^{(1)}(t) - x(t)\|_\infty < \epsilon$ for all $t = 1, 2, \ldots, k$. Furthermore, because the system (2.2.1) is time-invariant and $\mathfrak{X} = \mathfrak{X}_1$, for any solution $x(\cdot)$ of (2.2.1) and any $j = 1, 2, \ldots$, an element $x_b^{(j)}(\cdot)$ of Q exists such that

$$\|x_b^{(j)}(t) - x(t)\|_\infty < \epsilon$$
$$\forall t = (j-1)k + 1, (j-1)k + 2, \ldots, jk. \tag{2.3.8}$$

Let $f_j(x(\cdot))$ be the index of this element $x_b^{(j)}$ in Q. Now introduce the following coder–decoder pair

$$h(jk) := f_j(x(\cdot)); \tag{2.3.9}$$

$$\begin{pmatrix} \hat{x}((j-1)k+1) \\ \vdots \\ \hat{x}(jk-1) \\ \hat{x}(jk) \end{pmatrix} := \begin{pmatrix} x_b^{(j)}(1) \\ \vdots \\ x_b^{(j)}(k-1) \\ x_b^{(j)}(k) \end{pmatrix}. \tag{2.3.10}$$

It follows immediately from (2.3.8) that the condition (2.2.6) holds. Furthermore, by construction, the coder–decoder pair satisfies the communication constraint (2.2.2). This completes the proof of the theorem. □

Remark 2.3.10. Theorem 2.3.6 gives an "almost" necessary and sufficient condition for observability in the communication channel class $\mathfrak{C}_R$. Notice that in the critical case $R = H(F(\cdot,\cdot), \mathfrak{X}_1, \mathfrak{X}, \Omega)$, both possibilities can occur. Indeed, let $F(x, \cdot) \equiv x$ for any x; then it is obvious that $H(F(\cdot,\cdot), \mathfrak{X}_1, \mathfrak{X}, \Omega) = 0$. However, if $\mathfrak{X}_1 = \mathfrak{X} = \{x_0\}$, then the corresponding system is observable in the communication channel class $\mathfrak{C}_0$. On the other hand, if $\mathfrak{X}_1 = \mathfrak{X} = \{x_0, x_1\}$ where $x_0 \neq x_1$, then the corresponding system is not observable in the communication channel class $\mathfrak{C}_0$.

Definition 2.3.11. *The system* (2.2.1) *is said to be robustly stable if for any* $\epsilon > 0$, *an integer* $k \geq 1$ *exists such that*

$$\|x(t)\|_\infty < \epsilon \quad \forall t \geq k \tag{2.3.11}$$

for any solution $x(\cdot)$ *of the system* (2.2.1)*.*

Proposition 2.3.12. *Consider the system* (2.2.1)*, and assume that* Ω *is compact,* $F(\cdot,\cdot)$ *is continuous, and the system* (2.2.1) *is robustly stable. Then*

$$H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega) = 0.$$

Proof of Proposition 2.3.12. Let $\epsilon > 0$ be given and $k \geq 1$ be an integer such that (2.3.11) holds. Since $\mathfrak{X}_1, \Omega$ are compact and $F(\cdot,\cdot)$ is continuous, a finite $(k,\epsilon)-$spanning set Q exists. Let N be the cardinality of Q. For any $j > k$, introduce the set Q_j by extension of solutions of (2.2.1) from Q with arbitrary $\omega(t) \in \Omega$ for $t \geq k$. The cardinality of Q_j is N for any j. The condition (2.3.11) obviously implies that

$$\|x(t) - x_b(t)\|_\infty < 2\epsilon \quad \forall t \geq k$$

for any solution $x(\cdot)$ of (2.2.1), any j, and any $x_b(\cdot)$ from Q_j. Therefore, for any j, Q_j is a $(j,2\epsilon)-$spanning set. Hence, $q(j,2\epsilon) \leq N$ for any j, and by Definition 2.3.1, $H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega) = 0$. This completes the proof of the proposition. □

Definition 2.3.13. *Let* $x(\cdot)$ *be a solution of* (2.2.1)*. The system* (2.2.1) *is said to be locally reachable along the trajectory* $x(\cdot)$ *if a constant* $\delta > 0$ *and an integer* $N \geq 1$ *exist such that for any* $k \geq 1$ *and any* $a, b \in \mathbb{R}^n$ *such that*

$$\|x(k) - a\| \leq \delta\|x(k)\|, \quad \|x(k+N) - b\| \leq \delta\|x(k+N)\|,$$

a solution $\tilde{x}(\cdot)$ *of* (2.2.1) *exists with*

$$\tilde{x}(k) = a, \quad \tilde{x}(k+N) = b.$$

Definition 2.3.14. *A solution* $x(\cdot)$ *of* (2.2.1) *is said to be separated from the origin, if a constant* $\delta_0 > 0$ *exists such that*

$$\|x(t)\| \geq \delta_0 \quad \forall t \geq 1.$$

We will use the following assumptions.

Assumption 2.3.15. *The system* (2.2.1) *is locally reachable along a trajectory separated from the origin.*

Assumption 2.3.16. *The system* (2.2.1) *is locally reachable along a trajectory* $x(\cdot)$ *such that* $\|x(t)\|_\infty \to \infty$ *as* $t \to \infty$*.*

Theorem 2.3.17. *Consider the system* (2.2.1)*. The following two statements hold:*

(i) If Assumption 2.3.15 *is satisfied, then* $H(F(\cdot,\cdot),\mathfrak{X}_1,\mathfrak{X},\Omega) = \infty$*; hence, according to Theorem* 2.3.6*, the system* (2.2.1) *is not observable in the communication channel class* $\mathfrak{C}_R$ *with any* R*;*

(ii) If Assumption 2.3.16 *is satisfied, then for any coder–decoder pair of the form* (2.2.3)*,* (2.2.4) *with any* T, R

$$\sup_{t,x(\cdot)} \|x(t) - \hat{x}(t)\|_\infty = \infty,$$

where the supremum is taken over all times t *and all solutions* $x(\cdot)$ *of the system* (2.2.1)*.*

Proof of Theorem 2.3.17. Statement (i). Suppose that Assumption 2.3.15 holds, and let N be the integer from Definition 2.3.13. It follows from Assumption 2.3.15 that the system has a trajectory $x(\cdot)$ and a vector $a_0 \neq 0$ exists such that the following property holds: For any $c_1, c_2, c_3, \dots$ from the interval $[0, 1]$, a solution $\tilde{x}(\cdot)$ of (2.2.1) exists with the property $\tilde{x}(jN) = x(jN) + c_j a_0$ for $j = 1, 2, 3, \dots$. Now for any $M \geq 1$ and $j \geq 1$, consider the set S_M^j of such solutions over the interval $t = 1, 2, \dots, jN$ with c_i taking discrete values $0, \frac{1}{M}, \dots, \frac{M-1}{M}$. Then S_M^j is $\left(jN, \frac{\|a_0\|_\infty}{M}\right)$–separated and the cardinality of S_M^j is M^j. Hence, by Lemma 2.3.8,

$$\begin{aligned} H(F(\cdot,\cdot), \mathfrak{X}_1, \mathfrak{X}, \Omega) &\geq \lim_{M\to\infty} \limsup_{j\to\infty} \frac{1}{jN} \log_2(M^j) \\ &= \lim_{M\to\infty} \frac{1}{N} \log_2(M) = \infty. \end{aligned}$$

This completes the proof of this part of the theorem.

Statement (ii). Suppose that Assumption 2.3.16 holds, and let N be the integer from Definition 2.3.13. Furthermore, consider a coder–decoder pair of the form (2.2.3), (2.2.4) with some parameters l, T. It follows from Assumption 2.3.16 that the system has a trajectory $x(\cdot)$ and a vector $a_0 \neq 0$ and a sequence $\{d_j\}$ where $d_j > 0, d_{j+1} \geq d_j, \lim_{j\to\infty} d_j = \infty$ exist such that the following property holds: For any sequence $\{c_j\}$ where $c_j \in [0, d_j]$, a solution $\tilde{x}(\cdot)$ of (2.2.1) exists with the property $\tilde{x}(jNT) = x(jNT) + c_j a_0$ for $j = 1, 2, 3, \dots$. Now for some $M > l^N$ and any $i \geq 1$, consider the set S_M^i of such solutions over the interval $t = iNT, iNT+1, \dots, i^2NT$, with c_j taking discrete values $0, \frac{1}{M}d_j, \dots, \frac{M-1}{M}d_j$. Then S_M^i is $\left(i^2NT, \frac{d_i\|a_0\|_\infty}{M}\right)$–separated and the cardinality of S_M^i is $M^{(i^2-i)}$. On the other hand, by the time i^2NT, the channel can transmit l^{i^2N} various symbolic sequences. Since $M > l^N$, $M^{(i^2-i)} > l^{i^2N}$ for large i. It means that for large i, there will be at least two distinct elements of S_M^i coded by the same symbolic sequences. Furthermore, the set S_M^i is $\left(i^2NT, \frac{d_i\|a_0\|_\infty}{M}\right)$–separated and $d_i \to \infty$ as $i \to \infty$. The statement (ii) follows immediately from this. □

Remark 2.3.18. Notice that it is not surprising in Theorem 2.3.17 that topological entropy is infinite for a large class of systems with disturbances, since the number of unknowns in the system, i.e., initial state and disturbances, grows to infinity with time.

2.4 The Case of Linear Systems

In this section, we first consider a linear system without uncertainty:

$$x(t+1) = Ax(t), \quad x(1) \in \mathfrak{X}_1, \tag{2.4.1}$$

where $t = 1, 2, 3, \ldots$, $x(t) \in \mathbb{R}^n$ is the state, $\mathfrak{X}_1 \subset \mathbb{R}^n$ is a given compact set, and A is a given square matrix.

We will suppose that the following assumption holds.

Assumption 2.4.1. *The origin is an interior point of the set $\mathfrak{X}_1$: A constant $\delta > 0$ exists such that*

$$\|a\|_\infty < \delta \quad \Rightarrow a \in \mathfrak{X}_1.$$

Furthermore, let $S(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be the set of eigenvalues of the matrix A. Introduce the following value:

$$H(A) := \sum_{\lambda_i \in S(A)} \log_2(\max\{1, |\lambda_i|\}). \tag{2.4.2}$$

Topological entropy of linear systems without uncertainty is described by the following theorem.

Theorem 2.4.2. *Consider the system* (2.4.1)*, and suppose that Assumption* 2.4.1 *holds. Then, the topological entropy of the system* (2.4.1) *is equal to $H(A)$ where $H(A)$ is defined by* (2.4.2)*.*

To prove Theorem 2.4.2, introduce the following linear system:

$$x(t+1) = Ax(t), \quad x(1) \in \mathbb{R}^n. \tag{2.4.3}$$

The only difference between this system and the system (2.4.1) is that in (2.4.3), initial conditions $x(1)$ take values in the whole space $\mathbb{R}^n$.

We will need the following lemma.

Lemma 2.4.3. *Consider the system* (2.4.1)*. Let α be a given constant such that $\alpha > H(A)$, where $H(A)$ is defined by* (2.4.2)*, and let $\beta \geq 0$ be a given constant. Then for any $\epsilon > 0$, an integer k_0 exists such that for any $k \geq k_0$, a set $Q_{\epsilon,k} = \{x_1(\cdot), \ldots, x_N(\cdot)\}$ of N solutions of the system* (2.4.3) *for $t = 1, 2, \ldots, k$ exists with the properties $\frac{1}{k}\log_2(N) \leq \alpha$, and for any $x_a(\cdot) \in \mathfrak{X}_k$, an element $x^{(k)}(\cdot) \in Q_{\epsilon,k}$ exists such that*

$$\beta \sum_{t=1}^{k} \|x_a(t) - x^{(k)}(t))\|^2 + \max_{t=1,\ldots,k} \|x_a(t) - x^{(k)}(t)\|_\infty < \epsilon. \tag{2.4.4}$$

The proof of Lemma 2.4.3 is given in Sect. 2.7.

Proof of Theorem 2.4.2. Let H denote the entropy of the system (2.4.1). We prove that $H = H(A)$. First prove that if $\alpha > H(A)$, then $H \leq \alpha$. Lemma 2.4.3 with $\beta = 0$ immediately implies the existence of an $(k, \epsilon)-$spanning set of cardinality N where $\frac{1}{k}\log_2(N) \leq \alpha$ for any $\epsilon > 0$ and all large enough k. Now Definition 2.3.3 of the topological entropy implies that $H \leq \alpha$. Since α is any number that is greater than $H(A)$, we have proved that $H \leq H(A)$.

Now we prove by contradiction that $H \geq H(A)$. Indeed, assume that this is not true and $H < H(A)$. This inequality can hold only for positive $H(A)$. Therefore, the matrix A has at least one eigenvalue λ with $|\lambda| > 1$. By a linear change of state variable, the matrix A can be transformed to the form:

$$A = \begin{pmatrix} A^+ & 0 \\ 0 & A^- \end{pmatrix},$$

where all eigenvalues of A^+ lie outside of the closed unit disk, and all eigenvalues of A^- lie inside of the closed unit disk. Then $H(A^+) = H(A)$. Furthermore, let $x = \textbf{col}\ (x^+, x^-)$ be the corresponding partitioning of the state vector x, $m^+ > 0$ be the dimension of x^+, and $\mathfrak{X}_1^+ := \mathfrak{X}_1 \cap \{\textbf{col}\ (x^+, 0)\}$. Assumption 2.4.1 implies that a convex set $\mathfrak{Y} \subset \mathfrak{X}_1^+$ exists with $V_m(\mathfrak{Y}) > 0$, where $m := m^+$ and $V_m(\cdot)$ denotes the m-dimensional volume of a set. Moreover,

$$\begin{gathered} V_m(A^k\mathfrak{Y}) = V_m(A^{+k}\mathfrak{Y}) = \\ |detA^+|^k V_m(\mathfrak{Y}) = 2^{kH(A)}V_m(\mathfrak{Y}). \end{gathered} \tag{2.4.5}$$

On the other hand, the assumption $H < H(A)$ and Definition 2.3.3 of the topological entropy imply that an $\epsilon > 0$, a constant $\tilde{H}$, a time sequence $k_i \to \infty$, and a sequence of $(k_i, \epsilon)-$spanning sets of cardinality N_i exist such that

$$\frac{1}{k_i}\log_2(N_i) \leq \tilde{H} < H(A). \tag{2.4.6}$$

It follows from (2.4.5) and (2.4.6) that

$$\lim_{i\to\infty} \frac{1}{N_i} V_m(A^{k_i-1}\mathfrak{Y}) = \infty. \tag{2.4.7}$$

On the other hand, because we have an $(k_i, \epsilon)-$spanning set of cardinality N_i, any element of the set $A^{k_i-1}\mathfrak{Y}$ must belong to one of N_i balls with radius ϵ in $\|\cdot\|_\infty$ metric. This obviously contradicts to (2.4.7). This completes the proof of Theorem 2.4.2. □

The following corollary immediately follows from Theorem 2.4.2.

Corollary 2.4.4. *Consider the system* (2.4.1)*, and suppose that Assumption* 2.4.1 *holds. Then, the topological entropy of the system* (2.4.1) *is equal to* 0 *if and only if* $|\lambda| \leq 1$ *for any eigenvalue* λ *of the matrix A.*

Remark 2.4.5. Theorem 2.4.2 together with Theorem 2.3.6 give an "almost" necessary and sufficient condition for observability of the system (2.4.1) in the communication channel class $\mathfrak{C}_R$.

Remark 2.4.6. Another proof of Theorem 2.4.2 is given in [216] (see Theorem 7.12). Notice that, in fact, results similar to Theorem 2.4.2, but stated in different terms, were derived in [115,137,140]. Also, Theorem 2.4.2 is a reminder of the well-known result on topological entropy of algebraic automorphisms of torus; see, e.g., [3].

Now consider a linear, uncertain, discrete-time dynamic system of the form:

$$x(t+1) = [A + B\omega(t)]\, x(t), \quad x(1) \in \mathfrak{X}_1, \tag{2.4.8}$$

where $t = 1, 2, 3, \ldots$, $x(t) \in \mathbb{R}^n$ is the state, $\omega(t) \in \Omega$ is the uncertainty matrix, $\mathfrak{X}_1 \subset \mathbb{R}^n$ is a given compact set, $\Omega \subset \mathbb{R}^{r\times n}$ is a given set, and A, B are given matrices of corresponding dimensions.

We suppose that the following assumptions hold.

Assumption 2.4.7. *The matrix A has at least one eigenvalue λ outside the unit circle: $|\lambda| > 1$.*

Assumption 2.4.8. *The pair (A, B) is reachable (see, e.g., [12], p.94).*

Assumption 2.4.9. *The origin is an interior point of the set Ω: a $\delta > 0$ exits such that*

$$\|\omega\|_\infty < \delta \quad \Rightarrow \omega \in \Omega.$$

Here $\|\cdot\|_\infty$ is the induced matrix norm related to the vector norm (2.2.5)*.*

Now we are in a position to present the following corollary of Theorem 2.3.17.

Proposition 2.4.10. *Consider the system* (2.4.8)*. If Assumptions* 2.4.1 *and* 2.4.7*–*2.4.9 *hold, then for any coder–decoder pair of the form* (2.2.3)*,* (2.2.4) *with any T, R*

$$\sup_{t, x(\cdot)} \|x(t) - \hat{x}(t)\|_\infty = \infty,$$

where the supremum is taken over all times t and all solutions $x(\cdot)$ of the system (2.4.8)*.*

Remark 2.4.11. Proposition 2.4.10 shows that any state estimator with bit rate constraints for a linear unstable system is not robust. For example, all estimators from [204] will produce infinite error under any small parametric perturbation of the matrix A.

Proof of Proposition 2.4.10. This statement follows from the statement (ii) of Theorem 2.3.17. Assumptions 2.4.1 and 2.4.7 imply that the system (2.4.8) has a solution $x(\cdot)$ such that $\|x(t)\|_\infty$ tends to infinity as t tends to infinity and Assumptions 2.4.8 and 2.4.9 imply that the system is locally reachable along any its trajectory that tends to infinity. This completes the proof of Proposition 2.4.10. □

As an illustrative example, consider a linear, uncertain, discrete-time dynamical system of the form:

$$x(t+1) = [A + B\omega(t)]\, x(t) + b, \quad x(1) \in \mathfrak{X}_1, \tag{2.4.9}$$

where $t = 1, 2, 3, \ldots$, $x(t) \in \mathbb{R}^n$ is the state, $\omega(t) \in \Omega$ is the uncertainty matrix, $\mathfrak{X}_1 \subset \mathbb{R}^n$ is a given compact set, $b \in \mathbb{R}^n$ is a given vector, and A, B are given matrices of corresponding dimensions.

We will need the following assumptions.

Assumption 2.4.12. *The uncertainty $\omega(t)$ satisfies the standard norm bound condition (see, e.g., [148, 151, 174, 178, 180]): $\|w(t)\| \leq c$, where $c > 0$ is a given constant.*

Assumption 2.4.13. *The matrix A is stable; i.e., $|\lambda| < 1$ for any its eigenvalue λ.*

Assumption 2.4.14. *The following frequency domain condition holds:*

$$\max_{z \in \mathbb{C}: |z|=1} \|(zI - A)^{-1} B\| < \frac{1}{c}. \tag{2.4.10}$$

Proposition 2.4.15. *Consider the uncertain system* (2.4.9)*, and let $c > 0$ be a given constant. Suppose that Assumptions* 2.4.8 *and* 2.4.12–2.4.14 *are satisfied. Then the following two statements hold:*

(i) If $b = 0$, then the topological entropy of the system (2.4.9) *is equal to* 0*;*
(ii) If $b \neq 0$, then the topological entropy of the system (2.4.9) *is equal to* ∞.

Proof of Proposition 2.4.15. Statement (i). According to a discrete-time analog of the circle stability criterion from the theory of absolute stability (see, e.g., [144]), the frequency domain inequality (2.4.10) implies that the uncertain system (2.4.9) with $b = 0$ and the uncertainty satisfying Assumption 2.4.12 is robustly stable. Therefore, this part of Proposition 2.4.15 follows from Proposition 2.3.12.

Statement (ii). It is obvious that if $b \neq 0$ then any trajectory of the system (2.4.9) is separated from the origin. Furthermore, Assumption 2.4.9 immediately follows from Assumption 2.4.12. Assumptions 2.4.9 and 2.4.8 imply that the system is locally reachable along any trajectory separated from the origin. Therefore, Assumption 2.3.15 holds. Now the statement (ii) follows from Theorem 2.3.17. This completes the proof of this proposition. □

2.5 Stabilization via Communication Channels

In this section, we consider a linear, discrete-time controlled system without uncertainty of the form:

$$x(t+1) = Ax(t) + Bu(t), \quad x(1) \in \mathfrak{X}_1, \tag{2.5.1}$$

where $t = 1, 2, 3, \ldots, x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control input, $\mathfrak{X}_1 \subset \mathbb{R}^n$ is a given compact set, and A, B are given matrices of corresponding dimensions.

We consider the problem of stabilization of the linear system (2.5.1) via a digital communication channel with a bit-rate constraint. Our controller consists of two components. The first component is developed at the measurement location by taking the measured state $x(\cdot)$ and coding to the codeword $h(jT + 1)$. This component will be called "coder." Then the codeword $h(jT + 1)$ is transmitted via a limited capacity communication channel to the second component, which is called a "decoder-controller." The second component developed at a remote location takes the codeword $h(jT + 1)$ and produces the sequence of control inputs $u(jT+1), \ldots, u((j+1)T-1), u((j+1)T)$. This situation is illustrated in Fig. 2.2.

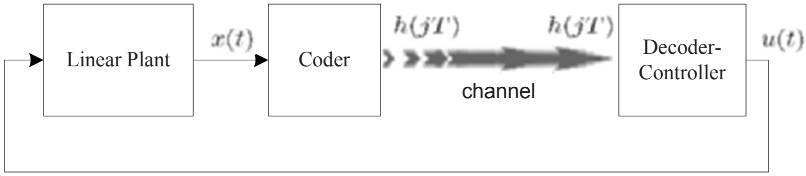


Fig. 2.2. Control via digital communication channel

Our digital communication channel carries one discrete-valued symbol $h(jT+1)$ at time $jT + 1$, selected from a coding alphabet $\mathfrak{H}$ of size l. Here $T \geq 1$ is a given integer period, and $j = 0, 1, 2, \ldots$.

This restricted number l of codewords $h(jT + 1)$ is determined by the transmission data rate of the channel. Let $R \geq 0$ be a given constant. We consider the class $\mathfrak{C}_R$ of such channels with any period T satisfying the transmission data rate constraint (2.2.2).

The coder and the decoder-controller are of the following form:

$$h(jT + 1) = \mathcal{F}_j \left(x(\cdot)|_1^{jT+1} \right); \tag{2.5.2}$$

$$\begin{pmatrix} u(jT+1) \\ \vdots \\ u((j+1)T-1) \\ u((j+1)T) \end{pmatrix} = \mathcal{U}_j \left(h_1, h_{T+1}, ..., h_{jT+1} \right). \tag{2.5.3}$$

We will need the following standard assumption.

Assumption 2.5.1. *The pair* (A, B) *is stabilizable (see, e.g., [63]).*

Definition 2.5.2. *The linear system* (2.5.1) *is said to be stabilizable in the communication channel class* $\mathfrak{C}_R$ *if a period* $T \geq 1$ *and a coder–decoder-controller pair*

(2.5.2), (2.5.3) *with a coding alphabet of size* l *satisfying the constraint* (2.2.2) *exist such that the closed-loop system* (2.5.1), (2.5.2), (2.5.3) *is stable in the following sense: For any* $\epsilon_0 > 0$, *an integer* $k \geq 1$ *exists such that*

$$\|x(t)\|_\infty < \epsilon_0 \quad \forall t \geq k \tag{2.5.4}$$

for any solution $[x(\cdot), u(\cdot)]$ *of the closed-loop system with initial condition* $x(1) \in \mathfrak{X}_1$.

Now we are in a position to present the main result of this section.

Theorem 2.5.3. *Consider the system* (2.5.1). *Let* $R \geq 0$ *be a given constant and* $H(A)$ *be the value* (2.4.2). *Suppose that Assumptions* 2.4.1 *and* 2.5.1 *are satisfied. Then the following two statements hold:*

(i) If $R < H(A)$, *then the system* (2.5.1) *is not stabilizable in the communication channel class* $\mathfrak{C}_R$;
(ii) If $R > H(A)$, *then the system* (2.5.1) *is stabilizable in the communication channel class* $\mathfrak{C}_R$.

The proof of Theorem 2.5.3 is given in Sect. 2.7.

Remark 2.5.4. Theorem 2.5.3 gives an “almost” necessary and sufficient condition for stabilizability of linear systems in the communication channel class $\mathfrak{C}_R$. Notice that in the critical case $R = H(A)$, both possibilities can occur. Indeed, first consider the system (2.5.1) with a stable matrix A. In this case, $H(A) = 0$ and the system is stabilizable in the communication channel class $\mathfrak{C}_0$ because the open-loop system is stable.

On the other hand, if we take the system (2.5.1) with a matrix A with all its eigenvalues on the unit circle, then $H(A) = 0$ according to Corollary 2.4.4. However, the system is obviously not stabilizable in the communication channel class $\mathfrak{C}_0$.

2.6 Optimal Control via Communication Channels

In this section, we address the problem of optimal control of the linear system (2.5.1) via a digital communication channel with a bit-rate constraint.

We will consider the following quadratic cost function associated with the linear system (2.5.1) and initial condition $x(1)$:

$$J_{1,\infty}[x(1)] := \sum_{t=1}^{+\infty} [x(t)^{\mathrm{T}} C^{\mathrm{T}} C x(t) + u(t)^{\mathrm{T}} G u(t)], \tag{2.6.1}$$

where C and $G = G^{\mathrm{T}}$ are given matrices of corresponding dimensions.

We will need the following assumptions that are standard for linear-quadratic optimal control problems.

Assumption 2.6.1. *The pair* (A, C) *has no unobservable nodes on the unit circle (see, e.g.,* [63]*).*

Assumption 2.6.2. *The matrix* G *is positive definite.*

In this section, we consider the following optimal control problem:

$$J_{1,\infty}[x(1)] \to \min . \tag{2.6.2}$$

If we do not have any limited capacity communication channel and the whole state $x(\cdot)$ is available to the controller, then the problem (2.5.1), (2.6.1), (2.6.2) is the standard linear-quadratic optimal control problem and its solution is well known (see, e.g., [63]). Under Assumptions 2.5.1, 2.6.1, and 2.6.2, for any initial condition $x(1)$, the optimal control is given by

$$u(t) = Kx(t), \tag{2.6.3}$$

where

$$K = -(G + B^{\mathrm{T}}PB)^{-1}B^{\mathrm{T}}PA \tag{2.6.4}$$

and the square matrix P is a solution of the discrete-time algebraic Riccati equation

$$A\left(P - PB(G + B^{\mathrm{T}}PB)^{-1}B^{\mathrm{T}}P\right)A + C^{\mathrm{T}}C - P = 0 \tag{2.6.5}$$

such that the matrix $A + BK$ is stable (has all its eigenvalues inside the unit circle). Furthermore, the optimal value of the cost function is given by

$$J^{\text{opt}}_{1,\infty}[x(1)] = x(1)^{\mathrm{T}}Px(1). \tag{2.6.6}$$

Definition 2.6.3. *The optimal control problem* (2.5.1)*,* (2.6.1)*,* (2.6.2) *is said to be solvable in the communication channel class* $\mathfrak{C}_R$ *if for any* $\epsilon > 0$*, a period* $T \geq 1$ *and a coder–decoder-controller pair* (2.5.2)*,* (2.5.3) *with a coding alphabet of size* l *satisfying the constraint* (2.2.2) *exist such that the following conditions hold:*

(i) The closed-loop system (2.5.1)*,* (2.5.2)*,* (2.5.3) *is stable in the following sense: For any* $\epsilon_0 > 0$*, an integer* $k \geq 1$ *exists such that* (2.5.4) *is satisfied for any solution* $[x(\cdot), u(\cdot)]$ *of the closed-loop system with initial condition* $x(1) \in \mathfrak{X}_1$*;*

(ii) For any solution $[x(\cdot), u(\cdot)]$ *of the closed-loop system with initial condition* $x(1) \in \mathfrak{X}_1$*,*

$$J_{1,\infty}[x(1)] \leq J^{\text{opt}}_{1,\infty}[x(1)] + \epsilon, \tag{2.6.7}$$

where $J^{\text{opt}}_{1,\infty}[x(1)]$ *is given by* (2.6.6)*.*

Now we are in a position to present the main result of this section.

Theorem 2.6.4. *Consider the system* (2.5.1) *and the cost function* (2.6.1)*. Let* $R \geq 0$ *be a given constant and* $H(A)$ *be the value* (2.4.2)*. Suppose that Assumptions* 2.4.1, 2.5.1, 2.6.1*, and* 2.6.2 *are satisfied. Then the following two statements hold:*

(i) If $R < H(A)$, then the optimal control problem (2.5.1), (2.6.1), (2.6.2) *is not solvable in the communication channel class $\mathfrak{C}_R$;*
(ii) If $R > H(A)$, then the optimal control problem (2.5.1), (2.6.1), (2.6.2) *is solvable in the communication channel class $\mathfrak{C}_R$.*

The proof of Theorem 2.6.4 is given in Sect. 2.7.

Remark 2.6.5. Theorem 2.6.4 gives an "almost" necessary and sufficient condition for solvability of the optimal control problem in the communication channel class $\mathfrak{C}_R$. Notice that in the critical case $R = H(A)$, both possibilities can occur. Indeed, consider the system (2.5.1) with a stable matrix A and nonzero B. In this case, $H(A) = 0$. If we take the cost function

$$J_{1,\infty}[x(1)] := \sum_{t=1}^{+\infty} \|u(t)\|^2,$$

then the optimal control input is zero, and the optimal control problem is solvable in the communication channel class $\mathfrak{C}_0$.

On the other hand, if we take the cost function

$$J_{1,\infty}[x(1)] := \sum_{t=1}^{+\infty} \|x(t)\|^2 + \|u(t)\|^2,$$

then the optimal control problem is obviously not solvable in the communication channel class $\mathfrak{C}_0$.

Remark 2.6.6. Notice that according to the results of Chap. 3, the rate of exponential decay of the closed-loop system cannot be greater than $R - H(A)$. However, the rate of exponential decay of a linear optimal closed-loop system can be greater than $R - H(A)$. Our coder–decoder-controller closely approximates the trajectory of the optimal linear system on the time interval $t = 1, 2, \ldots, T$ and guarantees just convergence to the origin for $t \to \infty$. Roughly speaking, the closeness to the optimal cost is determined mostly by transient response, whereas the decay rate is determined by behavior on infinity. That is why two closed-loop systems may have very close quadratic costs and completely different decay rates. It should also be pointed out that when we decrease ϵ we have to increase the period T.

Comment 2.6.7. Notice that the problem of linear-quadratic optimal control via a limited capacity communication channel but with an instantaneous data rate was explicitly solved for scalar plants in [143].

2.7 Proofs of Lemma 2.4.3 and Theorems 2.5.3 and 2.6.4

We first prove Lemma 2.4.3 from Sect. 2.4.

Proof of Lemma 2.4.3. Consider the set $\mathcal{A} = \{a_1, \dots, a_m\}$, where $1 < a_1 < \dots < a_m$ are defined by the following rule: $a_i \in \mathcal{A}$ if and only if $a_i > 1$ and the matrix A has an eigenvalue λ such that $|\lambda| = a_i$. In other words, $\mathcal{A}$ is the set of all possible magnitudes of unstable eigenvalues of A. Now we partition the set $S(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ of eigenvalues of the matrix A into groups $S_0, S_1, \dots, S_m$ as follows. The set S_0 consists of all eigenvalues λ such that $|\lambda| \le 1$. Furthermore, for any $i = 1, \dots, m$, the set S_i consists of all eigenvalues λ such that $|\lambda| = a_i$. Without any loss of generality, we consider real matrices A of the form

$$A = \begin{pmatrix} A_0 & 0 & \dots & 0 \\ 0 & A_1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & A_m \end{pmatrix}, \tag{2.7.1}$$

where $S(A_i) = S_i$ for $i = 0, 1, \dots, m$. Indeed, any matrix A can be transformed to the form (2.7.1) by a linear transformation, and the value $H(A)$ and the property stated by this lemma are obviously invariant under any linear transformation. Also, let $\mathfrak{Y}_1$ be the corresponding linear transformation of the set $\mathfrak{X}_1$. Obviously, $\mathfrak{Y}_1$ is also compact. Notice that $H(A) = H(A_0) + H(A_1) + \dots + H(A_m)$, $H(A_0) = 0$, and $H(A_i) > 0$ for any $i > 0$.

Furthermore, let $n_i > 0$ be the number of elements in S_i. For any i we take numbers c_i, b_i satisfying the following rules: $c_0 > b_0 > 1$; $c_i > b_i > a_i$ for $i = 1, \dots, m$; and $\sum_{i=0}^{m} n_i \log_2(c_i) = \alpha$. Such numbers obviously exist because $\alpha > H(A)$. Since the set $\mathfrak{Y}_1$ is compact, for some $d > 0$, $\mathfrak{Y}_1$ is a subset of the set

$$\mathfrak{Y}_d := \{x \in \mathbb{R}^n : \|x\|_\infty \le d\}. \tag{2.7.2}$$

Consider the partitioning of the state vector $x(t)$ corresponding to (2.7.1):

$$x(t) = \textbf{col}\,[y_0(t), y_1(t), \dots, y_m(t)]\,,$$

where $y_i(t) \in \mathbb{R}^{n_i}$. Furthermore, for any i, k introduce integers $N_{i,k}$ as follows: $N_{i,k}$ is the largest integer such that $N_{i,k} \le c_i^k$. Now consider the set Q_k of solutions of the system (2.4.3) defined by initial conditions from $\mathfrak{Y}_d$ of the following form: n_i components of the vector $y_i(1)$ take all possible values in the discrete set $\{-d, -d + \frac{2d}{N_{i,k}-1}, \dots, d - \frac{2d}{N_{i,k}-1}, d\}$. The cardinality N_k of the set Q_k can be estimated as follows:

$$\begin{aligned} N_k &= N_{0,k}^{n_0} \times N_{1,k}^{n_1} \times \dots \times N_{m,k}^{n_m} \\ &\le c_0^{kn_0} \times c_1^{kn_1} \times \dots \times c_m^{kn_m}. \end{aligned}$$

By our construction, $\sum_{i=0}^{m} n_i \log_2(c_i) = \alpha$; hence, $\log_2(N_k) \le \alpha \cdot k$ for any k.

Now let

$$x_a(t) = \textbf{col}\,[y_{a,0}(t), y_{a,1}(t), \dots, y_{a,m}(t)]$$

be any solution of (2.4.1) with the matrix A of the form (2.7.1). Then we can take a solution

$$x^{(k)}(t) = \mathbf{col}\left[y_0^{(k)}(t), y_1^{(k)}(t), \ldots, y_m^{(k)}(t)\right]$$

of (2.4.3) from Q_k such that

$$\|y_{a,i}(1) - y_i^{(k)}(1)\|_\infty \leq \frac{d}{N_{i,k} - 1} \quad \forall i.$$

Now we prove that the condition (2.4.4) holds if k is large enough. Indeed, because b_i is strictly greater than the absolute value of any eigenvalue of A_i, a constant $c > 0$ exists such that

$$\|y_i(t)\|_\infty \leq c b_i^{t-1} \|y_i(1)\|_\infty; \tag{2.7.3}$$
$$\|y_i(t)\| \leq c b_i^{t-1} \|y_i(1)\|_\infty \tag{2.7.4}$$

for any $i = 0, 1, \ldots, m$. Furthermore, from (2.7.4) we have

$$\sum_{t=1}^{k} \|x_a(t) - x^{(k)}(t))\|^2 = \sum_{t=1}^{k} \sum_{i=0}^{m} \|y_{a,i}(t) - y_i^{(k)}(t))\|^2$$
$$\leq c^2 \sum_{i=0}^{m} \sum_{t=1}^{k} b_i^{2(t-1)} \frac{d^2}{(N_{i,k} - 1)^2}. \tag{2.7.5}$$

Because $b_i < c_i$ and $\frac{N_{i,k}}{c_i^k} \to 1$ as $k \to \infty$, from (2.7.5) we obtain that

$$\lim_{k \to \infty} \sum_{t=1}^{k} \|x_a(t) - x^{(k)}(t))\|^2 = 0. \tag{2.7.6}$$

Moreover, from (2.7.3) we have

$$\max_{t=1,\ldots,k} \|x_a(t) - x^{(k)}(t)\|_\infty =$$
$$\max_{t=1,\ldots,k} \max_{i=0,\ldots,m} \|y_{a,i}(t) - y_i^{(k)}(t)\|_\infty \leq$$
$$\max_{t=1,\ldots,k} \max_{i=0,\ldots,m} b_i^{(t-1)} \frac{d}{N_{i,k} - 1}. \tag{2.7.7}$$

Because $b_i < c_i$ and $\frac{N_{i,k}}{c_i^k} \to 1$ as $k \to \infty$, from (2.7.7) we have

$$\lim_{k \to \infty} \max_{t=1,\ldots,k} \|x_a(t) - x^{(k)}(t)\|_\infty = 0. \tag{2.7.8}$$

Finally, (2.7.6) and (2.7.8) immediately imply (2.4.4). This completes the proof of the lemma. □

Proof of Theorems 2.5.3 and 2.6.4. It is obvious from Definitions 2.5.2 and 2.6.3 that the condition for solvability of the optimal control problem in the communication channel class $\mathfrak{C}_R$ is stronger than that for stabilizability in the same class.

Therefore, it is enough to prove statement (i) of Theorem 2.5.3 and statement (ii) of Theorem 2.6.4.

Statement (i) of Theorem 2.5.3. We prove this statement by contradiction. Indeed, assume that the linear system is stabilizable in the communication channel class $\mathfrak{C}_R$ with $R < H(A)$. From Theorem 2.4.2 we know that $H(A)$ is the topological entropy of the system (2.4.1). Therefore, it follows from Lemma 2.3.8 that a constant $\epsilon_* > 0$ exists such that

$$\limsup_{k' \to \infty} \frac{1}{k'} \log_2(s(k', \epsilon_*)) > R. \tag{2.7.9}$$

Consider a coder–decoder-controller pair (2.5.2), (2.5.3) such that the requirements of Definition 2.5.2 hold, and let $T > 0$ be its period. Furthermore, let $k \geq 1$ be an integer from the condition (2.5.4) of Definition 2.5.2 corresponding to $\epsilon_0 = \frac{1}{2}\epsilon_*$. The inequality (2.7.9) implies that an integer $j > 0$ and an $(jT, \epsilon_*)-$separated set S of cardinality N exist such that $\frac{\log_2 N}{jT} > R$. Moreover, we can take an $(jT, \epsilon_*)-$separated set S of cardinality N with $\frac{\log_2 N}{jT} > R$ such that for distinct points $x_a(\cdot), x_b(\cdot) \in S$, we have

$$\|x_a(t) - x_b(t)\|_\infty \geq \epsilon_* \quad \text{for some} \quad t = k, k+1, \ldots, jT. \tag{2.7.10}$$

Furthermore, let $\hat{S}$ be the set of all possible control sequences $u(1), u(2), \ldots, u(jT)$ produced by (2.5.2), (2.5.3). Then, the cardinality of $\hat{S}$ does not exceed l^j. Since $\frac{\log_2 N}{jT} > R$, the condition (2.2.2) implies that $l^j < N$. Because the condition (2.5.4) with $\epsilon_0 = \frac{1}{2}\epsilon_*$ must be satisfied for any solution of the closed-loop system (2.5.1)–(2.5.3) with some control input from $\hat{S}$ and the cardinality of $\hat{S}$ is less than the cardinality of S, two elements $x_a(\cdot), x_b(\cdot$ of S and an element $u_0(1), u_0(2), \ldots, u_0(jT)$ of $\hat{S}$ exist such that the condition (2.5.4) holds with $x(\cdot) = x_a(\cdot) + x_u(\cdot)$ and $x(\cdot) = x_b(\cdot) + x_u(\cdot)$, where $x_u(\cdot)$ is the solution of the system

$$x_u(t+1) = Ax_u(t) + Bu_0(t),$$

where $x_u(1) = 0$. This implies that $\|x_a(t) - x_b(t)\|_\infty < \epsilon_*$ for all $t = k, k+1, \ldots, jT$. However, the latter inequality contradicts to (2.7.10). This completes the proof of this part of the theorem.

Statement (ii) of Theorem 2.6.4. Assume that $R > H(A)$, and prove that the optimal control problem is solvable via a communication channel with capacity R. Let $\epsilon > 0$ be a given constant. We will build a coder–decoder-controller pair (2.5.2), (2.5.3) with the communication channel class $\mathfrak{C}_R$ such that the condition (2.6.7) holds.

Since the set $\mathfrak{X}_1$ is compact, a constant $D_0 > 0$ exists such that $\mathfrak{X}_1 \subset B_0 := \{x : \|x\|_\infty \leq D_0\}$. Consider the system:

$$x_a(t+1) = Ax_a(t), \quad x_a(1) \in B_0. \tag{2.7.11}$$

Introduce the following notation:

$$J_{1,T}^{\text{opt}}[x(1)] := \sum_{t=1}^{T} [x(t)^{\mathrm{T}} C^{\mathrm{T}} C x(t) + u(t)^{\mathrm{T}} G u(t)], \tag{2.7.12}$$

where the sum is taken for the solution of the optimal closed-loop system (2.5.1), (2.6.3)–(2.6.5) with the initial condition $x(1)$.

It is obvious that a constant $\delta > 0$ exists such that

$$|J_{1,T}^{\text{opt}}[x^{(1)}(1)] - J_{1,T}^{\text{opt}}[x^{(2)}(1)]| < \frac{\epsilon}{4} \tag{2.7.13}$$

for any $T > 1$ and for any $x^{(1)}(1), x^{(2)}(1)$ from B_0 such that $\|x^{(1)}(1) - x^{(2)}(1)\|_\infty \leq \delta$. Furthermore, a constant $c > 0$ exists such that

$$\max_{x(1) \in B_0} J_{1,\infty}^{\text{opt}}[x(1)] \leq c \tag{2.7.14}$$

for any solution of the closed-loop optimal linear system (2.5.1), (2.6.3)–(2.6.5).

Introduce the following constants:

$$\overline{a} := \frac{\epsilon}{2(c + \frac{\epsilon}{2})}, \quad b := \sqrt{\frac{\overline{a}}{1 + \overline{a}}}. \tag{2.7.15}$$

The closed-loop optimal linear system (2.5.1), (2.6.3)–(2.6.5) is exponentially stable. Hence, a time k_0 exists such that for any $k \geq k_0$ the inequality

$$\|x_{\text{opt}}(t + k)\|_\infty \leq \frac{b}{2} \|x_{\text{opt}}(t)\|_\infty \tag{2.7.16}$$

holds for any solution of the closed-loop optimal linear system (2.5.1), (2.6.3)–(2.6.5) and $t = 1, 2, \ldots$.

Now introduce

$$\epsilon_0 := \min \left\{ \delta, \frac{b}{2}, \frac{b}{2} D_0, \frac{\epsilon}{4} \right\} \tag{2.7.17}$$

and let α be any constant such that $H(A) < \alpha < R$. From (2.7.16) and Lemma 2.4.3, we obtain that an integer k exists, satisfying (2.7.16) and the inequality

$$k > \frac{R}{R - \alpha}, \tag{2.7.18}$$

and such that a set $Q_\epsilon = \{x_1(\cdot), \ldots, x_N(\cdot)\}$ of N solutions of the system (2.7.11) for $t = 1, 2, \ldots, k$ exists with the properties $\frac{1}{k} \log_2(N) \leq \alpha$, and for any solution $x_a(\cdot)$ of (2.7.11), an element $x_{\min}(\cdot) \in Q_\epsilon$ exists such that

$$\beta \sum_{t=1}^{k} \|x_a(t) - x_{\min}(t))\|^2 + \max_{t=1,\ldots,k} \|x_a(t) - x_{\min}(t)\|_\infty < \epsilon_0, \tag{2.7.19}$$

where we take $\beta := \|C^{\mathrm{T}} C\|$.

We will build a coder–decoder-controller with the period

$$T := k - 1. \tag{2.7.20}$$

Let $f_0(x(1))$ be the index of the element $x_{\min}(\cdot) \in Q_\epsilon$ such that (2.7.19) holds for the solution of (2.7.11) with the initial condition $x_a(1) = x(1)$. Moreover, let $\tilde{u}(\cdot)$ be the optimal control input in (2.5.1), (2.6.3)–(2.6.5) with the initial condition $x_{\text{opt}}(1) = x_{\min}(1)$. Now introduce the following coder–decoder pair for $t = 1, 2, \ldots, T$:

$$h(1) := f_0(x(1)); \tag{2.7.21}$$

$$\begin{pmatrix} u(1) \\ \vdots \\ u(T-1) \\ u(T) \end{pmatrix} := \begin{pmatrix} \tilde{u}(1) \\ \vdots \\ \tilde{u}(T-1) \\ \tilde{u}(T) \end{pmatrix}. \tag{2.7.22}$$

Furthermore, for any $j = 1, 2, \ldots$, introduce the constants $D_j > 0$ as follows:

$$D_j := b^j D_0. \tag{2.7.23}$$

Also, introduce the set $B_j := \{x : \|x\|_\infty \leq D_j\}$.

We now prove that $x(T+1) \in B_1$ for any solution of the closed-loop system (2.5.1), (2.7.21), (2.7.22) with the initial condition $x(1) \in B_0$. Indeed, any solution $x(t)$ of (2.5.1), (2.7.21), (2.7.22) can be represented as

$$x(t) = x_{\text{opt}}(t) + x_a(t), \tag{2.7.24}$$

where x_{opt} is the solution of the optimal linear system (2.5.1), (2.6.3)–(2.6.5) with the initial condition $x_{\text{opt}}(1) = x_{\min}(1)$ and $x_a(t)$ is the solution of the linear system (2.7.11) with the initial condition $x_a(1) = x(1) - x_{\text{opt}}(1)$. From (2.7.16) we have

$$\|x_{\text{opt}}(T+1)\|_\infty \leq \frac{b}{2} D_0. \tag{2.7.25}$$

Furthermore, since $\epsilon_0 \leq \frac{b}{2} D_0$ by (2.7.17), the inequality (2.7.19) implies that $\|x_a(T+1)\|_\infty \leq \frac{b}{2} D_0$. From this and (2.7.25) we obtain

$$\|x(T+1)\|_\infty \leq \|x_{\text{opt}}(T+1)\|_\infty + \|x_a(T+1)\|_\infty \leq b D_0.$$

Hence, $x(T+1) \in B_1$.

Now introduce the following coder–decoder-controller for $t = jT+1, jk+2, \ldots, (j+1)T$ and for $x(jT+1) \in B_j$ as follows:

$$h(jT+1) := f_0\left(\frac{1}{b^j} x(jT+1)\right); \tag{2.7.26}$$

$$\begin{pmatrix} u(jT+1) \\ \vdots \\ u((j+1)T-1) \\ u((j+1)T) \end{pmatrix} := b^j \begin{pmatrix} \tilde{u}(1) \\ \vdots \\ \tilde{u}(T-1) \\ \tilde{u}(T) \end{pmatrix}. \tag{2.7.27}$$

Here $j = 1, 2, \ldots$.

It is obvious that if $(x(\cdot), u(\cdot))$ is a solution of (2.5.1), (2.7.26), (2.7.27) for $t = jT+1, \ldots, (j+1)T$ with $x(jT+1) \in B_j$, then $(\frac{1}{b^j}x(\cdot), \frac{1}{b^j}u(\cdot))$ is a solution of (2.5.1), (2.7.21), (2.7.22) for $t = 1, \ldots, T$ with $x(1) \in B_0$. Since we have proved that $x(T+1) \in B_1$ for any solution of (2.5.1), (2.7.21), (2.7.22) with $x(1) \in B_0$, we obtain by mathematical induction that $x(jT+1) \in B_j$ for all j. Hence, the closed-loop system (2.5.1), (2.7.21), (2.7.22), (2.7.26), (2.7.27) is well defined for all $t = 1, 2, \ldots$.

Because the cardinality of the set Q_ϵ is N and $\frac{1}{k}\log_2 N \leq \alpha < R$, the condition (2.7.18) holds and $T = k-1$, the coder–decoder-controller (2.7.21), (2.7.22), (2.7.26), (2.7.27) is from the communication channel class $\mathfrak{C}_R$.

Now we prove that the closed-loop system (2.5.1), (2.7.21), (2.7.22), (2.7.26), (2.7.27) satisfies the condition (2.6.7) of Definition 2.6.3.

Introduce the following notation:

$$J_{i,r}[x(1)] := \sum_{t=i}^{r}[x(t)^{\mathrm{T}}C^{\mathrm{T}}Cx(t) + u(t)^{\mathrm{T}}Gu(t)], \tag{2.7.28}$$

where the sum is taken for the solution of the closed-loop system (2.5.1), (2.7.21), (2.7.22), (2.7.26), (2.7.27) with the initial condition $x(1)$. Since $\epsilon_0 \leq \delta$ by our definition (2.7.17), inequality (2.7.19) implies that

$$\|x(1) - x_{\min}(1)\|_\infty \leq \delta.$$

From this and (2.7.13) we obtain

$$|J_{1,T}^{\text{opt}}[x(1)] - J_{1,T}^{\text{opt}}[x_{\min}(1)]| < \frac{\epsilon}{4}. \tag{2.7.29}$$

Moreover, (2.7.17) and (2.7.19) also imply that

$$\sum_{t=1}^{T}\|C(x_a(t) - x_{\min}(t))\|^2 \leq \frac{\epsilon}{4}.$$

This and (2.7.29) imply that

$$|J_{1,T}[x(1)] - J_{1,T}^{\text{opt}}[x(1)]| \leq \frac{\epsilon}{2}. \tag{2.7.30}$$

Now we derive an upper estimate for $J_{jT+1,(j+1)T}[x(1)]$ for any $j = 1, 2, \ldots$. From (2.7.14) and (2.7.30) we obviously obtain that

$$J_{1,T}[x(1)] \leq c + \frac{\epsilon}{2}. \tag{2.7.31}$$

It is obvious that if $(x(\cdot), u(\cdot))$ is a solution of (2.5.1), (2.7.26), (2.7.27) for $t = jT+1, \ldots, (j+1)T$, then $(\frac{1}{b^j}x(\cdot), \frac{1}{b^j}u(\cdot))$ is a solution of (2.5.1), (2.7.21), (2.7.22) for $t = 1, \ldots, T$. Therefore, (2.7.31) implies that

$$J_{jT+1,(j+1)T}[x(1)] \leq (c + \frac{\epsilon}{2})b^{2j}. \tag{2.7.32}$$

From this and (2.7.15) we have

$$J_{T+1,\infty}[x(1)] \leq (c + \frac{\epsilon}{2}) \sum_{j=1}^{\infty} b^{2j} =$$

$$(c + \frac{\epsilon}{2}) \frac{b^2}{1 - b^2} = (c + \frac{\epsilon}{2}) \overline{a} = \frac{\epsilon}{2}. \quad (2.7.33)$$

Furthermore,

$$J_{1,\infty}[x(1)] - J_{1,\infty}^{\text{opt}}[x(1)] \leq |J_{1,T}[x(1)] - J_{1,T}^{\text{opt}}[x^{(1)}]| + J_{T+1,\infty}[x(1)]. \quad (2.7.34)$$

This and the estimates (2.7.30) and (2.7.33) immediately imply (2.6.7).

Finally, the Lyapunov stability condition (2.5.4) follows immediately from the property proved above that $x(jT+1) \in B_j$ for any solution of the closed-loop system (2.5.1), (2.7.21), (2.7.22), (2.7.26), (2.7.27) with the initial condition $x(1) \in B_0$. This completes the proof of the statement. □

3

Stabilization of Linear Multiple Sensor Systems via Limited Capacity Communication Channels

3.1 Introduction

In this chapter, we study a stabilization problem via quantized state feedback for a linear time-invariant partially observed system. We consider a multi-channel communication between multiple sensors and the controller, where each sensor is served by its own finite capacity channel, and there is no information exchange between the sensors. Furthermore, there is no feedback communication from the controller to the sensors, and the sensors have no direct access to control. The objective is to establish, first, the tightest lower bounds on the capacities of the channels for which the stabilization is possible and, second, the rate of exponential stability that is achievable for given capacities obeying those bounds. To this end, we obtain necessary and sufficient conditions for stabilizability.

Another crucial point is that we do not assume the channels to be perfect. Sensor signals may incur independent and time-varying delays and arrive at the controller out of order. There may be periods when the sensor is denied access to the channel. Transmitted data may be corrupted or even lost. However, we assume that the communication noise is compensated, and so ultimately it reveals itself only in the form of decay of the channel information capacity. For example, employing error correcting block codes [68, Chap. 12] means that the channel is partly engaged in transmission of redundant check symbols, which decreases the average amount of the primal messages carried from the sensors. We suppose that error correction is the function of the channel; i.e., the corresponding coder and decoder are given and considered as part of what is termed "channel" in this chapter. The key assumption is that the time-average number of bits per sample period that can be successfully transmitted across the channel during a time interval converges to what we call the transmission capacity as the length of the interval becomes large. The stabilizability region is given in terms of these capacities. Note that bounded communication delays do not influence them and, thus the region, although they affect the design of the stabilizing controller.

In the particular case where the channels are perfect and the system is detectable via each sensor, the conditions for stabilizability obtained in this chapter are in har-

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_3,

mony with those from Theorem 2.5.3. Thus we show that in this case, multiple communicating sensors and channels with separate capacity constraints may be treated as a single sensor and as a single channel with a united constraint, respectively. However, employing multiple sensors usually means that there are problems with detectability by means of a single sensor, and then the model with nondetecting sensors is often a good option.

Stabilization with limited information feedback was studied in the presence of transmission delays in [221]. Unlike the current chapter, the transmission time required to transfer one bit was assumed constant, a continuous-time linear plant and a network with the simplest topology and constant access channels were considered, and conditions for nonasymptotic stability but a weaker property called containability were established.

Multiple sensor systems were examined in [204]. Considered were perfect (i.e., noiseless and undelayed) channels under the assumption that the control is known at the sensor sites. The arguments from [204] also presuppose that the system is reducible to the real-diagonal form so that any "mode" is in a simple relation with any sensor. The latter means that the mode either does not affect the sensor outputs or can be completely determined from these outputs.[1] A Slepian-Wolf-type stabilization scheme was proposed and was shown to achieve stability under certain conditions on the channels data rates. The answer is given in terms of the controller parameters called the rate vectors. They are tuples of naturals, with each being the number of the quantizer levels with respect to a certain state coordinate. These conditions can be reformulated in terms of the capacities of the channels. The criterion for stabilizability via the above scheme reduces to solvability of some linear system of inequalities in integers. In [205], this result was extended to the case where the signals from multiple sensors are transmitted to a single controller over a network of independent perfect channels with a fixed topology.

In this chapter, we show that stabilizable systems exist, which, however, cannot be stabilized by means of the aforementioned control scheme. Moreover, insufficient are all schemes that merely have some features in common with that one (see Sect. 3.9).[2] This stresses that stabilizability should be tested in the class of all controllers with a given information pattern. Such an analysis is offered in this chapter, which gives rise to the additional job: to show that there is no gap between the necessary and sufficient conditions, where all and the specific controllers proposed in this chapter are concerned, respectively. To achieve this objective, not only time averaging but also convex duality techniques are employed. The final criterion is given in terms of only the plant and channel parameters.

The case of multiple sensors and actuators, where each sensor is directly linked with every actuator by a separate perfect channel, was studied in [139] for real-

[1] The arguments from [204] do not much require the diagonal form. However, the assumption that the system can be decomposed into independent subsystems each in simple relations with the sensors seems to be crucial.

[2] For example, stabilization may be due to a control scheme that is not recurrent but is cyclic (i.e., periodically varies in time).

diagonalizable systems. Separate necessary and sufficient conditions for stabilizability were obtained. In general, they are not tight [139]. In the case where the system is stabiliazible by every actuator and detectable by every sensor, a common necessary and sufficient criterion was established in [139].

In this chapter, we consider the case where the system is not necessarily reducible to a diagonal form. It is shown that nontrivial Jordan blocks may make it impossible to disintegrate the system into state-independent subsystems each in simple relations with the sensors. To treat this case, we propose sequential stabilization based on triangular decomposition of the system into state-dependent subsystems. They are stabilized separately and successively. In doing so, their interinfluence is interpreted as an exogenous disturbance. This disturbance can be treated as exponentially decaying since thanks to the triangular architecture of the system, disturbance of any given subsystem is generated by the preceding ones, which are supposed to be already stabilized according to the sequential stabilization approach. The controller design employs ideas related to those from [149] as well as [28, 135, 136, 204]. Other ideas concern an account for transmission delays and disturbances decaying at a known rate. No other characteristics of the disturbance (e.g., an upper bound) are assumed to be known. Apart from state-dependency, the subsystems are also dependent via control. Since it is common, the control aimed to stabilize a particular subsystem may disturb the others. We offer a method to cope with this problem.

The main results of this chapter were originally published in [115].

The body of the chapter is organized as follows. We first illustrate the problem statement and the main result by an example in Sect. 3.2. The general problem statement is given in Sect. 3.3. Section 3.4 offers basic definitions and assumptions. The main result is presented in Sect. 3.5. Its proof is given in Sects. 3.7 and 3.8, where necessary and sufficient conditions for stabilizability are, respectively, justified. In Sect. 3.6, the main result is applied to the example from Sect. 3.2. The concluding Sect. 3.9 comments on an important assumption adopted in this chapter.

3.2 Example

We first illustrate the class of problems to be studied by an example.

We consider a platoon composed of k vehicles moving along a line and enumerated from right to left. The dynamics of the platoon are uncoupled, and the vehicles are described by the equations

$$\dot{x}_i = v_i, \quad \dot{v}_i = u_i, \quad i = 1, \dots, k, \tag{3.2.1}$$

where x_i is the position of the ith vehicle, v_i is its velocity, and u_i is the control input. Each vehicle is equipped with a sensor giving the distance $y_i = x_i - x_{i-1}$ from it to the preceding one for $i \geq 2$ and the position $y_1 = x_1$ for $i = 1$. It is also served by a digital communication channel over which the measurement y_i is sent to the central controller. To this end, the sensor signals are sampled with a period $\Delta > 0$. This channel is delayed, nonstationary, and lossy and transmits on average

$\mathfrak{c}_i > 0$ bits per sample period. Employing the data that arrive over all channels, the central controller produces the control inputs for all vehicles at the sample times. The objective is to stabilize the platoon motion about a given constant-velocity trajectory: $v_i = v_i^0, x_i(t) = x_i^0 + v_i^0 t \ \forall i$. This situation is illustrated in Fig. 3.1 for $k = 4$.

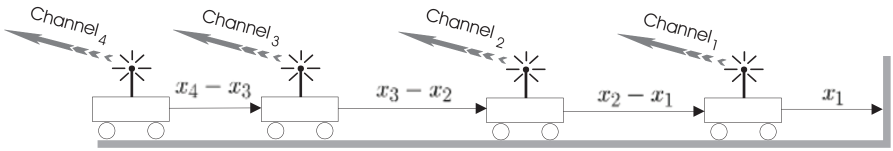


Fig. 3.1. Platoon of autonomous vehicles.

In this context, we pose the following questions:

(i) What is the minimum rate of the information transmission for which stabilization is possible?
(ii) Which rate of stability can be achieved for channels with given capacities $\mathfrak{c}_j$ and sample period Δ?

More precisely, we are interested in the rate μ at which the platoon is able to approach the desired trajectory:

$$|v_i(t) - v_i^0| \le K_{v,i}\mu^{t/\Delta}, \qquad |x_i(t) - x_i^0 - v_i^0 t| \le K_{x,i}\mu^{t/\Delta}.$$

As will be shown in Sect. 3.6, stabilization of the platoon is possible for any capacities $\mathfrak{c}_i > 0$ and at any rate

$$\mu > \mu^0 := \sqrt{2}^{\,-\mathfrak{c}_{\min}}, \quad \text{where} \quad \mathfrak{c}_{\min} := \min_{i=1,\dots,k} \mathfrak{c}_i.$$

At the same time, no rate $\mu < \mu^0$ is achievable.

Now consider another situation where the sensor system accommodated by each vehicle is able to give the distances to $l < k$ vehicles to the right, as well as to l vehicles to the left. Then the platoon motion remains stabilizable for any capacities $\mathfrak{c}_i$. However, the above threshold stability rate μ^0 is changed: $\mu^0 = \sqrt{2}^{\,-\mathfrak{c}_{k,l}}$. Here

$$\mathfrak{c}_{k,l} := \min\left\{\mathfrak{c}_{k,l}^{(1)}, \mathfrak{c}_{k,l}^{(2)}, \mathfrak{c}_{k,l}^{(3)}\right\} \quad \text{if} \quad 2l \geq k,$$

$$\text{and} \quad \mathfrak{c}_{k,l} := \min\left\{\mathfrak{c}_{k,l}^{(4)}, \mathfrak{c}_{k,l}^{(5)}, \mathfrak{c}_{k,l}^{(6)}\right\} \quad \text{if} \quad 2l < k, \quad \text{where}$$

$$\mathfrak{c}_{k,l}^{(1)} := \min_{i=1,\dots,k-l} \frac{1}{i}\sum_{j=1}^{i} \mathfrak{c}_j, \quad \mathfrak{c}_{k,l}^{(2)} := \frac{1}{k}\sum_{j=1}^{k} \mathfrak{c}_j, \quad \mathfrak{c}_{k,l}^{(3)} := \min_{i=l+1,\dots,k-1} \frac{1}{k-i}\sum_{j=i+1}^{k} \mathfrak{c}_j,$$

$$\mathfrak{c}_{k,l}^{(4)} := \min_{i=1,\dots,l+1} \frac{1}{i}\sum_{j=1}^{i} \mathfrak{c}_j, \quad \mathfrak{c}_{k,l}^{(5)} := \min_{i=l+2,\dots,k-l} \mathfrak{c}_i,$$

$$\mathfrak{c}_{k,l}^{(6)} := \min_{i=k-l,\dots,k-1} \frac{1}{k-i}\sum_{j=i+1}^{k} \mathfrak{c}_j. \quad (3.2.2)$$

In the situation at hand, every vehicle is equipped with several sensors, each producing data about its relative position with respect to some other vehicle. Due to the limited capacity of the channel, the entire cumulative sensor data cannot be communicated from the vehicle to the controller. This gives rise to two natural questions:

(iii) In which way should the bits of the message currently dispatched from every vehicle be distributed over data from different sensors?
(iv) Which information should be carried by the bits assigned to every sensor?

The objective of this chapter is to present a general theory that enables one to obtain answers to the questions like those in (i)–(iv).

3.3 General Problem Statement

We consider linear discrete-time multiple sensor systems of the form

$$x(t+1) = Ax(t) + Bu(t); \qquad x(0) = x_0; \qquad (3.3.1)$$

$$y_j(t) = C_j x(t), \qquad j = 1, \dots, k. \qquad (3.3.2)$$

Here $x(t) \in \mathbb{R}^n$ is the state; $u(t) \in \mathbb{R}^{n_u}$ is the control; and $y_j(t) \in \mathbb{R}^{n_{y,j}}$ is the output of the jth sensor. The system is unstable: There is an eigenvalue λ of the matrix A with $|\lambda| \geq 1$. The objective is to stabilize the plant: $x(t) \xrightarrow{t\to\infty} 0$.

We consider a remote control setup. Each sensor is served by its own communication channel capable of transmitting signals from a finite *alphabet* $\mathfrak{E}_j$. Over this channel, the jth *coder* sends a message $e_j(t) \in \mathfrak{E}_j$ based on the prior measurements

$$e_j(t) = \mathcal{E}_j[t, y_j(0), \dots, y_j(t)]. \qquad (3.3.3)$$

On the basis of the data $\overline{e}(t)$ received over all channels up to the current time t, the *decoder* selects the control

$$u(t) = \mathcal{U}\left[t, \overline{e}(t)\right]. \qquad (3.3.4)$$

In this situation illustrated by Fig. 3.2, the *controller* is constituted by the decoder

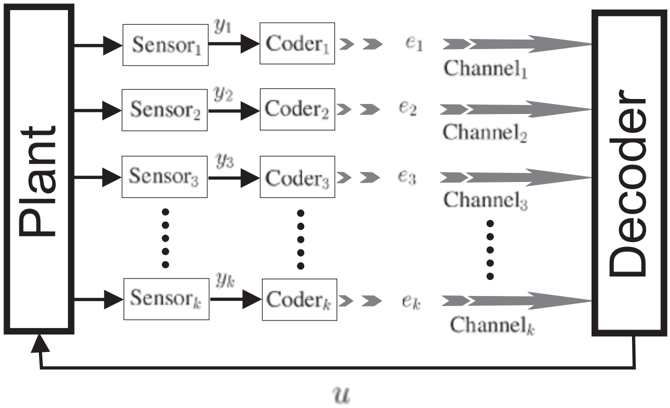


Fig. 3.2. Feedback control via communication channels.

and the set of coders each serving a particular sensor

$$\mathcal{C} := \left[\mathcal{E}_1(\cdot), \ldots, \mathcal{E}_k(\cdot), \mathcal{U}(\cdot)\right]. \tag{3.3.5}$$

Transmitted messages incur delays and may be lost: The message $e(t)$ dispatched at time t arrives at the decoder at the discrete time

$$t + \tau_j[t, e(t)] \geq t,$$

where $\tau_j[t, e(t)] := \infty$ if it is lost. So the data available to the decoder at time t are

$$\overline{e}(t) := \left[\overline{e}_1(t), \ldots, \overline{e}_k(t)\right], \quad \text{where} \quad \overline{e}_j(t) := \left[e_j(\theta_1), \ldots, e_j(\theta_{\sigma_j^t})\right] \tag{3.3.6}$$

are the data that arrived via the jth channel by the time t:

$$\{\theta_1 < \theta_2 < \cdots < \theta_{\sigma_j^t}\} = \{\theta = 0, 1, \ldots : \theta + \tau_j[\theta, e(\theta)] \leq t\}.$$

The main question to be considered is what is the minimum rate of the information exchange in the system for which stabilization is possible? In other words, we look for necessary and sufficient conditions for stabilizability expressed in terms of the channels *transmission capacities* $\mathfrak{c}_1, \ldots, \mathfrak{c}_k$, along with the plant-sensors parameters A, B, C_j. Roughly speaking, such a capacity is the average number of bits transmitted over the channel during the sample period, despite the losses and delays. The rigorous definition will be offered in Subsect. 3.4.1.

3.4 Basic Definitions and Assumptions

3.4.1 Transmission Capacity of the Channel

It should be remarked that there may be a difference between the number of bits that happen to reach the decoder thanks to occasional favorable circumstances and the number of bits that can be successfully transmitted under any circumstances. In fact, these numbers give rise to two concepts of capacity. The first and second of them are concerned in the necessary and sufficient conditions for stabilizability, respectively. To simplify matters, we postulate that these capacities coincide: The discrepancy between those numbers is considerably less than the time of a long experiment.

We also consider the case where there is an uncertainty about the channel. Specifically, its *regime of operation* given by the distribution of integer transmission delays $\tau_j(t, e)$ over time t and dispatched messages e may not be known in advance. However, we suppose that it satisfies certain assumptions, and the designer of the controller is aware of some lower and upper bounds for the number of bits transmitted across the channel during a time interval of a given duration.

To specify the details, we start with the following.

Definition 3.4.1. *We say that a message* $e(t)$ *is* transmitted within a time interval $[t_0 : t_1]$ *if it departs and arrives at times t and* $t + \tau_j[t, e(t)]$ *from this interval:*

$$t, t + \tau_j[t, e(t)] \in [t_0 : t_1].$$

The *length* or *duration* of a discrete time interval $[t_0 : t_1]$ is defined to be $t_1 - t_0$.

Assumption 3.4.2. *For each channel, two integer functions* $b_j^-(r)$ *and* $b_j^+(r)$ *of time* r *exist such that*

(i) no more than $b_j^+(r)$ *bits are brought by the transmissions that occur within any time interval* $[t_0 : t_1]$ *of length r;*

(ii) given a time interval of duration r, a way exists to transmit without losses and errors no less than $b_j^-(r)$ *bits of information within this interval;*

(iii) as the length r *of the interval increases, the averaged numbers* $b_j^+(r)/r$ *and* $b_j^-(r)/r$ *converge to a common limit*

$$\mathfrak{c}_j = \lim_{r\to\infty} \frac{b_j^-(r)}{r} = \lim_{r\to\infty} \frac{b_j^+(r)}{r} \tag{3.4.1}$$

called the transmission capacity *of the channel.*

Explanation 3.4.3. Claim (i) means that[3] the number of sequences $\overline{e}_j(t_1)$ from (3.3.6), where only messages dispatched after t_0 are admitted, does not exceed $2^{b_j^+(t_1-t_0)}$. In claim (ii), the "way" is constituted by encoding and decoding rules. The former translates b-bit words $\boldsymbol{\beta} = (\beta_1, \dots, \beta_b), \beta_i = 0, 1$ into sequences of messages $e \in \mathfrak{E}_j$ sent consecutively during the interval at hand. The decoder transforms the sequence of messages that arrived within this interval into a b-bit word $\boldsymbol{\beta}'$. The overall transmission must be errorless: $\boldsymbol{\beta}' = \boldsymbol{\beta}$.

We suppose that the designer of the controller is aware of these rules, along with the functions $b_j^-(r), b_j^+(r)$. A regime of the channel operation $\{\tau_j(\cdot)\}$ compatible with these data is said to be *possible*.

[3]For a given regime of the channel operation.

3.4.2 Examples

Now we offer examples of channels satisfying Assumption 3.4.2. (Other examples can be found in [111].) In this subsection, N_j denotes the size of the jth channel alphabet $\mathfrak{E}_j$.

Noiseless instantaneous channel. The channel is constantly accessible; any transmission is successful and instantaneous. Then

$$b_j^-(r) = \lfloor (r+1)\cdot \log_2 N_j \rfloor, \quad b_j^+(r) = \lceil (r+1)\cdot \log_2 N_j \rceil, \quad \mathfrak{c}_j = \log_2 N_j.$$

Explanation 3.4.4. We recall that the symbols $\lfloor a \rfloor$ and $\lceil a \rceil$ stand for the integer floor and ceiling of a real number a, respectively; i.e.,

$$\lfloor a \rfloor := \max\{k = 0, \pm 1, \pm 2, \ldots : k \leq a\}, \qquad \lceil a \rceil := \min\{k = 0, \pm 1, \pm 2, \ldots : k \geq a\}.$$

Periodic-access noiseless instantaneous channel. Within any time interval $[ih_j : (i+1)h_j - 1]$ with a given $h_j \geq 1$, the channel is open for transmission only at times t from the sliding window $t \in W_j + ih_j$, where $W_j \subset [0 : h_j - 1]$ is a given set with $d_j \leq h_j$ elements. Any transmission is successful and instantaneous.[4] Then

$$b_j^+(r) = \left\lceil d_j \left\lceil \frac{r+1}{h_j} \right\rceil \log_2 N_j \right\rceil, \; b_j^-(r) = \left\lfloor d_j \left\lfloor \frac{r}{h_j} \right\rfloor \log_2 N_j \right\rfloor, \; \mathfrak{c}_j = \frac{d_j}{h_j} \log_2 N_j.$$

Aperiodic-access noiseless instantaneous channel. The channel is open for transmission occasionally at times $t_1 < t_2 < \ldots$; any transmission is successful and instantaneous. The access rate stabilizes as the duration of the time interval increases:

$$\mu(t', t'') := (t'' - t')^{-1} \max\{\sigma - \eta : t' \leq t_\eta \leq t_\sigma \leq t''\} \to \mu \quad \text{as} \quad t'' - t' \to \infty.$$

Then

$$b_j^-(r) = \left\lfloor r \log_2 N_j \inf_{t \geq 0} \mu(t, t+r) \right\rfloor, \qquad b_j^+(r) = \left\lceil \log_2 N_j \left(r \cdot \sup_{t \geq 0} \mu(t, t+r) + 1 \right) \right\rceil, \qquad \mathfrak{c}_j = \mu \log_2 N_j.$$

Noiseless instantaneous channel with periodically varying alphabet. At time t, only a part $\mathfrak{E}_j(t) \subset \mathfrak{E}_j$ of the channel alphabet $\mathfrak{E}_j$ can be used.[5] This part varies periodically $\mathfrak{E}_j(t + h_j) = \mathfrak{E}_j(t)$ and contains $N_j(t) \leq N_j$ elements. Transmission of any message $e \in \mathfrak{E}_j(t)$ is successful and instantaneous $\tau(t, e) = 0$. In this case,

[4] This type of communication is typical for control networks [75, 153].

[5] This can be modeled by dropout $\tau_j(t, e) = \infty$ of messages $e \notin \mathfrak{E}_j(t)$.

$$b_j^+(r) = \left\lceil \left\lceil \frac{r+1}{h_j} \right\rceil \sum_{t=1}^{h_j} \log_2 N_j(t) \right\rceil, \quad b_j^-(r) = \left\lfloor \left\lfloor \frac{r}{h_j} \right\rfloor \sum_{t=1}^{h_j} \log_2 N_j(t) \right\rfloor,$$

$$\mathfrak{c}_j = \frac{1}{h_j} \sum_{t=1}^{h_j} \log_2 N_j(t).$$

Noiseless instantaneous channel with aperiodically varying alphabet. In the previous example, the available part of the channel alphabet varies not necessarily periodically. The average number of bits per unit time that can be communicated via the transmission scheme at hand stabilizes as the length of the time interval increases

$$\beta_j(t_0, t_1) := (t_1 - t_0)^{-1} \sum_{t=t_0}^{t_1} \log_2 N_j(t) \to \beta_j^\infty \quad \text{as} \quad t_1 - t_0 \to \infty.$$

In this case,

$$b_j^-(r) = \left\lfloor r \inf_{t \geq 0} \beta_j(t, t+r) \right\rfloor, \quad b_j^+(r) = \left\lceil r \sup_{t \geq 0} \beta_j(t, t+r) \right\rceil, \quad \mathfrak{c}_j = \beta_j^\infty.$$

Constant-access error-corrected instantaneous channel. The transmissions are instantaneous, and the channel is constantly accessible. It is noisy and lossy, but the errors are corrected by means of a block code [68, Ch.12]. Any block has length l_j, encodes $m_j < l_j$ information messages, and contains $l_j - m_j$ check symbols carrying no new information. Those messages will be received simultaneously when the last symbol of the block arrives at the controller. While a current block is transmitted, the next block is being formed. In this case, we have

$$b_j^+(r) = \left\lceil m_j \left\lceil \frac{r+1}{l_j} \right\rceil \log_2 N_j \right\rceil, \quad b_j^-(r) = \left\lfloor m_j \left\lfloor \frac{r+1}{l_j} \right\rfloor \log_2 N_j \right\rfloor,$$

$$\mathfrak{c}_j = \frac{m_j}{l_j} \log_2 N_j.$$

Delayed channels. Suppose that in any of the above examples, the transmission time $\tau_j(t, e) \in [0 : \tau_j^+]$. (In the examples concerning time-varying alphabets, this inclusion should hold only for $e \in \mathfrak{E}_j(t)$.) Then the corresponding formulas for $\mathfrak{c}_j$ and $b_j^+(t)$ remain true, whereas the function $b_j^-(r)$ is transformed by the rule

$$b_j^-(r) := \max\{b_j^-(r - \tau_j^+); 0\}.$$

Conclusion 3.4.5. Bounded time delays do not alter the transmission capacities in the above examples.

3.4.3 Stabilizable Multiple Sensor Systems

Now we introduce the concept of stabilizability examined in this chapter.

Definition 3.4.6. *We say that a controller* (3.3.5) uniformly and exponentially stabilizes *the system* at the rate $\mu \in (0,1)$ *if the corresponding trajectories obey the inequalities*

$$\|x(t)\| \leq K_x \mu^t, \quad \|u(t)\| \leq K_u \mu^t \qquad \forall t = 0,1,2,\dots \tag{3.4.2}$$

whenever $\|x_0\| \leq K_0$*. This must be true irrespective of the possible regime of the channels operation. The constants* K_x *and* K_u *must not depend on time* t *and this regime. The above requirements should be satisfied for arbitrary* K_0*, and the constants* K_x *and* K_u *may depend on* K_0*.*

Definition 3.4.7. *The system is said to be* uniformly and exponentially stabilizable at the rate $\mu \in (0,1)$ *if a controller* (3.3.5) *exists that uniformly and exponentially stabilizes the system at this rate.*

This controller may depend on μ, along with $b_j^+(\cdot), b_j^-(\cdot)$, and A, B, C_j.

Definition 3.4.8. *The system uniformly and exponentially stabilizable at some rate* $\mu \in (0,1)$ *is said to be* uniformly and exponentially stabilizable*. The infimum value of* μ *is called the* rate of exponential stabilizability.

3.4.4 Recursive Semirational Controllers

Formulas (3.3.3) and (3.3.4) describe the widest reasonable class of controllers: The only requirement to them is nonanticipation. Study of this class is instructive with respect to necessary conditions for stabilizability. At the same time, the complexity of the corresponding controllers is not bounded: They are formally permitted to carry out an asymptotically infinite amount of computations per sample period.

It will be shown that conditions necessary for stabilization by means of a nonanticipating controller are simultaneously sufficient for existence of a more realistic stabilizing controller that carries out a bounded (as time progresses) number of operations per unit time and employs a bounded memory. This subsection provides a formal description of the general class of the corresponding controllers.

We start with two preliminary technical definitions.

Definition 3.4.9. *A map* $\mathcal{F}(\cdot)$ *from a subset* D *of an Euclidean space* $\mathbb{R}^s = \{\omega\}$ *into a finite set* $\mathfrak{F}$ *is said to be* semialgebraic *if the following two statements hold:*

(i) The map acts by checking a given set of inequalities

$$l_\nu(\omega) \geq 0, \qquad \nu = 1,\dots,N^{\mathcal{F}} \tag{3.4.3}$$

with (multivariate) rational functions $l_\nu(\cdot)$ *(whose domains include* D*) and forming the sequence of the answers* $\boldsymbol{A} := [A_1,\dots,A_{N^{\mathcal{F}}}]$*, where* $A_\nu := 1$ *if the* ν*th inequality is satisfied and* 0 *otherwise;*

(ii) The value $\mathcal{F}(\omega)$ *is uniquely determined by the distribution* $\boldsymbol{A}$ *of the answers over the set of inequalities:*

$$\mathcal{F}(\omega) := \mathcal{G}(\boldsymbol{A}).$$

Explanation 3.4.10. If the set D is connected, the denominator of the rational function $l_\nu(\omega)$ from (3.4.3) does not change the sign as ω runs over D. So multiplying by this denominator transforms (3.4.3) into an inequality with a polynomial left hand side. This demonstrates that the level sets $\{\omega : \mathcal{F}(\omega) = f\}$ of the map $\mathcal{F}(\cdot)$ are semialgebraic [22], and explains the term introduced in Definition 3.4.9.

Definition 3.4.11. *A map $\mathcal{F}(\cdot)$ from the product $D \times \mathfrak{G}$ of a set $D \subset \mathbb{R}^s$ and a finite set $\mathfrak{G}$ into another finite set $\mathfrak{F}$ is said to be* semialgebraic *if the map $\mathcal{F}(\cdot, g)$ is semialgebraic for any $g \in \mathfrak{G}$.*

Definition 3.4.12. *A map $\mathcal{Z}(\cdot)$ from the product $D \times \mathfrak{G}$ of a set $D \subset \mathbb{R}^s$ and a finite set $\mathfrak{G}$ into an Euclidean space $\mathbb{R}^p$ is said to be* semirational *if this map is representable in the form*

$$\mathcal{Z}[z, g] = \mathcal{Z}_*[z, g, \mathcal{F}(z, g)] \qquad \forall z \in D, g \in \mathfrak{G},$$

where the map $\mathcal{F}(\cdot) : D \times \mathfrak{G} \to \mathfrak{F}$ into a finite set $\mathfrak{F}$ is semialgebraic and the function $\mathcal{Z}(\cdot, g, f)$ is rational for any $g \in \mathfrak{G}, f \in \mathfrak{F}$.

Explanation 3.4.13. A vector function is rational if all its scalar entries are ratios of two polynomials. These ratios should be well defined on D.

Remark 3.4.14. Definition 3.4.12 clearly concerns the case of a map $\mathcal{Z}(\cdot) : D \to \mathbb{R}^p$.

Definition 3.4.15. *The coder* (3.3.3) *is said to be* simple semirational r-step recursive *(where $r = 1, 2, \ldots$) if the following statements hold:*

(i) At any time $t = ir$, the entire code word composed by all code symbols

$$E_i := \big(e_j[ir], e_j[ir+1], \ldots, e_j[(i+1)r-1]\big)$$

that will be consecutively emitted into the channel during the forthcoming operation epoch $[ir : (i+1)r - 1)$ *is generated by the coder;*

(ii) This code word is produced via a recursion of the form:

$$E_i := \mathcal{E}_j(\omega_i), \qquad z_j[(i+1)r] := \mathcal{Z}_j[\omega_i] \in \mathbb{R}^{s_j}, \qquad z_j(0) := z_j^0, \quad \text{where}$$
$$\omega_i := \mathbf{col}\Big(z_j[ir], y_j[(i-1)r+1], \ldots, y_j[ir]\Big),$$
$$(y_j(t) := 0 \; \forall t < 0); \quad (3.4.4)$$

(iii) The functions $\mathcal{E}_j(\cdot)$ and $\mathcal{Z}_j(\cdot)$ are semialgebraic and semirational, respectively.

Explanation 3.4.16. It is assumed that the maps $\mathcal{E}_j(\cdot)$ and $\mathcal{Z}_j(\cdot)$ are defined on the set $D_j \times \mathbb{R}^{r \cdot n_{y,j}}$, where $D_j \subset \mathbb{R}^{s_j}$, the function $\mathcal{Z}_j(\cdot)$ takes values in D_j, and $z_j^0 \in D_j$.

Definition 3.4.17. *The decoder* (3.3.4) *is said to be* simple semirational r-step recursive *if the following statements hold:*

(i) At any time $t = ir$, the decoder generates the control program

$$U_i = \textbf{col}\left(u[ir], \ldots, u[(i+1)r-1]\right)$$

for the entire forthcoming operation epoch;

(ii) This program is produced via a recursion of the form:

$$U_i = \mathcal{U}\Big[z(ir), \overline{\boldsymbol{e}}^r(ir)\Big], \quad z[(i+1)r] = \mathcal{Z}\Big[z(ir), \overline{\boldsymbol{e}}^r(ir)\Big] \in \mathbb{R}^s, \quad z(0) = z^0.$$

Here $\overline{\boldsymbol{e}}^r(ir)$ is the data arrived at the decoder during the previous epoch:

$$\overline{\boldsymbol{e}}^r(ir) := \left[\overline{e}_1^r(ir), \ldots, \overline{e}_k^r(ir)\right], \qquad \overline{e}_j^r(ir) := \left[e_j(\theta_1^r), \ldots, e_j(\theta_{\sigma_j^r}^r)\right],$$
$$\text{where} \quad \{\theta_1^r < \theta_2^r < \cdots < \theta_{\sigma_j^r}^r\} = \{\theta : (i-1)r \leq \theta + \tau_j[\theta, e(\theta)] < ir\}.^6$$

(iii) The maps $\mathcal{U}(\cdot)$ and $\mathcal{Z}(\cdot) \in D$ are semirational functions of $z \in D$ and $\overline{e}^r$. Here $D \subset \mathbb{R}^s$ is some set, and $z^0 \in D$.

Remark 3.4.18. Since the functions $\mathcal{Z}_j(\cdot), \mathcal{E}_j(\cdot), \mathcal{U}(\cdot), \mathcal{Z}(\cdot)$ do not vary as time progresses, the coders and decoders introduced by these definitions perform a limited (as t runs over $t = 0, 1, \ldots$) number of operations per step.

Remark 3.4.19. The special coders and decoder that will be proposed in this chapter to stabilize the plant (3.3.1), (3.3.2) exhibit stable behaviour of the controller inner dynamical variables z_j and z.

Explanation 3.4.20. Definitions 3.4.15 and 3.4.17 are not intended to describe the entire class of controllers with limited algebraic complexities. Their purpose is to underscore critical features of the controller to be proposed that ensure its membership in this class, while omitting many details.

Remark 3.4.21. In fact, Definitions 3.4.15 and 3.4.17 permit both encoding and decoding rules to be altered finitely many times. For example, the recursion from (3.4.4) may be of the form

$$z_j[(i+1)r] := \begin{cases} \mathcal{Z}_j(\omega_i) & \text{if } i \geq k \\ \mathcal{Z}_{j,i}(\omega_i) & \text{if } i < k \end{cases}, \tag{3.4.5}$$

where the functions $\mathcal{Z}_j, \mathcal{Z}_{j,i}(\cdot)$ are semirational. To embed this case into the formalism from Definition 3.4.15, it is sufficient to add one more scalar component $\zeta \in \mathbb{R}$ to the vector z_j, which evolves as the counter: $\zeta_{i+1} := \zeta_i + 1, \zeta_0 := 0$. Corresponding to ζ is the set $\boldsymbol{A}(\zeta) = (A_1, \ldots, A_k)$ of the "answers" A_ν resulting from checking the set of the inequalities $\zeta_i \geq -1/2 + \nu, \nu = 1, \ldots, k$. It remains to note that (3.4.5) is reduced to the form from (3.4.4) by picking in (3.4.4)

[6] If this set is empty (which necessarily holds for $i = 0$), then $\overline{e}_j^r(ir) := \circledast$, where $\circledast$ is a special "void" symbol.

$$\mathcal{Z}_j[\omega,\zeta] := \begin{cases} \mathcal{Z}_{j,0}(\omega) & \text{if } \boldsymbol{A}(\zeta) = (0,\dots,0) \\ \mathcal{Z}_j(\omega) & \text{if } \boldsymbol{A}(\zeta) = (1,\dots,1) \\ \mathcal{Z}_{j,\nu}(\omega) & \text{otherwise} \end{cases},$$

where ν is the serial number of the last affirmative answer: $A_\nu = 1, A_{\nu+1} = 0$.

3.4.5 Assumptions about the System (3.3.1), (3.3.2)

Assumption 3.4.22. *The pair (A, B) is stabilizable.*

The next assumption concerns the subspaces that are not observed and detected, respectively, by a given sensor:

$$L_j^{-o} := \{x \in \mathbb{R}^n : C_j A^\nu x = 0 \qquad \forall \nu \in [0 : n-1]\},$$
$$L_j^{-} := M_{\text{unst}}(A) \cap L_j^{-o}. \quad (3.4.6)$$

Here $M_{\text{unst}}(A)$ is the unstable subspace of the matrix A, i.e., the invariant subspace related to the unstable part $\sigma^+(A) := \{\lambda \in \sigma(A) : |\lambda| \geq 1\}$ of the spectrum $\sigma(A)$.

Definition 3.4.23. *A spectral set $\sigma \subset \sigma(A)$ is said to be* elementary *if it consists of either one real eigenvalue or a couple of conjugate complex ones.*

Any such set is associated with one or more Jordan blocks in the real Jordan representation of the matrix A. These blocks are of the form

$$\begin{pmatrix} \Lambda & I & 0 & 0 & \cdots & 0 \\ 0 & \Lambda & I & 0 & \cdots & 0 \\ 0 & 0 & \Lambda & I & \cdots & 0 \\ 0 & 0 & 0 & \Lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I \\ 0 & 0 & 0 & 0 & \cdots & \Lambda \end{pmatrix}.$$

If σ consists of one real eigenvalue λ, all entries have the size 1×1 and $\Lambda = \lambda$. If σ is composed by a couple of conjugate complex eigenvalues $d(\cos\varphi \pm \imath \sin\varphi)$, the size of all entries is 2×2 and $\Lambda = d \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$.

The symbol M_σ stands for the invariant subspace of A related to σ. We also introduce the subspace $L_j^{-o} \cap M_\sigma$ $(= L_j^- \cap M_\sigma$ if $\sigma \subset \sigma^+(A))$ of states $x \in M_\sigma$ ignored by the jth sensor. All sensors altogether give rise to a whole variety $\{L_j^{-o} \cap M_\sigma\}_{j=1}^k$ of subspaces.

Assumption 3.4.24. *Consider an elementary subset $\sigma \subset \sigma^+(A)$ of the unstable part of the spectrum that gives rise to more than one real Jordan block. The following statement holds for any such a subset:*

(i) The variety $\{L_j^- \cap M_\sigma\}_{j=1}^k$ *of unobservable subspaces of* M_σ *has the atomic structure: The space* M_σ *can be decomposed into a direct sum*

$$M_\sigma = M_\sigma^1 \oplus \cdots \oplus M_\sigma^{m_\sigma} \tag{3.4.7}$$

of atom subspaces M_σ^i *so that any unobservable subspace* $L_j^- \cap M_\sigma$ *is the sum of several atoms*

$$L_j^-(\sigma) := L_j^- \cap M_\sigma = \oplus_{i \in I(j)} M_\sigma^i, \qquad \text{where} \quad I(j) \subset [1 : m_\sigma]. \tag{3.4.8}$$

Here the set $I(j)$ may be empty.[7] Assumption 3.4.24 is trivially satisfied whenever there is no elementary spectral set σ with the properties described in its preamble. In other words, the following claim holds.

Remark 3.4.25. Assumption 3.4.24 is valid whenever any elementary subset $\sigma \subset \sigma^+(A)$ of the unstable part of the spectrum gives rise to only one real Jordan block.

This clearly holds if the matrix A has no multiple unstable eigenvalues. As is well known, the last property is true for almost all (with respect to the Lebesgue measure) matrices A. So Assumption 3.4.24 is valid for most square matrices A.

At the same time, the matrix A has multiple unstable eigenvalues for interesting application examples. Among them, there is the simplest dynamical system $\ddot{x} = u$. Indeed, suppose that the control is constant on any sample period of duration Δ, denote $v := \dot{x}$, and consider the trajectory only at sample times:

$$x(\tau) := x(\tau\Delta), \quad v(\tau) := v(\tau\Delta), \quad u(\tau) := u(\tau\Delta + 0). \tag{3.4.9}$$

Then

$$x(\tau+1) = x(\tau) + \Delta \cdot v(\tau) + \frac{\Delta^2}{2} u(\tau), \quad v(\tau+1) = v(\tau) + \Delta \cdot u(\tau).$$

So the matrix $A = \left(\begin{smallmatrix} 1 & \Delta \\ 0 & 1 \end{smallmatrix}\right)$ has the eigenvalue 1 of multiplicity 2. Observe that this eigenvalue gives rise to only one Jordan block. So Assumption 3.4.24 is fulfilled, although the matrix has multiple unstable eigenvalues.

Remark 3.4.26. Assumption 3.4.24 also holds if sensor detects or ignores any of the aforementioned subspaces M_σ only completely:

$$\text{either} \quad M_\sigma \cap L_j^- = \{0\} \quad \text{or} \quad M_\sigma \subset L_j^- \tag{3.4.10}$$

for any "unstable" $\sigma \subset \sigma^+(A)$ elementary set σ and sensor j.

In this case, the decomposition (3.4.7) is trivial: $m_\sigma = 1, M_\sigma^1 = M_\sigma$.

It should be noted that the condition (3.4.10) is sufficient but not necessary for Assumption 3.4.24 to hold, as is demonstrated by the following.

[7]The direct sum over the empty set is defined to be the trivial subspace $\{0\}$.

Example 3.4.27. Consider the system whose dynamical matrix is a Jordan block

$$\begin{array}{l} x_1(t+1) = \lambda x_1(t) + \quad 0 \quad + b_1^{\mathsf{T}} u(t) \\ x_2(t+1) = \lambda x_2(t) + \; x_1(t) \; + b_2^{\mathsf{T}} u(t) \\ \qquad \vdots \\ x_d(t+1) = \lambda x_d(t) + x_{d-1}(t) + b_d^{\mathsf{T}} u(t) \end{array}, \quad \begin{array}{l} y_1(t) = x_1(t) \\ y_2(t) = x_2(t) \\ \qquad \vdots \\ y_d(t) = x_d(t) \end{array}, \quad |\lambda| > 1. \tag{3.4.11}$$

The system is served by d sensors. Their unobservable subspaces are, respectively,

$$L_1^- := \{x : x_1 = 0\}, \quad L_2^- = \{x : x_1 = x_2 = 0\}, \ldots, \\ \ldots, L_{d-1}^- = \{x : x_1 = \cdots = x_{d-1} = 0\}, \quad L_d^- = \{0\}. \tag{3.4.12}$$

So (3.4.10) does not hold for the only elementary set $\sigma = \{\lambda\}$ and any sensor except for the dth one. At the same time, Assumption 3.4.24 is true by Remark 3.4.25.

It should be noted that the statement (i) from Assumption 3.4.24 necessarily holds for any elementary set $\sigma \subset \sigma^+(A)$ giving rise to only one real Jordan block.[8] This observation, first, explains why such sets are not considered in Assumption 3.4.24, and, second, it demonstrates that the property (i) is natural.

A typical example of the situation forbidden by this assumption is as follows:

$$x(t+1) = \lambda x(t) + u(t) \in \mathbb{R}^2, \qquad \lambda > 1, \quad \text{where} \quad x = (x_1, x_2). \tag{3.4.13}$$

There are three sensors

$$y_1(t) = x_1(t), \quad y_2(t) = x_2(t), \quad y_3(t) = x_1(t) - x_2(t), \tag{3.4.14}$$

whose nondetectable subspaces equal the following lines, respectively,

$$L_1^- = \{x : x_1 = 0\}, \quad L_2^- = \{x : x_2 = 0\}, \quad L_3^- = \{x : x_1 = x_2\}$$

(see Fig. 3.3). There is only one elementary spectral set $\sigma = \{\lambda\}$, which gives rise to two Jordan blocks of size 1×1. It is easy to see that the plane $M_\sigma = \mathbb{R}^2$ cannot be decomposed into a direct sum (3.4.7) so that its special partial sub-sums give each of the three lines $L_i^-, i = 1, 2, 3$ from Fig. 3.3, as is required by Assumption 3.4.24.

Assumption 3.4.24 is technical and is imposed to simplify matters. The case where this assumption is violated will be addressed in Sect. 3.9.

3.5 Main Result

The stabilizability conditions to be presented are constituted by a set of inequalities. These inequalities can be enumerated by groups of sensors $J \subset [1 : k]$. The inequality depends on the group via the space of states nondetectable by this group:

[8] See Lemma 3.8.3 on p. 66.

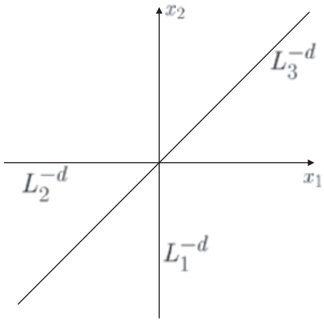


Fig. 3.3. The unobservable subspaces.

$$L(J) := \bigcap_{j \in J} L_j^- . \tag{3.5.1}$$

We recall that L_j^- is given by (3.4.6). For consistency, we assign the unstable subspace $M_{\text{unst}}(A)$ to the empty group.

Different groups J may produce a common space $L(J)$ and thus a common inequality. So it is beneficial to parametrize the inequalities not by the groups of sensors but by the nondetectable subspaces (3.5.1). To this end, we introduce the following.

Notation 3.5.1. *The set of all subspaces $L \subset \mathbb{R}^n$ of the form $L = L(J)$ except for $L = \{0\}$ is denoted by $\mathfrak{L} = \{L\}$. (Here J runs over all groups of sensors J.)*

As discussed, the size of $\mathfrak{L}$ may be less than the number of all such groups.

Now we are in a position to state the main result of the chapter.

Theorem 3.5.2. *Suppose that Assumptions* 3.4.2, 3.4.22, *and* 3.4.24 *(on pp.* 43 *and* 49 *) hold. Then the following two statements are equivalent:*

(i) The system (3.3.1), (3.3.2) *is uniformly and exponentially stabilizable;*[9]

(ii) For every subspace (3.5.1) $L \in \mathfrak{L}$ *constituted by all states nondetectable by a certain group of sensors, the following inequality holds:*

$$\log_2 |\det A|_L| < \sum_{j \notin J(L)} \mathfrak{c}_j, \qquad \text{where}$$
$$J(L) := \{j = 1, \dots, k : C_j x = 0 \quad \forall x \in L\} . \tag{3.5.2}$$

Here $A|_L$ is the operator A acting in its invariant subspace L, the sum is over the sensors that do not completely ignore the subspace L at hand, and $\mathfrak{c}_j$ is the transmission capacity (3.4.1) *of the jth channel.*

Now suppose that the equivalent claims (i) and (ii) are true. If the matrix A has no stable eigenvalues $\sigma^-(A) := \{\lambda \in \sigma(A) : |\lambda| < 1\} = \emptyset$, the rate μ^0 of exponential stabilizability of the system is given by

[9]See Definition 3.4.8 on p. 46.

$$\log_2 \mu^0 = \max_{L \in \mathfrak{L}} \frac{1}{\dim L} \left(\log_2 |\det A|_L| - \sum_{j \notin J(L)} \mathfrak{c}_j \right). \tag{3.5.3}$$

Moreover, this formula is true whenever $\sigma^-(A) \neq \emptyset$ *and the quantity in the right-hand side is no less than* $\max_{\lambda \in \sigma^-(A)} |\lambda|$.

The proof of this theorem will be given in Sects. 3.7 and 3.8. In Sect. 3.7, we prove its necessity part (i) ⇒ (ii). The converse (i) ⇐ (ii) is established in Sect. 3.8, where formula (3.5.3) is also justified.

Remark 3.5.3. The quantity $\log_2 |\det A|_L|$ from (3.5.2) and (3.5.3) equals the topological entropy[10] of the linear system

$$x(t+1) = Ax(t), \quad x(t) \in L, \quad x(1) \in \mathfrak{X}_1 \subset L,$$

where $\mathfrak{X}_1 \subset L$ is an arbitrary compact set for which 0 is an interior (in L) point.

This holds thanks to Theorem 2.4.2 (on p. 21), since the subspaces $L \in \mathfrak{L}$ are A-invariant and all eigenvalues λ of $A|_L$ are unstable $|\lambda| \geq 1$ by (3.4.6) and (3.5.1).

By Remark 3.5.3, the quantity $\log_2 |\det A|_L|$ represents the unit time increment of the number of bits required to describe the state of the open-loop ($u(\cdot) \equiv 0$) system (3.3.1) considered on the invariant subspace L. At the same time, the right-hand side of the inequality from (3.5.2) can be interpreted as the joint capacity of all channels except for those carrying no information about the state $x \in L$. (The latter channels serve the sensors that completely ignore such states $C_j x = 0 \; \forall x \in L$.) Thus the condition (3.5.2) means that the amount of information concerning the state $x \in L$ that the decoder may receive over all channels for the unit time exceeds the unit time growth of the number of bits required to describe the state to a given high accuracy. It should be noted here that some bits counted in (3.5.2) characterize the state $x \in L$ only partly. They correspond to any sensor whose outputs are not sufficient to reconstruct the entire state $x \in L$. Moreover all sensors may be of such a kind. Nevertheless, when inequalities (3.5.2) are taken for all subspaces $L \in \mathfrak{L}$, they constitute a sufficient and necessary criterion for stabilizability.

In general, the number of inequalities (3.5.2) does not exceed that of "unstable" invariant subspaces. It also does not exceed the number 2^k of all groups J of sensors. Moreover, the inequalities may be directly parameterized by these groups:

$$\log_2 \left|\det A|_{L(J)}\right| < \sum_{j \notin J} \mathfrak{c}_j. \tag{3.5.4}$$

Here $L(J)$ is given by (3.5.1) and J ranges over all subsets $J \subset [1:k]$ except those for which $L(J) = \{0\}$. As discussed, (3.5.4) may contain more inequalities than (3.5.2) does. Indeed, if $L(J_1) = L(J_2)$ for $J_1 \neq J_2$, then $L(J_1) = L(J_2) = L(J_1 \cup J_2)$ and so the sets $J_1, J_2, J_1 \cup J_2$ are served by a single inequality in (3.5.2). At the same time, they give rise to three inequalities in (3.5.4), with those for $J = J_1$ and $J = J_2$ being trivial consequences of the inequality with $J = J_1 \cup J_2$.

[10]See Definition 2.3.3 on p. 16.

Generally speaking, relations (3.5.2) are not independent. However, revealing "superfluous" inequalities is usually a harder task than direct verification of the entire inequality set (3.5.2).

Formula (3.5.3) holds only under special circumstances described in Theorem3.5.2. In general, the conditions for exponential stabilizability at the rate μ are structurally altered as μ passes any element ρ_ν of the set

$$\left\{|\lambda| : \lambda \in \sigma^-(A)\right\} = \left\{\rho_1 > \rho_2 > \ldots > \rho_p\right\}, \tag{3.5.5}$$

while continuously running over $[0, 1)$. The point is that the "modes" $x \in M_{\sigma^\nu}$ from the invariant subspace M_{σ^ν} related to the spectral set

$$\sigma^\nu := \left\{\lambda \in \sigma(A) : |\lambda| = \rho_\nu\right\}$$

are stabilized at any rate $\mu > \rho_\nu$ "for free." At the same time, their stabilization at a rate $\mu < \rho_\nu$ requires special efforts and so demands communication resources, which causes alteration of the stabilizability conditions as μ passes ρ_ν.

To extend formula (3.5.3) to the general case, we introduce the following analogs of the subspaces (3.5.1):

$$L^\nu(J) := M_{\sigma^\nu} \cap \bigcap_{j \in J} L_j^{-o},$$

where the unobservable subspace L_j^{-o} of the jth sensor is given by (3.4.6) (on p.49). For consistency, we assign $L^\nu(\emptyset) := M_{\sigma^\nu}$ and put $L^0(J) := L(J)$.

An exhaustive characterization of the rate of exponential stabilizability is offered by the following proposition. To simplify notations, its statement proceeds from the stabilizability criterion in the form (3.5.4).

Proposition 3.5.4. *Suppose that Assumption* 3.4.2 *(on p.* 43*) and (i) of Theorem* 3.5.2 *(on p.* 52 *) hold, the pair* (A, B) *is controllable, and Assumption* 3.4.24 *(on p.* 49*) is true for any elementary spectral set* $\sigma \subset \sigma(A)$. *Let* ν_* *denote the maximal integer* $\nu = 1, \ldots, p$ *for which the following set of inequalities holds:*

$$\left(\sum_{\alpha=0}^{\nu-1} \dim L^\alpha(J)\right) \log_2 \rho_\nu > \sum_{\alpha=0}^{\nu-1} \log_2 \left|\det A|_{L^\alpha(J)}\right| - \sum_{j \notin J} \mathfrak{c}_j$$
$$\forall J \in \Xi^\nu := \left\{J \subset [1:k] : L^0(J) \oplus L^1(J) \oplus \cdots \oplus L^\nu(J) \neq \{0\}\right\}, \tag{3.5.6}$$

and $\nu_* := 0$ *if such an integer does not exist.*[11]

The rate μ^0 *of the exponential stabilizability of the system is given by*

$$\log_2 \mu^0 = \max_{J \in \Xi^{\nu_*}} \frac{1}{\sum_{\alpha=0}^{\nu_*} \dim L^\alpha(J)} \left[\sum_{\alpha=0}^{\nu_*} \log_2 \left|\det A|_{L^\alpha(J)}\right| - \sum_{j \notin J} \mathfrak{c}_j\right]. \tag{3.5.7}$$

[11]We suppose that the last case holds whenever $\sigma^-(A) = \emptyset$.

The proof of this proposition will be given in Sect. 3.8.

Remark 3.5.5. It is easy to see that $\log_2 \left|\det A|_{L^\alpha(J)}\right| = \dim L^\alpha(J) \cdot \log_2 \rho_\alpha$ for $\alpha \geq 1$. It follows that, first, whenever (3.5.6) holds for some $\nu \geq 1$, it also is true for all lesser ν and, second, $\mu^0 \leq \rho_{\nu_*}$, where $\rho_0 := 0$.

3.5.1 Some Consequences from the Main Result

Remark 3.5.6. The conditions (3.5.2) imply that the system is detectable via the entire set of the sensors.

Indeed, otherwise (3.5.2) fails to be true for

$$L := \bigcap_{j=1}^{k} L_j^- \neq \{0\}$$

since then the sum in (3.5.2) is over the empty set, which is defined to be 0.

Remark 3.5.7. If the system is detectable by each sensor $L_j^- = \{0\}\ \forall j$, the set $\mathfrak{L}$ contains only one unstable space $M_{\text{unst}}(A)$ and so (ii) reduces to only one inequality,

$$\log_2 |A|_{M_{\text{unst}}(A)}| < \mathfrak{c} := \sum_{j=1}^{k} \mathfrak{c}_j.$$

The sum $\mathfrak{c}$ can be interpreted as the capacity of the channel composed of all channels at hand. At the same time, $\log_2 |A|_{M_{\text{unst}}(A)}|$ equals the topological entropy $H(A)$ of the open-loop plant (3.3.1) by Theorem 2.4.2 (on p. 21). So the inequality is in harmony with Theorem 2.5.3 (on p. 26) concerning the case of one perfect channel.

3.5.2 Complement to the Sufficient Conditions

The sufficiency part (ii) $\Rightarrow$ (i) of Theorem 3.5.2 can be enhanced by the following.

Proposition 3.5.8. *Suppose that Assumptions* 3.4.2, 3.4.22*, and* 3.4.24 *(on pp.* 43 *and* 49 *) hold and (ii) of Theorem* 3.5.2 *(on p.* 52*) is true. Then the system* (3.3.1)*,* (3.3.2) *is uniformly and exponentially stabilizable by means of simple semirational* r*-step recursive coders and a decoder.*[12]

Moreover, coders and decoders of such a kind fit to uniformly and exponentially stabilize the system at any rate μ *exceeding the rate* μ^0 *of exponential stabilizability* $\mu > \mu^0$*, which is given by* (3.5.3).

The proof of this proposition will be given in Sect. 3.8, where a stabilizing controller will be described explicitly. It will be shown that the controller exhibits stable behavior of its inner dynamical variables z_j and z from (3.4.4) and (ii) of Definition 3.4.17 (see p. 47).

Remark 3.5.9. The step r can be chosen common for all coders and the decoder.

[12] See Definitions 3.4.15 and 3.4.17 on p. 47.

3.5.3 Complements to the Necessary Conditions

The (i) $\Rightarrow$ (ii) part of Theorem 3.5.2 can be complemented by the following facts.

Remark 3.5.10. The implication (i) $\Rightarrow$ (ii) remains true even if Assumption 3.4.24 (on p. 49) is dropped.

This easily follows from the proof of this implication presented in Sect. 3.7.

If in (3.5.2) the nonstrict inequality sign is substituted in place of the strict one, the resultant inequalities form necessary conditions for the property that is weaker than stabilizability. The rigorous statement of this fact is given by the following.

Lemma 3.5.11. *Let a controller* (3.3.5) *exist that makes the trajectories of the closed-loop system bounded*

$$\sup_{\|x_0\|\leq K_0, t=0,1,\ldots} \|x(t)\| < \infty \qquad \forall K_0 > 0. \tag{3.5.8}$$

Then (ii) of Theorem 3.5.2 *holds with* $<$ *replaced by* $\leq$ *in* (3.5.2)*.*

Proof. We first show that the property (3.5.8) can be extended on the control:

$$\sup_{\|x_0\|\leq K_0, t=0,1,\ldots} \|u(t)\| < \infty \qquad \forall K_0 > 0. \tag{3.5.9}$$

To this end, we modify the decoder (3.3.4) by putting $\mathcal{U}(\cdot) := \pi\mathcal{U}(\cdot)$, where π is the orthogonal projection from $\mathbb{R}^n_u$ onto the orthogonal complement $M_u := (\ker B)^\perp$ to the kernel $\ker B := \{u : Bu = 0\}$. This modification does not alter the action of the controller on the plant since $B = B\pi$. So (3.5.8) is kept true. At the same time, a constant $c > 0$ exists such that $\|u\| \leq c\|Bu\|$ whenever $u \in M_u$. So

$$\|u(t)\| \leq c\|Bu(t)\| \overset{(3.3.1)}{=\!=} c\,\|x(t+1) - Ax(t)\| \leq c\,\|x(t+1)\| + c\,\|A\|\,\|x(t)\|\,.$$

Thus we see that (3.5.8) entails (3.5.9).

Now we pick $\mu \in (0, 1)$. The transformation

$$z(t) := \mu^t x(t), \quad v(t) := \mu^t u(t) \tag{3.5.10}$$

establishes a one-to-one correspondence between the trajectories $\{x(t), u(t)\}$ and $\{z(t), v(t)\}$ of the open-loop systems given by, respectively, (3.3.1) and the equation

$$z(t+1) = \mu Az(t) + \mu Bv(t), \qquad z(0) = x_0. \tag{3.5.11}$$

We equip the latter with the sensors $\widetilde{y}_j = C_j z, j \in [1 : k]$, the coders

$$e_j(t) = \mathcal{E}_j[t, \widetilde{y}_j(0), \mu^{-1}\widetilde{y}_j(1), \mu^{-2}\widetilde{y}_j(2), \ldots, \mu^{-t}\widetilde{y}_j(t)],$$

and the decoder

$$v(t) = \mu^t\mathcal{U}\left[t, \overline{\boldsymbol{e}}(t)\right].$$

Here $\mathcal{E}_j(\cdot)$ and $\mathcal{U}(\cdot)$ are the parts of the controller making the trajectories of the original system bounded. It is easy to see that (3.5.10) still holds for the closed-loop systems. Then Definition 3.4.6 (on p. 46) implies that the proposed coders and decoder uniformly and exponentially stabilize the system (3.5.11) at the rate μ. The proof is completed by applying the (i) $\Rightarrow$ (ii) part of Theorem 3.5.2 to the system (3.5.11) and letting $\mu \to 1 - 0$. $\square$

3.6 Application of the Main Result to the Example from Sect. 3.2

In this section, we justify the statements from Sect. 3.2. We recall that they concern a platoon of k vehicles described by (3.2.1). Each of them is equipped with a sensor giving the distance $y_i = x_i - x_{i-1}$ from it to the preceding vehicle for $i \geq 2$ and the position $y_1 = x_1$ for $i = 1$. It is also served by a communication channel with the transmission capacity $\mathfrak{c}_i > 0$ carrying signals to the controller with the sample period $\Delta > 0$. The objective is to stabilize the platoon motion about a given constant-velocity trajectory: $v_i = v_i^0, x_i(t) = x_i^0 + v_i^0 t \; \forall i$.

The substitution of the variables

$$v_i := v_i - v_i^0, \quad x_i := x_i - x_i^0 - v_i^0 t$$

keeps the dynamics equations unchanged and shapes the control goal into $x_i = 0, v_i = 0$. To put the problem into the discrete-time framework adopted in this chapter, we consider the trajectory only at sample times; i.e., we introduce the variables (3.4.9), where now x, v, and u are marked by the lower index $_i$. We also modify the sensor output by subtracting the sensor signal corresponding to the desired trajectory. Then

$$x_i(\tau+1) = x_i(\tau) + \Delta \cdot v_i(\tau) + \frac{\Delta^2}{2} u_i(\tau), \quad v_i(\tau+1) = v_i(\tau) + \Delta \cdot u_i(\tau),$$
$$y_i(\tau) = x_i(\tau) - x_{i-1}(\tau), \quad (3.6.1)$$

where $x_0 := 0$. Now

$$A = \begin{pmatrix} A_1 & 0 & 0 & \cdots & 0 \\ 0 & A_2 & 0 & \cdots & 0 \\ 0 & 0 & A_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A_k \end{pmatrix}, \qquad A_i = \mathcal{A} = \left(\begin{smallmatrix} 1 & \Delta \\ 0 & 1 \end{smallmatrix}\right).$$

The unique elementary spectral set $\sigma = \{1\}$ gives rise to k Jordan blocks. The nonobservable and nondetectable subspaces (3.4.6) coincide and equal

$$L_j^- = \{\mathfrak{x} := \textbf{col}\,(z_1, w_1, \ldots, z_k, w_k) : z_j = 0, w_j = 0\},$$
$$\text{where} \quad z_i := x_i - x_{i-1}, \quad w_i := v_i - v_{i-1},$$

and $x_0 := v_0 := 0$. Assumption 3.4.24 holds with

$$M_\sigma^i := \{\mathfrak{x} : z_j = 0, w_j = 0 \; \forall j \neq i\}, \qquad i = 1, \ldots, m_\sigma := k.$$

Assumption 3.4.22 is immediate from (3.6.1). Since $|\det A|_L| = 1$ for any invariant subspace L, Theorem 3.5.2 guarantees that

The platoon is uniformly and exponentially stabilizable under arbitrary transmission capacities $\mathfrak{c}_i > 0$.

To determine the rate of stabilizability μ^0, note that the states nondetectable by (maybe, empty) group $\mathfrak{J} \subset [1:k]$ of sensors constitute the subspace

$$L(\mathfrak{J}) := \{\mathfrak{x} : z_j = 0, w_j = 0 \ \forall j \in \mathfrak{J}\}.$$

Since $\dim L(\mathfrak{J}) = 2(k - |\mathfrak{J}|)$, where $|\mathfrak{J}|$ is the size of $\mathfrak{J}$, relation (3.5.3) shapes into

$$\log_2 \mu^0 = \max_{\mathfrak{J}} \frac{1}{2(k - |\mathfrak{J}|)} \left(- \sum_{j \notin \mathfrak{J}} \mathfrak{c}_j \right) = -\frac{1}{2} \mathfrak{c}_{\min},$$
$$\text{where} \quad \mathfrak{c}_{\min} := \min_{j=1,\ldots,k} \mathfrak{c}_j. \qquad (3.6.2)$$

It follows that

For the platoon at hand, the rate of the exponential stabilizability equals $\sqrt{2}^{\,-\mathfrak{c}_{\min}}$ *per sample period.*

Now consider the situation where the sensor system accommodated by each vehicle gives the distances to $l < k$ vehicles to the right, as well as to l vehicles to the left. (We assume an imaginary vehicle that is numbered by 0 and stays at the origin.) Then clearly

$$L_j^- = \Big\{\mathfrak{x} : z_i = 0, w_i = 0 \quad \forall i = \max\{j - l + 1, 1\}, \ldots, \min\{j + l, k\}\Big\}.$$

Assumption 3.4.24 remains true with the same subspaces M_σ^i, and the platoon evidently remains uniformly and exponentially stabilizable. What can be said about the rate of stabilizability? Now the collection $\mathfrak{L}$ introduced in Notation 3.5.1 (on p. 52) consists of spaces $L(\mathfrak{J})$ related to sets $\mathfrak{J}$, which along with any element $j \in \mathfrak{J}$ contain a certain interval of the form

$$[i - l + 1 : i + l] \cap [1 : k] \ni j, \qquad i = 1, \ldots, k.$$

Such sets are said to be *wide*. To proceed, we consider separately two cases.

1. Let $2l \geq k$. Then any two aforementioned intervals contain a common point. It follows that apart from $\mathfrak{J} = \emptyset$, the wide sets $\mathfrak{J}$ are only intervals of the form

$$[1 : i], \quad i \geq l + 1, \qquad \text{or} \qquad [i : k], \quad i \leq k - l + 1.$$

By retracing (3.6.2), we see that $\mu^0 = \sqrt{2}^{\,-\mathfrak{c}_{k,l}}$, where $\mathfrak{c}_{k,l}$ is given by (3.2.2) (on p. 41).

2. Now let $2l < k$. Then the sets

$$[1 : i - 1] \cup [i + 1 : k], \qquad i = l + 2, \ldots, k - l,$$
$$[i : k], \qquad i = 2, \ldots, l + 2,$$
$$[1 : i], \qquad i = k - l, \ldots, k - 1$$

are wide. By restricting the maximum in (3.6.2) to only these sets $\mathfrak{J}$, we see that $\mu^0 \geq \sqrt{2}^{-\mathfrak{c}_{k,l}}$, where $\mathfrak{c}_{k,l}$ is given by (3.2.2) (on p. 41). In fact, $\mu^0 = \sqrt{2}^{-\mathfrak{c}_{k,l}}$. To prove this, it suffices to show that

$$\frac{1}{k - |\mathfrak{J}|} \sum_{j \notin \mathfrak{J}} \mathfrak{c}_j \geq \mathfrak{c}_{k,l} \tag{3.6.3}$$

for any wide set $\mathfrak{J}$. To this end, we put

$$i_- := \min\{j : j \in \mathfrak{J}\} \quad \text{and} \quad i_+ := \max\{j : j \in \mathfrak{J}\}.$$

Then

$$i_- \leq l + 1 \Rightarrow [i_- : l + 1] \subset \mathfrak{J} \quad \text{and} \quad i_+ \geq k - l + 1 \Rightarrow [k - l + 1 : i_+] \subset \mathfrak{J}$$

by the definition of the wide set. Hence

$$\{j : j \notin \mathfrak{J}\} = \mathfrak{J}_1 \cup \cdots \cup \mathfrak{J}_s,$$

where the sets $\mathfrak{J}_\nu$ are pairwise disjoint and each of them has the form

$$\mathfrak{J}_\nu = \{i\}, \quad i = l + 2, \ldots, k - l, \qquad \text{or}$$
$$\mathfrak{J}_\nu = [1 : i], \quad i = 1, \ldots, l + 1, \qquad \text{or}$$
$$\mathfrak{J}_\nu = [i : k], \quad i = k - l + 1, \ldots, k.$$

By (3.2.2) (on p. 41),

$$\frac{1}{|\mathfrak{J}_\nu|} \sum_{j \in \mathfrak{J}_\nu} \mathfrak{c}_j \geq \mathfrak{c}_{k,l}.$$

This implies (3.6.3) as follows:

$$\frac{1}{k - |\mathfrak{J}|} \sum_{j \notin \mathfrak{J}} \mathfrak{c}_j = \frac{1}{k - |\mathfrak{J}|} \sum_{\nu=1}^{s} |\mathfrak{J}_\nu| \frac{1}{|\mathfrak{J}_\nu|} \sum_{j \in \mathfrak{J}_\nu} \mathfrak{c}_j \geq \mathfrak{c}_{k,l}.$$

3.7 Necessary Conditions for Stabilizability

3.7.1 An Extension of the Claim (i) from Theorem 2.5.3 (on p. 26)

This theorem deals with an instantaneous and lossless channel. Now we show that its statement (i) remains true for delayed and lossy channels considered in this chapter. For technical reasons, which will become clear soon, we extend the class of systems and consider ones of the following form

$$x(t + 1) = Ax(t) + \mathcal{B}[t, u(0), \ldots, u(t)], \tag{3.7.1}$$

where $\mathcal{B}(\cdot)$ is a given function. We also recall that the matrix A is unstable.

Lemma 3.7.1. *Suppose that there is only one channel* $k = 1$. *Then* $H(A) < \mathfrak{c} := \mathfrak{c}_1$ *whenever the system* (3.7.1) *is uniformly and exponentially stabilizable. Here* $H(A)$ *is the topological entropy* (2.4.2) *(see p.* 21*) of the open-loop system* (3.3.1)*.*

Proof. Suppose first that $H(A) > \mathfrak{c}$. By putting

$$v(t) := \mathcal{B}[t, u(0), \dots, u(t)], \qquad \mathfrak{X}_1 := \{x : \|x\| \leq 1\}, \qquad \text{and} \quad B := I,$$

we shape (3.7.1) into (2.5.1) (on p. 24) with $u(t) := v(t)$. From this point, the proof proceeds by merely retracing the arguments from the proof of (i) from Theorem 2.5.3 (see p. 30), where $T := 1$ and $R \in (\mathfrak{c}, H(A))$ are taken. However when the set $\widehat{S}$ of all possible control sequences $v(0), \dots, v(T)$ is examined, it should be noted that now its size $|\widehat{S}|$ does not exceed $2^{b_1^+(T)}$ by (i) of Assumption 3.4.2 (on p. 43). At the same time, $\lim_{T\to\infty} b_1^+(T)/T = \mathfrak{c} < R$ due to (3.4.1) (on p. 43). So $\frac{\log_2 |\widehat{S}|}{T} < R$ if T is large enough, which keeps all arguments from the proof of (i) of Theorem 2.5.3 true and so entails a contradiction to the stabilizability of the system.

Thus $H(A) \leq \mathfrak{c}$. To prove the strict inequality $H(A) < \mathfrak{c}$, we employ the hint similar to that from the proof of Lemma 3.5.11 (on p. 56). In other words, we consider the process $x_*(t) := \varkappa^t x(t)$. It clearly satisfies the equation of the form (3.7.1)

$$x_*(t+1) = \varkappa A x_*(t) + \varkappa^{t+1}\mathcal{B}[t, u(0), \dots, u(t)].$$

To keep this system stable, the constant $\varkappa > 1$ is chosen so that $\varkappa\mu < 1$, where μ is taken from (3.4.2) (on p. 46). Then by the foregoing, $H(\varkappa A) \leq \mathfrak{c}$. It remains to note that $H(\varkappa A) > H(A)$ owing to (2.4.2) (on p. 21). □

Lemma 3.7.1 evidently remains true if the regime of the channel operation (given by $\tau_1(\cdot)$) is known in advance.

3.7.2 An Auxiliary Subsystem

To prove (3.5.2) (on p. 52), we revert to the system (3.3.1) and pick a subspace $L \in \mathfrak{L}$ constituted by the states not detectable by a certain group of sensors. Then we restrict ourselves to trajectories $\{x(t)\}_{t=0}^{\infty}$ starting at $x_0 \in L$ and apply Lemma 3.7.1 to them. More precisely, we take into account that $\{x(t)\}$ may leave L due to controls, and we consider $x_L(t) := \pi x(t)$. Here π is a projector from $\mathbb{R}^n$ onto L. It is easy to check that the evolution of x_L is governed by the equation of the form (3.7.1):

$$x_L(t+1) = A|_L x_L(t) + \pi B u(t) + \sum_{i=0}^{t-1} (\pi A - A\pi) A^{t-1-i} B u(i),$$
$$x_L(t) \in L, \qquad x_L(0) = x_0. \quad (3.7.2)$$

Remark 3.7.2. The first equation simplifies if $\pi A = A\pi$. However, such a projector π exists if and only if an A-invariant subspace exists that is complementary to L, which is not true in general.

Notation 3.7.3. *For $I \subset [1:k]$, the symbol C_I denotes the block matrix that results from arranging the blocks C_i with $i \in I$ into a column.*

For any entities v_i enumerated by $i \in [1:k]$, the symbol v_I is defined likewise.

We interpret (3.7.2) as equations of an imaginary system and equip it with the sensor

$$y_L(t) = C_{J^c} x_L(t). \tag{3.7.3}$$

Here $J^c := \{j : j \notin J\}$ is the complement to the set

$$J = \{j : C_j x = 0 \ \forall x \in L\}$$

of sensors ignoring the subspace L. We also suppose that all channels with $j \notin J$ are commissioned to transmit y_L.

The sum in the right-hand side of (3.5.2) (on p. 52) equals the capacity of the union of these channels. Hence (3.5.2) follows from Lemma 3.7.1 applied to the system (3.7.2), (3.7.3). To complete the proof, it suffices to show that this system is stabilizable whenever the original one (3.3.1), (3.3.2) can be stabilized. In doing so, one must cope with the fact that the trajectory of the original closed-loop system (3.3.1)–(3.3.4) may leave the subspace L. So the observations (3.3.2) and the entries of (3.7.3) may differ. Moreover, the sensors omitted in (3.7.3) may see the state $x(t)$ for $t \geq 1$. It should be shown that they are yet useless and can be dropped.

3.7.3 Stabilizability of the Auxiliary Subsystem

Lemma 3.7.4. *Let the system* (3.3.1), (3.3.2) *be exponentially stabilized by some controller, and let the regime of the channels operation be known in advance. Then the system* (3.7.2), (3.7.3) *is also exponentially stabilizable.*

Proof. We first show that for $x_0 \in L$, the process in the original closed-loop system obeys the relations

$$\begin{aligned} e_J(t) &= \mathcal{E}'_j\left[t, \overline{e}_{J^c}(t-1)\right], \\ y_J(t) &= \mathcal{Y}'\left[t, \overline{e}_{J^c}(t-1)\right], \qquad y_{J^c}(t) = y_L(t) + \mathcal{Y}''\left[t, \overline{e}_{J^c}(t-1)\right]. \end{aligned} \tag{3.7.4}$$

We recall that the data $\overline{e}_j(t)$ that arrived via the jth channel by time t is given by (3.3.6) (on p. 42). The observation $y_L(t)$ is defined by (3.7.2) and (3.7.3) for the sequence of controls $u(t)$ identical to that driving the original system.

For $t = 0$, we have

$$x(0) \in L \Rightarrow y_J(0) = 0 \quad \text{and} \quad y_{J^c}(0) = y_L(0).$$

So (3.7.4) with $t = 0$ follows from (3.3.3). Now suppose that (3.7.4) with $t := \theta$ holds for all $\theta \leq t$. Then

$$\overline{e}_J(\theta) = \overline{\mathcal{E}}'\left[\theta, \overline{e}_{J^c}(t-1)\right]$$

and so

$$\overline{e}(\theta) = [\overline{e}_J(\theta), \overline{e}_{J^c}(\theta)] = \overline{\mathcal{E}}\big[\theta, \overline{e}_{J^c}(t)\big] \overset{(3.3.4)}{\Longrightarrow} u(\theta) = \mathcal{U}'\big[\theta, \overline{e}_{J^c}(t)\big] \tag{3.7.5}$$

for $\theta \leq t$. Now we invoke (3.3.1) and note that

$$x_0 \in L \Rightarrow A^{t+1}x_0 \in L \Rightarrow C_J A^{t+1} x_0 = 0 \quad \text{and} \quad (I - \pi)A^{t+1}x_0 = 0.$$

As a result, we see that

$$y_J(t+1) = \underbrace{C_J A^{t+1} x_0}_{=0} + \sum_{\theta=0}^{t} C_J A^{t-\theta} B u(\theta) =: \mathcal{Y}'\big[t+1, \overline{e}_{J^c}(t)\big],$$
$$y_{J^c}(t+1) - y_L(t+1) = C_{J^c}\big[x(t+1) - \pi x(t+1)\big] = \underbrace{C_{J^c}(I-\pi)A^{t+1}x_0}_{=0}$$
$$+ C_{J^c} \sum_{\theta=0}^{t} (I-\pi)A^{t-\theta} B u(\theta) =: \mathcal{Y}''\big[t+1, \overline{e}_{J^c}(t)\big];$$

i.e., the last two relations from (3.7.4) do hold with $t := t+1$. Then the first relation follows from (3.3.3).

It follows from (3.3.3) and (3.7.4) that the signal $e_{J^c}(t)$ is determined by the prior measurements from (3.7.3),

$$e_{J^c}(t) = \mathcal{E}_L[t, y_L(0), \ldots, y_L(t)].$$

Now we interpret this as the equation of the coder and the last relation from (3.7.5) (where $\theta := t$) as that of the decoder for the system (3.7.2), (3.7.3). By the foregoing, this coder–decoder pair generates the trajectory $\pi x(t), u(t), t = 0, 1, \ldots$, where $x(t), u(t)$ is the trajectory of the original closed-loop system. So the inequalities (3.4.2) (on p. 46) are inherited by the system (3.7.2), (3.7.3), which completes the proof. □

3.7.4 Proof of the Necessity Part (i) ⇒ (ii) of Theorem 3.5.2

As discussed, this implication is immediate from Lemmas 3.7.1 and 3.7.4.

3.8 Sufficient Conditions for Stabilizability

In this section, we suppose that the assumptions and (ii) of Theorem 3.5.2 (on p. 52) hold, and until stated otherwise, adopt one more assumption.

Assumption 3.8.1. *The system* (3.3.1) *has no stable* $|\lambda| < 1$ *eigenvalues* λ.

In the general case, a stabilizing controller will be obtained by applying that presented below to the unstable part of the system.

3.8.1 Some Ideas Underlying the Design of the Stabilizing Controller

To stabilize the system, we employ the scaled quantization scheme (see, e.g., [28,73, 135,136,149,184,204]). It was mainly developed for only one channel and is briefly as follows. Both coder and decoder compute a common upper bound δ of the current state norm $\|x\|_\infty := \max_i |x_i|$. They are also given a partition of the unit ball into m balls (cubes) with small radii $\leq \varsigma(m)$. The number m matches the channel capacity so that the serial number of the cube can be communicated to the decoder. The coder determines the current state from the observations and notifies the decoder which cube contains this state divided by δ. Since the decoder knows δ, it thus becomes aware of a ball B with the radius $\leq \delta\varsigma(m)$ containing the current state. Then it selects a control that drives the system from the center of this ball to zero. The ball itself is expanded because of the unstable dynamics of the system and transformed into a set $D_+(B)$ centered about zero: $D_+(B) \subset B_0^\alpha$. So the radius α can be taken as a new upper bound δ. Here $\alpha \leq \delta\varsigma(m)\mu$, where μ characterizes the expansion rate of the system. If $\varsigma(m)\mu < 1$, the bound δ is thus improved $\delta := \delta\varsigma(m)\mu < \delta$ and by continuing likewise, it is driven to zero $\delta \to 0$, along with the state x.

In the context of this paper, a problem with the above scheme is that no coder may be aware of the entire state x. So a natural idea [204] is to disintegrate the system (3.3.1) into subsystems each observable by some sensor. Then each subsystem can be stabilized by following the above lines, provided the stability condition $\varsigma(m)\mu < 1$ holds for it. In fact, this condition means that there is a way to communicate a sufficiently large amount of information from the subsystem to the decoder: the smaller the radius $\varsigma(m)$, the larger the size m of the partition, and so the larger the number of bits required to describe which of m cubes contains the state.

No channel in itself may meet the above stability condition. At the same time, this condition may be met if several channels are commissioned to transmit information about a given subsystem. Then each channel may carry only a part of this information, whereas the decoder assembles these parts, thus getting the entire message (see Fig. 3.4). Certainly, these channels should be chosen among those serving the sensors that observe the subsystem at hand.

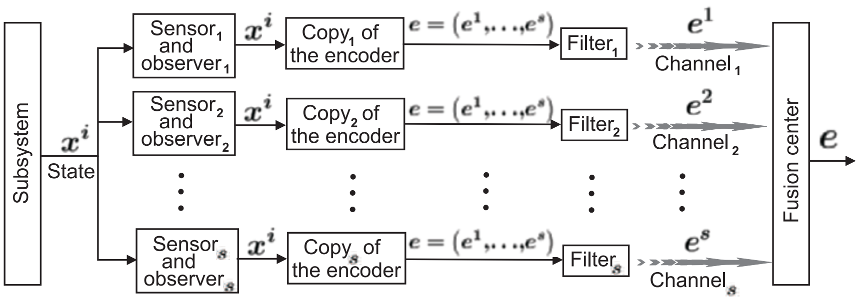


Fig. 3.4. Distribution of data over parallel channels.

Since a given sensor may observe several subsystems, the above scheme means that each channel must transmit a set of messages each concerning a particular subsystem (see Fig. 3.5a). As a result, each subsystem is served by a variety of channels, whereas every channel is fed by several subsystems (see Fig. 3.5b). This gives rise to

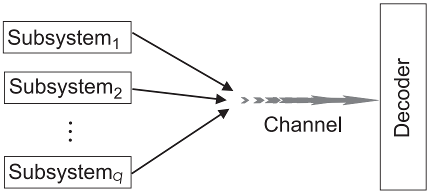


Fig. 3.5(a). Data transfer over a given channel.

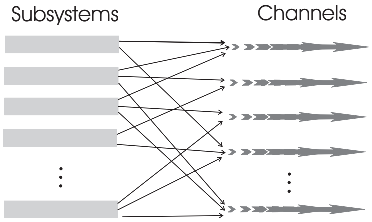


Fig. 3.5(b). Data transfer over all channels.

the question: Is it possible to distribute the required information about each particular subsystem over parallel channels in such a way that the total amount of information carried via every channel meets its capacity? It will be shown via convex duality arguments that the answer is in the affirmative whenever (ii) of Theorem 3.5.2 holds.

Another problem is how to employ a sensor observing the subsystem only partly. Certainly, this problem does not hold if there are no such sensors and subsystems: The state of each subsystem either is completely determined from or does not affect the outputs of any given sensor. Decomposition into a set of such subsystems is possible. However, in general, these subsystems are dependent. The reasons for this are twofold. First, the control is common. Second, Jordan blocks may entail an unavoidable interinfluence between the states of the subsystems.

To illustrate the last claim, we invoke the system (3.4.11) (on p. 51) whose dynamical matrix is the standard Jordan block. The unobservable subspaces of the sensors are given by (3.4.12). There are no other invariant proper subspaces.[13] So this system cannot be decomposed into subsystems with independent open-loop dynamics. At the same time, any sensor except for the last one observes the state x only partly. So to exclude such a partial vision, disintegration into state-dependent subsystems is unavoidable.

To deal with them, we employ *sequential stabilization*. We define the sth subsystem as that described in the sth row from (3.4.11). Its state is x_s. Then we stabilize the first subsystem, which is independent of the others. This makes x_1 exponentially decaying. In the equations of the second subsystem, we interpret x_1 as an exogenous disturbance. By constructing a device stabilizing this subsystem under any exponentially vanishing disturbances, we make x_2 exponentially decaying. The entire system is stabilized by continuing likewise.

These arguments, however, do not take into account that the control affects all subsystems, and some of them may be unstabilizable (although the entire system

[13]See the proof of Lemma 3.8.3 on p. 66.

is controllable). For example, the subsystems with $s \geq 2$ are unstabilizable if $b_1 = 1, b_2 = \cdots = b_d = 0$ in (3.4.11). This obstacle can be easily overcome via increasing the sample period.

Indeed, let us pick $r = 1, 2, \ldots$. The state $x_i := x(ir)$ evolves as follows:

$$x_{i+1} = A^r x_i + \mathfrak{B}\boldsymbol{U}^i, \tag{3.8.1}$$

where

$$\boldsymbol{U}^i := \textbf{col}\left[u(ir), u(ir+1), \ldots, u(ir+r-1)\right] \quad \text{and}$$

$$\mathfrak{B}\boldsymbol{U} := \sum_{j=0}^{r-1} A^{r-1-j} B u_j \tag{3.8.2}$$

is the state to which the control program $\boldsymbol{U} = \textbf{col}\left[u_0, \ldots, u_{r-1}\right]$ drives the system at time $t = r$ from $x(0) = 0$. Since the system (3.3.1) with no stable modes is controllable by Assumption 3.4.22 (on p. 49), the operator $\mathfrak{B}$ is onto if $r \geq n$. Related to the decomposition of the system state $x = \textbf{col}\,(x^1, \ldots, x^d)$ into the states x^i of subsystems is a block partition

$$\mathfrak{B}\boldsymbol{U} = \textbf{col}\left[\mathfrak{B}_1\boldsymbol{U}, \ldots, \mathfrak{B}_d\boldsymbol{U}\right].$$

Since all operators $\mathfrak{B}_s$ have full rank, any subsystem is controllable. Moreover, any control action $y = \mathfrak{B}_s\boldsymbol{U}$ in the sth subsystem can be implemented by a control $\boldsymbol{U}$ that does not disturb the other subsystems: $\mathfrak{B}_j\boldsymbol{U} = 0 \; \forall j \neq s$.

3.8.2 Plan of Proving the Sufficiency Part of Theorem 3.5.2

Summarizing, we adopt the following plan.

Step 1. We decompose the system so that, first, for any given sensor, the state of each subsystem either does not affect or is determined from the sensor outputs and, second, the decomposition is triangular. The latter permits us to employ the sequential stabilization approach.

Step 2. We increase the sample period and, for each subsystem, offer a class of controllers stabilizing it under any exponentially decaying disturbance. In doing so, we assume that the coder is aware of the current state at any sample time $t = ir$, and there is a way to communicate as much information as desired from the coder to the decoder.

Step 3. Within the above class, we point out the controller that requires a nearly minimum bit-rate of such a communication.

Step 4. We show that if all subsystems are equipped with these controllers, the entire system is stabilized.

Step 5. We obtain conditions under which the entire set of these controllers can be implemented[14] by means of real channels and sensors. These conditions are not constructive and require that a linear system of inequalities be solvable in integers.

[14] In particular, the required information traffic can be arranged.

Step 6. By employing convex duality arguments, we show that these conditions are equivalent to (ii) of Theorem 3.5.2.

Step 7. We drop Assumption 3.8.1 and show that (ii) of Theorem 3.5.2 suffices to stabilize the system with both unstable and stable eigenvalues (modes).

3.8.3 Decomposition of the System

Now we perform step 1 of the above plan. In other words, we represent the system as a set of subsystems interacting in a special manner. The main result is as follows.

Proposition 3.8.2. *Suppose that Assumption* 3.8.1 *(on p.* 62*) holds. Then after a proper one-to-one linear transformation and partition of the state*

$$x = \mathbf{col}\,(x^1, \dots, x^d) \tag{3.8.3}$$

into several blocks $x^s \in \mathbb{R}^{n_s}$ interpreted as the states of subsystems, the following statements hold:

(i) The unobservable subspace (3.4.6) *(see p.* 49*)* $L_j^{-o} = L_j^-$ *of any sensor is composed of several blocks:*

$$L_j^- = \{x : x^s = 0 \; \forall s \in O_j\}, \qquad \text{where} \quad O_j \subset [1:d];$$

(ii) The block representation of the dynamics equations (3.3.1) *is lower triangular:*

$$x^s(t+1) = \sum_{i=1}^{s} A_{si} x^i(t) + B_s u(t), \qquad s = 1, \dots, d. \tag{3.8.4}$$

By (i), the states x^s of subsystems $s \notin O_j$ do not affect the outputs of the jth sensor, whereas the states x^s with $s \in O_j$ are uniquely determined from these outputs.

The remainder of the subsection is devoted to the proof of Proposition 3.8.2. We start with two technical facts.

Lemma 3.8.3. *The claim (i) of Assumption* 3.4.24 *(on p.* 49*) holds for any elementary spectral set σ.*

Proof. In view of Assumption 3.4.24, it suffices to prove the lemma assuming that the set σ gives rise to only one real Jordan block. We put $M := M_\sigma, A_\sigma := A|_M$ and note that $\det[\lambda I - A_\sigma] = \varphi(\lambda)^p$, where the polynomial φ is irreducible over the field of real numbers. By employing the basis in M reducing A_σ to the real Jordan form, it is easy to see that the formula $\mathcal{L}(\nu) := \ker \left[\varphi(A_\sigma)\right]^\nu$ produces $(p+1)$ distinct

$$\{0\} = \mathcal{L}(0) \subset \mathcal{L}(1) \subset \cdots \subset \mathcal{L}(p) = M$$

invariant subspaces and $\dim \mathcal{L}(\nu) = \nu \deg \varphi$. We are going to show that there are no other invariant subspaces.

Indeed let L be such a subspace and ψ be the minimal annihilating polynomial of L. Then ψ is a divisor of φ^p and so $\psi = \varphi^\nu, \nu = 0, \dots, p$. Hence

$$L \subset \ker \left[\varphi(A_\sigma)\right]^\nu = \mathcal{L}(\nu).$$

At the same time, Theorem 2 of [61, p. 180] implies that $\dim L = \deg \psi$. Thus $\dim L = \nu \deg \varphi = \dim \mathcal{L}(\nu)$, and so $L = \mathcal{L}(\nu)$.

As a result, we see that all invariant subspaces $L_j^- \cap M_\sigma$ are among $\mathcal{L}(0), \dots, \mathcal{L}(p)$. It remains to pick $M_\sigma^1 := \mathcal{L}(1), m_\sigma := p$, and for $i = 2, \dots, p$, choose M_σ^i so that $\mathcal{L}(i-1) \oplus M_\sigma^i = \mathcal{L}(i)$. □

The next lemma plays the key role in the proof of Proposition 3.8.2.

Lemma 3.8.4. *In Assumption* 3.4.24 *(on p.* 49*), the atoms* $M_\sigma^i, i = 1, \dots, m_\sigma$*, can be chosen so that all partial direct sums of the form*

$$M_\sigma^1 \oplus \cdots \oplus M_\sigma^i, \qquad i = 1, \dots, m_\sigma,$$

are A*-invariant.*

Proof. We consider the set of atoms with the minimal size m_σ. We also introduce the undetectable subspaces $L_j := L_j^- \cap M_\sigma$ of M_σ, then we form all their intersections

$$L^\cap = L_{j^1} \cap \cdots \cap L_{j^p},$$

and then we form all algebraic sums (not necessarily direct) of such intersections

$$L^\Sigma = L_{i^1}^\cap + \cdots + L_{i^r}^\cap.$$

Here p, r and the subspaces $L_{j^\nu}, L_{i^\mu}^\cap$ are chosen arbitrarily. Let $\mathfrak{M}$ denote the set of all L^Σ's. It is clear that

1) Any space $L \in \mathfrak{M}$ is invariant and decomposable into a direct sum of several atoms;
2) $L \in \mathfrak{M} \Rightarrow L \cap L_j \in \mathfrak{M} \quad \forall j$;
3) $M_\sigma \in \mathfrak{M}$;
4) $L', L'' \in \mathfrak{M} \Rightarrow L' + L'' \in \mathfrak{M}$;
5) The set $\mathfrak{M}$ is finite.

Now we pick a minimal element $L_{\min}$ among $L \in \mathfrak{M}, L \neq \{0\}$, i.e., such that

$$L \subset L_{\min} \ \&\ L \in \mathfrak{M} \ \&\ L \neq \{0\} \Rightarrow L = L_{\min}.$$

By trying here $L := L_{\min} \cap L_j$, we see that either $L_{\min} \subset L_j$ or $L_{\min} \cap L_j = \{0\}$. Hence any L_j contains either all atoms constituting $L_{\min}$ or none of them. So these atoms can be replaced by their sum in (i) of Assumption 3.4.24. Since the number of all atoms is minimal, only one atom is concerned: $L_{\min} = M_\sigma^\nu$. By permuting the atoms, we set $\nu = 1$. Then the claim of the lemma does hold for $i = 1$ by 1).

Now let $L_{\min}$ denote a minimal element among $L \in \mathfrak{M}$ such that $L \supset M_\sigma^1$ and $L \neq M_\sigma^1$. By 2) and 4),

$$L := M_\sigma^1 + L_j \cap L_{\min} \in \mathfrak{M}.$$

So the minimum property yields that either $L = M_\sigma^1$ or $L_{\min} \subset L$. In terms of the decomposition from 1)

$$L_{\min} = M_\sigma^1 \oplus M_{\min} \in \mathfrak{M}$$

(where $M_{\min}$ is the sum of several atoms), this means that either $M_{\min} \cap L_j = \{0\}$ or $M_{\min} \subset L_j$. Like above, this implies that $M_{\min}$ consists of only one atom M_σ^ν. By permuting the atoms, we set $\nu = 2$, thus making the claim of the lemma true for $i = 2$ by 1). The proof is completed by continuing likewise. □

Proof of Proposition 3.8.2. We decompose the spectrum

$$\sigma(A) = \sigma^1 \cup \cdots \cup \sigma^p$$

into the union of disjoint elementary sets. Then

$$\mathbb{R}^n = M_{\sigma^1} \oplus \cdots \oplus M_{\sigma^p},$$

and any invariant subspace L_j^- is decomposed

$$L_j^- = L_j^-(1) \oplus \cdots \oplus L_j^-(p)$$

into the invariant subspaces $L_j^-(\nu) := L_j^- \cap M_{\sigma^\nu}$. So it suffices to show that, for any ν, linear coordinates in M_{σ^ν} and their block partition exist for which any subspace $L_j^-(\nu), j = 1, \ldots, k$, is the direct sum of several "blocks" and the operator $A|_{M_{\sigma^\nu}}$ has a lower triangular form with respect to this partition. These blocks z^i are in fact given by Lemma 3.8.4:

$$z^i \in M_{\sigma^\nu}^{m_{\sigma^\nu} - i + 1}.$$

More precisely, it suffices to pick a basis in each subspace $M_{\sigma^\nu}^i, i = 1, \ldots, m_{\sigma^\nu}$, unite them to produce a basis in M_{σ^ν}, and then consider the coordinates with respect to this basis and their partition that corresponds to the partition $z = z^1 + \cdots + z^{m_{\sigma^\nu}}$ of z into $z^i \in M_{\sigma^\nu}^{m_{\sigma^\nu} - i + 1}$. □

3.8.4 Separate Stabilization of Subsystems

In this subsection, we perform step 2 (see p. 65) of the plan from Subsect. 3.8.2.

Introducing Subsystems

We pick an integer parameter r and focus attention only on the states at times $\tau_i = i \cdot r$. The evolution of these states is given by (3.8.1) (on p. 65), which evidently inherits the lower triangular structure from (3.8.4),

$$x^s_{i+1} = \sum_{\nu=1}^{s} A^{(r)}_{s\nu} x^\nu_i + \mathfrak{B}_s \boldsymbol{U}_i \tag{3.8.5}$$

for $s = 1, \dots, d$. Here

$$x^s_i := x^s(\tau_i), \qquad \boldsymbol{U}_i = \textbf{col}\left[u(\tau_i), \dots, u(\tau_i + r - 1)\right],$$

and the diagonal coefficients from (3.8.5) are the rth powers of the matching coefficients from (3.8.4), $A^{(r)}_{ss} = A^r_{ss}$.

The *sth subsystem* is described by the following equations:

$$x^s_{i+1} = A^r_{ss} x^s_i + \mathfrak{B}_s \boldsymbol{U}_i + \xi_{s,i}, \qquad i = 0, 1, \dots. \tag{3.8.6}$$

Here in accordance with (3.8.5),

$$\xi_{s,i}(t) \equiv 0 \quad \text{for} \quad s = 1 \quad \text{and} \qquad \xi_{s,i}(t) = \sum_{\nu=1}^{s-1} A^{(r)}_{s\nu} x^\nu_i \quad \text{otherwise.} \tag{3.8.7}$$

In this subsection, we ignore this rule and interpret $\xi_{s,i}(t)$ as an exogenous disturbance. This permits us to study each subsystem independently of the others. We also suppose that the disturbance decays at a known rate ρ_ξ:

$$|\xi_{s,i}| \le K_\xi \rho^i_\xi, \qquad \rho_\xi \in [0, 1), \quad i = 0, 1, \dots \tag{3.8.8}$$

with K_ξ unknown, and offer a controller that stabilizes the sth subsystem under all such disturbances. In doing so, we assume that the current state x^s_i is measured online. The proposed controller uses only finitely many bits of information about x^s_i.

Remark 3.8.5. The controller will be mainly based on the ideas from [28, 73, 77, 135, 136, 149, 184, 201, 204]. Major distinctions concern two points. First, we take into account exogenous disturbances decaying at a known rate. Second, we consider the case where the transmission of the above bits to the decoder takes some time (specifically, r units of time). This implies complements to the stabilization scheme, e.g., the need to quantize not the current state but the state prognosis.

We first introduce components of which the coder and the decoder will be assembled.

Quantizer

To communicate a continuous sensor data over a discrete (digital) channel, an analog-to-digital converter is required.

Definition 3.8.6. *An m-level* quantizer $\mathfrak{Q}^s$ *in $\mathbb{R}^{n_s}$ is a partition of the closed unit ball $\overline{B^1_0} \subset \mathbb{R}^{n_s}$ with respect to some norm $\|\cdot\|$ into m disjoint sets $Q_1, \dots, Q_m$ each equipped with a* centroid $q^{Q_i} \in Q_i$.

Such a quantizer converts any vector $x^s \in Q_i$ into its *quantized value* $\mathfrak{Q}^s(x^s) := q^{Q_i}$ and any vector $x^s \notin \overline{B^1_0}$ outside the unit ball into an *alarm symbol* $\mathfrak{Q}^s(x^s) := \maltese$.

Definition 3.8.7. *The quantizer is said to be* r-contracted *(for the* s*th subsystem) if*

$$A_{ss}^{r}\left(Q - q^{Q}\right) \subset \rho_{\mathfrak{Q}^s} B_0^1 \; \forall Q = Q_i,\; i = 1, \ldots, m, \text{ where } \quad \rho_{\mathfrak{Q}^s} \in (0,1). \quad (3.8.9)$$

The constant $\rho_{\mathfrak{Q}^s} \in (0,1)$ *is called the* contraction rate.

Definition 3.8.8. *The quantizer is said to be* polyhedral *if the quantizer map* $\mathfrak{Q}^s(\cdot)$ *is semialgebraic*[15] *with linear functions* $l_i(x^s) = a_i^{\mathrm{T}} x^s + b_i, i \in [1 : M^s]$ *in* (3.4.3) *(on p.* 46*).*

In other words, such a quantizer acts by checking the linear inequalities $l_i(x^s) \geq 0, i \in [1 : M^s]$ and forming the tuple $\boldsymbol{A}(x^s)$ whose ith entry is 1 (yes) if the inequality is satisfied and 0 (no) otherwise. The quantizer output is uniquely determined by the distribution of answers (yes, no) over the set of inequalities, i.e., by this tuple.

Remark 3.8.9. Not only nonstrict but also strict inequalities can be considered here. This holds since the results of checking the strict $l_i(x^s) > 0$ and nonstrict $-l_i(x^s) \geq 0$ inequalities, respectively, are uniquely determined from each other by the negation.

Remark 3.8.10. For polyhedral quantizers, any level domain Q_i (along with $Q_{\maltese} := \{x^s : \mathfrak{Q}^s(x^s) = \maltese\}$) is the union of a finite number of convex polyhedra.

Note also that for such quantizers, the ball from Definition 3.8.6 is a convex polytop.

Example 3.8.11. One of the simplest examples of a quantizer is given by the uniform partition of the square with the side length 2 into $m = N^2$ congruent subsquares (see Fig. 3.6, where $N = 12$). The norm in Definition 3.8.6 is given by $\|x\|_\infty = \max\{|x_1|, |x_2|\}, x = \textbf{col}\,(x_1, x_2)$. The centroid is the center of the corresponding subsquare. This quantizer is polyhedral and served by $2(N+1)$ linear inequalities:

$$x_1 \geq -1 + (i-1)\frac{2}{N},\; i = 1, \ldots, N, \qquad x_1 > 1,$$
$$x_2 \geq -1 + (i-1)\frac{2}{N},\; i = 1, \ldots, N, \qquad x_2 > 1. \quad (3.8.10)$$

In this case, it is convenient to split the tuple $\boldsymbol{A}(x)$ of answers into two subtuples $\boldsymbol{A}_1(x)$ and $\boldsymbol{A}_2(x)$ related to the first and second subsets of inequalities, respectively. If all entries of either $\boldsymbol{A}_1(x)$ or $\boldsymbol{A}_2(x)$ equal each other (all are 1 or all are 0), the quantizer output is $\maltese$. Otherwise

$$\mathfrak{Q}^s(x) = \textbf{col}\left(-1 + [i_1 - 1]\frac{2}{N} + \frac{1}{N}, -1 + [i_2 - 1]\frac{2}{N} + \frac{1}{N}\right),$$

where i_ν (for $\nu = 1, 2$) is the maximal serial number of the inequality for which the affirmative answer 1 is written in the tuple $\boldsymbol{A}_\nu(x)$.

[15] See Definition 3.4.9 on p. 46.

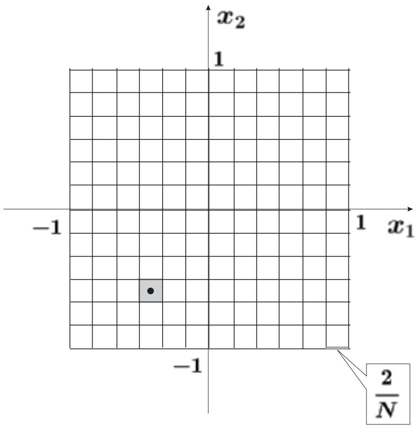


Fig. 3.6. Uniform square quantizer.

Explanation 3.8.12. In (3.8.10), the strict inequalities are taken for $i = N$ to make the quantizer's *effective domain* $\{x : \mathfrak{Q}^s(x) \neq \maltese\}$ closed, as is required by Definition 3.8.6.

Remark 3.8.13. Example 3.8.11 is easily generalized on the case of the general Euclidean space $\mathbb{R}^{n_s}$. Then the uniform partition of the cube with side length 2 into N^{n_s} congruent subcubes is performed. The centroid is the center of the corresponding subcube. This quantizer is still polyhedral.

Remark 3.8.14. The last example entails that for any r, a polyhedral r-contracted quantizer exists.

Indeed, the set $Q - q^Q, Q = Q_i$ from (3.8.9) is the cube with side length $2/N$ centered about zero. Equivalently, this is the ball of radius $1/N$ with respect to the norm

$$\|x^s\|_\infty := \max\left\{|x_1^s|, \dots, |x_{n_s}^s|\right\}, \qquad x^s = \mathbf{col}\left(x_1^s, \dots, x_{n_s}^s\right). \tag{3.8.11}$$

So to ensure (3.8.9), it suffices to pick $N > \|A_{ss}^r\|_\infty$, where $\|\cdot\|_\infty$ is the matrix norm matching (3.8.11): $\|\mathcal{A}\|_\infty := \max_{\|x^s\|_\infty = 1} \|\mathcal{A}x^s\|_\infty$.

Assumption 3.8.15. *In the remainder of this subsection, we suppose that a polyhedral r-contracted quantizer is given (for any r).*

As will be shown, the number of its levels determines the communication bit rate required for stabilization. So an r-contracted quantizer with the minimal number of levels is of special interest. Such a quantizer will be offered in Subsect. 3.8.5.

Deadbeat Stabilizer

This is a linear transformation $\mathcal{N}$ of an initial state x_0^s into a control program $\boldsymbol{U}$ that drives the unperturbed $\xi_{s,i} \equiv 0$ subsystem (3.8.6) to zero,

$$0 = x_1^s (= x^s(r)) = A_{ss}^r x_0^s + \mathfrak{B}_s \boldsymbol{U} \quad \text{for} \quad \boldsymbol{U} := \mathcal{N} x_0^s \quad \text{and any} \quad x_0^s. \qquad (3.8.12)$$

A particular deadbeat stabilizer with advanced properties[16] will be proposed in Subsect. 3.8.6.

Assumption 3.8.16. *In the remainder of this subsection, we suppose that a deadbeat stabilizer is given.*

Parameters

Apart from r, the controller employs two more parameters ρ and γ chosen so that

$$r > n, \qquad \gamma > \|A_{ss}\|^r, \qquad \text{and} \qquad 1 > \rho > \max\{\rho_\xi, \rho_{\mathfrak{Q}^s}\}, \qquad (3.8.13)$$

where A_{ss}, ρ_ξ, and $\rho_{\mathfrak{Q}^s}$ are taken from (3.8.4), (3.8.8), and (3.8.9), respectively.

Description of the Coder and Decoder

Both coder and decoder compute controls $\boldsymbol{U}_i^c$, $\boldsymbol{U}_i^d$ and upper bounds δ_i^c, δ_i^d for the state norm $\|x_i^s\|$, respectively. Actually, acting upon the plant is the control $\boldsymbol{U}_i^d$. The initial bounds are common: $\delta_0^c = \delta_0^d = \delta_0 > 0$. (The inequality $\delta_0 \geq \|x_0^s\|$ may be violated.) At any time $\tau_i = ir$, the coder selects a finite-bit message based on x_i^s and sends it to the decoder. We suppose that this message arrives by time τ_{i+1}.

Specifically, the coder and decoder operate as follows.

The sth coder (at the times $t = \tau_i, i = 1, 2, \ldots$)

c.1) Proceeding from the knowledge of the current state x_i^s, computes the prognosis $\widehat{x}_{i+1}^s$ of the state at $t = \tau_{i+1}$,

$$\widehat{x}_{i+1}^s := A_{ss}^r x_i^s + \mathfrak{B}_s \boldsymbol{U}_i^c; \qquad (3.8.14)$$

c.2) Employs the r-contracted quantizer $\mathfrak{Q}^s$ to compute the quantized value q_i of the scaled state at $t = \tau_{i+1}$,

$$\boldsymbol{\varepsilon}_i := \left[\delta_i^c\right]^{-1} \widehat{x}_{i+1}^s, \qquad q_i := \mathfrak{Q}^s\left[\boldsymbol{\varepsilon}_i\right]; \qquad (3.8.15)$$

c.3) Encodes this quantized value q_i for transmission and sends it to the decoder;

c.4) Computes the next control program by means of the deadbeat stabilizer $\mathcal{N}$ and corrects the upper bound,

$$\boldsymbol{U}_{i+1}^c := \mathcal{N}\left[\delta_i^c \overset{\star}{q}_i\right], \qquad \delta_{i+1}^c := \delta_i^c \times \langle q_i \rangle_{\rho,\gamma}, \qquad \text{where} \qquad (3.8.16)$$

$$\overset{\star}{q} := \begin{cases} q & \text{if } q \neq \maltese, \\ 0 & \text{otherwise}, \end{cases} \qquad \langle q \rangle_{\rho,\gamma} := \begin{cases} \rho \text{ if } q \neq \maltese, \\ \gamma \text{ otherwise}, \end{cases}. \qquad (3.8.17)$$

[16] These properties are beneficial when all subsystems are stabilized simultaneously.

The sth decoder (at the times $t = \tau_i, i = 2, 3, \ldots$)

d.1) Decodes the newly received data and thus acquires q_{i-1};
d.2) Computes the current control program and corrects the upper bound,

$$\boldsymbol{U}_i^d := \mathcal{N}\big[\delta_i^d \overset{\star}{q}_{i-1}\big], \qquad \delta_{i+1}^d := \delta_i^d \times \langle q_{i-1} \rangle_{\rho,\gamma}. \tag{3.8.18}$$

Remark 3.8.17. For definiteness, the initial control programs $\boldsymbol{U}_0^c, \boldsymbol{U}_0^d, \boldsymbol{U}_1^c, \boldsymbol{U}_1^d$ are taken to be zero.

Explanation 3.8.18. We introduced separate controls $\boldsymbol{U}_i^c, \boldsymbol{U}_i^d$ and bounds δ_i^c, δ_i^d to stress that the coder and decoder compute them independently. However, it easily follows from (3.8.16), (3.8.18) and induction on i that they in fact coincide,

$$\delta_i^d = \delta_{i-1}^c, \qquad \boldsymbol{U}_i^d = \boldsymbol{U}_i^c, \qquad i = 1, 2, \ldots. \tag{3.8.19}$$

Remark 3.8.19. The second relation from (3.8.19) implies that the error in the state prognosis (3.8.14) is equal to the disturbance from (3.8.6),

$$\widehat{x}_{i+1}^s = x_{i+1}^s - \xi_{s,i}. \tag{3.8.20}$$

Explanation 3.8.20. To be communicated across the channel, the quantized value q_i should be encoded into a code word $\big(e_s[ir], e_s[ir+1], \ldots, e_s[(i+1)r-1]\big) \in \mathfrak{E}_s^r$, whose symbols are consecutively emitted into the sth channel during the forthcoming operation epoch $[ir : (i+1)r)$.

Remark 3.8.21. Let m denote the number of the quantizer levels.[17] By **c.3)**, the coder sends on average $\frac{\log_2(m+1)}{r}$ bits per unit time to the decoder.

In Subsect. 3.8.5, the minimum of this ratio over r-contracted quantizers will be studied.

Observation 3.8.22. *As follows from Definitions* 3.4.15 *and* 3.4.17 *(on p.* 47*), the proposed coder and decoder are simple semirational one-step recursive.*

We recall that the sample period was increased so that now one step is equivalent to r former ones.

Remark 3.8.23. The inner dynamical variables $z^s (= z_s)$ and z of the coder and decoder (see Definitions 3.4.15 and 3.4.17) can be defined as $z_i^s := (\delta_i^c, \boldsymbol{U}_i^c), z_i := \delta_i^d$.

Stabilizing Properties of the Coder and Decoder

These properties are revealed by the following main result of the subsection.

Proposition 3.8.24. *Suppose that Assumption* 3.8.1 *(on p.* 62*) and* (3.8.13) *hold, and the disturbance in the sth subsystem* (3.8.6) *satisfies* (3.8.8)*. Then the above coder–decoder pair uniformly and exponentially stabilizes this subsystem:*

$$\|x_i^s\| \le \overline{K}_x \rho^i, \quad \|\boldsymbol{U}_i^d\| \le \overline{K}_u \rho^i \quad \forall i \ge 0 \text{ whenever } \quad \|x_0^s\| \le K_0. \tag{3.8.21}$$

Here $\rho < 1$ is the parameter of the controller from (3.8.13)*, and the constants $\overline{K}_x, \overline{K}_u$ may depend on K_ξ from* (3.8.8) *and K_0.*

[17] Hence with regard to the extra alarm symbol, the quatizer output can take $m + 1$ values.

Proof of Proposition 3.8.24

The remainder of the subsection is devoted to this proof, which is broken into the string of several lemmas. We start with rough estimates of concerned variables.

Lemma 3.8.25. *The following inequalities hold for all* $i \geq 1, h \geq 0$, *and* $p \geq h$:

$$\delta_0 \rho^{i-1} \leq \delta_i^c \leq \delta_0 \gamma^{i-1}, \quad \|\boldsymbol{U}_i^d\| \leq \|\mathfrak{N}\| \delta_{i-1}^c,$$
$$\|x_p^s\| \leq a_s^{r(p-h)} \left[\|x_h^s\| + K_\xi' \rho_\xi^h \right] + K_\gamma \gamma^p |\mathfrak{I}(p,h)|, \quad (3.8.22)$$

where $K_\xi' := K_\xi / (a_s^r - 1)$ *and* a_s *is an arbitrary constant such that*

$$a_s > \|A_{ss}\|, \qquad (3.8.23)$$

$|\mathfrak{I}(p,h)|$ *is the size of the set*

$$\mathfrak{I}(p,h) := \{ j = h, \ldots, p-1 : j \geq 2 \ \& \ q_{j-1} \neq \maltese \},$$

and the constant K_γ *does not depend on* x_0, K_0, h, p, K_ξ.

Remark 3.8.26. Due to Assumption 3.8.1 (on p. 62), $\|A_{ss}\| \geq 1$. Hence $a_s > 1$ by (3.8.23) and so the constant K_ξ' is well defined.

Explanation 3.8.27. If $\|A_{ss}\| > 1$, the lemma remains true with $a_s := \|A_{ss}\|$. The constant a_s is introduced for uniformity of the formulas concerning the cases $\|A_{ss}\| > 1$ and $\|A_{ss}\| = 1$, respectively.

Proof of Lemma 3.8.25. The first formula is immediate from (3.8.16) and (3.8.17) since $\rho < 1 < \gamma$ by (3.8.13). The second one results from (3.8.18) and (3.8.19) since $\| \overset{\star}{q} \| \leq 1$ due to (3.8.17). To prove the last formula, we first note that

$$\boldsymbol{U}_j^d = 0 \qquad \forall j \notin \mathfrak{I}(p,h), \quad h \leq j \leq p-1,$$

by (3.8.17) and (3.8.18). Hence

$$\|x_p^s\| \overset{(3.8.6)}{=\!=} \left\| A_{ss}^{r(p-h)} x_h^s + \sum_{j=h}^{p-1} A_{ss}^{r(p-1-j)} \left[\mathfrak{B}_s \boldsymbol{U}_j^d + \xi_{s,j} \right] \right\|$$
$$\overset{(3.8.8)}{\leq} \|A_{ss}\|^{r(p-h)} \|x_h^s\| + \|\mathfrak{B}_s\| \|\mathfrak{N}\| \delta_0 \sum_{j \in \mathfrak{I}(p,h)} \underbrace{\|A_{ss}\|^{r(p-1-j)}}_{\leq \gamma^{p-1-j} \text{by (3.8.13)}} \gamma^{j-2}$$
$$+ K_\xi \underbrace{\rho_\xi^j}_{\leq \rho_\xi^h \text{by (3.8.8)}} \sum_{j=h}^{p-1} \underbrace{\|A_{ss}\|^{r(p-1-j)}}_{\leq a_s^{r(p-1-j)} \text{ by (3.8.23)}}$$
$$\leq a_s^{r(p-h)} \|x_h^s\| + \|\mathfrak{B}_s\| \|\mathfrak{N}\| \delta_0 |\mathfrak{I}(p,h)| \gamma^{p-3} + K_\xi \rho_\xi^h \frac{a_s^{r(p-h)} - 1}{a_s^r - 1},$$

which yields the last formula from (3.8.22). □

To justify stability, it suffices to show that δ_i^c are true bounds for the state prognosis

$$\|\widehat{x}_{i+1}^s\| \leq \delta_i^c \quad \text{for all large} \quad i.$$

Indeed, then $\|\boldsymbol{\varepsilon}_i\| \leq 1 \; \forall i \approx \infty$ by (3.8.15). Hence (3.8.16) and (3.8.17) ensure that the bound δ_i^c and thus $\widehat{x}_{i+1}^s$ decay exponentially $\delta_{i+1}^c = \rho\delta_i^c$ for $i \approx \infty$. Then so does the state x_{i+1}^s thanks to (3.8.8) and (3.8.20); i.e., the system is stable. We start by showing that even if the bound δ_i^c is incorrect for some i, it becomes correct later.

Lemma 3.8.28. *For any K_0, K_ξ, and i_0, an integer $p_0 \geq i_0$ exists such that the bound δ_i^c is correct $\|\widehat{x}_{i+1}^s\| \leq \delta_i^c$ for at least one index $i \in [i_0 : p_0]$ whenever $\|x_0^s\| \leq K_0$ and* (3.8.8) *holds.*

Proof. With regard to (3.8.13), we pick a constant a_s in (3.8.23) so that

$$a_s^r < \gamma. \tag{3.8.24}$$

The symbol $\overline{K}$ (with possible indices) is used to denote a constant that depends on K_0, K_ξ but not x_0^s and $\xi_{s,i}$. By putting $h := 0$ and the estimate $|\mathfrak{I}(p,h)| \leq p - h$ into the last inequality from (3.8.22), we see that

$$\|x_{p+1}^s\| \leq \overline{K}^{(p)} \quad \text{whenever } \|x_0^s\| \leq K_0 \text{ and (3.8.8) holds.} \tag{3.8.25}$$

Now suppose that the bound δ_i^c is incorrect for all i from some interval $[i_0 : i_1]$ with the left end i_0. To estimate i_1, we note that $\mathfrak{I}(p,h) = \emptyset$ for $h := i_0 + 1$ and $p := i_1 + 1$ due to (3.8.15). So the last inequality from (3.8.22) yields

$$\begin{aligned}
\|x_{i_1+1}^s\| &\leq a_s^{r(i_1-i_0)}\left[\overline{K}^{\,i_0} + K'_\xi \rho_\xi^{i_0+1}\right]; \|\widehat{x}_{i_1+1}^s\| \overset{(3.8.20)}{=\!=} \|x_{i_1+1}^s - \xi_{s,i_1}\| \\
&\overset{(3.8.8)}{\leq} a_s^{r(i_1-i_0)}\left[\overline{K}^{\,i_0} + K'_\xi \rho_\xi^{i_0+1}\right] + K_\xi \rho_\xi^{i_1} \\
&\overset{\rho_\xi<1}{\leq} a_s^{r(i_1-i_0)}\left[\overline{K}^{\,i_0} + K'_\xi\right] + K_\xi \\
&\overset{1<a_s}{\leq} \left(\overline{K}^{\,i_0} + K'_\xi + K_\xi\right) a_s^{r(i_1-i_0)}.
\end{aligned}$$

At the same time, (3.8.16) and (3.8.17) entail that $\delta_{i+1}^c = \gamma\delta_i^c$ for $i \in [i_0 : i_1]$. So

$$\delta_{i_1}^c = \gamma^{i_1-i_0}\delta_{i_0}^c \geq \gamma^{i_1-i_0}\rho^{i_0-1}\delta_0,$$

where the last inequality is based on (3.8.22). Since the bound $\delta_{i_1}^c$ is incorrect, it follows that

$$1 \leq \frac{\|\widehat{x}_{i_1+1}^s\|}{\delta_{i_1}^c} \leq \left(\frac{a_s^r}{\gamma}\right)^{i_1-i_0} \frac{\overline{K}^{\,i_0} + K'_\xi + K_\xi}{\rho^{i_0-1}\delta_0}.$$

By invoking (3.8.24), we conclude that $i_1 \leq i_0 + \nu$, where

$$\nu := \left\lfloor \frac{\log_2 \left(\overline{K}^{\,i_0} + K'_\xi + K_\xi\right) - (i_0 - 1)\log_2 \rho - \log_2 \delta_0}{\log_2 \gamma - r \log_2 a_s} \right\rfloor .$$

So one may pick $p_0 := i_0 + 1 + \max\{\nu, 0\}$. The claim of the lemma remains true with the same p_0 if the interval $[i_0 : i_1]$ does not exist, because then the bound $\delta^c_{i_0}$ is correct. □

The next lemma in fact completes the proof of Proposition 3.8.24.

Lemma 3.8.29. *Suppose that* $\|x^s_0\| \le K_0$ *and* (3.8.8) *holds. Whenever the bound* δ^c_i *becomes correct* $\|\widehat{x}^s_{i+1}\| \le \delta^c_i$*, it is kept correct afterward, provided that* $i \ge i_0$*. Here* i_0 *is taken so that*

$$\frac{\rho_{\mathfrak{Q}^s}}{\rho} + \frac{\|A_{ss}\|^r K_\xi}{\delta_0} \left(\frac{\rho_\xi}{\rho}\right)^i < 1 \qquad \forall i \ge i_0. \tag{3.8.26}$$

Remark 3.8.30. Such an i_0 exists due to the last inequality from (3.8.13).

Proof of Lemma 3.8.29. By (3.8.15), $\|\boldsymbol{\varepsilon}_i\| \le 1$. So (3.8.16) and (3.8.17) imply that

$$\boldsymbol{\varepsilon}_i \in Q, \quad q_i = q^Q \quad \text{for some} \quad Q \in \{Q_1, \dots, Q_m\};$$
$$\boldsymbol{U}^c_{i+1} = \mathcal{N}\left[\delta^c_i q_i\right], \quad \delta^c_{i+1} = \rho \delta^c_i, \tag{3.8.27}$$

where Q_j are the level sets of the quantizer. By (3.8.12), the third relation yields

$$\delta^c_i A^r_{ss} q_i + \mathfrak{B}_s \boldsymbol{U}^c_{i+1} = 0.$$

Hence

$$\left(\delta^c_{i+1}\right)^{-1} \left\|\widehat{x}^s_{i+2}\right\| \overset{(3.8.14)}{=\!=} \left(\delta^c_{i+1}\right)^{-1} \left\|A^r_{ss} x^s_{i+1} + \mathfrak{B}_s \boldsymbol{U}^c_{i+1}\right\|$$
$$= \left(\delta^c_{i+1}\right)^{-1} \left\|A^r_{ss} x^s_{i+1} - \delta^c_i A^r_{ss} q_i\right\| \overset{(3.8.20)}{=\!=} \left(\delta^c_{i+1}\right)^{-1} \left\|A^r_{ss}\left[\widehat{x}^s_{i+1} + \xi_{s,i}\right] - \delta^c_i A^r_{ss} q_i\right\|$$
$$\overset{(3.8.15)}{\le} \frac{\delta^c_i}{\delta^c_{i+1}} \left\|A^r_{ss}\left[\boldsymbol{\varepsilon}_i - q_i\right]\right\| + \left(\delta^c_{i+1}\right)^{-1} \left\|A^r_{ss} \xi_{s,i}\right\| .$$

It follows from (3.8.9) and the first two relations in (3.8.27) that $\left\|A^r_{ss}\left[\boldsymbol{\varepsilon}_i - q_i\right]\right\| \le \rho_{\mathfrak{Q}^s}$. We proceed by invoking (3.8.8) and the last relation from (3.8.27), along with the first inequality from (3.8.22),

$$\left(\delta^c_{i+1}\right)^{-1} \left\|\widehat{x}^s_{i+2}\right\| \le \frac{\rho_{\mathfrak{Q}^s}}{\rho} + \|A_{ss}\|^r K_\xi \frac{\rho^i_\xi}{\delta^c_{i+1}} \le \frac{\rho_{\mathfrak{Q}^s}}{\rho} + \frac{\|A_{ss}\|^r K_\xi}{\delta_0} \left(\frac{\rho_\xi}{\rho}\right)^i \overset{(3.8.26)}{<} 1.$$

Thus the bound δ^c_{i+1} is true, which completes the proof. □

Proof of Proposition 3.8.24. Consider the number p_0 from Lemma 3.8.28, where i_0 is taken from Lemma 3.8.29. By these lemmas, the bound δ^c_i is true $\|\widehat{x}^c_{i+1}\| \le \delta^c_i$

whenever i $\geq p_0$. Then $\delta^c_{i+1} = \rho\delta^c_i \; \forall i \geq p_0$ thanks to (3.8.15), (3.8.16), and (3.8.17). With regard to the first relation from (3.8.22), we see that

$$\frac{\delta^c_i \leq \delta_0\gamma^i \quad \text{for} \quad i \leq p_0, \qquad \delta^c_i = \delta^c_{p_0}\rho^{i-p_0} \leq \delta_0\big(\gamma/\rho\big)^{p_0}\rho^i \quad \text{for} \quad i \geq p_0}{\Downarrow}$$
$$\delta^c_i \leq \overline{K}_\delta\rho^i \quad \forall i, \quad \text{where} \quad \overline{K}_\delta := \delta_0\big(\gamma/\rho\big)^{p_0}. \tag{3.8.28}$$

This and the second formula from (3.8.22) give the second inequality in (3.8.21). To prove the first one, we note that

$$\big\|x^s_{i+1}\big\| \overset{(3.8.20)}{=\!=} \big\|\widehat{x}^s_{i+1} + \xi_{s,i}\big\| \overset{(3.8.8)}{\leq} \delta^c_i + K_\xi\rho^i_\xi \overset{(3.8.13)}{\leq} \overline{K}^0_x\rho^{i+1} \quad \forall i \geq p_0,$$
$$\text{where} \quad \overline{K}^0_x := \rho^{-1}\left(\overline{K}_\delta + K_\xi\right).$$

For $i \leq p_0 + 1$, inequality (3.8.25) yields

$$\|x^s_i\| \leq \overline{K}^{(i-1)} \leq \overline{K}' := \max\left\{\max_{j=1,\dots,p_0} \overline{K}^{(j)}; K_0\right\}.$$

Thus the first inequality in (3.8.21) does hold with $K_x := \max\{\overline{K}^0_x; \overline{K}'\rho^{-p_0-1}\}$.

□

In conclusion, we observe that the coder and decoder inner dynamical variables z^s_i and z_i exponentially decay to zero thanks to Remark 3.8.23, (3.8.28), and the second formulas from (3.8.19) and (3.8.22).

3.8.5 Contracted Quantizer with the Nearly Minimal Number of Levels

Now we in fact perform step 3 (on p. 65) from the plan described in Subsect. 3.8.2.

Let m denote the number of quantizer levels. Due to Remark 3.8.21 (on p. 73), the r-contracted[18] quantizer with the minimal value of the ratio $\frac{\log_2(m+1)}{r}$ is of special interest. For a given sample period r, such a quantizer is that with the minimum number of levels. We start with obtaining a simple lower bound on this number.

Lemma 3.8.31. *For any r-contracted quantizer,[19] the following inequality holds:*

$$m > |\det A_{ss}|^r. \tag{3.8.29}$$

Proof. Due to (3.8.9),

$$|\det A_{ss}|^r\boldsymbol{V}(Q_i) = |\det A_{ss}|^r\boldsymbol{V}[Q_i - q^{Q_i}] = \boldsymbol{V}\big[A^r_{ss}(Q_i - q^{Q_i})\big]$$
$$\leq \rho^{n_s}_{\mathfrak{Q}^s}\boldsymbol{V}\big[B^1_0\big] < \boldsymbol{V}\big[B^1_0\big].$$

Summing over $i = 1, \dots, m$ results in (3.8.29). □

[18]See Definition 3.8.7 on p. 70.

[19]Strictly speaking, we restrict ourselves to quantizers with measurable level domains.

In fact, the lower bound (3.8.29) is tight. To show this we introduce the following.

Notation 3.8.32. *The symbols* $\lesssim$ *and* $\eqsim$ *stand for inequality and equality up to a polynomial factor. In other words,*

$$f(r) \lesssim g(r) \Leftrightarrow f(r) \leq \varphi(r)g(r) \qquad \forall r = 1, 2, \dots,$$

where $\varphi(r)$ *is a polynomial in* r*, and*

$$f(r) \eqsim g(r) \Leftrightarrow f(r) \lesssim g(r) \,\&\, g(r) \lesssim f(r).$$

When $f(r)$ *and* $g(r)$ *depend on some other variables, the polynomial is assumed to be independent of them.*

Proposition 3.8.33. *For any natural* r*, an* r*-contracted polyhedral*[20] *quantizer* $\mathfrak{Q}^s$ *exists with the number of levels*

$$m_r^s \lesssim |\det A_{ss}|^r. \tag{3.8.30}$$

Remark 3.8.34. To implement the controller introduced in Subsect. 3.8.4, it is required to communicate on average $R := \frac{\log_2(m+1)}{r}$ bits per unit time from the coder to the decoder by Remark 3.8.21. Lemma 3.8.31 and Proposition 3.8.33 imply that

$$R \geq \log_2 |\det A_{ss}| + \alpha_r, \qquad \text{where} \quad \alpha_r \to 0 \quad \text{as} \quad r \to \infty,$$

and for large r, the rate R can be made close to the asymptotic bound $\log_2 |\det A_{ss}|$. So the quantizer from that proposition can be viewed as almost optimal.

The proof of Proposition 3.8.33 employs the following technical observations.

Lemma 3.8.35. *The following two statements hold:*

(i) Whenever the claim of Proposition 3.8.33 *is true for some matrix* A_{ss}*, it is valid for any similar matrix* $A'_{ss} = UA_{ss}U^{-1}$*;*
(ii) Whenever this claim is true for two square matrices A'_{ss} *and* A''_{ss}*, it is also true for the block matrix* $A_{ss} = \begin{pmatrix} A'_{ss} & 0 \\ 0 & A''_{ss} \end{pmatrix}$.

Proof. Statement (i). Given r, consider a polyhedral r-contracted (for A_{ss}) quantizer with $m_r^s \leq \varphi(r)|\det A_{ss}|^r$ levels, where $\varphi(\cdot)$ is a polynomial. We also consider the corresponding ball B_0^1 and norm $\|\cdot\|$ from Definition 3.8.6 (on p. 69). Evidently, $U^r B_0^1$ is the unit ball with respect to the norm $\|x^s\|_* := \|U^{-r}x^s\|$. So the partition $U^rQ_1, \dots, U^rQ_{m_r^s}$ of $U^rB_0^1$, where the centroid of U^rQ_i is defined to be U^rq_i, is a quantizer. Since $\left(A'_{ss}\right)^r = U^rA_{ss}^rU^{-r}$, (3.8.9) for A'_{ss} is immediate from this formula written for A_{ss}. Consider the linear functions $l_i(x^s)$ from Definition 3.8.8 (on p. 70) applied to the original quantizer. Then it is easy to see that the new quantizer is also polyhedral and served by the functions $l_i(U^{-r}x^s)$. The proof is

[20] See Definition 3.8.8 on p. 70.

completed by observing that these quantizers have a common number of levels and $\det A_{ss} = \det A'_{ss}$.

Statement (ii). Given r, the matrices A'_{ss} and A''_{ss} can be supplied with polyhedral quantizers

$$\mathfrak{Q}_1^s = \left[Q'_1 \sim q'_1, \ldots, Q'_{m'_r} \sim q'_{m'_r}\right] \quad \text{and} \quad \mathfrak{Q}_2^s = \left[Q''_1 \sim q''_1, \ldots, Q''_{m''_r} \sim q''_{m''_r}\right],$$

respectively, such that $m_r^\nu \leq \varphi^\nu(r)|\det A_{ss}^\nu|^r$ for both $\nu = \prime$ and $\nu = \prime\prime$. Here $\varphi'(r)$ and $\varphi''(r)$ are polynomials. Let x be vectors of dimension matching the size of A_{ss}. We partition them $x = \textbf{col}\,(x', x'')$ in accordance with the block partition of the matrix. We also introduce the norm $\|x\| = \max\{\|x'\|', \|x''\|''\}$, where $\|\cdot\|'$ and $\|\cdot\|''$ are the norms from Definition 3.8.6 (on p. 69) applied to the first and second of the above quantizers, respectively. It is easy to see that the sets

$$Q_{ij} := \left\{x : x' \in Q'_i, x'' \in Q''_j\right\}, \qquad i = 1, \ldots, m'_r, j = 1, \ldots, m''_r$$

equipped with the centroids $q_{ij} := \textbf{col}\,(q'_i, q''_j)$ form a quantizer $\mathfrak{Q}^s$. Formulas (3.8.9) written for A'_{ss} and A''_{ss}, respectively, imply (3.8.9) for A_{ss} (with $\rho_{\mathfrak{Q}^s} := \max\{\rho_{\mathfrak{Q}_1^s}, \rho_{\mathfrak{Q}_2^s}\} < 1$). By considering the sets of linear functions $l_i(x'), i = 1, \ldots, M_1^s$ and $l_j(x''), j = 1, \ldots, M_2^s$ from Definition 3.8.8 applied to $\mathfrak{Q}_1^s$ and $\mathfrak{Q}_2^s$, respectively, it is easy to see that the quantizer $\mathfrak{Q}^s$ is polyhedral and served by the union of these sets with $l_i(x')$ and $l_j(x'')$ interpreted as functions of $x = \textbf{col}\,(x', x'')$. For the number m_r^s of the levels of the quantizer $\mathfrak{Q}^s$, we have

$$m_r^s = m'_r \cdot m''_r \leq \underbrace{\varphi'(r)\varphi''(r)}_{\varphi(r)}|\det A'_{ss}|^r|\det A''_{ss}|^r = \varphi(r)|\det A_{ss}|^r. \qquad \square$$

Proof of Proposition 3.8.33. By employing the real Jordan form of A_{ss} and Lemma 3.8.35, the proof is reduced to the case where the matrix is a real Jordan block. Let n_s denote its size, λ its eigenvalue, and $\omega := |\lambda|$. Then (see, e.g., [213, Lemma 3.1, p. 64]) a polynomial $\varphi(\cdot)$ exists such that

$$\Xi(r) := \omega^{-r}\varphi(r)^{-1}A_{ss}^r \to 0 \quad \text{as} \quad r \to \infty.$$

So $\|\Xi(r)\| < \rho < 1$ for $r \approx \infty$. Here $\|\cdot\|$ is the operator norm associated with the norm $\|z\|_\infty := \max_i |z_i|$ in $\mathbb{R}^{n_s} = \{z = \textbf{col}\,(z_1, \ldots, z_{n_s})\}$. Balls with respect to this norm are geometrically cubes. Multiplying $\varphi(r)$ by a sufficiently large scalar factor makes the inequality $\|\Xi(r)\| < \rho$ true for all r. Now consider the uniform quantizer $\mathfrak{Q}^s$ (see Example 3.8.11 and Remark 3.8.13 on pp. 70 and 71) partitioning the cube B_0^1 into $m_r^s := k^{n_s}$ congruent subcubes Q_i with the side length $\frac{2}{k}$, where $k := \lceil \omega^r\varphi(r)\rceil$. The centroid q^{Q_i} is the center of Q_i. Then

$$\|\Xi(r)\| < \rho \Rightarrow \Xi(r)[Q_i - q^{Q_i}] \subset \rho[Q_i - q^{Q_i}]$$
$$= \frac{\rho}{k}B_0^1 \Rightarrow A_{ss}^r[Q_i - q^{Q_i}] \subset \rho\frac{\omega^r\varphi(r)}{k}B_0^1 \subset \rho B_0^1.$$

Thus the quantizer is r-contracted. The proof is completed by observing that

$$m_r^s = k^{n_s} \leq [\omega^r \varphi(r) + 1]^{n_s} \leq 2^{n_s - 1}\Big([\omega^r \varphi(r)]^{n_s} + 1\Big)$$
$$= 2^{n_s-1}\Big(|\det A_{ss}|^r \varphi(r)^{n_s} + 1\Big) \overset{1 \leq |\det A_{ss}|}{\leq} 2^{n_s-1}|\det A_{ss}|^r \Big[\varphi(r)^{n_s} + 1\Big]. \square$$

In Proposition 3.8.33, the contraction rate $\rho_{\mathfrak{Q}^s}$ of the concerned quantizer is not specified. At the same time, this rate can be made geometrically decreasing in r provided that the number of levels m_r^s is slightly increased.

Lemma 3.8.36. *Let $\eta > H(A) = \log_2 |\det A_{ss}|$ be given. Then for any $r = 1, 2, \ldots$, an r-contracted polyhedral quantizer $\mathfrak{Q}^s$ exists with the contraction rate $\rho_{\mathfrak{Q}^s} = \varkappa^{2r}$ and $m_r^s \lesssim 2^{r\eta}$ levels, where $\varkappa = \varkappa_{\eta, A_{ss}} \in (0,1)$ does not depend on r.*

Proof. In the above proof of Proposition 3.8.33, one should alter the choice of k by $k := \lceil \alpha^r \omega^r \varphi(r) \rceil$. Here $\alpha > 1$ is a parameter to be adjusted. This evidently provides the rate of contraction $\rho_{\mathfrak{Q}^s} \leq \rho\alpha^{-r} \leq \alpha^{-r}$ and gives rise to a quantizer with $m_r^s \approx \alpha^{n_s r}\omega^{n_s r} = 2^{r[H(A) + n_s \log_2 \alpha]}$ levels. Then the statement of the lemma results from properly adjusting the value of $\alpha > 1$. $\square$

Remark 3.8.37. Proposition 3.8.33 and Lemma 3.8.36 hold for an arbitrary square matrix A_{ss} with no stable $|\lambda| < 1$ eigenvalues λ.

3.8.6 Construction of a Deadbeat Stabilizer

When all subsystems are equipped with the proposed controllers, the stability of a particular subsystem may be violated by the controllers serving other subsystems since the control is common. To avoid this, it suffices to choose the controls in such a way that they influence only the subsystem for which they are intended. For the basic (unit) sample period, this may be impossible. Now we show that this can be done if the sample period r is properly increased: The controls generated for a given subsystem do not affect the states of the other ones at times $t = i \cdot r, i = 0, 1, \ldots$.

Common controls give rise to another trouble. The sth coder will be implemented at the sites of all sensors observing the sth subsystem. To compute the states $x_i^s = x^s(\tau_i)$ used by this coder, not only the observations but also controls must be known at these sites. However, the sth coder may know the control only partly. It is aware of its own summand in the overall control, which is the sum of the controls generated for all subsystems. At the same time, it cannot determine the summands based on the modes x^j invisible at its site. To overcome this obstacle, it suffices to note that the controls must be known for only n times t preceding τ_i. So it suffices to ensure that all controllers produce zero controls at these times.

Now we show that deadbeat stabilizers with the above properties do exist.

Lemma 3.8.38. *Whenever $r > n$, a deadbeat stabilizer $\mathcal{N}$ for the sth subsystem exists. Moreover, it can be chosen so that it generates control programs $\boldsymbol{U} =$*

col $(u_0, \ldots, u_{r-1})$ *vanishing since the time* $t = n$*, i.e.,* $u_n = \cdots = u_{r-1} = 0$*, and does not disturb the other subsystems, i.e.,*

$$\mathfrak{B}_j \boldsymbol{U} = 0 \quad \text{for} \quad j \neq s, \quad \boldsymbol{U} = \mathcal{N}x^s, \quad \text{and any} \quad x^s.$$

Proof. By (3.8.12), a deadbeat stabilizer is the right inverse to the operator $\mathfrak{D}\boldsymbol{U} := -A_{ss}^{-r}\mathfrak{B}_s\boldsymbol{U}$. In (3.8.1), $\mathfrak{B}\boldsymbol{U}$ is the state to which the control program $\boldsymbol{U}$ drives the system (3.3.1) at time $t = r$ from $x(0) = 0$. By Assumption 3.8.1 (on p. 62), this system has no stable modes. So it is controllable thanks to Assumption 3.4.22 (on p. 49). It follows that the operator $\mathfrak{B}$ is onto. Moreover, $\mathfrak{B}|_M$ is onto, where

$$M := \{\boldsymbol{U} : u_n = \cdots = u_{r-1} = 0\}.$$

Indeed for any x, it suffices to pick the control program $u_0, \ldots, u_{n-1}$ that drives the system from 0 at $t = 0$ to $A_{ss}^{n-r}x$ at $t = n$ and to extend it by zeros to form $\boldsymbol{U} \in M$. Then evidently $\mathfrak{B}\boldsymbol{U} = x$. Now consider x such that in (3.8.3) all blocks are zeros except for $x^s \in \mathbb{R}^{n_s}$. Since this block can be chosen arbitrarily, it follows that the operator $\mathfrak{B}_s$ maps $L := \{\boldsymbol{U} \in M : \mathfrak{B}_j\boldsymbol{U} = 0 \, \forall j \neq s\}$ onto $\mathbb{R}^{n_s}$. So evidently does $\mathfrak{D}$. It remains to define $\mathcal{N}$ as the right inverse to $\mathfrak{D}|L$. □

3.8.7 Stabilization of the Entire System

Now we perform step 4 (on p. 65) from the plan described in Subsect. 3.8.2. We revert to considering all subsystems in their actual relationship. In particular, this means that the disturbance in (3.8.6) is given by (3.8.7). We also pick $r > n$ and suppose that the following assumptions hold for any s:

A.1) The block $x^s(\tau_i)$ of the state can be determined at any time $\tau_i = i \cdot r$ at a certain site *(called the* s*th site)*;

A.2) There is a way to transfer the quantized value q_i^s generated by the sth coder at the step **c.3)** (on p.72) from the sth site to the decoder site during the time interval $[\tau_i : \tau_{i+1})$.

Explanation 3.8.39. In A.2), considered is the site where the actual decoder (see Fig. 3.2 on p. 42) should be situated.

Architecture of the Stabilizing Controller

Now we suppose that

SC.1) The sth coder from Subsect. 3.8.4 (starting on p. 68) is implemented at the sth site from A.1) for every s;

SC.2) For all s, the sth decoder from Subsect. 3.8.4 is implemented at the site from Explanation 3.8.39.

Remark 3.8.40. Assumption A1) makes SC.1) possible since the sth coder is driven by the sequence of states $x_i^s = x^s(\tau_i)$. Since the sth decoder is driven only by the sequence of quantized values $q_i^s, i = 0, 1, \ldots$, SC.2) is possible thanks to A.2).

Each decoder produces its own sequence of controls

$$\boldsymbol{U}_i^d = \textbf{col}\left[u^s(ir), u^s(ir+1), \dots, u^s(ir+r-1)\right].$$

These sequences are summed over all decoders to produce the control sequence acting upon the plant:

$$u(t) := u^1(t) + u^2(t) + \dots + u^d(t).$$

Specifying the Coders and Decoders

To complete their description, a quantizer, deadbeat stabilizer, and parameters r, γ, ρ from (3.8.13) should be chosen for each coder.

The parameter $r > n$ has already been picked.

For any subsystem s, the quantizer and deadbeat stabilizer are taken from Proposition 3.8.33 and Lemma 3.8.38, respectively.

The parameter $\gamma = \gamma_s$ is chosen to satisfy the second relation from (3.8.13).

As for the third relation, it is indefinite under the circumstances since ρ_ξ from (3.8.8) is not given. So now we pick the parameter $\rho = \rho_s$ in another way. It is chosen successively for $s = 1, 2, \dots, d$ and so that

$$1 > \rho_1 > \rho_{\mathfrak{Q}^1}, \quad 1 > \rho_2 > \max\{\rho_{\mathfrak{Q}^2}; \rho_1\}, \quad 1 > \rho_3 > \max\{\rho_{\mathfrak{Q}^3}; \rho_2\}, \dots, \\ \dots, 1 > \rho_d > \max\{\rho_{\mathfrak{Q}^d}; \rho_{d-1}\}, \quad (3.8.31)$$

where $\rho_{\mathfrak{Q}^s}$ is taken from (3.8.9).

Stabilizing Properties of the Proposed Control Scheme

They are described by the following.

Proposition 3.8.41. *Let assumptions* A.1) *and* A.2) *be true. Then the controller proposed in this subsection uniformly and exponentially stabilizes*[21] *the entire system* (3.3.1) *at the rate* $\mu = \rho_d^{1/r}$.

The remainder of the subsection is devoted to the proof of this proposition. We preface it with a simple technical fact.

Lemma 3.8.42. *Suppose that a trajectory of the system* (3.3.1) *satisfies the estimates*

$$\|x_i\| \le \overline{K}_x \rho^i, \quad \|\boldsymbol{U}_i\| \le \overline{K}_u \rho^i, \qquad i = 0, 1, 2, \dots, \quad (3.8.32)$$

where

$$x_i := x(\tau_i), \tau_i := i \cdot r, \rho \in [0,1), \boldsymbol{U}_i := \textbf{col}\left[u(\tau_i), u(\tau_i + 1), \dots, u(\tau_i + r - 1)\right].$$

Then (3.4.2) *(on p.46) holds, where* $\mu := \rho^{1/r}$ *and the constants* K_x, K_u *are determined by* $\overline{K}_x, \overline{K}_u$, *and* ρ *(for a given system).*

[21] See Definition 3.4.6 on p. 46.

Proof. Whenever $t \in [\tau_i : \tau_{i+1})$, we have

$$\rho^i = \mu^{\tau_i} = \mu^{\tau_i - t}\mu^t \le \mu^{-r}\mu^t = \rho^{-1}\mu^t.$$

So

$$\|u(t)\| \le \|\boldsymbol{U}_i\| \le \overline{K}_u \rho^i \le \overline{K}_u \rho^{-1}\mu^t;$$

i.e., the second inequality from (3.4.2) does hold. We denote $\chi := 1 + \|A\|$. Then

$$\begin{aligned}
\|x(t)\| &= \left\| A^{t-\tau_i} x(\tau_i) + \sum_{j=\tau_i}^{t-1} A^{t-1-j} Bu(j) \right\| \\
&\le \|A\|^{t-\tau_i}\|x(\tau_i)\| + \sum_{j=\tau_i}^{t-1} \|A\|^{t-1-j}\|B\|\|u(j)\| \\
&\le \left[\|A\|^{t-\tau_i} + \|B\| \sum_{j=\tau_i}^{t-1} \|A\|^{t-1-j} \right] \times \left[\|x(\tau_i)\| + \|\boldsymbol{U}_i\| \right] \\
&\le \left[\chi^r + \|B\| \sum_{j=\tau_i}^{\tau_{i+1}-1} \chi^{\tau_{i+1}-1-j} \right] \left[\overline{K}_x + \overline{K}_u \right] \rho^i,
\end{aligned}$$

where $\rho^i \le \rho^{-1}\mu^t$. The index substitution $j := \tau_i + \nu$ in the last sum proves that (3.4.2) is true. □

Proof of Proposition 3.8.41. Suppose that $\|x(0)\| \le K_0$, where K_0 is given. The controls u^s with $s \ge 2$ do not disturb the first block $x_i^1 := x^1(\tau_i)$ of the state at times $\tau_i = i \cdot r$ since the deadbeat stabilizers are taken from Lemma 3.8.38. So this block $x_i^1, i = 0, 1, \ldots$, evolves just as in the first subsystem (3.8.6) driven by the first coder and decoder and perturbed by the noise $\xi_{1,i}$, which is zero by (3.8.7). Then Proposition 3.8.24 and the first inequality from (3.8.31) imply that the first subsystem $s = 1$ is uniformly exponentially stabilized (3.8.21) at the rate $\rho := \rho_1$. This and (3.8.7) imply that the noise $\xi_{2,i}$ in the second subsystem (3.8.6) (where $s = 2$) exponentially decays (3.8.8) at the rate ρ_1.

Now we retrace the above arguments with respect to this subsystem and employ the second relation from (3.8.31). As a result, we establish that this subsystem is stabilized at the rate ρ_2; i.e., (3.8.21) holds for $s = 2$ and $\rho := \rho_2$. By continuing likewise, we see that for any s, inequalities (3.8.21) are true with $\rho := \rho_s$ and proper constants $\overline{K}_x, \overline{K}_u$ (depending on s) whenever $\|x(0)\| \le K_0$. Since $\rho_d \ge \rho_s \, \forall s$ by (3.8.31), it follows that (3.8.32) holds with $\rho := \rho_d$ and some constants $\overline{K}_x, \overline{K}_u$ depending on K_0. Lemma 3.8.42 and Definition 3.4.6 (on p. 46) complete the proof. □

3.8.8 Analysis of Assumptions A1) and A2) on p. 81

Our next goal is to show that these assumptions stated in the previous subsection are satisfied whenever $r > 2n$ and (3.5.2) (on p. 52) holds. This in fact will complete the proof of Theorem 3.5.2. In this subsection, we perform the first step to this end.

We start with assumption A1). By (i) of Proposition 3.8.2 (on p. 66), the unobservable subspace (3.4.6) $L_j^{-o} = L_j^-$ of the jth sensor is composed of several blocks $x^s, s \notin O_j$ of the state (3.8.3). These blocks do not affect its outputs y_j, whereas all other blocks $x^s, s \in O_j$, can be determined from these outputs.

Lemma 3.8.43. *Whenever $r > 2n$, assumption* A1) *holds. For any s, the site of any sensor j with $O_j \ni s$ can be taken as the sth site in* A1).

Proof. We recall that the deadbeat stabilizers were taken from Lemma 3.8.38. So they produce control programs with zeros at any place $i \geq n$:

$$\boldsymbol{U} = \textbf{col}\,(u_0, \ldots, u_{r-1}), \qquad u_i = 0 \quad \forall i \geq n.$$

For $r > 2n$, this means that the corresponding controls $u(t), t = 0, 1, \ldots$, vanish $u(t) = 0$ for at least n times t preceding each $\tau_i = i \cdot r, i = 0, 1, \ldots$. The proof is completed by invoking the remarks from the paragraph prefacing Lemma 3.8.38. □

Remark 3.8.44. As is well known, the current state x^s can be generated as a linear function of n previous observations.

Now we turn to analysis of A.2). We recall that in A.2), the value q_i^s is given by an m_r^s-level quantizer $\mathfrak{Q}^s$. Description of such a value (which may equal ✠) requires $b_s = \lceil \log_2(m_r^s + 1) \rceil$ bits. This number may exceed the capacity of the channel that serves any particular sensor j observing the block x^s. So we employ all such channels. Specifically, the following scheme of transmission q_i^s to the decoder site is used for each subsystem $s = 1, \ldots, d$:

T.1) The sth coder is implemented at the sites of all sensors j observing the state x^s, i.e., such that $s \in O_j$;
T.2) By employing a common encoding rule, the value q_i^s produced at each of these sites is then transformed into a b_s-bit sequence $\boldsymbol{\beta}_i^s = (\beta_1, \beta_2, \ldots, \beta_{b_s})$ of binary digits $\beta_\nu = 0, 1$;
T.3) By applying a common rule, this sequence $\boldsymbol{\beta}$ is split up into several subsequences $\boldsymbol{\beta}_i^{s,j}$ each associated with one of the concerned sensors j, i.e., such that $s \in O_j$;
T.4) Each of these sensors j sends only its own subsequence $\boldsymbol{\beta}_i^{s,j}$ over the attached channel to the decoder site;
T.5) At the decoder site, the required value q_i^s is reconstructed by reversing the rules from T.2) and T.3).

We assume that the rules from T.2) and T.3) do not change as i progresses and are known at the decoder site. Furthermore, the rule from T.2) is lossless: The value q_i^s can be reconstructed from $\boldsymbol{\beta}$. This makes T.5) possible.

Remark 3.8.45. The claims T.1)–T.4) can be interpreted as if the sth site from A.1) is distributed over the sites of all sensors j such that $s \in O_j$.

Notation 3.8.46. *We denote by b_{sj} the number of bits in $\boldsymbol{\beta}_i^{s,j}$ whenever $s \in O_j$, and put $b_{sj} := 0$ otherwise.*

The above scheme means that several binary words $\boldsymbol{\beta}_i^{s,j}, s \in O_j$ must be transmitted over the common jth channel during any time interval $[\tau_i : \tau_{i+1})$ of duration $r-1$. By (ii) of Assumption 3.4.2 (on p. 43), this is possible if the total length of these words does not exceed $b_j^-(r-1)$. Summarizing, we arrive at the following lemma.

Lemma 3.8.47. *Assumption* A.2) *is satisfied whenever nonnegative integer numbers* $b_{sj}, s = 1, \ldots, d$*,* $j = 1, \ldots, k$ *exist such that the following relations hold:*

$$\sum_{j=1}^{k} b_{sj} = b_s = \lceil \log_2(m_r^s + 1) \rceil \; \forall s, \quad \sum_{s=1}^{d} b_{sj} \le b_j^-(r-1) \; \forall j,$$
$$\text{and} \quad b_{sj} = 0 \quad \text{whenever} \quad s \notin O_j. \tag{3.8.33}$$

Here k and d are the numbers of sensors and subsystems, respectively; m_r^s is the number of levels for the r-contracted quantizer taken from Proposition 3.8.33*; and* $b_j^-(\cdot)$ *and* O_j *are taken from* (ii) *of Assumption* 3.4.2 *(on p.* 43*) and* (i) *of Proposition* 3.8.2 *(on p.* 66*), respectively.*

3.8.9 Inconstructive Sufficient Conditions for Stabilizability

These conditions are immediate from Proposition 3.8.41 and Lemmas 3.8.43 and 3.8.47. Obtaining them amounts to carrying out step 5 (on p. 65) from Subsect. 3.8.2.

Proposition 3.8.48. *Suppose that the following system of relations:*

$$\log_2 |\det A_{ss}| < \sum_{j=1}^{k} \alpha_{sj} \; \forall s, \quad \sum_{s=1}^{d} \alpha_{sj} < \mathfrak{c}_j \; \forall j, \quad \alpha_{sj} \ge 0 \; \forall s, j,$$
$$\alpha_{sj} = 0 \quad \text{whenever} \quad s \notin O_j \tag{3.8.34}$$

is solvable in real numbers α_{sj}. Here A_{ss} is taken from (ii) *of Proposition* 3.8.2 *(on p.* 66*) and $\mathfrak{c}_j$ is the transmission capacity* (3.4.1) *of the jth channel. Then the system* (3.3.1)*,* (3.3.2) *is uniformly and exponentially stabilizable.*[22]

Proof. It suffices to show that for all large r, the system (3.8.33) is solvable in nonnegative integers b_{sj}. Indeed such an r can be clearly chosen so that $r > 2n$. Then Lemmas 3.8.43 and 3.8.47 ensure that assumptions A.1) and A.2) (on p. 81) hold. Then Proposition 3.8.41 completes the proof.

We note first that in (3.8.33), the first relation can be replaced by the inequality

$$\sum_{j=1}^{k} b_{sj} \ge \lceil \log_2(m_r^s + 1) \rceil. \tag{3.8.35}$$

[22]See Definition 3.4.8 on p. 46.

Indeed, if after this the system is solvable, then a solution for the original relation can be obtained by properly decreasing the non-negative integers b_{sj}. Specifically, they are decreased to satisfy the first relation from (3.8.33), which may only enhance the second relation and keep the third relation true.

We are going to show that a solution is given by $b_{sj} := \lfloor r \cdot \alpha_{sj} \rfloor$, provided $r \approx \infty$. Indeed the third relation in (3.8.33) follows from the last one in (3.8.34). Furthermore,

$$\frac{1}{r}\sum_{j=1}^{k} b_{sj} \xrightarrow{r\to\infty} \sum_{j=1}^{k} \alpha_{sj} \overset{(3.8.34)}{=\!=} \log_2 |\det A_{ss}| + \varkappa_s, \text{ where } \varkappa_s > 0,$$
$$\frac{1}{r}\lceil \log_2(m_r^s + 1)\rceil \leq \frac{1}{r}\left[\log_2(m_r^s + 1) + 1\right]$$
$$\overset{(3.8.30)}{\leq} \frac{1}{r}\Big[\log_2\big(\varphi_s(r)|\det A_{ss}|^r + 1\big) + 1\Big] \xrightarrow{r\to\infty} \log_2 |\det A_{ss}|,$$

where $\varphi_s(r)$ is a polynomial in r. It follows that (3.8.35) does hold for all $r \approx \infty$. Likewise,

$$\frac{1}{r}\sum_{s=1}^{d} b_{sj} \xrightarrow{r\to\infty} \sum_{s=1}^{d} \alpha_{sj} \overset{(3.8.34)}{=\!=} \mathfrak{c}_j - \eta_j, \quad \text{where} \quad \eta_j > 0,$$
$$\frac{b_j^-(r-1)}{r} \xrightarrow{(3.4.1)} \mathfrak{c}_j \text{ as } r \to \infty.$$

Thus the second relation from (3.8.33) is also true for all $r \approx \infty$. □

3.8.10 Convex Duality and a Criterion for the System (3.8.34) to be Solvable

Now we perform step 6 (on p. 66) in the proof of the sufficiency part of Theorem 3.5.2 by justifying the following claim.

Proposition 3.8.49. *The system* (3.8.34) *is solvable in real numbers* α_{sj} *if and only if* (ii) *of Theorem* 3.5.2 *(on p.* 52*) holds.*

Then by invoking Proposition 3.8.48, we arrive at the following corollary.

Corollary 3.8.50. *Let Assumption* 3.8.1 *hold. Then* (ii) *of Theorem* 3.5.2 *implies* (i).

We preface the proof of Proposition 3.8.49 with a useful reformulation of (ii) from Theorem 3.5.2 in terms of the decomposition from Proposition 3.8.2 (on p. 66).

Lemma 3.8.51. *Along with the sets* O_j *from* (i) *of Proposition* 3.8.2*, consider all their unions* $O = \bigcup_{j\in J} O_j$*, where* J *ranges over all groups of sensors. (The union of the empty group of sets* O_j *is included and interpreted as the empty set.) Then* (ii) *of Theorem* 3.5.2 *is true if and only if for any such a union* $O \neq [1:d]$*,*

$$\sum_{s\notin O} \log_2 |\det A_{ss}| < \sum_{j:O_j \not\subset O} \mathfrak{c}_j. \tag{3.8.36}$$

Proof. Due to (i) of Proposition 3.8.2, the sets (3.5.1) $L = \bigcap_{j \in J} L_j^-$ have the form

$$L = \{x : x^s = 0 \; \forall s \in O\}, \qquad \text{where} \quad O = \bigcup_{j \in J} O_j.$$

So (3.8.4) implies $\det A|_L = \prod_{s \notin O} \det A_{ss}$. Hence the left-hand sides in (3.5.2) (on p. 52) and (3.8.36) coincide. The proof is completed by observing that so do the right-hand ones since in (3.5.2) $J(L) = \{j : O_j \subset O\}$ owing to (3.4.6), (3.5.1), and (i) of Proposition 3.8.2. □

Proof of Proposition 3.8.49. *Necessity.* Let (3.8.34) have a solution α_{sj}. Then

$$\sum_{s \notin O} \log_2 |\det A_{ss}| \overset{(3.8.34)}{<} \sum_{s \notin O} \sum_{j=1}^{k} \alpha_{sj} = \sum_{j=1}^{k} \sum_{s \notin O} \alpha_{sj}$$
$$\overset{(3.8.34)}{=\!=} \sum_{j : O_j \not\subset O} \sum_{s \notin O} \alpha_{sj} \leq \sum_{j : O_j \not\subset O} \sum_{s=1}^{d} \alpha_{sj} \overset{(3.8.34)}{<} \sum_{j : O_j \not\subset O} \mathfrak{c}_j;$$

i.e., (3.8.36) holds. By Lemma 3.8.51, so does (3.5.2) (on p. 52).

Sufficiency. Now suppose that (ii) of Theorem 3.5.2 is true. By Lemma 3.8.51, this means that (3.8.36) holds for the union O of any sets O_j, provided $O \neq [1 : d]$. It should be shown that (3.8.34) is solvable in real numbers α_{sj}.

Suppose the contrary. Then the following convex polyhedra in the space of matrices $\alpha = (\alpha_{sj})$ are disjoint:

$$C_1 := \left\{ \alpha : \log_2 |\det A_{ss}| < \sum_{j=1}^{k} \alpha_{sj} \; \forall s \right\},$$
$$C_2 := \left\{ \alpha : \sum_{s=1}^{d} \alpha_{sj} < \mathfrak{c}_j \; \forall j, \alpha_{sj} \geq 0 \; \forall s, j, \alpha_{sj} = 0 \text{ if } s \notin O_j \right\}.$$

Hence they can be separated by a hyperplane: A nonzero matrix $\gamma = (\gamma_{sj})$ exists such that

$$\inf_{\alpha \in C_1} \sum_{s,j} \gamma_{sj} \alpha_{sj} \geq \sup_{\alpha \in C_2} \sum_{s,j} \gamma_{sj} \alpha_{sj}. \tag{3.8.37}$$

The definition of C_1 implies that

$$\inf_{\alpha \in C_1} \sum_{s,j} \gamma_{sj} \alpha_{sj} = \sum_{s=1}^{d} \inf_{(\alpha_j) : \sum_{j=1}^{k} \alpha_j > \log_2 |\det A_{ss}|} \sum_{j=1}^{k} \gamma_{sj} \alpha_j.$$

Every infimum on the right is that of a linear functional over a half-space of (α_j) bounded by a hyperplane with the normal vector $(1, \ldots, 1)$. This infimum is finite only if the functional is generated by a vector colinear with the normal vector. So $\gamma_{sj} = \theta_s \; \forall j$ for some $\theta_s \geq 0$ and $\sum_s \theta_s > 0$. It follows that

$$\inf_{\alpha \in C_1} \sum_{s,j} \gamma_{sj} \alpha_{sj} = \sum_{s=1}^{d} \theta_s \log_2 |\det A_{ss}|.$$

At the same time, the definition of C_2 implies that

$$\sup_{\alpha \in C_2} \sum_{s,j} \gamma_{sj} \alpha_{sj} = \sup_{\substack{\alpha_{sj} \geq 0, \sum_s \alpha_{sj} < \mathfrak{c}_j, \\ s \notin O_j \Rightarrow \alpha_{sj} = 0}} \sum_{sj} \theta_s \alpha_{sj} = \sum_{j=1}^{k} \max_{\alpha_s \geq 0, \sum_s \alpha_s \leq \mathfrak{c}_j} \sum_{s \in O_j} \theta_s \alpha_s$$
$$= \sum_{j=1}^{k} \mathfrak{c}_j \max_{s \in O_j} \theta_s.$$

By (3.8.37), the cone

$$K := \{\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d : \theta_s \geq 0\}$$

contains a nonzero solution of the inequality

$$\sum_{s=1}^{d} \theta_s \log_2 |\det A_{ss}| \geq \sum_{j=1}^{k} \mathfrak{c}_j \max_{s \in O_j} \theta_s. \tag{3.8.38}$$

This cone can be partitioned into a finite number of convex polyhedral subcones such that the right-hand side of (3.8.38) is linear on any subcone. It follows that (3.8.38) must be satisfied on some extreme ray of some subcone. Any of them is bounded by a finite number of hyperplanes, each described by an equation of the form either $\theta_\nu = 0$ or $\theta_\mu = \theta_\nu$, where $\nu \neq \mu$ and $\nu, \mu \in O_j$ for some j. This implies [147, p. 104] that the extreme ray is described by a finite system of such equations, which determines its solution uniquely up to multiplication by a scalar. It is easy to see that the solution of such a system looks as follows: $\theta_s = \theta$ whenever $s \notin \mathcal{O}$, and $\theta_s = 0$ otherwise. Here $\mathcal{O} \subset [1:d]$ is some set, $\mathcal{O} \neq [1:d]$. For vectors on the above extreme ray, we have $\theta > 0$, and (3.8.38) shapes into

$$\sum_{s \notin \mathcal{O}} \log_2 |\det A_{ss}| \geq \sum_{j: O_j \not\subset \mathcal{O}} \mathfrak{c}_j.$$

Changing

$$\mathcal{O} := O := \bigcup_{j: O_j \subset \mathcal{O}} O_j$$

does not alter the right-hand side, and possibly increases the left-hand one, which keeps the inequality true, in violation of (3.8.36). The contradiction obtained proves that the system (3.8.34) is solvable in real numbers α_{sj}. □

3.8.11 Proof of the Sufficiency Part of Theorem 3.5.2 for Systems with Both Unstable and Stable Modes

Now we do the final step 7 (on p.66) from Subsect. 3.8.2 by showing that in Corollary 3.8.50, Assumption 3.8.1 can be dropped.

Proposition 3.8.52. *The statement* (ii) *of Theorem* 3.5.2 *implies* (i).

Proof. Consider the system (3.3.1) (on p. 41) with both unstable and stable modes that satisfies (ii). It is clear that it suffices to stabilize only its unstable part

$$x_+(t+1) = A_+x_+(t) + \pi_+Bu(t), \quad x_+(0) := \pi_+x_0 \in L_+,$$
$$y_+(t) = Cx_+(t). \quad (3.8.39)$$

Here $L_+ := M_{\text{unst}}(A)$ and $L_- := M_{\text{st}}(A)$ are the invariant subspaces of A related to the unstable and stable parts of its spectrum, π_+ and π_- are the projectors onto L_+ parallel to L_- and vice versa, respectively, and $A_\pm := A|_{L_\pm}$. Thanks to the second relation from (3.4.6) (on p. 49), (ii) still holds for the system (3.8.39). By the foregoing, this system can be uniformly and exponentially stabilized by some controller. While constructing it, we employed the parameter $r > 2n$. Now we apply this controller to the primal system (3.3.1). In doing so, the proof of possibility of A.1) (on p. 81) from Subsect. 3.8.8 (starting on p. 83) should be revisited. Indeed the sth coder can be implemented at the jth sensor site (where $s \in O_j$) only if $x^s(\tau_i), \tau_i := i \cdot r$ can be determined there. Formerly this was done on the basis of the past measurements from (3.8.39). Now we must employ the measurements (3.3.2) (on p. 41). This is possible due to (3.4.6) (on p. 49) and (i) of Proposition 3.8.2 (on p. 66) since the dynamics of the system (3.3.1) (on p. 41) is free $u(t) = 0$ at least n time steps before τ_i.

By Definition 3.4.6 (on p. 46), a constant $\mu \in [0,1)$ exists such that whenever a constant K_0 is given and $\|x_0\| \leq K_0$, the following relations hold:

$$\|\pi_+x(t)\| \leq K_x^+\mu^t, \quad \|u(t)\| \leq K_u\mu^t \qquad \forall t = 0,1,2,\dots.$$

The evolution of $x_-(t) := \pi_-x(t)$ is described by the first two equations from (3.8.39), where the index $_+$ is switched to $_-$. Since the operator A_- is stable and the controls $u(t)$ exponentially decay, so do the states $\|x_-(t)\| \leq K_x^-\rho^t$. Here $\rho \in (0,1)$ does not depend on K_0. Since

$$\|x(t)\| = \|x_-(t) + x_+(t)\| \leq \|x_-(t)\| + \|x_+(t)\|,$$

increasing $\mu := \max\{\mu, \rho\}$ yields (3.4.2) (on p. 46). Definitions 3.4.6– 3.4.8 (on p. 46) complete the proof. □

3.8.12 Completion of the Proof of Theorem 3.5.2 and the Proofs of Propositions 3.5.4 and 3.5.8

Proof of Theorem 3.5.2 (on p. 52). It was shown in Sect. 3.7 that (i) ⇒ (ii). The converse (ii) ⇒ (i) is given by Proposition 3.8.52. So it remains to justify (3.5.3) (on p. 53). To this end, we note that the transformation

$$z(t) := \mu^{-t}x(t), \quad v(t) := \mu^{-t}u(t)$$

establishes a one-to-one correspondence between the trajectories $\{x(t), u(t)\}$ and $\{z(t), v(t)\}$ of the systems given by, respectively, (3.3.1) (on p. 41) and the equation

$$z(t+1) = \mu^{-1} A z(t) + \mu^{-1} B v(t).$$

We equip the latter with the sensors $\widetilde{y}_j = C_j z, j \in [1 : k]$. It easily follows from Definitions 3.4.6 and 3.4.7 (on p. 46) that the initial system is uniformly and exponentially stabilizable at a rate $\mu' \in (0, \mu)$ if and only if the second one is uniformly and exponentially stabilizable. By applying the (i) $\Leftrightarrow$ (ii) part of Theorem 3.5.2 to the second system in the case where $\sigma^-(A) = \emptyset$, we get

$$- \dim L \cdot \log_2 \mu + \log_2 |\det A|_L| < \textstyle\sum_{j \notin J(L)} \mathfrak{c}_j \ \forall L \in \mathfrak{L},$$
$$\Updownarrow$$
$$\log_2 \mu > \max_{L \in \mathfrak{L}} \tfrac{1}{\dim L} \left(\log_2 |\det A|_L| - \textstyle\sum_{j \notin J(L)} \mathfrak{c}_j \right).$$

To arrive at (3.5.3), we note that the rate of exponential stabilizability μ^0 is the infimum of all such μ.

The last claim of Theorem 3.5.2 (on p. 52) follows from Proposition 3.5.4, which will be justified next. □

Proof of Proposition 3.5.4 (on p. 54). When applied to the second system in the general case, the criterion (3.5.4) (on p. 53) for stabilizability takes the form

$$- \sum_{\alpha=0}^{\nu} \dim L^\alpha(J) \log_2 \mu + \sum_{\alpha=0}^{\nu} \log_2 \left|\det A|_{L^\alpha(J)}\right| < \sum_{j \notin J} \mathfrak{c}_j \quad \forall J \in \Xi^\nu, \quad (3.8.40)$$

where $\nu = \nu_\mu = 0, \dots, p$ is determined by the inequalities $\rho_{\nu+1} < \mu \le \rho_\nu$ and $\rho_{p+1} := 0$. We recall that $\rho_0 := 1$ and $\rho_\nu, \nu = 1, \dots, p$ are defined in (3.5.5) (on p. 54). It follows that the rate μ^0 of the exponential stabilizability of the primal system is the infimum of all μ's satisfying (3.8.40). The arguments from Remark 3.5.5 (on p. 55) show that this infimum lies in the interval $\left[\rho_{\nu_*+1}, \rho_{\nu_*}\right]$. Hence (3.5.7) (on p. 54) is straightforward from (3.8.40). □

Proof of Proposition 3.5.8 (on p. 55). This proposition is straightforward from Observation 3.8.22 and Remark 3.8.44. □

3.9 Comments on Assumption 3.4.24

Now we explain why this assumption has such a big impact on the controller design. We also briefly discuss ideas underlying such a design in the case where this assumption does not hold.

To start with, we illuminate the role of Assumption 3.4.24 (on p. 49).

Definition 3.9.1. *A "subsystem" arising from* (3.8.3) *(on p.* 66*) is said to be in a* simple relation *with the* j*th sensor if*

- *either it does affect the output of this sensor at all*
- *or the state of this subsystem can be uniquely determined from the sensor outputs.*

The simplest case in stabilization of a multiple sensor system is where the system can be decomposed into independent subsystems each in a simple relation with any sensor. For example, this holds if all eigenvalues of the system are different. However, this is impossible in general. As was shown, a nontrivial Jordan block may form a barrier to decomposition into independent subsystems. The example (3.4.13), (3.4.14) (on p. 51) proves that it may be still worse: The system cannot be disintegrated into (even dependent) subsystems each in simple relations with the sensors. Assumption 3.4.24 in fact describes when this worst case does not occur.

So to deal with the general case where this assumption may be violated, one should cope with the situation where some sensor partly observes some subsystem: Its state cannot be determined at the site of this sensor although the sensor signals contain information about this state. Then an additional problem arises: How do we use this information in the coding and decoding scheme for stabilization purposes? As will be shown, the answer requires the revision of some basic principles on which the design of such schemes was based up to now.

3.9.1 Counterexample

To provide details, we pick natural $\mathfrak{c}$ and real λ numbers such that

$$\lambda \approx \sqrt{2}^{3\mathfrak{c}}, \qquad \lambda < \sqrt{2}^{3\mathfrak{c}}, \tag{3.9.1}$$

and we revert to the example (3.4.13), (3.4.14) (on p. 51):

$$x(t+1) = \lambda x(t) + u(t) \in \mathbb{R}^2,$$
$$y_1(t) = x_1(t), \quad y_2(t) = x_2(t), \quad y_3(t) = x_1(t) - x_2(t), \tag{3.9.2}$$

where $x = \mathbf{col}\,(x_1, x_2)$. There are three sensors each served by an undelayed and lossless channel of capacity $\mathfrak{c}$ bits per unit time; i.e., $\mathfrak{c}_1 = \mathfrak{c}_2 = \mathfrak{c}_3 = \mathfrak{c}$. The necessary conditions for stabilizability (3.5.2) (on p. 52) take the form of the second relation from (3.9.1) and are satisfied.

As was remarked in Subsect. 3.4.5, Assumption 3.4.24 does not hold: One of the sensors observes a certain subsystem only partly for any decomposition of the system (3.9.2).

For example, consider the natural decomposition $x = \mathbf{col}\,(x_1, x_2)$, where x_1 and x_2 are interpreted as the states of the subsystems. They are in simple relations with the first and second sensors. However, they are not in such relations with the third one. Indeed the state x_i influences its outputs $y_3 = x_1 - x_2$ but cannot be determined on the basis of them. Moreover, the only linear coordinate (i.e., function) of the state that can be determined on the site of the third sensor is its output y_3 (up to a scalar factor). Likewise, the first and second sensors permit us to find only x_1 and x_2, respectively. This conclusion holds for any decomposition.

In the remainder of this section, we justify the following two claims:

1. The system (3.9.2) is stabilizable;
2. It cannot be stabilized by a controller with the following features F.1)–F.5).

Before specifying them, we note that they are characteristic for most of the relevant controllers based on the design ideas presented in the literature (see, e.g., [28, 89, 90, 135–138, 149, 184, 202, 204, 221] and the literature therein). We also recall that for the system (3.9.2), the controller consists of three coders and a decoder.

F.1) Not only the "mode" y_i but also its upper (maybe incorrect) bound δ_i is determined at the sensor site;
F.2) The state x and these bounds in fact constitute the state of the closed-loop system, which is time-invariant;
F.3) At the decoder site, the information about the "mode" $y_i(t)$ comes to its quantized scaled value $e_i(t) = \mathfrak{Q}_i[\delta_i(t)^{-1} y_i(t)]$ given by a static quantizer $\mathfrak{Q}_i(\cdot)$ with convex level sets and the number m_i of levels matching $m_i + 1 \leq 2^{\mathfrak{c}}$ the channel capacity[23];
F.4) The next bound $\delta_i(t+1)$ is determined from $\delta_i(t)$ and the knowledge of whether $e_i(t) = \maltese$;
F.5) Whenever all bounds are true $\delta_i(t) \geq |y_i(t)| \, \forall i$, they remain true afterward.

It is clear that these features mainly concern the coding algorithm.

In the remainder of this subsection, we prove claim 2. Claim 1 (i.e., stabilizability of the system) will be justified in the next subsection.

Lemma 3.9.2. *Let* $\mathfrak{c} \geq 2$ *and a controller satisfying* F.1)–F.5) *be given. Then the closed-loop system* (3.9.2) *is not stable. Moreover,*

$$\limsup_{t\to\infty} \sup_{x_0 \in B_0^{\chi}} \|x(t)\| = \infty \tag{3.9.3}$$

for all $\chi > 0$ *and initial bounds* $\delta_i^0 > 0$.

Proof. By F.2) and F.4), $\delta_i(t+1) = \mathcal{D}_i[\delta_i(t)]$ whenever $e_i(t) \neq \maltese$. We are going to estimate $\mathcal{D}_i(\cdot)$ from below. Due to F.3), any quantizer $\mathfrak{Q}_i$ is related to a partition of the interval $[-1, 1]$ into m_i subintervals (level sets)

$$\Delta_1^{(i)}, \ldots, \Delta_{m_i}^{(i)}.$$

Let $\alpha_j^{(i)}$ and $\beta_j^{(i)}$ denote the left and right end points of $\Delta_j^{(i)}$, respectively. Since $m_i \leq 2^{\mathfrak{c}} - 1$, one of them has the length

$$\beta_{j^i}^{(i)} - \alpha_{j^i}^{(i)} > 2 \cdot 2^{-\mathfrak{c}}.$$

Now we pick $\delta > 0$, set the initial bounds $\delta_1(0) = \delta_2(0) = \delta, \delta_3(0) = 2\delta$, and note that all initial states from the segment

[23] We recall that m_i is the number of the quantizer outputs different from $\maltese$. So the total number of outputs is $m_i + 1$.

$$S := \{x_0 = \mathbf{col}\,(\delta\alpha^{(1)}_{j^1} + \theta, \delta\alpha^{(2)}_{j^2} + \theta) : 0 < \theta \leq 2\delta 2^{-\mathfrak{c}}\}$$

give rise to common outputs for each quantizer $i = 1, 2, 3$ at $t = 0$. So they give rise to a common control $u(0) = \mathbf{col}\,(u_1, u_2)$. For all these states, the above initial bounds are correct. Then F.5) ensures that for $i = 1, 2$

$$\delta_i(1) = \mathfrak{D}_i[\delta_i(0)] = \mathfrak{D}_i[\delta] \geq |y_i(1)| = |\lambda x_i(0) + u_i|.$$

Here $\lambda x_i(0) + u_i$ runs over an interval of length $2\lambda\delta 2^{-\mathfrak{c}}$ as x_0 ranges over S. Hence

$$\mathfrak{D}_i(\delta) \geq \lambda\delta 2^{-\mathfrak{c}} \tag{3.9.4}$$

for $i = 1, 2$. This inequality is extended on $i = 3$ by putting

$$\delta_1(0) = \delta_3(0) = \delta, \qquad \delta_2(0) = 2\delta,$$
$$S := \{x_0 = (\delta\alpha^{(1)}_{j^1} + \theta, \delta\alpha^{(1)}_{j^1} - \delta\alpha^{(3)}_{j^3}) : 0 \leq \theta \leq 2\delta 2^{-\mathfrak{c}}\}$$

and retracing the above arguments.

Now we suppose that (3.9.3) violates for some $\delta_i^0 > 0$ and $\chi > 0$

$$c := \sup_{x_0 \in B_0^\chi, t = 0, 1, \ldots} \|x(t)\| < \infty.$$

By decreasing χ, one can ensure that $\chi < \min\{\delta_1^0, \delta_2^0, \frac{1}{2}\delta_3^0\}$. Then for all initial states $x_0 \in B_0^\chi$, the bounds δ_i are correct for $t = 0$. Thanks to F.4) and F.5), they remain correct for all t and common for all $x_0 \in B_0^\chi$. Then (3.9.4) yields

$$\delta_i(t) \geq \chi \left(\frac{\lambda}{2^{\mathfrak{c}}}\right)^t.$$

Here $\lambda \approx \sqrt{2}^{3\mathfrak{c}}$ by (3.9.1). So $\delta_i(t) \to \infty$ as $t \to \infty$. As a result, the interval $[-2c, 2c]$ is covered by at most two intervals of the form $\delta_i(t)\Delta_j^{(i)}$ for each $i = 1, 2, 3$, provided t is large enough. Since

$$\|x(t)\| \leq c \Rightarrow |y_i(t)| \leq 2c, \qquad i = 1, 2, 3,$$

this and F.3) mean that for $x_0 \in B_0^\chi$ each $\mathfrak{Q}_i$ in fact acts as a binary quantizer (i.e., that with only two outputs) at any large time. Thus the decoder receives in fact no more than one bit of information about the processes with $x_0 \in B_0^\chi$ via each channel. By treating three channels as one and invoking Lemma 3.5.11 (on p. 56), we arrive at the inequality $\lambda^2 \leq 2^3 \Leftrightarrow \lambda \leq 2^{3/2}$. At the same time, $\mathfrak{c} \geq 2$ and so $\lambda \approx \sqrt{2}^{3\mathfrak{c}} > 2^{3/2}$. The contradiction obtained proves the lemma. □

3.9.2 Stabilizability of the System (3.9.2)

Now we show that despite Lemma 3.9.2, the system (3.9.2) is yet stabilizable. The stabilizing controller will lack the properties F.2) and F.3). It will employ a 2-periodic quantization scheme applied to not only scaled but also shifted observations.

We offer only a sketch of the proof. This is because our objective is to highlight the design ideas, whereas (3.9.2) is of little interest by its own right.

We equip every sensor with a coder (see Fig. 3.7). At any time, the jth coder selects a code-symbol $e_j(t)$ from the $\mathfrak{c}$-bits channel alphabet and sends it over the jth channel to the decoder. Furthermore, this coder computes recursively an upper bound $\delta_j(t)$ for the current observation

$$-\delta_j(t) \leq y_j(t) < \delta_j(t). \tag{3.9.5}$$

This computation is driven by the sequence of messages sent from the jth coder to the decoder, so that the decoder be able to compute $\delta_j(t)$ by itself.

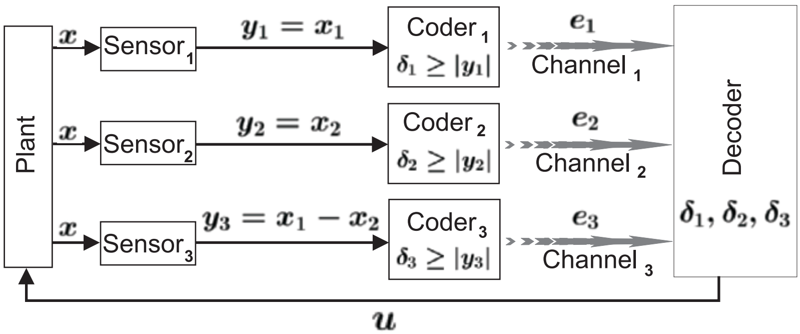


Fig. 3.7. Stabilization of the system (3.9.2).

We start to consider the stabilization process since the moment (treated as $t = 0$) when all bounds are correct. This may be due to the knowledge of such bounds for the initial state. Otherwise this may be achieved during a preparatory stage of the stabilization process via successive multiplying δ_j by a sufficiently large factor, like in Subsect. 3.8.4 (see p. 68). As will be shown, these bounds are kept true afterward by the algorithm to be proposed. It will be designed so that the following relations between the bounds δ_j are also kept true:

$$\begin{array}{lll} \delta_1(t) = \delta_2(t), \quad \delta_3(t) = \delta_1(t) + \delta_2(t) & \text{at odd times } t \\ \delta_2(t) = \delta_1(t) + \delta_3(t), \quad \delta_3(t) = \frac{2\delta_1(t)}{N} & \text{at even times } t \end{array}, \quad (N := 2^{\mathfrak{c}}). \tag{3.9.6}$$

Description of the Controller

The coders and decoder act in different ways at odd and even times t.

At odd times, the state $x(t)$ lies in the following square by (3.9.5):

$$x(t) \in M(t) := \{x : -\delta_j(t) \leq y_j < \delta_j(t), j = 1, 2\}. \tag{3.9.7}$$

1) By acting independently on and via their own observation and channel, respectively, the **first and second** ($j = 1, 2$) **coders** in fact apply the uniform N^2-level

quantization scheme (see Example 3.8.11 on p. 70) to the square (3.9.7) and notify the decoder which of the level domains contains the state (see Fig. 3.8a). Formally, for $j = 1, 2$, the jth coder determines which interval

$$[\mu_{i'}^{(j)}, \mu_{i'+1}^{(j)}), \quad \text{where} \quad \mu_i^{(j)} := i\frac{2\delta_j(t)}{N}, \quad i' = 0, \dots, N-1,$$

contains

$$\overline{y}_j(t) := y_j(t) + \delta_j(t)$$

and notifies the decoder about its serial number $i' = i^{(j)}$;

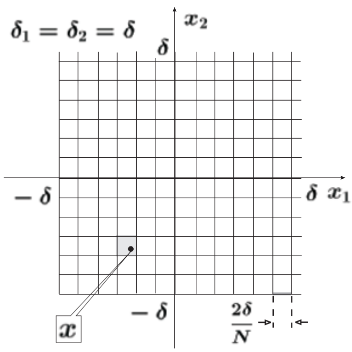


Fig. 3.8(a). Uniform square quantizer.

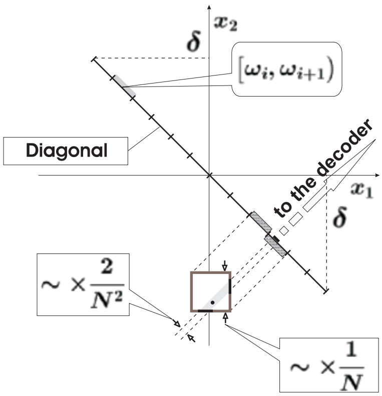


Fig. 3.8(b). Location of the state.

2) At the same time, the **third coder** ($j = 3$) does the following:
   a) finds which of $2N$ intervals

$$[\omega_i, \omega_{i+1}), \qquad \text{where} \quad \omega_i := i\frac{\delta_j(t)}{N}, \tag{3.9.8}$$

   contains $\overline{y}_j(t)$;
   b) uniformly partitions this interval into $N/2$ subintervals

$$[\omega_i^{(\nu)}, \omega_i^{(\nu+1)}), \qquad \omega_i^{(\nu)} := \omega_i + 2\nu\frac{\omega_{i+1} - \omega_i}{N} \tag{3.9.9}$$

   and notifies the decoder which of them $\nu = 0, \dots, N/2 - 1$ contains $\overline{y}_j(t)$, and
   c) uses the remaining bit to make the decoder aware of whether i is odd or even.

Explanation 3.9.3. In fact, the third sensor displays the orthogonal projection of the point $x = \mathbf{col}\,(x_1, x_2)$ onto the line $x_1 = -x_2$. The embedding square (3.9.7) is projected onto its diagonal. The set of the intervals (3.9.8) can be interpreted as the uniform partition of the diagonal into $2N$ segments (see Fig. 3.8b). It is easy to check that the shadowed square from Fig. 3.8a (the quantizer level domain) is projected onto the union of two neighboring segments. So the decoder can determine this union from the data provided by the first and second coders. The bit from c) enables the decoder to select the segment (3.9.8) that contains the projection from the above two neighboring ones.

Hence except for one bit, the remaining bits available to the third coder can be used to improve the precision in location of the projection within the segment (3.9.8), which is done by b).

Remark 3.9.4. In the more formal analytical way, the above arguments look as follows. It is easy to check that

$$(i^{(1)} + i_*^{(2)})\frac{\delta_3(t)}{N} < \overline{y}_3(t) < (i^{(1)} + i_*^{(2)} + 2)\frac{\delta_3(t)}{N}, \qquad i_*^{(2)} := N - i^{(2)} - 1.$$

So either $i = i^{(1)} + i_*^{(2)}$ or $i = i^{(1)} + i_*^{(2)} + 1$. Hence i can be found from the data $i^{(1)}, i^{(2)}$ given by the first and second sensors, and the bit from c).

3). Based on the data $i^{(1)}$ from the first sensor, the **decoder** finds the strip

$$\{x : y_1 \in -\delta_1(t) + [\mu^{(1)}_{i^{(1)}}, \mu^{(1)}_{i^{(1)}+1})\}$$

that contains $x(t)$. By finding i on the basis of the data from the first and second sensors, along with the bit from 2.c), and using the data ν described in 2.b), it finds another such strip (which is $N/2$-times "narrower" than the first one)

$$\{x : y_3 \in -\delta_3(t) + [\omega_i^{(\nu)}, \omega_i^{(\nu+1)})\}.$$

Then the decoder selects a control driving the system from the center of the intersection of the strips (the shadowed domain in Fig. 3.9a) to zero.

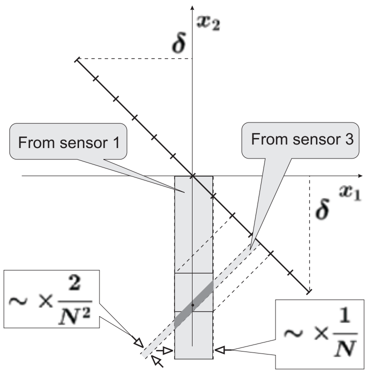


Fig. 3.9(a). Location of the state: the view of the decoder.

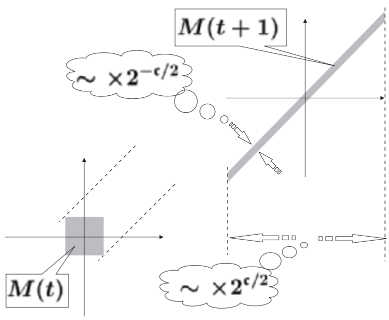


Fig. 3.9(b). Domain transformation at an odd step.

Explanation 3.9.5. In fact, the decoder might locate the state better by computing the shadowed domain from Fig. 3.8b. We make use of the domain from Fig. 3.9a since this ensures stability by employing simpler formulas.

Remark 3.9.6. By invoking (3.9.2) and (3.9.7), it is easy to check that the selected control drives the system to the state

$$x(t+1) \in M(t+1) := \{x : -\varepsilon' \leq y_1 < \varepsilon', \quad -\varepsilon'' \leq y_3 < \varepsilon''\},$$

where

$$\varepsilon' := \lambda\delta_1(t)N^{-1} \overset{(3.9.1)}{\approx} \delta_1(t)2^{3/2\mathfrak{c}}2^{-\mathfrak{c}} = \delta_1(t)2^{\mathfrak{c}/2} \overset{\mathfrak{c}\geq 2}{>} \delta_1(t),$$
$$\varepsilon'' := 2\lambda\delta_1(t)N^{-2} \overset{(3.9.1)}{\approx} \underbrace{2\delta_1(t)}_{=\delta_3(t) \text{ by } (3.9.6)} 2^{-\mathfrak{c}/2} < \delta_3(t).$$

Thus, for one step, the domain $M(t)$ locating the state $M(t) \ni x(t)$ is stretched in one direction and tightened in the other (see Fig. 3.9b).

It is easy to see that the area of the set $M(t)$ progresses by multiplying by $\frac{2\lambda^2}{2^{3\mathfrak{c}}}$ per odd step. So it decreases if $\lambda < \frac{1}{\sqrt{2}}2^{3/2\mathfrak{c}}$. This is a bit more restrictive than the condition $\lambda < 2^{3/2\mathfrak{c}}$ from (3.9.1). This gap can be discarded by increasing the sample period to r time units. Indeed, this "transforms" λ into λ^r, $\mathfrak{c}$ into $r\mathfrak{c}$, and the above condition $\lambda < 2^{\frac{3}{2}\mathfrak{c}-\frac{1}{2}}$ into $\lambda < 2^{\frac{3}{2}\mathfrak{c}-\frac{1}{2r}} \approx 2^{3/2\mathfrak{c}}$ for $r \approx \infty$.

4). Both the jth coder and the decoder define the **next number** δ_j as the upper bound for y_j when x ranges over $M(t+1)$; i.e.,

$$\delta_1(t+1) := \lambda N^{-1}\delta_1(t),$$
$$\delta_3(t+1) := \lambda N^{-2}\delta_3(t),$$
$$\delta_2(t+1) := \lambda\delta_2(t)N^{-1}\left[1 + \frac{2}{N}\right].$$

Observation 3.9.7. *This keeps* (3.9.6) *true at the next (even) time* $t := t+1$.

Remark 3.9.8. For odd steps, the first and second coders use the available $\mathfrak{c}$ bits in order to increase the accuracy of the state description by $2\mathfrak{c}$ bits per step. Due to the state dynamics, this is insufficient even to keep the accuracy at a given level. Indeed the required number of bits is no less than the topological entropy of the system at hand, which equals $2\log_2\lambda \approx 3\mathfrak{c}$ by (3.9.1). This insufficiency can be compensated by the bits available to the third sensor. However this sensor cannot distinguish between points lying on lines parallel to $x_1 = x_2$. So the corresponding bits can aid in refining the state description only in the perpendicular direction.

This explains why the state location domain is tightened in this direction and stretched along the above line $x_1 = x_2$. Another conclusion is that repeating of the above scheme (where the first and second sensors supply a primary "rough" information about the state and the third sensor is used to refine it) will stretch the state location domain along the line $x_1 = x_2$ further and will not result in stabilization. So the roles of the sensors should be interchanged. This is done at the even steps. Then the first and third sensors supply primary data, whereas the second one refines it.

At even times, the state $x(t)$ lies in the following parallelogram by (3.9.5):

$$x(t) \in M(t) := \{x : -\delta_i(t) \leq y_i < \delta_i(t), i = 1, 3\}, \tag{3.9.10}$$

where $\delta_3(t) = \frac{2\delta_1(t)}{N}$ and $\delta_2(t) = \delta_1(t) + \delta_3(t)$ by (3.9.6) (see Fig. 3.10a).

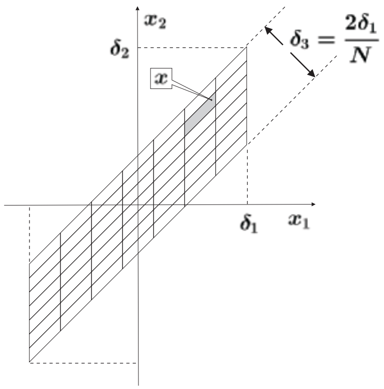


Fig. 3.10(a). Location of the state: the view of the first and third sensors.

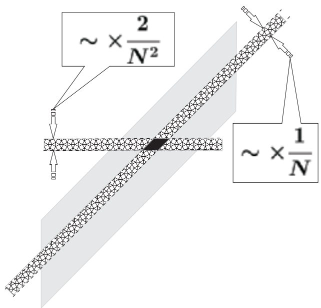


Fig. 3.10(b). Location of the state: the view of the decoder.

5). The operation 1) is carried out by the first and third **coders** $j = 1, 3$, and 2) is done by the second one $j = 2$ with ω_i altered: $\omega_i := i\frac{2\delta_2(t)}{N}$.

Explanation 3.9.9. By acting independently on and via their own observation and channel, respectively, the first and third coders in fact apply the N^2-level quantization scheme to the parallelogram (3.9.10) and notify the decoder which of the sub-parallelograms contains the state (see the shadowed parallelogram in Fig. 3.10a).

By projecting this sub-parallelogram on the line $x_1 = 0$, the decoder can determine a segment of length $\Delta := \frac{2\delta_2(t)}{N}$ containing the measurement $y_2(t) = x_2(t)$ from the second sensor. This segment intersects at most two intervals (3.9.8) of the same length Δ. The interval (3.9.8) found by the second coder is among these two ones. So the decoder can uniquely restore this interval by using the bit from c).

In the more formal analytical way, the above arguments look as follows. It is easy to check that $i^{(1)}, i^{(3)} \in [0 : N - 1]$ and

$$\frac{-\delta_2(t) + \frac{2\delta_2(t)}{N}\left[i^{(1)} + \frac{2}{N+2}\left(N-1-i^{(3)}-i^{(1)}\right)\right] < y_2(t) < -\delta_2(t) + \frac{2\delta_2(t)}{N}\left[i^{(1)} + \frac{2}{N+2}\left(N-1-i^{(3)}-i^{(1)}\right)+1\right]}{\Downarrow}$$

$$\begin{aligned} \text{either} \quad & -\delta_2(t) + \tfrac{2\delta_2(t)}{N}\left\lfloor a(i^{(1)}, i^{(3)})\right\rfloor < y_2(t) \\ & < -\delta_2(t) + \tfrac{2\delta_2(t)}{N}\left(\left\lfloor a(i^{(1)}, i^{(3)})\right\rfloor + 1\right) \\ \text{or} \quad & -\delta_2(t) + \tfrac{2\delta_2(t)}{N}\left(\left\lfloor a(i^{(1)}, i^{(3)})\right\rfloor + 1\right) \le y_2(t) \\ & < -\delta_2(t) + \tfrac{2\delta_2(t)}{N}\left(\left\lfloor a(i^{(1)}, i^{(3)})\right\rfloor + 2\right) \end{aligned}$$

where

$$a(i^{(1)}, i^{(3)}) := \frac{2}{N+2}\left(N-1-i^{(3)}-i^{(1)}\right) + i^{(1)}.$$

It follows that the serial number $i = 1, 2$ of the interval $[\omega_i, \omega_{i+1})$ (with $\omega_i := i\frac{2\delta_2(t)}{N}$) containing $y_2(t)$ equals either $\left\lfloor a(i^{(1)}, i^{(3)})\right\rfloor$ or $\left\lfloor a(i^{(1)}, i^{(3)})\right\rfloor + 1$. So it can be found by the decoder on the basis of the data $i^{(1)}, i^{(3)}$ from the first and third sensors and the information from 2.c).

So by using the data from b), the decoder can also determine the corresponding subinterval (3.9.9) of length $\frac{4\delta_2(t)}{N^2}$.

Remark 3.9.10. The definition of the quantity ω_i from 3) is altered at even steps since the projections of the quantizer level domains from Fig. 3.10a on the line $x_1 = 0$ are arranged in a less regular way, as compared with Fig. 3.8a. In the case from Fig. 3.8a, all projections can be obtained via successive displacement of one of them by exactly the half length. In the case from Fig. 3.10a, the displacement should be by $\frac{2}{N+2}\times$ length.

6) The **decoder** finds a domain containing $x(t)$ by intersecting two strips:

$$\left\{x : y_3 \in -\delta_3(t) + [\mu^{(3)}_{i^{(3)}}, \mu^{(3)}_{i^{(3)}+1})\right\}, \qquad \left\{x : y_2 \in -\delta_2(t) + [\omega_i^{(\nu)}, \omega_i^{(\nu+1)})\right\}.$$

Then the decoder selects a control driving the system from the center of this domain to zero (see Fig. 3.10b).

Remark 3.9.11. By invoking (3.9.2), it is easy to check that the selected control drives the system to the state

$$x(t+1) \in M'(t+1) := \left\{x : -\varepsilon' \le y_3 < \varepsilon', \quad -\varepsilon'' \le y_2 < \varepsilon''\right\},$$

where

$$\begin{aligned} \varepsilon' := \lambda\delta_3(t)N^{-1}, \quad \varepsilon'' &:= 1/2[\omega_i^{(\nu+1)} - \omega_i^{(\nu)}] \\ &= \frac{2\lambda\delta_2(t)}{N^2} = \lambda\frac{N+2}{N^2}\delta_3(t) \le 3\lambda\delta_3(t)N^{-1}. \end{aligned}$$

The set $M'(t+1)$ is covered by the square

$$M(t+1) := \left\{ x : |x_1|, |x_2| \leq 3\lambda \frac{\delta_3(t)}{N} \right\}.$$

7) Both the jth coder and the decoder define the **next number** δ_j so that $\delta_1 = \delta_2$ become the half length of the edge of $M(t+1)$ and $\delta_3 = \delta_1 + \delta_2$; i.e.,

$$\delta_1(t+1) := \frac{6\lambda}{N^2}\delta_1(t),$$

$$\delta_2(t+1) := \frac{6\lambda}{N(N+2)}\delta_2(t),$$

$$\delta_3(t+1) := 6\lambda\delta_3(t)N^{-1}.$$

Thus the description of the stabilization process is completed.

Observation 3.9.12. *For two steps, the square $M(t) \ni x(t)$ with the edge $2\delta_1(t)$ (where t is odd) is transformed into the square $M(t+2)$ with the edge*

$$2\delta_1(t) \times \frac{6\lambda^2}{N^3}.$$

So the system is stabilized if

$$\frac{6\lambda^2}{N^3} < 1 \Leftrightarrow \lambda < \frac{1}{\sqrt{6}} 2^{3/2\mathfrak{c}}. \tag{3.9.11}$$

Remark 3.9.13. The condition (3.9.11) is a bit worse than the necessary condition for stabilizability $\lambda < 2^{3/2\mathfrak{c}}$ from (3.9.1). This gap can be discarded by increasing the sample period to r time units, where r is large enough. Indeed, this "transforms" λ into λ^r, $\mathfrak{c}$ into $r\mathfrak{c}$, and the sufficient condition (3.9.11) for stabilizability into

$$\lambda^r < 6^{-\frac{1}{2}} 2^{\frac{3}{2} r\mathfrak{c}} \Leftrightarrow \lambda < 2^{\frac{3}{2}\mathfrak{c}} 6^{-\frac{1}{2r}} \approx 2^{3/2\mathfrak{c}} \quad \text{for} \quad r \approx \infty.$$

Thus we see that even for a very simple system, violation of Assumption 3.4.24 (on p. 49) complicates the coding–decoding scheme.

4

Detectability and Output Feedback Stabilizability of Nonlinear Systems via Limited Capacity Communication Channels

4.1 Introduction

In the previous chapter, the problem of stabilizability via limited capacity communication channels was studied for linear systems. In this chapter, we consider the problem of stabilizability for nonlinear networked systems with a globally Lipschitz nonlinearity that is common in absolute stability and robust control theories; see, e.g., [144, 151, 222]. Furthermore, we obtain a criterion of detectability of a nonlinear system via limited capacity communication channels. Several results on state feedback stabilization of uncertain plants via communication channels were obtained in [150]. The problem of local stabilization of singular points of nonlinear networked control systems was addressed in [140, 141]. Unlike [140, 141, 150], we consider a much more difficult case of output feedback stabilization. A criterion for output feedback stabilization of networked nonlinear systems was given in [41]. However, the systems considered in [41] are required to be transformable into some triangular form, whereas the nonlinear systems studied in this chapter are not assumed to satisfy such a strong requirement. The results of this chapter are given in terms of an algebraic Riccati inequality that originates in the theory of robust and H^∞ control; see, e.g., [148, 151, 174, 178–180].

It should be pointed out that unlike Chaps. 2 and 3, which considered discrete-time systems, this chapter deals with a continuous-time plant.

The remainder of the chapter is organized as follows. Section 4.2 addresses the problem of detectability of a nonlinear system via a digital communication channel. The output feedback stabilizability problem is studied in Sect. 4.3. As an illustration, Sect. 4.4 presents simulation results on the output feedback control of a flexible joint robotic system.

The main results of the chapter were originally published in [170]. The illustrative example from Sect. 4.4 was first presented in the paper [172], which also contains an extension of the results of [170]. Moreover, further extensions of the results of this chapter to the case of nonlinear systems with monotonic nonlinearities were obtained in [32].

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_4,

4.2 Detectability via Communication Channels

In this section, we consider nonlinear continuous-time dynamical systems of the form:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + B_1 f(z(t));\\ z(t) &= Kx(t);\\ y(t) &= Cx(t),\end{aligned} \tag{4.2.1}$$

where $x(t) \in \mathbb{R}^n$ is the state; $z(t) \in \mathbb{R}^q$ is a linear output; $y(t) \in \mathbb{R}^k$ is the measured output; A, B_1, K, and C are given matrices of the corresponding dimensions, and $f(z(t)) \in \mathbb{R}^p$ is a given continuous nonlinear vector function. We also assume that initial conditions of the system (4.2.1) lie in a known bounded set $\mathfrak{X}_0$:

$$x(0) \in \mathfrak{X}_0. \tag{4.2.2}$$

We assume that the vector function $f(\cdot)$ satisfies the following globally Lipschitz condition:

$$\|f(z_1) - f(z_2)\|^2 \le \|z_1 - z_2\|^2 \qquad \forall z_1, z_2. \tag{4.2.3}$$

The requirement (4.2.3) is a special case of a typical sector-type constraint from the absolute stability theory; see, e.g., [144, 151, 222]. A simple common example of such a constraint is a scalar nonlinearity satisfying conditions $f(0) = 0$ and

$$-1 \le \frac{f(z_1) - f(z_2)}{z_1 - z_2} \le 1 \qquad \forall z_1, z_2.$$

In our detectability problem, a sensor measures the state $x(t)$ and is connected to the controller, which is at the remote location. Moreover, the only way of communicating information from the sensor to that remote location is via a digital communication channel that carries one discrete-valued symbol $h(jT)$ at time jT, selected from a coding alphabet $\mathfrak{H}$ of size l. Here $T > 0$ is a given period, and $j = 1, 2, 3, \ldots$.

This restricted number l of code words $h(jT)$ is determined by the transmission data rate of the channel. For example, if μ is the number of bits that our channel can transmit at any time instant, then $l = 2^\mu$ is the number of admissible code words. We assume that the channel is a perfect noiseless channel and that there is no time delay.

We consider the problem of estimation of the state $x(t)$ via a digital communication channel with a bit-rate constraint. Our state estimator consists of two components. The first component is developed at the measurement location by taking the measured output $y(\cdot)$ and coding to the codeword $h(jT)$. This component will be called "coder." Then the codeword $h(jT)$ is transmitted via a limited capacity communication channel to the second component, which is called "decoder." The second component developed at the remote location takes the codeword $h(jT)$ and produces the estimated state $\hat{x}(t)$. This situation is illustrated in Fig. 2.1 (on p. 15).

The coder and the decoder are of the following forms:

Coder:

$$h(jT) = \mathcal{F}_j \left(y(\cdot)|_0^{jT} \right); \tag{4.2.4}$$

Decoder:

$$\hat{x}(t)|_{jT}^{(j+1)T} = \mathcal{G}_j \left(h(T), h(2T), ..., h((j-1)T), h(jT) \right). \tag{4.2.5}$$

Here $j = 1, 2, 3, \ldots$.

Definition 4.2.1. *The system* (4.2.1)*, is said to be detectable via a digital communication channel of capacity* l *if a coder–decoder pair* (4.2.4)*,* (4.2.5) *exists with a coding alphabet of size* l *such that*

$$\lim_{t \to \infty} \|x(t) - \hat{x}(t)\|_\infty = 0 \tag{4.2.6}$$

for any solution of (4.2.1)*,* (4.2.2)*, and* (4.2.3)*. A coder–decoder pair* (4.2.4)*,* (4.2.5) *satisfying condition* (4.2.6) *is said to be* detecting.

4.2.1 Preliminary Lemmas

We will consider the following pair of Riccati algebraic inequalities

$$(A - \alpha I)^{\mathrm{T}} X + X(A - \alpha I) + K^{\mathrm{T}} K + X B_1 B_1^{\mathrm{T}} X < 0; \tag{4.2.7}$$

$$Y A + A^{\mathrm{T}} Y + Y B_1 B_1^{\mathrm{T}} Y + K^{\mathrm{T}} K - \alpha_1 C^{\mathrm{T}} C < 0, \tag{4.2.8}$$

where I is the identity square matrix, and $\alpha > 0$ and $\alpha_1 > 0$ are given numbers.

In this subsection, we prove two preliminary lemmas.

Lemma 4.2.2. *Suppose that for some* $\alpha > 0$*, a solution* $X > 0$ *of the Riccati inequality* (4.2.7) *exists. Then a time* $T_0 > 0$ *exists such that for any* $T \geq T_0$ *and any two solutions* $x_1(\cdot), x_2(\cdot)$ *of the system* (4.2.1)*,* (4.2.2)*,* (4.2.3)*, the following inequality holds:*

$$\|x_1(t+T) - x_2(t+T)\|_\infty \leq e^{\alpha T} \|x_1(t) - x_2(t)\|_\infty \tag{4.2.9}$$

for all $t \geq 0$.

Proof of Lemma 4.2.2. Let

$$\begin{aligned} \tilde{x}(t) &:= e^{-\alpha t}(x_1(t) - x_2(t)); \\ \phi(t) &:= e^{-\alpha t}(f(Kx_1(t)) - f(Kx_2(t))). \end{aligned} \tag{4.2.10}$$

Then $\tilde{x}(\cdot), \phi(\cdot)$ obviously satisfy the equation

$$\begin{aligned} \dot{\tilde{x}}(t) &= (A - \alpha I)\tilde{x}(t) + B_1 \phi(t); \\ \tilde{z}(t) &= K\tilde{x}(t), \end{aligned} \tag{4.2.11}$$

and the constraint

$$\|\phi(t)\|^2 \leq \|\tilde{z}(t)\|^2. \tag{4.2.12}$$

Then according to the strict bounded real lemma (see, e.g., Lemma 3.1.2 of [151] the system (4.2.11), (4.2.12) is quadratically stable. This implies that a time $T_0 > 0$ exists such that for any $T \geq T_0$ and any solution $\tilde{x}(\cdot)$ of the system (4.2.11), (4.2.12), the following inequality holds:

$$\|\tilde{x}(t+T)\|_\infty \leq \|\tilde{x}(t)\|_\infty \quad \forall t \geq 0. \tag{4.2.13}$$

The properties (4.2.13) and (4.2.10) immediately imply (4.2.9). This completes the proof of Lemma 4.2.2. □

Now consider the following state estimator that will be a part of our proposed coder:

$$\begin{aligned} \dot{\tilde{x}}(t) &= (A - GC)\tilde{x}(t) + Gy(t) + B_1 f(\tilde{z}(t)); \\ \tilde{z}(t) &= K\tilde{x}(t), \quad \tilde{x}(0) = 0. \end{aligned} \tag{4.2.14}$$

Furthermore, we introduce the gain G by

$$G := \frac{\alpha_1}{2} Y^{-1} C^{\mathrm{T}}, \tag{4.2.15}$$

where $Y > 0$ is a solution of (4.2.8).

Lemma 4.2.3. *Suppose that for some $\alpha_1 > 0$, a solution $Y > 0$ of the Riccati inequality* (4.2.8) *exists. Then a time $T_0 > 0$ and a constant $\alpha_0 > 0$ exist such that*

$$\|x(t+T) - \tilde{x}(t+T)\|_\infty \leq e^{-\alpha_0 T}\|x(t) - \tilde{x}(t)\|_\infty \tag{4.2.16}$$

for any $t \geq 0, T > T_0$ and any solution of (4.2.1)–(4.2.3)*,* (4.2.14)*, and* (4.2.15)*.*

Proof of Lemma 4.2.3 Let

$$\begin{aligned} \xi(t) &:= x(t) - \tilde{x}(t)); \quad \zeta(t) := K\xi(t); \\ \phi(t) &:= f(z(t)) - f(\tilde{z}(t)). \end{aligned} \tag{4.2.17}$$

Then $\xi(\cdot), \phi(\cdot)$ obviously satisfy the equation

$$\dot{\xi}(t) = (A - GC)\xi(t) + B_1\phi(t) \tag{4.2.18}$$

and the constraint

$$\|\phi(t)\|^2 \leq \|\zeta(t)\|^2. \tag{4.2.19}$$

Since $Y > 0$ is a solution of (4.2.8) and G is defined by (4.2.15), the matrix Y is also a positive-definite solution of the Riccati inequality

$$(A - GC)^{\mathrm{T}}Y + Y(A - GC) + K^{\mathrm{T}}K + YB_1B_1^{\mathrm{T}}Y < 0.$$

Therefore, according to the strict bounded real lemma (see, e.g., Lemma 3.1.2 of [151]), the system (4.2.17), (4.2.18), (4.2.19) is quadratically stable. Now the statement of the lemma immediately follows from quadratic stability. This completes the proof of Lemma 4.2.3. □

4.2.2 Uniform State Quantization

Our proposed coder–decoder pair uses uniform quantization of the states $\tilde{x}$ of the system (4.2.14) in which the same number of bits is used to quantize each state variable. The corresponding quantizer was introduced in Remark 3.8.13 (on p. 71). For the convenience of the reader, now we recall the basic formulas describing this quantizer. Furthermore, now we consider the case where the effective quatization domain is an arbitrary cube, whereas only the cube with side length 2 was considered in Remark 3.8.13.

To quantize the state space of the estimator (4.2.14), let $a > 0$ be a given constant and consider the set:

$$B_a := \{x \in \mathbb{R}^n : \|x\|_\infty \leq a\}.$$

The state space of the system (4.2.14) is quantized by dividing the set B_a into q^n hypercubes, where q is a specified integer. Indeed, for each $i \in \{1, 2, \dots, n\}$, we divide the corresponding component of the vector $\tilde{x}_i$ into q intervals as follows:

$$\begin{aligned} I_1^i(a) &:= \left\{\tilde{x}_i : -a \leq \tilde{x}_i < -a + \frac{2a}{q}\right\}; \\ I_2^i(a) &:= \left\{\tilde{x}_i : -a + \frac{2a}{q} \leq \tilde{x}_i < -a + \frac{4a}{q}\right\}; \\ &\vdots \\ I_q^i(a) &:= \left\{\tilde{x}_i : a - \frac{2a}{q} \leq \tilde{x}_i \leq a\right\}. \end{aligned} \tag{4.2.20}$$

Then for any $\tilde{x} \in B_a$, unique integers $i_1, i_2, \dots, i_n \in \{1, 2, \dots, q\}$ exist such that

$$\tilde{x} \in I_{i_1}^1(a) \times I_{i_2}^2(a) \times \dots \times I_{i_n}^n(a).$$

Also, corresponding to the integers $i_1, i_2, \dots, i_n$, we define the vector

$$\eta(i_1, i_2, \dots, i_n) := \begin{bmatrix} -a + a(2i_1 - 1)/q \\ -a + a(2i_2 - 1)/q \\ \vdots \\ -a + a(2i_n - 1)/q \end{bmatrix}. \tag{4.2.21}$$

This vector is the center of the hypercube $I_{i_1}^1(a) \times I_{i_2}^2(a) \times \dots \times I_{i_n}^n(a)$ containing the original point $\tilde{x}$.

Note that the regions $I_{i_1}^1(a) \times I_{i_2}^2(a) \times \dots \times I_{i_n}^n(a)$ partition the region B_a into q^n regions; e.g., for $n = 2$ and $q = 3$, the region B_a would be divided into nine regions as shown in Fig. 4.1.

In our proposed coder–decoder pair, each of these regions will be assigned a code word and the coder will transmit the code word corresponding to the current state of the system (4.2.14) $\tilde{x}(jT)$. The transmitted code word will correspond to the integers $i_1, i_2, \dots, i_n$. In order that the communication channel be able to accomplish this

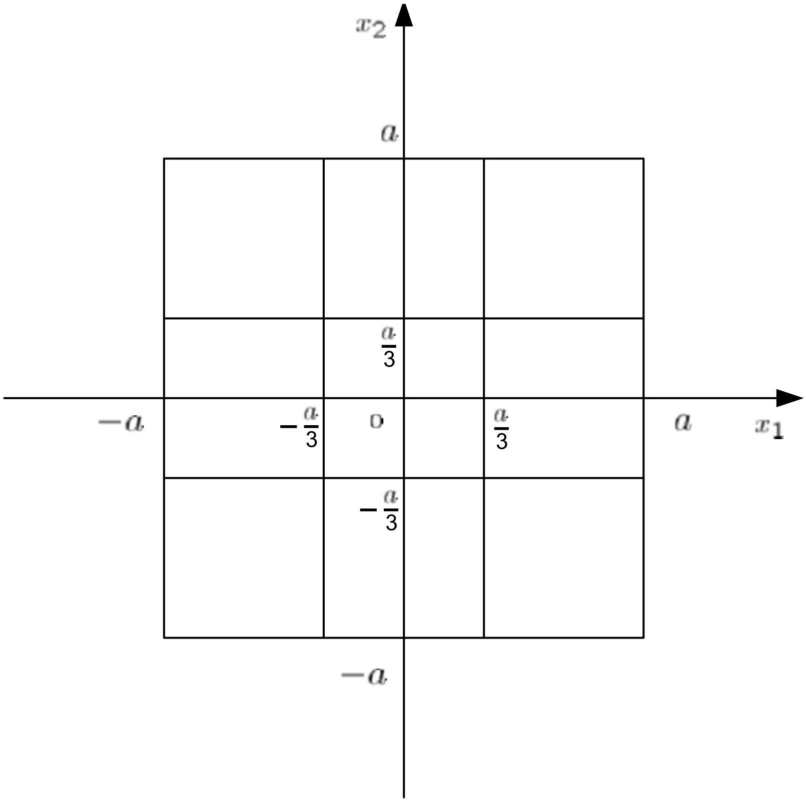


Fig. 4.1. Uniform quantization of the state space.

transmission, the number q^n of quantizer outputs should not exceed the size l of the channel alphabet $\mathfrak{H}$

$$q^n \leq l. \tag{4.2.22}$$

The above quantization of the state space of the system (4.2.14) depends on the scaling parameter $a > 0$. In our proposed coder–decoder pair, this parameter will be the quantization scaling $a(jT)$, where $j = 1, 2, \ldots$.

We now suppose that the assumptions of Lemmas 4.2.2 and 4.2.3 are satisfied. Then let $T > 0$ be a time such that the conditions (4.2.9) and (4.2.16) hold. Furthermore, introduce

$$m_0 := \sup_{x_0 \in \mathfrak{X}_0} \|x_0\|_\infty; \quad a(T) := (e^{\alpha T} + e^{-\alpha_0 T}) m_0;$$
$$a(jT) := e^{\alpha T} \frac{a((j-1)T)}{q} +$$
$$(e^{(\alpha - \alpha_0 (j-1))T} + e^{-\alpha_0 j T}) m_0 \quad \forall j = 2, 3, \ldots. \tag{4.2.23}$$

Introduce now our proposed coder–decoder pair:

Coder:

$$h(jT) = \{i_1, i_2, \ldots, i_n\} \tag{4.2.24}$$

for

$$(\tilde{x}(jT) - \hat{x}(jT-0) \in I^1_{i_1}(a(jT)) \times I^2_{i_2}(a(jT)) \times \ldots \times I^n_{i_n}(a(jT)) \subset B_{a(jT)};$$

Decoder:

$$\begin{aligned} \hat{x}(0) = 0; \quad \hat{z}(t) = K\hat{x}(t), \\ \dot{\hat{x}}(t) = A\hat{x}(t) + B_1 f(\hat{z}(t)) \quad \forall t \neq jT; \\ \hat{x}(jT) = \hat{x}(jT-0) + \eta(i_1, i_2, \ldots, i_n) \\ \text{for } h(jT) = \{i_1, i_2, \ldots, i_n\} \quad \forall j = 1, 2, \ldots. \end{aligned} \tag{4.2.25}$$

We recall that $\nu(t-0)$ denotes the limit of the function $\nu(\cdot)$ at the point t from the left. Notice that our decoder is described by a differential equation with jumps.

Also notice that equations (4.2.25) are a part of both coder and decoder, and it follows immediately from (4.2.9), (4.2.16) and initial condition $\tilde{x}(0) = 0$ that

$$\tilde{x}(T) \in B_{a(T)}; \quad \tilde{x}(jT) - \hat{x}(jT-0) \in B_{a(jT)} \tag{4.2.26}$$

for all $j = 2, 3, \ldots$ and for any solution of (4.2.1), (4.2.2), and (4.2.3).

The main requirement for our coding–decoding scheme is as follows:

$$q > e^{\alpha T}. \tag{4.2.27}$$

Now we are in a position to present the main result of this section.

Theorem 4.2.4. *Suppose that for some $\alpha > 0$, a solution $X > 0$ of the Riccati inequality* (4.2.7) *exists and for some $\alpha_1 > 0$, a solution $Y > 0$ of the Riccati inequality* (4.2.8) *exists. Furthermore, suppose that for some $T > 0$ satisfying conditions* (4.2.9), (4.2.16) *and some positive integer q, the inequality* (4.2.27) *holds. Suppose also that the size l of the channel alphabet meets the requirement* (4.2.22)*. Then the coder–decoder pair* (4.2.14), (4.2.15), (4.2.23), (4.2.24), (4.2.25) *is detecting for the system* (4.2.1), (4.2.2), (4.2.3).

Proof of Theorem 4.2.4. Condition (4.2.27) implies that

$$\lim_{j\to\infty} a(jT) = 0.$$

Hence

$$\lim_{j\to\infty} \Big[\eta(i_1, i_2, \ldots, i_n) - (\hat{x}(jT) - \tilde{x}(jT))\Big] = 0,$$

where $h(jT) = \{i_1, i_2, \ldots, i_n\}$. This implies that

$$\lim_{j\to\infty} (\hat{x}(jT) - \tilde{x}(jT)) = 0.$$

From this and Lemma 4.2.3 we obtain that

$$\lim_{j\to\infty} (x(jT) - \hat{x}(jT)) = 0$$

for any solution of the system (4.2.1), (4.2.2), (4.2.3). Detectability now immediately follows from this and (4.2.25). This completes the proof of Theorem 4.2.4. □

4.3 Stabilization via Communication Channels

In this section, we consider a nonlinear continuous-time dynamic system of the form:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1 f(z(t)) + B_2 u(t); \\ z(t) &= Kx(t); \\ y(t) &= Cx(t), \end{aligned} \tag{4.3.28}$$

where $x(t) \in \mathbb{R}^n$ is the state; $u(t) \in \mathbb{R}^m$ is the control input; $z(t) \in \mathbb{R}^q$ is a linear output; $y(t) \in \mathbb{R}^k$ is the measured output; A, B_1, B_2, K, and C are given matrices of the corresponding dimensions, and $f(z(t)) \in \mathbb{R}^p$ is a given continuous nonlinear vector function.

We also assume that initial conditions of the system (4.3.28) lie in a known bounded set $\mathfrak{X}_0$ (4.2.2). Furthermore, we suppose that the nonlinearity $f(z)$ satisfies the constraint (4.2.3).

We consider the problem of output feedback stabilization of the nonlinear system (4.3.28), (4.2.2), (4.2.3) via a digital communication channel with a bit-rate constraint. Our controller consists of two components. The first component is developed at the measurement location by taking the measured output $y(\cdot)$ and coding to the codeword $h(jT)$. This component will be called "coder." Then the codeword $h(jT)$ is transmitted via a limited capacity communication channel to the second component, which is called "decoder-controller." The second component developed at a remote location takes the codeword $h(jT)$ and produces the control input $u(t)$ where $t \in [jT, (j+1)T)$. This situation is illustrated in Fig. 2.2 (on p. 25), where now a nonlinear plant is considered.

The coder and the decoder are of the following forms:
Coder:

$$h(jT) = \mathcal{F}_j \left(y(\cdot)|_0^{jT} \right); \tag{4.3.29}$$

Decoder-Controller:

$$u(t)|_{jT}^{(j+1)T} = \mathcal{G}_j \left(h(T), h(2T), ..., h((j-1)T), h(jT) \right). \tag{4.3.30}$$

Here $j = 1, 2, 3, \ldots$.

Definition 4.3.1. *The system* (4.3.28) *is said to be stabilizable via a digital communication channel of capacity l if a coder–decoder-controller pair* (4.3.29), (4.3.30) *exists with a coding alphabet of size l such that*

$$\lim_{t\to\infty} \|x(t)\|_\infty = 0; \quad \lim_{t\to\infty} \|u(t)\|_\infty = 0 \tag{4.3.31}$$

for any solution of the closed-loop system (4.3.28), (4.2.2), (4.2.3). *A coder–decoder pair* (4.3.29), (4.3.30) *satisfying condition* (4.3.31) *is said to be* stabilizing.

We will need the following Riccati algebraic inequality:

$$A^{\mathrm{T}}R + RA + K^{\mathrm{T}}K + R\,(B_1B_1^{\mathrm{T}} - \alpha_2 B_2 B_2^{\mathrm{T}})\,R < 0 \tag{4.3.32}$$

and a related state feedback controller

$$u(t) = -\frac{\alpha_2}{2} B_2^{\mathrm{T}} R x(t). \tag{4.3.33}$$

Lemma 4.3.2. *Suppose that for some $\alpha_2 > 0$, a solution $R > 0$ of the Riccati inequality* (4.3.32) *exists. Then the closed-loop system* (4.3.28), (4.2.3), (4.3.33) *is globally asymptotically stable; i.e.,*

$$\lim_{t\to\infty} \|x(t)\|_\infty = 0. \tag{4.3.34}$$

Proof of Lemma 4.3.2 The system (4.3.28), (4.3.33) can be rewritten as

$$\dot{x}(t) = (A - \frac{\alpha_2}{2} B_2 B_2^{\mathrm{T}} R)x(t) + B_1 f(z(t)) \tag{4.3.35}$$

and the constraint (4.2.19). Since $R > 0$ is a solution of (4.3.32), it is also a solution of the Riccati inequality

$$\begin{aligned}(A - \frac{\alpha_2}{2} B_2 B_2^{\mathrm{T}} R)^{\mathrm{T}} R + R(A - \frac{\alpha_2}{2} B_2 B_2^{\mathrm{T}} R)& \\ + K^{\mathrm{T}}K + RB_1B_1^{\mathrm{T}}R < 0.&\end{aligned}$$

Therefore, according to the strict bounded real lemma (see, e.g., Lemma 3.1.2 of [151] the system (4.3.35), (4.3.33), (4.2.19) is quadratically stable. Now the statement of the lemma immediately follows from quadratic stability. This completes the proof of Lemma 4.3.2. □

Now consider the following state estimator that will be a part of our proposed coder:

$$\begin{aligned}\dot{\tilde{x}}(t) &= (A - GC)\tilde{x}(t) + Gy(t) \\ &\quad + B_1 f(\tilde{z}(t)) + B_2 u(t); \\ \tilde{z}(t) &= K\tilde{x}(t), \quad \tilde{x}(0) = 0,\end{aligned} \tag{4.3.36}$$

where the gain G is introduced by (4.2.15).

We now suppose that the assumptions of Lemmas 4.2.2 and 4.2.3 are satisfied. Then let $T > 0$ be a time such that conditions (4.2.9) and (4.2.16) hold. Furthermore, introduce $a(jT)$ by (4.2.23).

Introduce now our proposed coder–decoder-controller pair:

Coder:

$$h(jT) = \{i_1, i_2, \ldots, i_n\} \tag{4.3.37}$$

for $(\tilde{x}(jT) - \hat{x}(jT-0)) \in I^1_{i_1}(a(jT)) \times I^2_{i_2}(a(jT)) \times \ldots \times I^n_{i_n}(a(jT)) \subset B_{a(jT)}$;

Decoder-Controller:

$$\begin{aligned} \hat{x}(0) = 0; \quad \hat{z}(t) &= K\hat{x}(t); \\ \dot{\hat{x}}(t) = A\hat{x}(t) + B_1 f(\hat{z}(t)) &+ B_2 u(t) \quad \forall t \neq jT; \\ \hat{x}(jT) &= \hat{x}(jT-0) + \eta(i_1, i_2, \ldots, i_n) \\ \text{for } h(jT) = \{i_1, i_2, \ldots, i_n\} \quad &\forall j = 1, 2, \ldots; \\ u(t) &= -\frac{\alpha_2}{2} B_2^{\mathrm{T}} R\hat{x}(t). \end{aligned} \tag{4.3.38}$$

Similar to the coder–decoder pair proposed for the detectability problem in Sect. 4.2, the equations (4.3.38) are a part of both coder and decoder-controller. It then follows immediately from (4.2.9), (4.2.16) and initial condition $\tilde{x}(0) = 0$ that

$$\tilde{x}(T) \in B_{a(T)}; \quad \tilde{x}(jT) - \hat{x}(jT-0) \in B_{a(jT)} \tag{4.3.39}$$

for all $j = 2, 3, \ldots$ and for any solution of (4.3.28), (4.2.2), and (4.2.3).

Now we are in a position to present the main result of this section.

Theorem 4.3.3. *Suppose that for some $\alpha > 0$, a solution $X > 0$ of the Riccati inequality* (4.2.7) *exists; for some $\alpha_1 > 0$, a solution $Y > 0$ of the Riccati inequality* (4.2.8) *exists; and for some $\alpha_2 > 0$, a solution $R > 0$ of the Riccati inequality* (4.3.32) *exists. Furthermore, suppose that for some $T > 0$ satisfying conditions* (4.2.9), (4.2.16) *and some positive integer q, the inequality* (4.2.27) *holds. Suppose also that the size l of the channel alphabet meets the requirement* (4.2.22)*. Then the coder–decoder-controller pair given by* (4.3.36), (4.2.15), (4.2.23), (4.3.37), *and* (4.3.38) *is stabilizing for the system* (4.3.28), (4.2.2), (4.2.3).

Proof of Theorem 4.3.3 Condition (4.2.27) implies that $\lim_{j\to\infty} a(jT) = 0$. Hence

$$\lim_{j\to\infty} \Big[\eta(i_1, i_2, \ldots, i_n) - (\tilde{x}(jT) - \hat{x}(jT))\Big] = 0,$$

where $h(jT) = \{i_1, i_2, \ldots, i_n\}$. This implies that

$$\lim_{j\to\infty} (\hat{x}(jT) - \tilde{x}(jT)) = 0.$$

From this and Lemma 4.2.3 we obtain that

$$\lim_{j\to\infty}(x(jT) - \hat{x}(jT)) = 0$$

for any solution of the system (4.3.28), (4.2.2), (4.2.3). This and Lemma 4.3.2 implies the stability of the closed-loop system. This completes the proof of Theorem 4.3.3.

□

4.4 Illustrative Example

We consider the output feedback control problem of a one-link manipulator with a flexible joint via limited capacity digital communication channels. A model of the dynamics of the manipulator can be obtained from, e.g., [156, 197] and is given as follows:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1 f(z(t)) + B_2 u; \\ z(t) &= Kx(t); \\ y(t) &= Cx(t), \end{aligned} \tag{4.4.40}$$

where

$$\begin{aligned} x &= \begin{bmatrix} \theta_m \\ \omega_m \\ \theta_1 \\ \omega_1 \end{bmatrix}; \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -48.6 & -1.25 & 48.6 & 0 \\ 0 & 0 & 0 & 1 \\ 19.5 & 0 & -19.5 & 0 \end{bmatrix}; \\ B_1 &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ -3.33 \end{bmatrix}; \quad B_2 = \begin{bmatrix} 0 \\ 21.6 \\ 0 \\ 0 \end{bmatrix}; \quad f(z(t)) = \sin(z(t)); \\ K &= \begin{bmatrix} 0 \ 0 \ 1 \ 0 \end{bmatrix}; \quad C = \begin{bmatrix} 0 \ 0 \ 1 \ 0 \end{bmatrix}; \end{aligned} \tag{4.4.41}$$

and θ_m is the angular position of the motor, ω_m is the angular velocity of the motor, θ_1 is the angular position of the link, and ω_1 is the angular velocity of the link. The control u is the torque delivered by the motor.

We assume that only the angular position of the link θ_1, i.e., y, can be measured. To measure θ_1, a visual sensing scheme can be adopted, as shown in Fig. 4.2. A camera is mounted at a distance away from the manipulator and is connected to a processor that captures and processes the images from the camera to determine the angular position of the link θ_1. The measured θ_1 is then fed to the coder, which contains a state estimator, to generate the code words. By using any digital communication channel that has a sufficient bandwidth, the code words are transmitted to the decoder-controller, which then controls the motor to generate the appropriate torque u. In this setting, the sensor, i.e., the camera and the image processor, and the coder can be remotely located far away from the manipulator.

By choosing $\alpha = 6$, $\alpha_1 = 5$, and $\alpha_2 = 5$, we see that solutions X, Y, and R exist that satisfy the Riccati inequalities (4.2.7), (4.2.8), and (4.3.32), respectively.

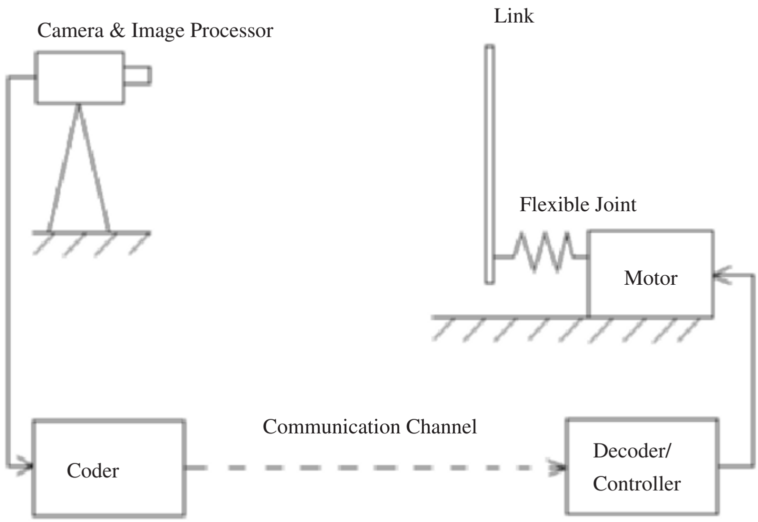


Fig. 4.2. Schematic of the vision-based control system.

The respective T_0 values in Lemmas 4.2.2 and 4.2.3 are 0.94 and 0.91, and the value α_0 in (4.2.16) is 3.1. Therefore, we choose the sampling period $T = 1.0$ second so that both inequalities (4.2.9) and (4.2.16) hold. As for the value q, we picked $q = 2001$ and this value satisfies the inequality (4.2.27); i.e., $q > e^{\alpha T}$.

By using (4.2.15) and (4.3.33), the corresponding state estimator gain and control law are

$$\begin{aligned} G &= \begin{bmatrix} 0.85 \; 1.13 \; 0.04 \; 1.39 \end{bmatrix}^{\mathrm{T}} \times 10^3; \\ u(t) &= \begin{bmatrix} -2.05 \; -0.36 \; -1.28 \; -0.90 \end{bmatrix} x(t). \end{aligned} \tag{4.4.42}$$

For simulation purposes, the initial value of x is chosen as $x(0) = [1 \;\; 1 \;\; 1 \;\; 1]^{\mathrm{T}}$, and the value m_0 is defined as $m_0 = 5$. Finally, a simulation result is shown in Figs. 4.3 and 4.4, and it agrees with the conditions (4.3.31).

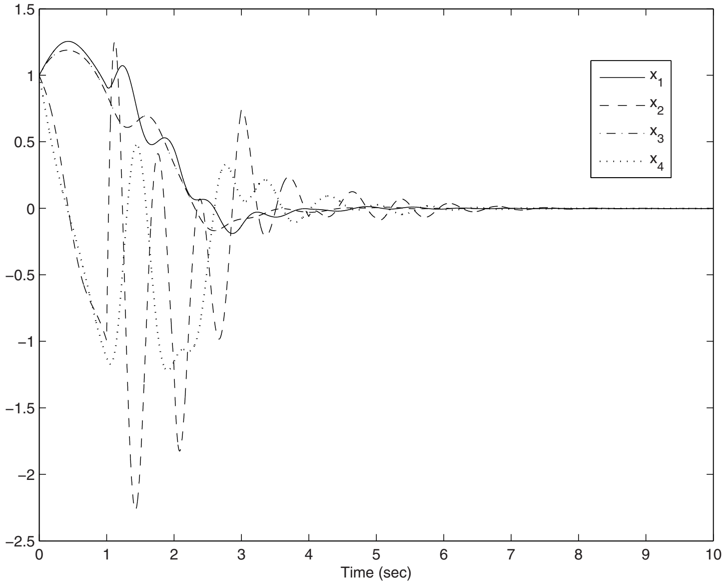


Fig. 4.3. Evolution of the state $x(t)$.

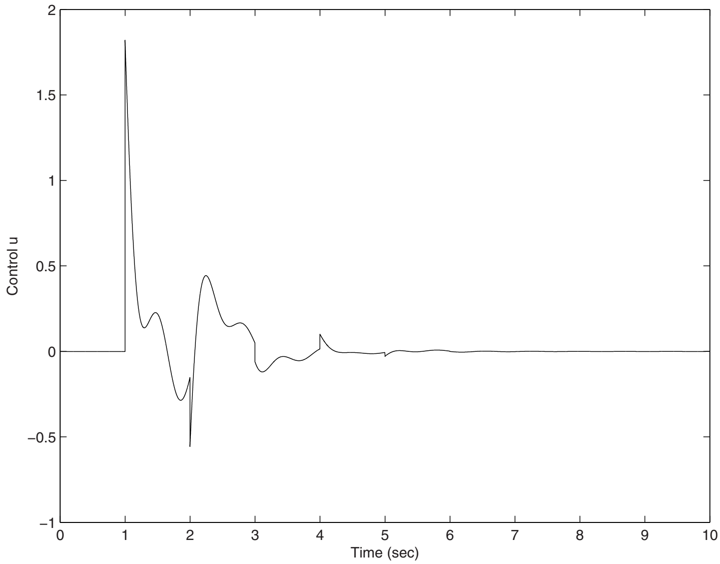


Fig. 4.4. Evolution of the control input $u(t)$.

5

Robust Set-Valued State Estimation via Limited Capacity Communication Channels

5.1 Introduction

In this chapter, we consider the problem of robust state estimation via a limited capacity digital communication channel. Many recent advances in the area of robust control system design assume that the system to be controlled is modeled as an uncertain system; e.g., see [46, 148, 151]. There are many different types of uncertain system models. The class of time-varying uncertain systems considered in this chapter contains uncertainty that is defined by a certain integral quadratic constraint. This class of uncertain systems originated in the work of Yakubovich on absolute stability theory (see, e.g., [222, 223]) and is a particularly rich uncertainty class allowing for nonlinear, time-varying, dynamic uncertainties. Furthermore, a number of new robust control system design methodologies have recently been developed for uncertain systems with integral quadratic constraints; e.g., see [148, 151, 174, 178–180, 182]. In this chapter, we adopt the approach to the robust state estimation problem proposed in [181] (also, see [148, 183]). Reference [181] builds on the deterministic interpretation of Kalman filtering presented in [21]. This deterministic approach to Kalman filtering also forms the launching point for the results of this chapter. As in the previous chapter, we consider the case of continuous-time plants.

In [21], the following deterministic state estimation problem is considered: Given output measurements from a time-varying linear system with noise inputs subjected to an $\mathbf{L}_2$ norm bound, find the set of all states consistent with these measurements. Such a problem is referred to as a *set-valued state estimation problem.* The solution to this problem was found to be an ellipsoid in the state space that is defined by the standard Kalman filter equations. Thus, the results of [21] give an alternative interpretation of the standard Kalman filter. In [148, 181, 183] the results of [21] were extended to the case of uncertain systems with integral quadratic constraints. In this chapter, we employ the set-valued approach to the state estimation problem and the deterministic interpretation of Kalman filtering from [148, 181] in the situation where state estimation is to be performed via a limited capacity communication channel.

In such a context, state estimation results for linear systems with bounded or Gaussian noise were obtained in [45, 133, 220]. Some state estimation scheme rel-

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_5,

evant to the problem of stabilization via a limited capacity communication channel was proposed in [70]. Also, in [70], the case of a linear system without uncertainty over an infinite time interval was considered.

The remainder of the chapter is organized as follows. In Sect. 5.2, we introduce the class of uncertain systems under investigation. Section 5.3 presents the statement of the optimal robust state estimation problem and some preliminary results on robust set-valued state estimation originally published in [181]. In Sect. 5.4, we give a solution to an optimal state estimation problem via limited capacity digital communication channels for the class of uncertain systems under consideration. Furthermore, in Sect. 5.5, we propose a suboptimal state estimation algorithm that is computationally nonexpansive and easily implementable in real time. Finally, in Sect. 5.6, proofs of two lemmas on set-valued state estimation can be found.

The main results presented in this chapter originally appeared in the paper [100]. An analog of these results for systems without uncertainty was published before that in [184], where some of the main ideas of the chapter were introduced. Also, some versions of technical results on robust state estimation that are used in this chapter were first obtained in [174, 176, 181, 183]. The uncertain system framework of this chapter was presented in the monographs [148, 151, 174]. In [99], the results of this chapter were successfully applied to the problem of precision missile guidance based on radar/video sensor fusion. An extension of the results of the chapter can be found in [33].

5.2 Uncertain Systems

In designing robust state estimation systems, one must specify the class of uncertainties against which the state estimator is to be robust. The most common approach in control engineering is to begin with a plant model that not only models the nominal plant behavior but also models the type of uncertainties that are expected. Such a plant model is referred to as an uncertain system.

There many various types of uncertain system models; e.g., see [46]. In this chapter, we deal with uncertain plants with uncertainties satisfying so-called integral quadratic constraints.

5.2.1 Uncertain Systems with Integral Quadratic Constraints

Consider the time-varying uncertain system defined over the finite time interval $[0, NT]$:

$$\begin{aligned}\dot{x}(t) &= A(t)x(t) + B(t)w(t);\\ z(t) &= K(t)x(t);\\ y(t) &= C(t)x(t) + v(t),\end{aligned} \tag{5.2.1}$$

where $N > 0$ is an integer; $T > 0$ is a given constant; $x(t) \in \mathbb{R}^n$ is the state; $w(t) \in \mathbb{R}^p$ and $v(t) \in \mathbb{R}^l$ are the uncertainty inputs; $z(t) \in \mathbb{R}^q$ is the *uncertainty* *output* and $y(t) \in \mathbb{R}^l$ is the measured output; and $A(\cdot), B(\cdot), K(\cdot)$, and $C(\cdot)$ are bounded piecewise continuous matrix functions.

The uncertainty in (5.2.1) is described by an equation of the form:

$$\begin{bmatrix} w(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} \phi_1(t, z(t)) \\ \phi_2(t, z(t)) \end{bmatrix}, \tag{5.2.2}$$

where the following integral quadratic constraint is satisfied.

Let $X_0 = X_0^T > 0$ be a given matrix, $x_0 \in \mathbb{R}^n$ be a given vector, $d > 0$ be a given constant, and $Q(\cdot) = Q(\cdot)^{\mathrm{T}}$ and $R(\cdot) = R(\cdot)^{\mathrm{T}}$ be given bounded piecewise continuous matrix weighting functions satisfying the following condition: A constant $\delta > 0$ exists such that $Q(t) \geq \delta I$, $R(t) \geq \delta I$ for all t . For a given finite time interval $[0, s]$, we will consider the uncertainty inputs $w(\cdot)$ and $v(\cdot)$ and initial conditions $x(0)$ such that

$$(x(0) - x_0)^{\mathrm{T}} X_0 (x(0) - x_0) + \int_0^s \left(w(t)^{\mathrm{T}} Q(t) w(t) + v(t)^{\mathrm{T}} R(t) v(t) \right) dt \leq d + \int_0^s \|z(t)\|^2 \, dt. \tag{5.2.3}$$

The uncertainty input in (5.2.2) can be regarded as the feedback interconnection with the uncertainty output $z(t)$ as shown in Fig. 5.1. The inequality (5.2.3) gives some constraint on the sizes of the uncertainty inputs $w(\cdot), v(\cdot)$ and on the size of uncertainty in the initial condition.

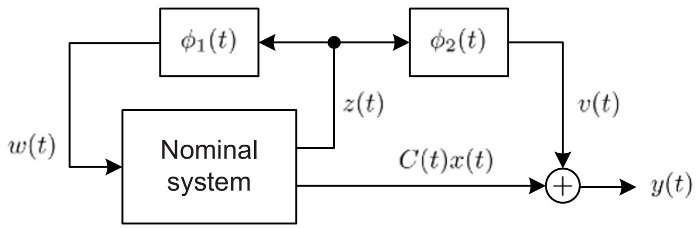


Fig. 5.1. Uncertain system.

This class of uncertainties was introduced in the theory of absolute stability by Yakubovich (e.g., see [222]), and extensively studied in the theory of robust control (e.g., see [148, 151, 174, 178–181]).

5.2.2 Uncertain Systems with Norm-Bounded Uncertainty

An important class of uncertain systems satisfying the integral quadratic constraint (5.2.3) consists of linear uncertain systems with uncertainties satisfying a standard norm-bounded constraint. In this case, the uncertain system is described by the relationships

$$\dot{x} = [A(t) + B(t)\Delta_1(t)K(t)]\,x(t);$$
$$y = [C(t) + \Delta_2(t)K(t)]x\,(t)\,, \qquad (5.2.4)$$

where $\Delta_1(t)$, $\Delta_2(t)$ are uncertainty matrices such that

$$\left\| \begin{matrix} \Delta_1(t) \\ \Delta_2(t) \end{matrix} \right\| \leq 1 \qquad (5.2.5)$$

for all t.

Also, the initial conditions are required to satisfy the following inequality:

$$(x\,(0) - x_0)^{\mathrm{T}}\, X_0\,(x\,(0) - x_0) \leq d. \qquad (5.2.6)$$

The block-diagram of such a system is shown in Fig. 5.2.

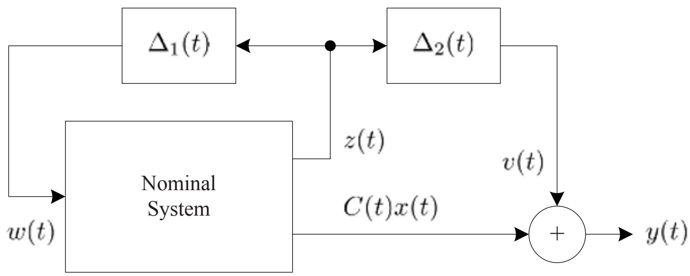


Fig. 5.2. Uncertain system with norm-bounded uncertainty.

To verify that such uncertainty is admissible for the uncertain system (5.2.1), (5.2.3), let $w(t) := \Delta_1(t)K(t)x(t)$, $v(t) := \Delta_2(t)K(t)x(t)$, and $z(t) := K(t)x(t)$. Then the condition (5.2.3) is satisfied with $Q(\cdot) \equiv I$ and $R(\cdot) \equiv I$.

5.2.3 Sector-Bounded Nonlinearities

This class of uncertainties arose from the celebrated theory of absolute stability; e.g., see [144, 222, 224]. Consider the time-invariant uncertain system (5.2.1) with scalar uncertainty input w and uncertainty output z, and the uncertainty is described by the equation

$$w(t) = \phi(z(t)), \qquad (5.2.7)$$

where $\phi(\cdot) : \mathbb{R} \to \mathbb{R}$ is an uncertain nonlinear mapping.

This system is represented in the block diagram shown in Fig. 5.3.

We will suppose that the uncertain nonlinearity $\phi(\cdot)$ satisfies the following sector bound (e.g., see [144]):

$$0 \leq \frac{\phi(z)}{z} \leq k, \qquad (5.2.8)$$

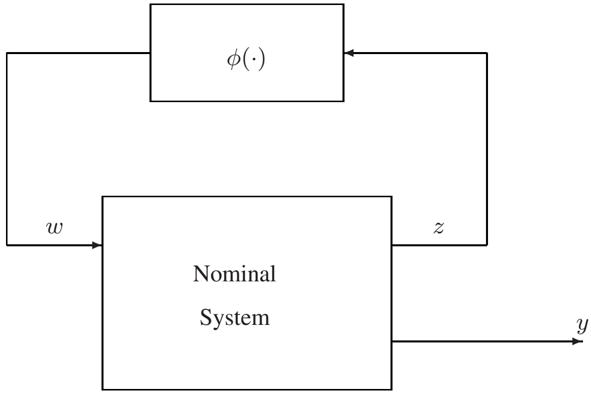


Fig. 5.3. Uncertain system with a single nonlinear uncertainty.

where $0 < k \leq \infty$ is a given constant associated with the system. Using the change of variables $\tilde{z} = k/2z$, $\widetilde{w} = \phi(2/k\tilde{z}) - \tilde{z}$, this system is transformed into the system

$$\dot{x}(t) = (A + \frac{k}{2}BK)x(t) + B\hat{w}(t);$$
$$\hat{z}(t) = \frac{k}{2}Kx(t).$$

The bound (5.2.8) on the uncertainty in this system then becomes a standard bound on the norm of the uncertainty input:

$$|\hat{w}(t)| \leq |\hat{z}(t)|. \tag{5.2.9}$$

This observation motivates us to think of sector bounded uncertainty as a special case of norm-bounded uncertainty.

Remark 5.2.1. Notice that the nonlinear system with a nonlinearity satisfying the globally Lipschitz condition (4.2.3) (on p. 102) that was studied in Chap. 4 is a special case of the system with norm-bounded nonlinearity.

5.3 State Estimation Problem

We consider the set-valued state estimation problem for the system (5.2.1), (5.2.3) that can be stated as follows. Let $y(t) = y_0(t)$ be a fixed measured output of the uncertain system (5.2.1), and let $s \in [0, NT]$ be given time. Then find the corresponding set $X_s\left[x_0, y_0\left(\cdot\right)|_0^s, d\right]$ of all possible states $x(s)$ at time s for the uncertain system (5.2.1), with the uncertainty input and initial states satisfying the constraint (5.2.3). The state estimator will be defined by the differential equation

$$\dot{\hat{x}}(s) = [A(s) + P(s)[K(s)^{\mathrm{T}}K(s) - C(s)^{\mathrm{T}}R(s)C(s)]]\,\hat{x}(s) \\ + P(s)C(s)^{\mathrm{T}}R(s)y(s), \qquad \hat{x}(0) = x_0, \quad (5.3.10)$$

where $P(s)$ is the solution of the following Riccati differential equation:

$$\dot{P}(s) = A(s)P(s) + P(s)A(s)^{\mathrm{T}} \\ + P(s)\left[K(s)^{\mathrm{T}}K(s) - C(s)^{\mathrm{T}}R(s)C(s)\right]P(s) + B(s)Q(s)^{-1}B(s)^{\mathrm{T}}, \\ P(0) = X_0^{-1}. \quad (5.3.11)$$

Definition 5.3.1. *The uncertain system* (5.2.1)*,* (5.2.3) *is said to be* robustly observable *on* $[0, NT]$*, if for any vector* $x_0 \in \mathbb{R}^n$*, any time* $s \in [0, NT]$*, any constant* $d > 0$*, any fixed measured output* $y(t) = y_0(t)$*, the set* $X_s[x_0, y_0(\cdot)|_0^s, d]$ *is bounded.*

The following result offers an exhaustive criterion for the system to be robustly observable.

Lemma 5.3.2. *Let* $X_0 = X_0^{\mathrm{T}} > 0$ *be a given matrix, and let* $Q(\cdot) = Q(\cdot)^{\mathrm{T}}$ *and* $R(\cdot) = R(\cdot)^{\mathrm{T}}$ *be given matrix functions such that* $Q(\cdot) \geq \delta I, R(\cdot) \geq \delta I$*, where* $\delta > 0$*. Consider the system* (5.2.1) *and the constraint* (5.2.3) *. Then the following statements hold:*

(i) The system (5.2.1)*,* (5.2.3) *is robustly observable on* $[0, NT]$ *if and only if the solution* $P(\cdot)$ *to the Riccati equation* (5.3.11) *is defined and positive-definite on the interval* $[0, NT]$ *;*
(ii) If the system (5.2.1)*,* (5.2.3) *is robustly observable on* $[0, NT]$*, then for any vector* $x_0 \in \mathbb{R}^n$*, any* $s \in [0, NT]$*, any constant* $d > 0$*, and any fixed measured output* $y(t) = y_0(t)$*, the set* $X_s[x_0, y_0(\cdot)|_0^s, d]$ *of all possible states* $x(s)$ *at time* s *is given by*

$$X_s[x_0, y_0(\cdot)|_0^s, d] \\ = \left\{ x_s \in \mathbb{R}^n : \begin{array}{r} (x_s - \hat{x}(s))^{\mathrm{T}} P(s)^{-1} (x_s - \hat{x}(s)) \\ \leq d + \rho_s[y_0(\cdot)] \end{array} \right\}, \quad (5.3.12)$$

where

$$\rho_s[y_0(\cdot)] \\ := \int_0^s \left[\|K(t)\hat{x}(t)\|^2 - (C(t)\hat{x}(t) - y_0(t))^{\mathrm{T}} R(t)(C(t)\hat{x}(t) - y_0(t)) \right] dt. \quad (5.3.13)$$

The proof of Lemma 5.3.2 is given in Sect. 5.6.

Remark 5.3.3. The state estimator defined by (5.3.10) and (5.3.11) is of the same form as the state estimator, which occurs in the output feedback H^∞ control problem; e.g., see [91, 130].

In our state estimation problem, the state $x(s)$ should be estimated. Now consider the case where the only way of communicating information from the measured output y is via a digital communication channel with the following bit-rate constraint. This channel can transmit codesymbols $h = h_k$ at the time instants of the form kT. The channel gives a limited finite number of admissible codesymbols h. This restricted number of codesymbols h is determined by the data rate of the channel. If $\mathcal{N}$ is the number of admissible codesymbols, then $\mathcal{N} = 2^l$, where l is the number of bits that our channel can transmit at any time kT. We assume that the channel is a perfect noiseless channel and that there is no time delay in the channel. In this chapter, we propose algorithms for state estimation via a digital communication channel with the above bit-rate constraint. Our algorithms consist of two components. The first component is developed at the measurement location by taking the measured output signal y and coding it into the codesymbol h_k at time kT. This component will be called *coder*. Then the codesymbol h_k is transmitted via a limited capacity communication channel to the second component, which is called *decoder*. The decoder takes the codesymbols $h_i, i \le k$ and produces the estimated state $\bar{x}(kT)$ that is an approximation of the center of all possible states of the system at time kT. This situation is illustrated in Fig. 5.4.

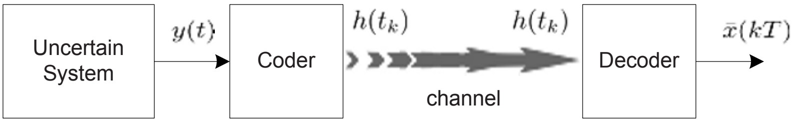


Fig. 5.4. State estimation via digital communication channel.

The coder and decoder are defined by the following equations:

Coder:

$$h\left(kT\right) = \mathcal{F}_k\left(x_0, d, y\left(\cdot\right)|_0^{kT}\right); \tag{5.3.14}$$

Decoder:

$$\bar{x}\left(kT\right) = \mathcal{G}_k\left(x_0, d, h_0, h_1, \ldots, h_k\right). \tag{5.3.15}$$

We recall that $\|x\|_\infty := \max_i |x_i|$ for $x = \textbf{col}\,[x_1, x_2, \ldots, x_n]$.

Definition 5.3.4. *Consider the system* (5.2.1)*,* (5.2.3) *with a given* $d > 0$*. Let* $\epsilon > 0$ *be a given constant, and let* $\mathcal{N} > 0$ *and* $0 < M \le N$ *be given integers. The coder–decoder pair* (5.3.14)*,* (5.3.15) *is said to solve the state estimation problem via a digital communication channel with the admissible number of codesymbols* $\mathcal{N}$ *on the time interval* $[MT, NT]$ *with the precision level* ϵ *if*

$$\|\hat{x}(kT) - \bar{x}(kT)\|_\infty \le \epsilon \tag{5.3.16}$$

for all $x_0, y\left(\cdot\right), k = M, M+1, \ldots, N$*. Here* $\hat{x}(kT)$ *is defined by* (5.3.10)*.*

Moreover, if the constant ϵ *in* (5.3.16) *is the infimum over all precision levels that can be achieved by coder–decoder pairs of the form* (5.3.14)*,* (5.3.15) *with the* *admissible number of codesymbols $\mathcal{N}$, the coder–decoder pair is said to be optimal for the state estimation problem via a digital communication channel with the admissible number of codesymbols $\mathcal{N}$ on the time interval $[MT, NT]$.*

Our problem is to design an optimal coder–decoder pair.

5.4 Optimal Coder–Decoder Pair

In this section, we construct an optimal coder–decoder pair for the problem of state estimation with bit-rate constraints. To formulate the main results of the section, we need to introduce the concept of strong robust observability that is close to the concept of uniform robust observability from Chap. 15.

Definition 5.4.1. *Consider the uncertain system* (5.2.1)*,* (5.2.3)*. Let $\hat{X}_s[d]$ be the set of all possible states $\hat{x}(s)$ at time $s \in (0, NT]$, defined by* (5.3.10) *with all possible initial conditions $x_0 \in R^n$ and all possible measured outputs $y_0(\cdot)$ of the system* (5.2.1)*,* (5.2.3)*. The uncertain system* (5.2.1)*,* (5.2.3) *is said to be* strongly, robustly observable *on $[0, NT]$ if it is robustly observable and for any time $s \in (0, NT]$ and constant $d > 0$, the set $\hat{X}_s[d]$ is bounded.*

Consider the following Riccati equation:

$$
\begin{aligned}
-\dot{X}(s) = {} & \hat{A}(s)^{\mathrm{T}} X(s) + X(s)\hat{A}(s) \\
& - X(s)\hat{B}(s)R^{-1}(s)\hat{B}(s)^{\mathrm{T}} X(s) - K(s)^{\mathrm{T}} K(s), \qquad X(0) = 0,
\end{aligned} \tag{5.4.17}
$$

where

$$\hat{A}(t) := A(t) + P(t)K(t)^{\mathrm{T}} K(t) \tag{5.4.18}$$

and

$$\hat{B}(t) := -P(t)C(t)^{\mathrm{T}} R(t). \tag{5.4.19}$$

Lemma 5.4.2. *The uncertain system* (5.2.1)*,* (5.2.3) *is strongly, robustly observable on $[0, NT]$ if and only if the solution $P(\cdot)$ to the Riccati equation* (5.3.11) *is defined and positive-definite on $[0, NT]$ and the solution $X(\cdot)$ to the Riccati equation* (5.4.17) *is defined on $[0, NT]$ and positive-definite on $(0, NT]$. Furthermore, if the system* (5.2.1)*,* (5.2.3) *is strongly, robustly observable, then the set $\hat{X}_s[d]$ is given by*

$$\hat{X}_s[d] = \{\hat{x}_s \in \mathbb{R}^n : \hat{x}_s^{\mathrm{T}} X(s)\hat{x}_s \le d\}. \tag{5.4.20}$$

The proof of Lemma 5.4.2 is given in Sect. 5.6.

Definition 5.4.3. *Let $\mathfrak{X} \subset \mathbb{R}^n$ be a convex bounded set and $\mathcal{N} > 0$ be a given integer. Furthermore, assume that $\mathfrak{X}$ is partitioned into $\mathcal{N}$ nonintersecting subsets $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_{\mathcal{N}}$ equipped with points $\bar{\eta}_1 \in \mathcal{I}_1, \ldots, \bar{\eta}_{\mathcal{N}} \in \mathcal{I}_{\mathcal{N}}$. The collection $\mathcal{V}[\mathfrak{X}, \mathcal{N}] = \{\mathcal{I}_1, \ldots, \mathcal{I}_{\mathcal{N}}, \bar{\eta}_1, \ldots, \bar{\eta}_{\mathcal{N}}\}$ is called an l_∞* Voronoi structure *on the set $\mathfrak{X}$ if $\|x - \bar{\eta}_i\|_\infty \le \|x - \bar{\eta}_s\|_\infty$ for any $i = 1, \ldots, \mathcal{N}$, any $x \in \mathcal{I}_i$ and any $s \neq i$.*

Notation 5.4.4. *Let $\mathcal{V}[\mathfrak{X}, \mathcal{N}]$ be an l_∞ Voronoi structure. Then*

$$D(\mathcal{V}[\mathfrak{X}, \mathcal{N}]) := \sup_{i=1,\ldots,\mathcal{N},\ x\in\mathcal{I}_i} \|x - \bar{\eta}_i\|_\infty. \tag{5.4.21}$$

Definition 5.4.5. *Let $\mathfrak{X}$ be a given convex bounded set and $\mathcal{N} > 0$ be a given integer. Then $\mathcal{V}[\mathfrak{X}, \mathcal{N}]$ is said to be an* optimal l_∞ Voronoi structure *on $\mathfrak{X}$ if $D(\mathcal{V}[\mathfrak{X}, \mathcal{N}]) \leq D(\mathcal{V}^*[\mathfrak{X}, \mathcal{N}])$ for any other l_∞ Voronoi structure $\mathcal{V}^*[\mathfrak{X}, \mathcal{N}]$.*

It is obvious that an optimal l_∞ Voronoi structure exists on any convex bounded set. Constructing optimal Voronoi structures has been the subject of much research in the field of vector quantization theory and computational geometry; see, e.g., [23,62].

The Coder–Decoder Pair

Let $\mathcal{V}_k[\hat{X}_{kT}[d], \mathcal{N}] = \{\mathcal{I}_1^k, \ldots, \mathcal{I}_\mathcal{N}^k, \bar{\eta}_1^k, \ldots, \bar{\eta}_\mathcal{N}^k\}$ be the l_∞ optimal Voronoi structure on the set $\hat{X}_{kT}[d]$ defined by (5.4.20), where $k = M, M+1, \ldots, N$. We are now in a position to describe our proposed optimal coder–decoder pair of the form (5.3.14), (5.3.15) associated with these structures.

Let $\bar{x}(0) = x_0$. For any $k = M, M+1, \ldots, N$, we consider (5.3.10) on $[(k-1)T, (kT)]$. Then, for all $k = M, M+1, \ldots, N$, we define our coder–decoder pair as follows:

Coder:

$$h(k) = i \quad \text{if} \quad \hat{x}(k) \in \mathcal{I}_i^k \subset \hat{X}_{kT}[d]; \tag{5.4.22}$$

Decoder:

$$\bar{x}(kT) = \bar{\eta}_i^k \quad \text{if} \quad h(k) = i. \tag{5.4.23}$$

Now we are in a position to present the main result of this chapter.

Theorem 5.4.6. *Consider the uncertain system* (5.2.1)*,* (5.2.3)*. Let $d > 0$ be a given constant, and $\mathcal{N} > 0$ and $0 < M \leq N$ be given integers. Suppose that this system is strongly, robustly observable and that the sets $\hat{X}_{kT}[d]$ are defined by* (5.4.20)*.*

Then the coder–decoder pair (5.4.22)*,* (5.4.23) *is optimal for the state estimation problem via a digital communication channel with the admissible number of codesymbols $\mathcal{N}$ on the time interval $[MT, NT]$.*

Proof. According to Theorem 5.4.2, strong, robust observability implies that $\hat{X}_{kT}[d]$ is the set of all possible $\hat{x}(kT)$. Furthermore, it follows from Definition 5.4.5 that the coding scheme that is based on a Voronoi optimal structure gives an optimal state estimate. This completes the proof of this theorem. □

5.5 Suboptimal Coder–Decoder Pair

In the previous section, we proposed an optimal coder–decoder pair. The result was given in terms of Voronoi structures. However, this method is computationally expensive. In this section, we describe another state estimation algorithm that is not optimal but well suited to real time implementation.

In this section, we propose a coder–decoder pair that uses the uniform quantization of the estimated states introduced in the two previous chapters. Also, the quantization parameters may be updated at every instant time kT.

To quantize the state space of (5.3.10), let $a(k) > 0$ be a given number and consider the set $B_{a(k)} := \{x(kT) \in \mathbb{R}^n : |x_i(kT)| \leq a(k) \ \ \forall i\}$. We propose to quantize each state component $x_i(kT)$ by means of $q(k)$ intervals, where $q(k)$ is a specified integer. Therefore, for n-dimensional state, the set $B_{a(k)}$ will be divided into $q(k)^n$ hypercubs. For each $i \in \{1, 2, \ldots, n\}$, we divide the corresponding component of the state vector $x_i(kT)$ into $q(k)$ intervals as follows:

$$
\begin{aligned}
I_1^i(a(k)) &:= \left\{ x_i(kT) : -a(k) \leq x_i(kT) < -a(k) + \frac{2a(k)}{q(k)} \right\}; \\
I_2^i(a(k)) &:= \left\{ x_i(kT) : -a(k) + \frac{2a(k)}{q(k)} \leq x_i(kT) < -a(k) + \frac{4a(k)}{q(k)} \right\}; \\
&\vdots \\
I_{q(k)}^i(a(k)) &:= \left\{ x_i(kT) : a(k) - \frac{2a(k)}{q(k)} \leq x_i(kT) \leq a(k) \right\}.
\end{aligned}
\tag{5.5.24}
$$

Then for any $\hat{x}(kT) \in B_{a(k)}$, unique integers

$$
i_1 \in \{1, 2, \ldots, q(k)\}, \ i_2 \in \{1, 2, \ldots, q(k)\}, \ldots, \ i_n \in \{1, 2, \ldots, q(k)\}
$$

exist such that

$$
\hat{x}(kT) \in I_{i_1}^1(a(k)) \times I_{i_2}^2(a(k)) \times \ldots \times I_{i_n}^n(a(k)). \tag{5.5.25}
$$

Also, corresponding to the integers $i_1, i_2, \ldots, i_n$, we define the vector

$$
\bar{\eta}(i_1, i_2, \ldots, i_n) := \begin{bmatrix} -a(k) + a(k)(2i_1 - 1)/q(k) \\ -a(k) + a(k)(2i_2 - 1)/q(k) \\ \vdots \\ -a(k) + a(k)(2i_n - 1)/q(k) \end{bmatrix}. \tag{5.5.26}
$$

This vector is the center of the hypercube

$$
I_{i_1}^1(a(k)) \times I_{i_2}^2(a(k)) \times \ldots \times I_{i_n}^n(a(k))
$$

containing the original point $\hat{x}(kT)$.

Note that the regions $I_{i_1}^1(a(k)) \times I_{i_2}^2(a(k)) \times \ldots \times I_{i_n}^n(a(k))$ partition $B_{a(k)}$ into $q(k)^n$ regions.

In the proposed coder–decoder pair, each region will be represented by a codesymbol, and the codesymbol corresponding to the vector $\hat{x}(kT)$ will be transmitted via a limited capacity channel.

The Coder–Decoder Pair

We are now in a position to describe our proposed coder–decoder pair of the form (5.3.14), (5.3.15).

Let $\bar{x}(0) = x_0$. For any $k = M, M+1, \ldots, N$, we consider (5.3.10) on $[(k-1)T, (kT)]$. Then, for all $k = M, M+1, \ldots, N$, we define our coder–decoder pair as follows:

Coder:

$$h(k) := \{i_1, i_2, \ldots, i_n\}$$
$$\text{for} \quad \hat{x}(k) \in I^1_{i_1}(a(k)) \times I^2_{i_2}(a(k)) \times \ldots \times I^n_{i_n}(a(k)) \subset B_{a(k)}; \quad (5.5.27)$$

Decoder:

$$\bar{x}(kT) = \bar{\eta}\,(i_1, i_2, \ldots, i_n) \text{ for } h(k) = \{i_1, i_2, \ldots, i_n\}. \quad (5.5.28)$$

Here the vector $\bar{\eta}\,(i_1, i_2, \ldots, i_n)$ is defined in (5.5.26).

The block-diagram of the uncertain system with a suboptimal coder–decoder pair is shown in Fig. 5.5.

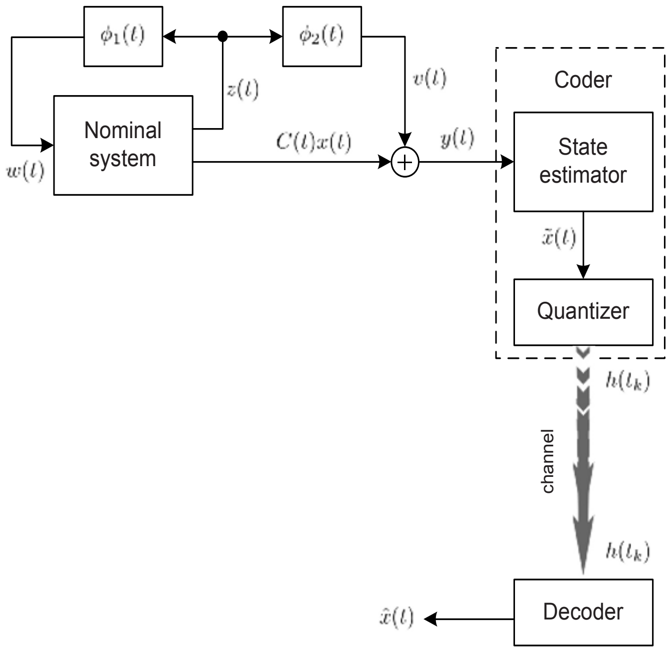


Fig. 5.5. Uncertain system with the suboptimal coder–decoder pair.

Remark 5.5.1. It should be pointed out that by our construction, the state estimate $\widehat{x}(k)$ necessarily belongs to $B_{a(k)}$.

Now we are ready to present the main result of the section. This will require the following notation.

Notation 5.5.2. *Let* $Z(\cdot) := X(\cdot)^{-1}$*, where* $X(\cdot)$ *is the solution to* (5.4.17)*. Let* $z_{ij}(\cdot)$ *be the corresponding element of the matrix* $Z(\cdot)$*. Then we define a constant* $c_i(k) > 0$ *by the following equation:*

$$c_i(k) := \sqrt{z_{ii}(kT)} \quad \text{for all} \quad i = 1, 2, \ldots, n, \quad k = M, M+1, \ldots, N. \tag{5.5.29}$$

Note that $z_{ii}(kT) > 0$ *for all* $i, t > 0$ *because* $Z(t)$ *is positive-definite.*

Theorem 5.5.3. *Consider the uncertain system* (5.2.1)*,* (5.2.3)*. Let* $d > 0$ *,* $\epsilon > 0$ *be given constants,* $\mathcal{N} > 0$ *and* $0 < M \leq N$ *be given integers. Suppose that this system is strongly, robustly observable and*

$$\frac{c_i(k)\sqrt{d}}{q(k)} \leq \epsilon, \tag{5.5.30}$$

for all $i = 1, 2, \ldots, n$ *and* $k = M, M+1, \ldots, N$*. Then the coder–decoder pair* (5.5.27)*,* (5.5.28) *with*

$$a(k) := \max_{i=1,\ldots,n} \{c_i(k)\sqrt{d}\} \tag{5.5.31}$$

solves the state estimation problem via a digital communication channel with the admissible number of codesymbols $q(k)^n \leq \mathcal{N}$ *on the time interval* $[MT, NT]$ *with the precision level* ϵ*.*

Proof. We will prove that $\hat{x}(kT) \in B_{a(k)}$ and inequality (5.3.16) holds for any $k = M, M+1, \ldots, N$. Indeed, let $x(\cdot)$ be a solution of the system (5.2.1) with some uncertainty input $w(\cdot)$ and $v(\cdot)$ satisfying (5.2.3), and let $\hat{x}(\cdot)$ be the corresponding estimated state. It follows from the strong, robust observability and Theorem 5.4.2 that the ellipsoid (5.4.20) is the set of all possible estimated states $\hat{x}(s)$. Thus we have that

$$\|\hat{x}(kT)\|_\infty \leq \max_{i=1,\ldots,n} \{c_i(k)\sqrt{d}\}. \tag{5.5.32}$$

Therefore, $\hat{x}(kT) \in B_{a(k)}$ with $a(k)$ defined by (5.5.31). Now the condition (5.3.16) follows from (5.5.30) with a suitable $q(k)$. This completes the proof of the theorem.

Remark 5.5.4. Notice that in [100], a slightly more general result was obtained. A coder–decoder pair from [100] uses non-uniform quantization of the estimated states in which different numbers of bits may be used to quantize various state variables. In this chapter, for the sake of simplicity, we consider a coder–decoder pair that is based on uniform quantization.

5.6 Proofs of Lemmas 5.3.2 and 5.4.2

Proof of Lemma 5.3.2. Statement (i): Necessity. In this case, we must establish the existence of a positive-definite solution to the Riccati equation (5.3.11). This will be achieved by showing that the cost function in a corresponding linear quadratic optimal control problem is bounded from below.

Given a measured output $y_0(\cdot)$, we have by the definition of $X_s[x_0, y_0(\cdot)|_0^s, d]$, that $x_s \in X_s[x_0, y_0(\cdot)|_0^s, d]$ if and only if there exist vector functions $x(\cdot), w(\cdot)$ and $v(\cdot)$ satisfying equation (5.2.1) and such that $x(s) = x_s$, the constraint (5.2.3) holds, and

$$y_0(t) = C(t)x(t) + v(t) \tag{5.6.33}$$

for all $t \in [0, s]$. Substitution of (5.6.33) into (5.2.3) implies that $x_s \in X_s[x_0, y_0(\cdot)|_0^s, d]$ if and only if an input $w(\cdot) \in \mathbf{L}_2[0, s]$ exists such that

$$J[x_s, w(\cdot)] \leq d, \tag{5.6.34}$$

where $J[x_s, w(\cdot)]$ is defined by

$$\begin{aligned} J[x_s, w(\cdot)] &\triangleq (x(0) - x_0)^{\mathrm{T}} X_0 (x(0) - x_0) \\ &\quad + \int_0^s \begin{pmatrix} w(t)^{\mathrm{T}} Q(t) w(t) - x(t)^{\mathrm{T}} K(t)^{\mathrm{T}} K(t) x(t) \\ +(y_0(t) - C(t)x(t))^{\mathrm{T}} R(t)(y_0(t) - C(t)x(t)) \end{pmatrix} dt \end{aligned} \tag{5.6.35}$$

and $x(\cdot)$ is the solution to (5.2.1) with the input $w(\cdot)$ and boundary condition $x(s) = x_s$.

Now consider the functional (5.6.35) with $x_0 = 0$ and $y_0(\cdot) \equiv 0$. In this case, J is a homogeneous quadratic functional with an end-point cost term. Also, consider the set $X_s[0, 0, 1]$ corresponding to $x_0 = 0$, $y_0(\cdot) \equiv 0$ and $d = 1$. Since $X_s[0, 0, 1]$ is bounded, a constant $h_s > 0$ exists such that all vectors $x_s \in \mathbb{R}^n$ with $\|x_s\| = h_s$ do not belong to the set $X_s[0, 0, 1]$. Hence,

$$J[x_s, w(\cdot)] > 1 \tag{5.6.36}$$

for all $x_s \in \mathbb{R}^n$ such that $\|x_s\| = h_s$ and for all $w(\cdot) \in \mathbf{L}_2[0, s]$. Since, J is a homogeneous quadratic functional, we have $J[ax_s, aw(\cdot)] = a^2 J[x_s, w(\cdot)]$ and (5.6.36) implies that

$$\inf_{w(\cdot) \in \mathbf{L}_2[0,s]} J[x_s, w(\cdot)] > 0 \tag{5.6.37}$$

for all $s \in [0, T]$ and all $x_s \neq 0$.

The problem of minimizing the functional from (5.6.37) subject to the constraint defined by the system (5.2.1) is a linear quadratic optimal control problem in which time is reversed. In this linear quadratic optimal control problem, a sign indefinite quadratic cost function is considered. Using a known result from the linear quadratic optimal control theory, we conclude that the condition (5.6.37) implies that there exists a solution $\Xi(\cdot)$ to the Riccati equation

$$\begin{aligned} -\dot{\Xi}(s) = \Xi(s)A(s) + A(s)^{\mathrm{T}}\Xi(s) + \Xi(s)B(s)Q(s)^{-1}B(s)^{\mathrm{T}}\Xi(s) \\ + K(s)^{\mathrm{T}}K(s) - C(s)^{\mathrm{T}}R(s)C(s) \end{aligned} \tag{5.6.38}$$

with the initial condition $\Xi(0) = X_0$; e.g., see p. 23 of [34]. Furthermore, this solution is positive-definite on $[0, T]$. From this, it follows that the solution to the

Riccati equation (5.3.11) is given by $P(\cdot) := \Xi(\cdot)^{-1}$. This completes the proof of this part of the lemma.

Statement (ii) and the sufficiency part of statement (i). We have shown above that $x_s \in X_s[x_0, y_0(\cdot)|_0^s, d]$ if and only if an input $w(\cdot) \in \mathbf{L}_2[0,s]$ exists such that the condition (5.6.34) holds for the functional (5.6.35). Now consider the following minimization problem:

$$\min_{w(\cdot)\in\mathbf{L}_2[0,s]} J[x_s, w(\cdot)], \tag{5.6.39}$$

where the minimum is taken over all $x(\cdot)$ and $w(\cdot)$ connected by (5.2.1) with the boundary condition $x(s) = x_s$. This problem is a linear quadratic optimal tracking problem in which the system operates in reverse time. The solution to this tracking problem is well known (e.g., see [87]). Indeed if the solution to the Riccati equation (5.3.11) exists, then the matrix function $\Xi(\cdot) = P(\cdot)^{-1} > 0$ is the solution to the Riccati equation (5.6.38) with initial condition $\Xi(0) = X_0$. From this, it follows that the minimum in (5.6.39) is achieved for any x_0 and any $y_0(\cdot)$. Furthermore as in [21], we can write

$$\min_{w(\cdot)\in\mathbf{L}_2[0,s]} J[x_s, w(\cdot)] = (x_s - \hat{x}(s))^{\mathrm{T}}\Xi(s)(x_s - \hat{x}(s)) - \rho_s,$$

where ρ_s is defined as in (5.3.13) and $\hat{x}(s)$ is the solution to (5.3.10) with the initial condition $\hat{x}(0) = x_0$. From this we can conclude that the set

$$X_s[x_0, y_0(\cdot)|_0^s, d] = \{x_s \in \mathbb{R}^n : \min_{w(\cdot)\in\mathbf{L}_2[0,s]} J[x_s, w(\cdot)] \le d\}$$

is given by (5.3.12). This completes the proof of the lemma. □

Proof of Lemma 5.4.2. Necessity. It is obvious that strong robust observability implies robust observability. Therefore, according to Lemma 5.3.2, the solution $P(\cdot)$ to the Riccati equation (5.3.11) is defined and positive-definite on $[0, NT]$. We now prove that the solution $X(\cdot)$ to the Riccati equation (5.4.17) is defined on $[0, NT]$ and positive-definite on $(0, NT]$. It is obvious from (5.3.12) that the set $X_s\left[x_0, y_0(\cdot)|_0^s, d\right]$ is not empty if and only if

$$d + \rho_s\left[y_0(\cdot)\right] \ge 0, \tag{5.6.40}$$

where $\rho_s\left[y_0(\cdot)\right]$ is defined by (5.3.13). Thus we can compose the following constraint is satisfied by all solutions to the linear system (5.3.10):

$$\int_0^s \left[-\|K(t)\hat{x}(t)\|^2 + (C(t)\hat{x}(t) - y_0(t))^{\mathrm{T}} R(t)\left(C(t)\hat{x}(t) - y_0(t)\right)\right] dt \le d. \tag{5.6.41}$$

By using the linear substitution

$$\hat{y}(t) := C(t)\hat{x}(t) - y_0(t), \tag{5.6.42}$$

the constraint (5.6.41) can be rewritten as

$$\int_0^s \left[\hat{y}(t)^{\mathrm{T}} R\hat{y}(t) - \|K(t)\hat{x}(t)\|^2\right] dt \le d, \tag{5.6.43}$$

where this constraint is satisfied by all solutions to the following linear system:

$$\dot{\hat{x}}(t) = \hat{A}(t)\hat{x}(t) + \hat{B}(t)\hat{y}(t), \;\; \hat{x}(0) = x_0, \tag{5.6.44}$$

where $\hat{A}(t)$ and $\hat{B}(t)$ are defined by (5.4.18) and (5.4.19), respectively.

In this case, we must prove that a positive-definite solution to the Riccati equation (5.4.17) exists. By the definition of $\hat{X}_s[d]$, we have that $\hat{x}_s \in \hat{X}_s[d]$ if and only if there exist vector functions $\hat{x}(\cdot)$ and $y_0(\cdot)$ satisfying the system equation (5.3.10) and such that $\hat{x}(s) = \hat{x}_s$ and the constraint (5.6.41) holds. If we treat $y_0(\cdot)$ as the control input of the system (5.3.10), it implies that $\hat{x}_s \in \hat{X}_s[d]$ if and only if $y_0(\cdot) \in \mathcal{L}_2[0, s]$ exists such that

$$\check{J}[\hat{x}_s, y_0(\cdot)] \leq d, \tag{5.6.45}$$

where $\check{J}[x_s, y_0(\cdot)]$ is defined by

$$\check{J}[\hat{x}_s, y_0(\cdot)] = \int_0^s \left[\begin{array}{r} (C(t)\hat{x}(t) - y_0(t))^{\mathrm{T}} R(t) \left(C(t)\hat{x}(t) - y_0(t)\right) \\ - \|K(t)\hat{x}(t)\|^2 \end{array} \right] dt. \tag{5.6.46}$$

By using the linear substitution (5.6.42), we can rewrite (5.6.46) as

$$\hat{J}[\hat{x}_s, \hat{y}(\cdot)] := \check{J}[\hat{x}_s, y_0(\cdot)] = \int_0^s \left[\hat{y}(t)^{\mathrm{T}} R(t)\hat{y}(t) - \|K(t)\hat{x}(t)\|^2 \right] dt. \tag{5.6.47}$$

Now $\hat{x}(\cdot)$ is the solution to (5.6.44) with the control input $\hat{y}(\cdot)$ and boundary condition $\hat{x}(s) = \hat{x}_s$. From (5.6.47), $\hat{J}$ is a homogeneous quadratic functional. Now consider the set $\hat{X}_s[1]$ corresponding to $d = 1$. Since $\hat{X}_s[1]$ is bounded, a constant $h_s > 0$ exists such that all vectors $\hat{x}_s \in \mathbb{R}^n$ with $\|\hat{x}_s\| = h_s$ do not belong to the set $\hat{X}_s[1]$. Hence,

$$\hat{J}[\hat{x}_s, \hat{y}(\cdot)] > 1 \tag{5.6.48}$$

for all $\hat{x}_s \in \mathbb{R}^n$ such that $\|\hat{x}_s\| = h_s$ and for all $\hat{y}(\cdot) \in \mathcal{L}_2[0, s]$. Since $\hat{J}$ is a homogeneous quadratic functional, we have that $\hat{J}[\alpha\hat{x}_s, \alpha\hat{y}(\cdot)] = \alpha^2 \hat{J}[\hat{x}_s, \hat{y}(\cdot)]$ and (5.6.48) implies that

$$\inf_{\hat{y}(\cdot) \in \mathcal{L}_2[0,s]} \hat{J}[\hat{x}_s, \hat{y}(\cdot)] > 0 \tag{5.6.49}$$

for all $\hat{x}_s \neq 0$. The problem of minimizing the functional from (5.6.49) subject to the constraint defined by the system (5.6.44) is a linear quadratic optimal control problem in which time is reversed. In this linear quadratic optimal control problem, a sign indefinite quadratic cost function is considered. By using a known result from the linear quadratic optimal control theory, we conclude that the condition (5.6.49) implies that there exists a solution $X(\cdot)$ to the Riccati equation

$$-\dot{X}(s) = \hat{A}(s)^{\mathrm{T}} X(s) + X(s)\hat{A}(s) - X(s)\hat{B}(s)R^{-1}(s)\hat{B}(s)^{\mathrm{T}} X(s) - K(s)^{\mathrm{T}} K(s),$$

with the initial condition $X(0) = 0$, for all $s \in (0, NT]$, such that $X(s) > 0$; e.g., see p. 23 of [34]. This completes the proof of this part of the lemma.

Sufficiency. We have already shown that $\hat{x}_s \in \hat{X}_s[d]$ if and only if $y_0(\cdot) \in \mathcal{L}_2[0, s]$ exists such that (5.6.45) is satisfied. Now consider the following optimization problem:

$$\inf_{\hat{y}(\cdot)\in\mathcal{L}_2[0,s]} \hat{J}[\hat{x}_s, \hat{y}(\cdot)],$$

where the infimum is taken over all solutions to (5.6.44) with the boundary condition $\hat{x}(s) = \hat{x}_s$. This optimal control problem is the standard linear quadratic optimal control problem with a sign indefinite cost function. Using the standard result from the theory of linear quadratic optimal control (e.g., see [34]), we obtain that

$$\inf_{\hat{y}(\cdot)\in\mathcal{L}_2[0,s]} \int_0^s \left[\hat{y}(t)^{\mathrm{T}} R(t)\hat{y}(t) - \|K(t)\hat{x}(t)\|^2\right] dt = \hat{x}_s^{\mathrm{T}} X(s)\hat{x}_s. \tag{5.6.50}$$

From (5.6.50) and (5.6.43), we have that

$$\hat{x}_s^{\mathrm{T}} X(s)\hat{x}_s \leq d.$$

From this we can conclude that the set $\hat{X}_s[d]$ is given by

$$\hat{X}_s[d] = \left\{\hat{x}_s \in \mathbb{R}^n : \hat{x}_s^{\mathrm{T}} X(s)\hat{x}_s \leq d\right\}.$$

This completes the proof of the lemma. □

6

An Analog of Shannon Information Theory: State Estimation and Stabilization of Linear Noiseless Plants via Noisy Discrete Channels

6.1 Introduction

In this chapter, we continue to consider the problems of state estimation and stabilization for discrete-time linear partially observed time-invariant systems. We still examine a remote control setup, where the sensors and decision maker (either observer or controller) are physically distant and connected by a noisy digital communication link. The critical feature of this chapter as compared with the previous ones is the account for channel errors by adopting the stochastic discrete memoryless channel model from the classic information theory. The objective is to examine how the channel quantization effects and errors limit the capacity for stabilization and reliable state estimation. The focus is on stabilizability and observability with probability 1. In other words, the estimation/stabilization error should be made small along (almost) any trajectory.[1]

In this chapter, we confine ourselves to consideration of the case where the uncertainties in the system model can be neglected. This is a natural initial step in developing the theory. Its generalizations on the case of systems with additive exogenous disturbances will be offered in Chaps. 7 and 8.

The main results of this chapter have points of similarity with the classic Shannon's noisy channel coding theorem [188] and are partly based on it. This theorem states that it is possible to ensure an errorless communication of information across the channel with as large a probability as desired if and only if the source produces data at the rate less than the fundamental characteristic of the noisy channel, introduced by Shannon [188] and called the *capacity* $\mathfrak{c}$. Similarly it is shown in this chapter that in order that the system be stabilizable/observable with as large a probability as desired or almost surely, it is sufficient and almost necessary that the unit

[1] A natural alternative approach deals with mth moment observability/stabilizability; see, e.g., [72, 77, 101–103, 132–138, 164–166, 192]. With the strong law of large numbers in mind, this approach can be viewed as aimed at making the time-average estimation/stabilization error small along (almost) any trajectory. This formally permits the error at a given time to be large but requires that large errors occur with a small frequency.

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_6,

time increment h (in bits) of the open-loop system state uncertainty be less than the channel capacity $\mathfrak{c}$. This increment is given by the topological entropy of the system. So the results of this chapter can also be viewed as natural extensions of the results of Chaps. 2 and 3 on the case of a noisy channel. We also specify the necessity of the bound $\mathfrak{c}$ by showing that whenever it is trespassed $h > \mathfrak{c}$, any estimation/stabilization algorithm almost surely exponentially diverges. We also show that the inequality $\mathfrak{c} \geq h$ is necessary for certain weaker forms of observability/stabilizability. For example, it holds whenever the time-average estimation/stabilization error is kept bounded with a nonzero probability.

By following the lines of the classic information theory, we examine the role of a possible communication feedback. This feedback holds where there is a way to communicate data in the direction opposite to that of the feedforward communication channel. As is known, even perfect (i.e., undelayed and noiseless) and complete (i.e., notifying the informant about the entire result of the transmission across the channel) feedback does not increase the rate at which the information can be communicated over the feedforward channel with as small probability of error as desired [44, 190]. In this chapter, we show that the communication feedback similarly does not alter the stabilizability/observability domain. However, such a feedback aids to improve the performance of the state estimator.

Specifically, we show that in the absence of a feedback communication, the estimation error can be made decaying to zero with as large a probability as desired by a proper design of the observer. However, this is achieved at the expense of using code words whose lengths grow as the estimation process progresses.[2] At the same time, the increasing code words lengths mean that the memories of the coder and decoder should increase accordingly, and the per-step complexity of the observer may be unlimited. At the same time, we show that these disadvantages can be discarded if a perfect and complete feedback link is available. The main result of the chapter concerning the state estimation problem states that then it is possible to design an almost surely converging observer of limited complexity that employs fixed-length code words, provided that $\mathfrak{c} > h$. This is achieved via complete synchronization of the coder and decoder, within which the coder duplicates the state estimate generated by the decoder by employing the data received over the feedback link.

Unlike state estimation, stabilization needs far less feedback communication. We first show that to design an almost surely stabilizing controller with limited complexity employing fixed-length code words, a feedback link of arbitrarily small capacity is sufficient, and a complete synchronization of the coder and decoder is not needed. Second, we demonstrate that in fact such a feedback communication requires no special means (like a special feedback link) since it can be implemented by means of control. The feedback communication can be arranged thanks to the fact that on the one hand, the decoder-controller influences the motion of the system and on the other hand, the sensor observes this motion and feeds the coder by the observation. So the controller can encode a message by imparting the motion of a certain specific

[2] Despite of this, the observer produces an asymptotically exact state estimate online, i.e., with no delay. In other words, the estimate of the state at time t is generated at time t.

feature. In its turn, the coder can receive the message by observing the motion and detecting this feature.

The main results of the chapter were originally published in [107, 108, 120].

In [201], the necessity of the inequality $\mathfrak{c} \geq h$ for almost sure observability/stabilizability was established for channels more general than discrete and memoryless. The sufficiency of the strict inequality $\mathfrak{c} > h$ was justified for a particular example of a discrete memoryless channel: the erasure channel with a perfect and complete feedback.

Observability and stabilizability of unstable linear plants over noisy discrete channels were also addressed in, e.g., [77, 101–103, 164–166, 192, 229]. The relevant problem of design of optimal sequential quantization schemes for uncontrolled Markov processes was addressed in [24, 204]. In [77, 164–166, 192], the focus is on scalar noisy linear systems and mean-square (and more generally, mth moment) observability/stabilizability. Both sufficient and necessary criteria for such an observability/stabilizability were given in [164–166] in terms of a new parametric notion of the channel capacity (called anytime capacity) introduced in [164]. An encoder–decoder pair for estimating the state of a scalar noisy linear system via a noisy binary symmetric channel was proposed in [77]. It was shown by simulation that the estimation error is bounded. Another such a pair was constructed in [192] for such a channel with a perfect feedback. Conditions ensuring that the mathematical expectation of the estimation error is bounded were obtained. In [229], the focus is on stabilizability by means of memoryless controllers in the case where the channels transmitting both observations and controls are noisy and discrete, and the plant is scalar, linear, and stochastic. A stochastic setting for the stabilization problem was investigated in [138], where an important fundamental result on minimum data rates was obtained. However, the paper [138] deals with a noiseless deterministic channel. In [101, 103], the robust rth moment stability of uncertain scalar linear plants with bounded additive disturbances was examined in the case where the signals are communicated over the so-called *truncation channel*. It transmits binary code words with dropping a random number of concluding (i.e., least important) bits. This particular example of the discrete memoryless channel generalizes the classic erasure channel and is motivated by certain wireless communication applications in [101]. Minimum data rates for stabilization and state estimation via such channels were also studied in many other papers in the area. With no intent to be exhaustive, we mention [28, 48, 70, 73, 133, 135–137, 149, 184, 202, 204, 220, 221, 227, 228]. Optimization problems for noisy Gaussian channels with power constraints and perfect finite alphabet channels were studied in, e.g., [127, 203, 204, 206].

The observers and controllers considered in this chapter are based on quantizers with adjusted sensitivity [28, 204] implemented in the multi-rate fashion [149]. Such a quantizer can be regarded as a cascade of a multiplier by an adjustable factor and an analog-to-digital converter. To be transmitted across the channel, the outputs of this converter are encoded by using low error block codes, whose existence is ensured by the classic Shannon channel coding theorem. This is in the vein of the classic source-channel separation principle [20, 60, 68]. In the context of estimating the state of an unstable plant via a noisy channel, a similar approach was considered in [164] in the form of the following separation of the source and channel. At first, a coder–decoder pair is designed under the assumption that the channel is perfect. Then another coder–decoder pair is constructed to carry the outputs of the first coder reliably across the channel. Various issues concerning stabilization and state estimation by means of quantizers with adjusted sensitivity were addressed in, e.g., [28, 70, 88, 90, 204].

The view of the control loop as a link transmitting information was in fact concerned in [47] in its "a posteriori" component. This means that the control loop does transmit information, although its contents may not be not clearly specified a priori. The "constructive" part of the same view is the idea that the control signals can be employed as carriers of a priori prespecified information from the decoder-controller to the coder. In some details, this issue was addressed in [166, 201] (see also [167] for a general discussion).

The body of the chapter is organized as follows. Sections 6.2, 6.5 and 6.4, 6.6 contain the statements of the state estimation and stabilization problems and present their solutions, respectively. Section 6.3 establishes the notation and offers basic definitions and assumptions. The proofs of the main results are scattered over Sects. 6.7–6.11.

6.2 State Estimation Problem

We consider unstable discrete-time invariant linear systems of the form:

$$x(t+1) = Ax(t), \qquad x(0) = x_0, \qquad y(t) = Cx(t). \tag{6.2.1}$$

Here $x(t) \in \mathbb{R}^n$ is the state and $y(t) \in \mathbb{R}^{n_y}$ is the measured output. The instability means that there is an eigenvalue λ of the matrix A with $|\lambda| \geq 1$. The initial state x_0 is a random vector. The objective is to estimate the current state on the basis of the prior measurements.

We consider the case where this estimate is required at a remote location. The only way to communicate information from the sensor to this location is via a given random noisy discrete channel. So to be transmitted, measurements should be first translated into a sequence of symbols e from the finite *input alphabet* $\mathfrak{E}$ of the channel. This is done by a special system's component, referred to as the *coder*. Its outputs e are then transmitted across the channel and transformed by some sort of random disturbance or noise into a sequence of channel's outputs s from a finite *output alphabet* $\mathfrak{S}$. By employing the prior outputs s, the *decoder(-estimator)* produces an estimate $\widehat{x}$ of the current state x. In this situation illustrated in Fig. 6.1, an *observer* is constituted by a coder–decoder pair.

The decoder is described by an equation of the form:

$$\widehat{x}(t) = \mathcal{X}\left[t, s(0), s(1), \ldots, s(t)\right]. \tag{6.2.2}$$

We consider two classes of coders each giving rise to a particular problem setup.

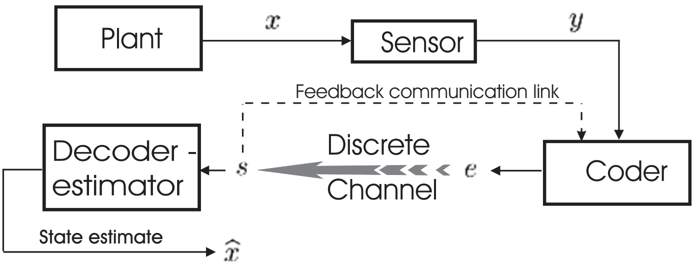


Fig. 6.1. Estimation via a limited capacity communication channel.

Coders with a Communication Feedback

The first class is related to feedback communication channels [190]: The result $s(t)$ of the current transmission across the "feedforward" channel becomes known at the coder site by the next time $t+1$. The coders from this class are said to be *with a feedback* and given by an equation of the form:

$$e(t) = \mathcal{E}[t, y(0), \dots, y(t), s(0), \dots, s(t-1)] \in \mathfrak{E}. \tag{6.2.3}$$

Coders without a Communication Feedback

The second class deals with the case where no such feedback is available. The corresponding coders are said to be *without a feedback* and given by an equation of the form:

$$e(t) = \mathcal{E}[t, y(0), \dots, y(t)] \in \mathfrak{E}. \tag{6.2.4}$$

Objective of State Estimation

The information received by the decoder is limited to a finite number of bits at any time. So the decoder is hardly able to restore the state with the infinite exactness $\widehat{x}(t) = x(t)$ for a finite time. In view of this, we pursue a more realistic objective of detecting the unstable modes of the system and accept that an observer succeeds if

$$\|x(t) - \widehat{x}(t)\| \to 0 \qquad \text{as} \qquad t \to \infty. \tag{6.2.5}$$

Definition 6.2.1. *The coder–decoder pair is said to* detect *or* track *the state whenever* (6.2.5) *is true, and to* keep the estimation error (or time-average error) bounded *if the following weaker properties hold, respectively:*

$$\overline{\lim_{t\to\infty}} \|x(t) - \widehat{x}(t)\| < \infty, \qquad \overline{\lim_{t\to\infty}} \frac{1}{t} \sum_{\theta=0}^{t-1} \|x(\theta) - \widehat{x}(\theta)\| < \infty. \tag{6.2.6}$$

Remark 6.2.2. The estimate $\widehat{x}(t)$ of the current state $x(t)$ should be produced at the current time t; no delay is permitted.

The main question to be discussed is as follows:

How low can the data rate of the channel be made before the construction of a coder–decoder pair detecting the state becomes impossible?

Explanation 6.2.3. Since the initial state is random, the process in the system is stochastic. So the objective (6.2.5) or (6.2.6) may be achieved for some elementary random events and fail to hold for others.

In this chapter, we focus on the cases where "detecting" means either "detecting with as large probability as desired" or "detecting almost surely."

Comments on Communication Feedback

The communication feedback enables the coder (6.2.3) to be aware of the actions of the decoder via duplicating them in accordance with (6.2.2). This gives the coder the ground to try to compensate for the previous channel errors. However, this feedback does not increase the rate at which the information can be communicated across the channel with as small probability of error as desired [44, 188]. At the same time, it may increase the rate at which information can be transmitted with the zero probability of error [189]. The feedback also may increase the reliability function [233] and simplify the coding and decoding operations [214]. For further discussion of this issue and a detailed survey, we refer the reader to [214].

A discussion of the role of communication feedback in control and state estimation was offered in, e.g., [201, 202, 204, 206, 219].

The perfect feedback communication link may model a situation where there is a bi-directional data exchange between the coder and the decoder, and the transmission power of the decoder essentially exceeds that of the coder, so that the feedback communication errors are negligible. Examples concern data exchange between a satellite and Earth surface station, or underwater autonomous sensors and the base station. For control systems, feedback communication of data from the decoder to the coder does not require special means (like a special link) since transmission of as much information as desired can be arranged by means of control (see Sect. 6.11).

6.3 Assumptions, Notations, and Basic Definitions

6.3.1 Assumptions

In this chapter, the noisy communication channel between the coder and the decoder is interpreted as a stochastic map transforming input channel symbols into output ones $e \mapsto s$. We suppose that the previous transmissions across the channel do not affect the current one, and the channel is time-invariant. These properties are summarized by the following.

Assumption 6.3.1. *Given a current channel input $e(t)$, the current output $s(t)$ is statistically independent of the other inputs and outputs $e(j), s(j), j < t$, and the probability of receiving s given that e is sent does not depend on time:*

$$W(s|e) := \boldsymbol{P}[s(t) = s|e(t) = e], \qquad s \in \mathfrak{S}, e \in \mathfrak{E}.$$

Remark 6.3.2. This means that we consider a discrete memoryless channel [38, 50, 68, 188, 190].

Remark 6.3.3. The above model incorporates the effect of message dropout by including a special "void" symbol $\circledast$ in the output alphabet $\mathfrak{S}$. Then $s(t) = \circledast$ means that the message $e(t)$ is lost by the channel.

Other assumptions concern mainly the system (6.2.1).

Assumption 6.3.4. *The system* (6.2.1) *does not affect the operation of the channel: Given an input* $e(t)$*, the output* $s(t)$ *is statistically independent of the initial state* x_0 *(along with the other channel inputs and outputs* $e(j), s(j), j < t$*).*

Assumption 6.3.5. *The initial state* x_0 *has a probability density* $p_0(x)$.

Assumption 6.3.6. *The pair* (A, C) *is detectable.*

To state the results of the chapter, we need the notion of the Shannon's capacity of the channel. To recall it, we start with the following.

6.3.2 Average Mutual Information

Let $F \in \mathfrak{F} = \{f\}$ and $G \in \mathfrak{G} = \{g\}$ be two random quantities defined on a common probability space and with respective probability distributions $\boldsymbol{P}(df)$ and $\boldsymbol{P}(dg)$.[3] The *average mutual information* between F and G is defined to be (see, e.g., [38, 50, 65, 68, 152, 188])

$$I(F, G) = \int \boldsymbol{P}(df, dg) \log_2 \frac{\boldsymbol{P}(df, dg)}{\boldsymbol{P}(df) \otimes \boldsymbol{P}(dg)}. \tag{6.3.1}$$

Here $\boldsymbol{P}(df, dg)$ is the joint distribution of F and G, and $\frac{\boldsymbol{P}(df,dg)}{\boldsymbol{P}(df)\otimes\boldsymbol{P}(dg)}$ stands for the density of this distribution with respect to the product measure $\boldsymbol{P}(df) \otimes \boldsymbol{P}(dg)$. Formula (6.3.1) holds if this density does exist; i.e., the joint distribution is absolutely continuous with respect to the product measure. Otherwise $I(F, G) := \infty$.

We recall that $\log_2 0 := -\infty$ and $\pm\infty \cdot 0 := 0$.

Remark 6.3.7. It can be shown that (either finite or infinite) integral from (6.3.1) exists, and $I(F, G) \in [0, +\infty]$ [65].

The *entropy* of a random quantity is defined as self-information: $H(F) := I(F, F)$. It can be interpreted as an average amount of information (in bits) that one receives when the result f of a random experiment becomes known [38, 50, 65, 68,

[3] For technical reasons, we suppose that $\mathfrak{F}$ and $\mathfrak{G}$ are separable metric spaces endowed with the Borel σ-algebras, and the quantities F and G are measurable as maps from the probability space into $\mathfrak{F}$ and $\mathfrak{G}$, respectively.

188]. An equivalent contrapositive statement is that the entropy is an average amount of information that one lacks before the experiment to uniquely and correctly foresee its result. In other words, the entropy can be viewed as a measure of uncertainty. In both cases, it is tacitly assumed that available is an apriorial statistical knowledge about the experiment in the form of the probability distribution $\boldsymbol{P}(df)$.

The *conditional entropy* $H_g(F) = H_{G=g}(F)$ is defined as the entropy produced by the conditional distribution of F given $G = g$, and the *averaged conditional entropy* (*equivocation*) $H(F|G) := \boldsymbol{E}H_G(F)$. The latter can be viewed as an average uncertainty about the result of the experiment F provided that the result of G is known.

If the sets $\mathfrak{F}, \mathfrak{G}$ are finite, then the above formulas can be rewritten in terms of the probability mass functions

$$p_F(f) := \boldsymbol{P}(F = f), \quad p_G(g) := \boldsymbol{P}(G = g), \quad p_{FG}(f, g) := \boldsymbol{P}(F = f \wedge G = g)$$

as follows:

$$\begin{aligned}
H(F) &:= -\sum_{f \in \mathfrak{F}} p_F(f) \log_2 p_F(f), \\
H_{G=g}(F) &= -\sum_{f \in \mathfrak{F}} \boldsymbol{P}(F = f|G = g) \log_2 \boldsymbol{P}(F = g|G = g) \\
&= -\sum_{f \in \mathfrak{F}} \frac{p_{F,G}(f, g)}{p_G(g)} \log_2 \frac{p_{F,G}(f, g)}{p_G(g)}, \\
H(F|G) &= -\sum_{f \in \mathfrak{F}, g \in \mathfrak{G}} p_{F,G}(f, g) \log_2 \frac{p_{F,G}(f, g)}{p_G(g)}.
\end{aligned} \tag{6.3.2}$$

$$\begin{aligned}
I(F, G) &= \sum_{f \in \mathfrak{F}, g \in \mathfrak{G}} p_{F,G}(f, g) \log_2 \frac{p_{F,G}(f, g)}{p_F(f) p_G(g)} \\
&= H(F) - H(F|G) = H(G) - H(G|F).
\end{aligned}$$

Remark 6.3.8. By the last two relations, $I(F, G)$ can be viewed as the number of bits by which the uncertainty about the result of one of two experiments decreases when the result of the other experiment becomes known. This supports the name of $I(F, G)$ as a mutual information.

6.3.3 Capacity of a Discrete Memoryless Channel

The Shannon capacity of such a channel is defined to be the maximum mutual information between the input and output of the channel (see, e.g., [38, 50, 65, 68, 188, 190]):

$$\mathfrak{c} = \max_{\boldsymbol{P}_E} I(E, S). \tag{6.3.3}$$

Here the maximum is over all probability distributions $\boldsymbol{P}_E$ on the input channel alphabet $\mathfrak{E} = \{e\}$. Whereas $\boldsymbol{P}_E$ is interpreted as the probability distribution of a random channel input E, the joint distribution of the channel input E and output S is taken to be that of (e, s) when s results from sending e over the channel:

$$\boldsymbol{P}_{E,S}(e, s) := W(s|e)p_E(e).$$

6.3.4 Recursive Semirational Observers

In the above problem statement, the complexity of an observer is not limited: It is permitted to carry out an asymptotically infinite amount of computations per sample period. At the same time, it will be shown that the conditions necessary for observability by means of such an observer are "almost sufficient" for existence of a more realistic observer, which performs only finitely many operations per step.[4] This observer consists of a semirational coder and decoder. They are basically defined in Subsect. 3.4.4 (starting on p. 46) up to formalities caused by the difference in the situations considered in this chapter and Chap. 3, respectively. (This difference concerns mainly the channel model.) In this subsection, we adjust the corresponding definitions from Subsect. 3.4.4 to the context of the current chapter. Moreover, with the needs of Chap. 7 in mind, we extend these definitions by permitting the current outputs of the coder and decoder to depend on the messages received during not only the previous but also the current operation epoch. This feature is addressed by dropping the attribute "simple" in the definitions.

Definition 6.3.9. *The feedback coder* (6.2.3) *(on p.* 135*) is said to be* semirational r-step recursive *(where $r = 1, 2, \ldots$) if the following statements hold:*

(i) The coder starts working at $t = r$ and at any time $t \in \big[ir : (i+1)r\big)$ (where $i = 1, 2, \ldots$) generates the current output by equations of the form:

$$e(t) := \mathcal{E}_{t-ir}\Big\{z_c[ir], y[(i-1)r+1], \ldots, y[t], s[(i-1)r], \ldots, s[t-1]\Big\},$$
$$z_c[ir] := \mathcal{Z}_c\Big(z_c[(i-1)r], Y_i, S_i\Big) \in \mathbb{R}^s, \quad z_c(0) := z_c^0, \quad (6.3.4)$$

where $i = 1, 2, \ldots$ and

$$Y_i := \textbf{col}\Big(y[(i-1)r+1], \ldots, y[ir]\Big),$$
$$S_i := \Big(s[(i-1)r], \ldots, s[ir-1]\Big) \quad (6.3.5)$$

are the sequences of messages arrived at the coder from the sensor and decoder, respectively, during the previous operation epoch;

(ii) The functions $\mathcal{Z}_c(\cdot)$ and $\mathcal{E}_0(\cdot), \ldots, \mathcal{E}_{r-1}(\cdot)$ are semirational and semialgebraic, respectively.[5]

[4] This holds in the presence of a feedback communication link.

[5] See Definitions 3.4.9 and 3.4.12 on pp. 46 and 47, respectively.

The coder (6.2.4) *(on p.* 135*) without a feedback is said to be* semirational r-step recursive *if it meets the requirements (i) and (ii) with the arguments of the forms $s(\theta)$ and S_i dropped from the right-hand sides in* (6.3.4)*.*

A particular case of the situation from this definition is where, like in Definition 3.4.15 (on p. 47), the entire code word composed by all code symbols

$$E_i := \big(e[ir], e[ir+1], \ldots, e[(i+1)r-1]\big)$$

that will be emitted into the channel during the forthcoming operation epoch $[ir : (i+1)r-1]$ is generated by the coder at the beginning $t = ir$ of the epoch via an equation of the form: $E_i = \mathcal{E}\big[z(ir), Y_i\big]$ with a semialgebraic function $\mathcal{E}(\cdot)$.

Definition 6.3.10. *The decoder* (6.2.2) *(on p.* 134*) is said to be* semirational r-step recursive *if at any time $t \in \big[ir : (i+1)r\big), i = 1, 2, \ldots,$ the current estimate is generated by equations of the form:*

$$\widehat{x}(t) := \mathcal{X}_{t-ir}\Big\{z_d[ir], s[(i-1)r], \ldots, s[t-1]\Big\},$$
$$z_d[ir] := \mathcal{Z}_d\big(z_d[(i-1)r], S_i\big) \in \mathbb{R}^\sigma, \quad z_d(0) = z_d^0, \quad (6.3.6)$$

where the functions $\mathcal{Z}_d(\cdot)$ and $\mathcal{X}_0(\cdot), \ldots, \mathcal{X}_{r-1}(\cdot)$ are semirational and S_i is given by (6.3.5)*.*

A particular case of the situation from this definition is where the entire sequence of estimates

$$\widehat{X}_i = \textbf{col}\,\big(\widehat{x}[ir], \ldots, \widehat{x}[(i+1)r-1]\big) \quad (6.3.7)$$

for the forthcoming operation epoch is generated via an equation of the form: $\widehat{X}_i = \mathcal{X}\big(z_d[ir], S_i\big)$ with a semirational function $\mathcal{X}(\cdot)$.

As in Remark 3.4.18 (on p. 48), it should be noted that the coders and decoders from the classes introduced by these definitions perform a limited (as t runs over $t = 0, 1, \ldots$) number of operations per step. Explanation 3.4.20 and Remark 3.4.21 (on p. 48) also extend to these coders and decoders.

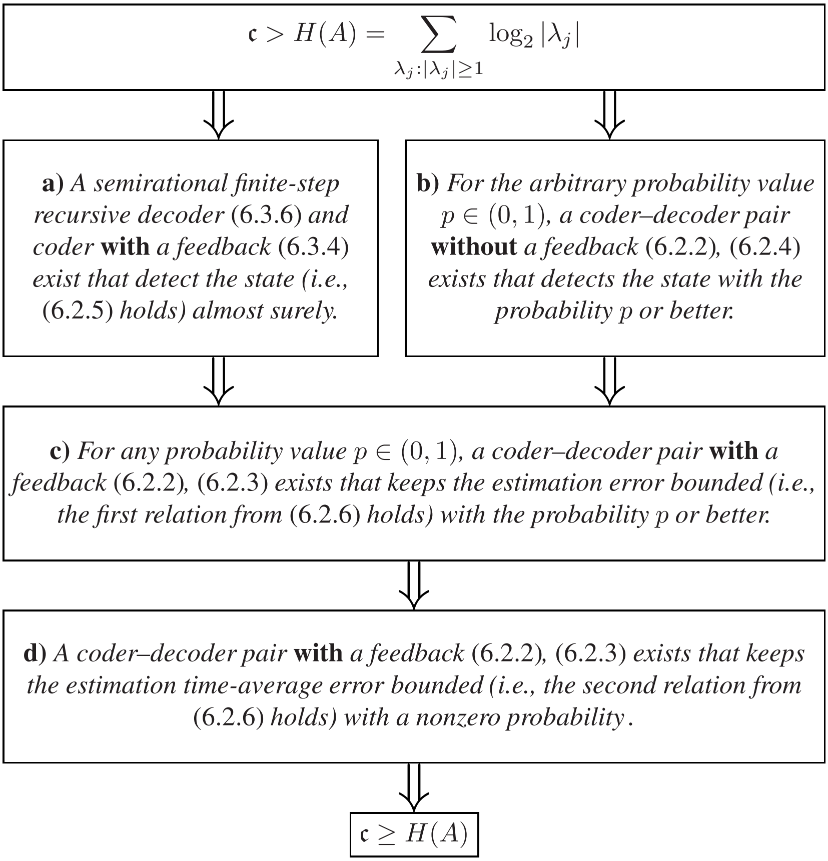
6.4 Conditions for Observability of Noiseless Linear Plants

These conditions are given by the following theorem, which constitutes the main result of this chapter with respect to the state estimation problem.

Theorem 6.4.1. *Suppose that Assumptions* 6.3.1 *and* 6.3.4–6.3.6 *hold. Denote by*

$$\lambda_1, \ldots, \lambda_n$$

the eigenvalues of the system (6.2.1) *repeating in accordance with their algebraic multiplicities, and by $\mathfrak{c}$ the capacity* (6.3.3) *of the communication channel. Then the following implications are true:*



The proof of this theorem will be given in Sects. 6.7–6.10.

Explanation 6.4.2. The implications **a)**$\vee$ **b)** $\Rightarrow$ **c)** $\Rightarrow$ **d)** are evident. They are mentioned to stress that the claims **a)**, **b)**, **c)**, and **d)** are included in the chain of implications with approximately identical extreme terms. Thus these statements are "almost equivalent," and the inequality $\mathfrak{c} > H(A)$ is sufficient and "almost necessary" for the system (6.2.1) to be observable via the noisy communication channel.

Remark 6.4.3. The implication $\mathfrak{c} > H(A) \Rightarrow$ a) means that whenever the above sufficient condition is met and a feedback communication is available, a reliable state estimation can be accomplished by an observer with a limited complexity, which performs a limited number of operations per step.

Remark 6.4.4. The corresponding observer (coder and decoder) will be explicitly constructed in Subsect. 6.9.1 (see p. 168). However, the scheme (coding and decoding rules) for transmission of information across the channel will not be described in detail. The point is that the proposed observer employs block codes transmitting data at a given rate below the channel capacity $\mathfrak{c}$ with a given probability of error. Classic information theory guarantees existence of such a code. Moreover, invention of such codes is the standard long-standing task in information sciences. It is supposed that a relevant solution should be borrowed to construct the observer.

Thus in the case where a perfect feedback communication link is available, the chapter demonstrates that whenever almost sure observability holds, it can be ensured by realistic observers with bounded (as time progresses) algebraic complexity and memory consumption per step, which are based on classic block coding–decoding schemes of communication.

Remark 6.4.5. By Theorem 2.4.2 (on p. 21), the quantity $H(A)$ is the topological entropy of the linear system (6.2.1). Hence for such systems, Theorem 6.4.1 can be viewed as an extension of Theorem 2.3.6 (on p. 16) to the case of a noisy channel and linear plant.

Comment 6.4.6. The paper [201] proves that the inequality $\mathfrak{c} \geq \eta(A)$ is necessary for existence of a coder–decoder pair tracking the state almost surely via noisy channels more general than discrete memoryless. The sufficiency of the inequality $\mathfrak{c} > \eta(A)$ for almost sure observability was justified in [201] for a particular discrete memoryless channel: the erasure channel with a finite alphabet.

Comparison with Shannon's Channel Coding Theorem

In spirit, Theorem 6.4.1 resembles the celebrated Shannon's channel coding theorem [50, 60, 68, 188]. Indeed, the latter states that whenever the source produces information at the rate $R < \mathfrak{c}$ bits per unit time, the success, i.e., errorless transmission, can be ensured with as large a probability as desired. If conversely $R > \mathfrak{c}$, this is impossible. Here the means to ensure success are the rules to encode and decode information before and after transmission, respectively. Theorem 6.4.1 asserts just the same provided the "success" is understood as asymptotic tracking (6.2.5) of the state, the "means" are the coder and decoder-estimator, and R is replaced by $H(A)$.

This analogy is enhanced by the similarity between the quantities R and $H(A)$. Each of them can be interpreted as the unit-time increment of the number of bits required to describe the entity that the receiver wants to know. Indeed in the context of Shannon's theorem, this entity is abstract information generated by a source at the rate R, and the interpretation is apparent. In the case considered in this chapter, this entity is the state, and the interpretation follows from Remark 3.5.3 (on p. 53).

Another point of similarity between Theorem 6.4.1 and the classic information theory concerns the communication feedback. Whereas the classic theory states that this feedback does not increase the rate R at which the information can be transmitted with as small a probability of error as desired, Theorem 6.4.1 shows that the feedback does not extend the class of systems for which state tracking (6.2.5) is possible with as small a probability of failure as desired. However, this concerns tracking by means of arbitrary non-anticipating observers. At the same time, Theorem 6.4.1 states that feedback allows for tracking the state by means of an observer with a limited computational power and not only with as large a probability as desired but also almost surely.

6.5 Stabilization Problem

Now we consider a controlled version of the plant (6.2.1). In other words, we deal with unstable linear discrete-time invariant linear systems of the form:

$$x(t+1) = Ax(t) + Bu(t), \qquad x(0) = x_0, \qquad y(t) = Cx(t), \tag{6.5.1}$$

where $u(t) \in \mathbb{R}^{n_u}$ is the control. The objective is to design a controller that asymptotically stabilizes the system:

$$x(t) \to 0 \quad \text{and} \quad u(t) \to 0 \quad \text{as} \quad t \to \infty.$$

We examine a remote control setup: The site where the control is produced is physically distant from the sensor site. The only way to communicate data from the second site to the first one is via a given discrete memoryless channel. Based on the prior observations, the coder selects a message e from the input channel alphabet $\mathfrak{E}$ and emits e into the channel. In the channel, this message is transformed into a symbol s from the output channel alphabet $\mathcal{S}$. Proceeding from the messages s up to the current time t, the *decoder(-controller)* selects a control $u(t)$ (see Fig. 6.2).

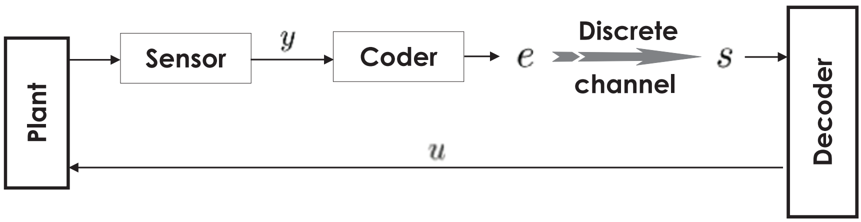


Fig. 6.2. Stabilization via a limited capacity communication channel.

Remark 6.5.1. In this situation, the *controller* is assembled of the coder and decoder.

In fact, the decoder is still given by an equation of the form (6.2.2). However, now its output is not a state estimate but the control:

$$u(t) = \mathfrak{U}\left[t, s(0), s(1), \ldots, s(t)\right]. \tag{6.5.2}$$

We also still consider two classes of coders, each giving rise to a particular problem statement. The first class corresponds to the case where a feedback communication link from the decoder to the coder is available (see Fig. 6.3). The second class is considered when there is no such feedback link (see Fig. 6.2). The coders from these two classes are still described by equations of the forms (6.2.3) and (6.2.4), respectively.

Remark 6.5.2. As will be shown in the next section, the feedback communication link is of much lesser importance for the stabilization problem than for the state estimation one. Only in order to demonstrate this explicitly, we continue to consider coders with a communication feedback.

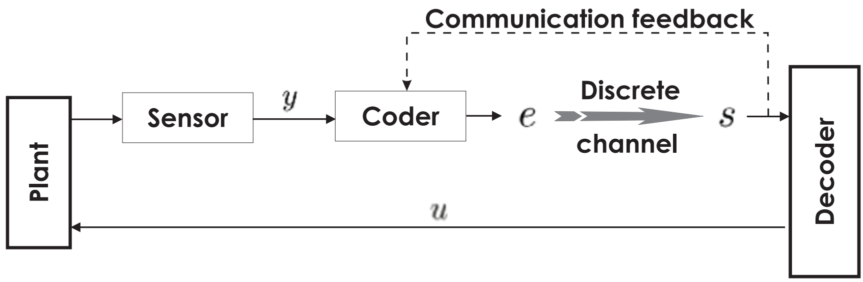


Fig. 6.3. Stabilization under a communication feedback.

Definition 6.5.3. *A coder–decoder pair is said to* stabilize *the system if*

$$\|x(t)\| \to 0 \quad \text{and} \quad \|u(t)\| \to 0 \qquad \text{as} \qquad t \to \infty, \tag{6.5.3}$$

and to keep the stabilization error (or time-average error) bounded *if the following weaker properties hold, respectively:*

$$\overline{\lim_{t\to\infty}} \|x(t)\| < \infty, \qquad \overline{\lim_{t\to\infty}} \frac{1}{t}\sum_{\theta=0}^{t-1} \|x(\theta)\| < \infty. \tag{6.5.4}$$

The main question to be studied is as follows:

What is the tightest bound on the data rate of the channel above which a stabilizing coder–decoder pair exists?

Explanation 6.5.4. Since the initial state is random, the process in the system is stochastic. So the objective (6.5.3) or (6.5.4) may be achieved for some elementary random events and fail to hold for others.

In this chapter, we focus on the cases where "stabilizing" means either "stabilizing with as large probability as desired" or "stabilizing almost surely."

Assumptions

Apart from Assumptions 6.3.1 and 6.3.4–6.3.6, now one more assumption is adopted.

Assumption 6.5.5. *The pair (A, B) is stabilizable.*

Recursive Semirational Controllers

In the above problem statement, the controller is formally permitted to carry out an asymptotically infinite amount of computations per sample period. At the same time, it will be shown that the conditions necessary for stabilizability by means of such a controller are "almost sufficient" for existence of a more realistic controller, which performs only a limited number of operations per step.

Remark 6.5.6. Such a controller exists irrespective of whether a feedback communication link is available, unlike the state estimation problem.

The aforementioned more realistic controller consists of a semirational coder and decoder. Such a coder is introduced by Definition 6.3.9 (on p. 139). As for the decoder, Definition 6.3.10 (on p. 140) serving the state estimation problem should be slightly modified since now the decoder output is the control.

Definition 6.5.7. *The decoder* (6.5.2) *is said to be* r-step semirational recursive *if at any time* $t \in \big[ir : (i+1)r\big), i = 1, 2, \ldots$, *the current control is generated by equations of the form:*

$$u(t) := \mathfrak{U}_{t-ir}\Big\{z_d[ir], s[(i-1)r], \ldots, s[t-1]\Big\},$$

$$z_d[ir] := \mathfrak{Z}_d\big(z_d[(i-1)r], S_i\big) \in \mathbb{R}^\sigma, \quad z_d(0) = z_d^0, \quad (6.5.5)$$

where the functions $\mathfrak{Z}_d(\cdot)$ *and* $\mathfrak{U}_0(\cdot), \ldots, \mathfrak{U}_{r-1}(\cdot)$ *are semirational*[6] *and* S_i *is given by* (6.3.5) *(on p.* 139*).*

A particular case of the situation from this definition is where like in Definition 3.4.17 (on p. 47), the control program

$$\boldsymbol{U}_i = \textbf{col}\,\big(u[ir], \ldots, u[(i+1)r-1]\big) \quad (6.5.6)$$

for the entire forthcoming operation epoch is generated via an equation of the form:

$$\boldsymbol{U}_i = \mathfrak{U}\big(z_d[ir], S_i\big)$$

with a semirational function $\mathfrak{U}(\cdot)$.

6.6 Conditions for Stabilizability of Noiseless Linear Plants

6.6.1 The Domain of Stabilizability Is Determined by the Shannon Channel Capacity

Theorem 6.6.1. *Suppose that Assumptions* 6.3.1*,* 6.3.4–6.3.6*, and* 6.5.5 *hold. Denote by* $\lambda_1, \ldots, \lambda_n$ *the eigenvalues of the system* (6.5.1) *repeating in accordance with their algebraic multiplicities, and by* $\mathfrak{c}$ *the Shannon capacity* (6.3.3) *of the communication channel. Then the following implications are true:*

$$\boxed{\mathfrak{c} > H(A) = \sum_{\lambda_j : |\lambda_j| \geq 1} \log_2 |\lambda_j|}$$

$$\Downarrow$$

[6] See Definition 3.4.12 on p. 47.

$\Downarrow$

a) *A semirational finite-step recursive decoder* (6.5.5) *and coder* (3.4.4) *(on p.* 47*)* **without** *a communication feedback exist that stabilize the system almost surely.*

$\Downarrow$

b) *A coder* **with** *a communication feedback* (6.2.3) *and a decoder* (6.5.2) *exist that stabilize the system almost surely.*

$\Downarrow$

c) *For the arbitrary probability value* $p \in (0, 1)$*, a coder* **with** *a communication feedback* (6.2.3) *and a decoder* (6.5.2) *exist that stabilize the system with the probability* p *or better.*

$\Downarrow$

d) *A coder* **with** *a communication feedback* (6.2.3) *and a decoder* (6.5.2) *exist that keep the time-average stabilization error bounded with a nonzero probability.*

$\Downarrow$

$\mathfrak{c} \geq H(A)$

The proof of this theorem will be given in Sects. 6.7 and 6.11.

Explanation 6.6.2. The evident implications a) $\Rightarrow$ b) $\Rightarrow$ c) $\Rightarrow$ d) are included in the statement of the theorem due to the reasons commented on in Explanation 6.4.2.

Remark 6.6.3. The implications d) $\Rightarrow$ $\mathfrak{c} \geq H(A)$ and $\mathfrak{c} > H(A) \Rightarrow$ a) mean that the condition $\mathfrak{c} \geq H(A)$ necessary for stabilizability with a nonzero probability in the weak sense (6.5.4) by means of a controller with unlimited complexity, which employs a perfect communication feedback, is simultaneously “almost sufficient” for stabilizability with probability 1 in the strong sense (6.5.3) and in the absence of any communication feedback by means of a realistic controller with limited computational power.

Explicit constructions of the coder and decoder constituting this more realistic controller will be offered in Subsects. 6.11.1 and 6.11.2 (see p. 185). This controller is still concerned by the notes from Remark 6.4.4 (on p. 141) that address the coding–decoding scheme of information transmission across the channel.

Thus in particular, this chapter demonstrates that whenever almost sure stabilizability holds, it can be ensured by realistic controllers with bounded (as time progresses) algebraic complexity and memory consumption per step, which are based on classic block coding decoding schemes of communication. This is true irrespective of whether a communication feedback is available.

Remark 6.6.4. As was remarked in Sect. 6.4, the quantity $H(A)$ is the topological entropy of the linear system (6.5.1) by Theorem 2.4.2 (on p. 21). Hence Theorem 6.6.1 can be interpreted as an extension of Theorem 2.5.3 (on p. 26) to the case of a noisy channel.

Remark 6.6.5. In the case of the noiseless channel ($\mathcal{E} = \mathcal{S}$ and $W(e|e) = 1$), Theorem 6.6.1 is also in harmony with Theorem 3.5.2 (on p. 52), where an undelayed and lossless channel is considered.

Comment 6.6.6. In [201], the implication b) $\Rightarrow \mathfrak{c} \geq H(A)$ was proved for channels more general than discrete memoryless. The sufficiency of the inequality $\mathfrak{c} > \eta(A)$ for almost sure stabilizability was justified in [201] for a special discrete memoryless channel: the erasure channel with a finite alphabet.

The comments from Sect. 6.4 on the similarity between Theorem 6.4.1 and the classic Shannon's channel coding theorem equally concern Theorem 6.6.1.

6.7 Necessary Conditions for Observability and Stabilizability

In this section, we prove the d) $\Rightarrow \mathfrak{c} \geq H(A)$ parts of both Theorems 6.4.1 and 6.6.1. So Assumptions 6.3.1 and 6.3.4–6.3.6 (and Assumption 6.5.5 in the case of Theorem 6.6.1) are supposed to hold throughout the section. The proof is accomplished via justifying the following two stronger statements.

Proposition 6.7.1. *Let* $\mathfrak{c} < H(A)$*. Then the state cannot be observed with a bounded error: For any coder* (6.2.3) *and decoder* (6.2.2)*, the following two claims hold:*

(i) The estimation error is almost surely unbounded

$$\overline{\lim_{t\to\infty}} \|x(t) - \widehat{x}(t)\| = \infty \qquad \text{a.s.} \tag{6.7.1}$$

(ii) This divergence is as fast as exponential. Specifically, pick $\alpha > 1$ *so that*

$$\log_2 \alpha < \frac{H(A) - \mathfrak{c}}{\dim(x)}. \tag{6.7.2}$$

Then

$$\overline{\lim_{t\to\infty}} \alpha^{-t} \|x(t) - \widehat{x}(t)\| = \infty \qquad \text{a.s.} \tag{6.7.3}$$

The second statement is similar and concerns the stabilization problem.

Proposition 6.7.2. *Let* $\mathfrak{c} < H(A)$*. Then the plant cannot be stabilized with a bounded error: For any coder* (6.2.3) *and decoder* (6.5.2)*, the following two claims hold:*

(i) The state is almost surely unbounded

$$\overline{\lim_{t\to\infty}} \|x(t)\| = \infty \qquad \text{a.s.} \tag{6.7.4}$$

(ii) This divergence is as fast as exponential. Specifically, pick $\alpha > 1$ *so that* (6.7.2) *holds. Then*

$$\overline{\lim_{t\to\infty}} \alpha^{-t}\|x(t)\| = \infty \qquad \text{a.s.} \tag{6.7.5}$$

Remark 6.7.3. Propositions 6.7.1 and 6.7.2 entail that d) $\Rightarrow \mathfrak{c} \geq H(A)$ in both Theorems 6.4.1 (on p. 140) and 6.6.1 (on p. 145).

For both theorems, the arguments underlying this remark are similar. For the definiteness, we focus on the case of Theorems 6.4.1. Let **d)** be true. Suppose that $\mathfrak{c} < H(A)$. Then by (6.7.3), random times $0 < \tau_1 < \tau_2 < \ldots$ exist such that

$$\|x(\tau_i) - \widehat{x}(\tau_i)\| \geq \alpha^{\tau_i} \quad \text{for all} \quad i \quad \text{a.s.}$$

Then

$$\frac{1}{\tau_i + 1}\sum_{\theta=0}^{\tau_i} \|x(\theta) - \widehat{x}(\theta)\| \geq \frac{\|x(\tau_i) - \widehat{x}(\tau_i)\|}{\tau_i + 1} \geq \frac{\alpha^{\tau_i}}{\tau_i + 1} \to \infty \text{ as } i \to \infty \text{ a.s.,}$$

in violation of d). The contradiction obtained proves that $\mathfrak{c} \geq H(A)$. □

The remainder of the section is devoted to the proofs of Propositions 6.7.1 and 6.7.2. We start by revealing relations between them, as well as their parts (i) and (ii). This will show that the proofs ultimately reduce to proving (i) of Proposition 6.7.1.

6.7.1 Proposition 6.7.2 Follows from Proposition 6.7.1

This is justified by the following simple observation.

Lemma 6.7.4. *Consider a controller consisting of a coder* (6.2.3) *and decoder* (6.5.2)*. Then there exist another coder and decoder-estimator of the form* (6.2.2)

$$e(t) = \mathcal{E}_{un}[t, y_{un}(0), \ldots, y_{un}(t), s(0), \ldots, s(t-1)],$$
$$\widehat{x}_{un}(t) = \mathcal{X}\left[t, s(0), s(1), \ldots, s(t)\right], \tag{6.7.6}$$

which generate an estimate $\widehat{x}_{un}(t)$ *of the state of the uncontrolled system* (6.2.1)

$$x_{un}(t+1) = Ax_{un}(t), \qquad x_{un}(0) = x_0, \qquad y_{un}(t) = Cx_{un}(t) \tag{6.7.7}$$

and produce the estimation error identical to the stabilization error of the original coder–decoder pair:

$$\|x_{un}(t) - \widehat{x}_{un}(t)\| = \|x(t)\|. \tag{6.7.8}$$

Proof. Let in (6.7.6) the decoder generate the estimate via the recursion

$$\widehat{x}_{\text{un}}(t+1) = A\widehat{x}_{\text{un}}(t) - Bu(t), \quad u(t) := \mathcal{U}\left[t, s(0), s(1), \dots, s(t)\right], \quad \widehat{x}_{\text{un}}(0) = 0,$$

where $\mathcal{U}(\cdot)$ is taken from (6.5.2), and let the coder be defined by the formula

$$e(t) := \mathcal{E}[t, y_{\text{un}}(0) - C\widehat{x}_{\text{un}}(0), \dots, y_{\text{un}}(t) - C\widehat{x}_{\text{un}}(t), s(0), \dots, s(t-1)],$$

where $\mathcal{E}(\cdot)$ is taken from (6.2.3). This formula presupposes that the coder also computes the estimate $\widehat{x}_{\text{un}}(t)$.

Now we consider the process $\{x(t), u(t)\}_{t=0}^{\infty}$ generated in the system (6.5.1) by the original coder and decoder. Arguing by induction on t, it is easy to see that, first, both coder–decoder pairs give rise to common sequences $\{e(t)\}$, $\{s(t)\}$, $\{u(t)\}$, and second, $y(t) = y_{\text{un}}(t) - C\widehat{x}_{\text{un}}(t)$ and (6.7.8) does hold. □

Corollary 6.7.5. *Proposition* 6.7.2 *follows from Proposition* 6.7.1.

6.7.2 Relationship between the Statements (i) and (ii) of Proposition 6.7.1

This relation is revealed by the following.

Lemma 6.7.6. *The statement (ii) of Proposition* 6.7.1 *follows from (i).*

Proof. Note that (6.7.3) results from applying (6.7.1) to the process

$$x_*(t) := \alpha^{-t}x(t), \quad \widehat{x}_*(t) := \alpha^{-t}\widehat{x}(t), \quad e(t), \quad s(t).$$

This is possible since the process is generated by (6.2.1), (6.2.2), and (6.2.3), where

$$A := \alpha^{-1}A, \qquad \mathcal{X}_*[t, \cdot] := \alpha^{-t}\mathcal{X}[t, \cdot],$$

and $[y(0), \dots, y(t)]$ is replaced by

$$x_*(0), \alpha x_*(1), \dots, \alpha^t x_*(t).$$

The condition $H(\alpha^{-1}A) > \mathfrak{c}$ holds since

$$H(\alpha^{-1}A) = \sum_{\lambda_j} \max\{\log_2(\alpha^{-1}|\lambda_j|), 0\} = \sum_{\lambda_j} \left[\max\{\log_2|\lambda_j|, \log_2\alpha\} - \log_2\alpha\right]$$
$$\geq \sum_{\lambda_j} \max\{\log_2|\lambda_j|, 0\} - n\log_2\alpha = H(A) - n\log_2\alpha > \mathfrak{c},$$

where $n = \dim(x)$ and the last inequality follows from (6.7.2). □

Corollary 6.7.5 and Lemma 6.7.6 permit us to focus on proving (i) of Proposition 6.7.1. In doing so, we employ the concepts described in the next subsection.

6.7.3 Differential Entropy of a Random Vector and Joint Entropy of a Random Vector and Discrete Quantity

Differential Entropy

To describe a random vector $V \in \mathbb{R}^s = \{v\}$ with the known probability density $p_V(\cdot)$, the infinite number of bits is required. So strictly speaking, its entropy is infinite. At the same time, approximately

$$h(V) + sb + \log_2 \boldsymbol{V}\left(B_0^1\right) \tag{6.7.9}$$

bits suffice to describe $V \in \mathbb{R}^s$ to the b-bit accuracy [40]. Here the quantity $h(V)$ characterizes the vector at hand, is called the *differential entropy* of V, and is defined as

$$h(V) := -\boldsymbol{E} \log_2 p_V(V) = -\int_{\mathbb{R}^s} p_V(v) \log_2 p_V(v)\, dv. \tag{6.7.10}$$

Remark 6.7.7. The differential entropy can take either negative or infinite values.

Remark 6.7.8. If the accuracy is high $b \approx \infty$ and $h(V) \in \mathbb{R}$, the second addend in (6.7.9) dominates the others. At the same time, this addend is common for all random vectors. In view of this, the differential entropy is not so much absolute as a comparative measure of uncertainty. Indeed, $h(V_2) - h(V_1)$ is approximately equal to the difference in the numbers of bits required to describe V_2 and V_1, respectively, to any common accuracy.

Now consider a random quantity $F \in \mathfrak{F}$ assuming a finite number of values $|\mathfrak{F}| < \infty$. Note that the conditional distribution of V given $F = f \in \mathfrak{F}$ has a density $p_V(\cdot|f)$.[7] The *conditional differential entropy* $h_f(V) = h_{F=f}(V)$ is defined as the differential entropy produced by the conditional probability density of V given $F = f$, and the *averaged conditional differential entropy*

$$h(V|F) := \boldsymbol{E} h_F(V) = -\sum_{f \in \mathfrak{F}} \boldsymbol{P}(F = f) \int_{\mathbb{R}^s} p_V(v|f) \log_2 p_V(v|f)\, dv.$$

Explanation 6.7.9. We put $p_V(\cdot|f) \equiv 0$ whenever $\boldsymbol{P}(F = f) = 0$.

Joint Entropy of a Random Vector and Discrete Quantity

Observe first that the following formula is immediate from (6.3.2):

$$H(F, G) = H(G|F) + H(F).$$

[7] Indeed for any measurable set $M \subset \mathbb{R}^s$ and f such that $\boldsymbol{P}(F = f) > 0$, one has $\boldsymbol{P}(V \in M|F = f) = \boldsymbol{P}(F = f)^{-1}\boldsymbol{P}(V \in M \wedge F = f) \leq \boldsymbol{P}(F = f)^{-1}\boldsymbol{P}(V \in M) = \boldsymbol{P}(F = f)^{-1} \int_M p_V(v)\, dv$. So $\boldsymbol{V}(M) = 0 \Rightarrow \boldsymbol{P}(V \in M|F = f) = 0$; i.e., the conditional distribution is absolutely continuous with respect to the Lebesgue measure.

Here both random quantities F, G assume only finitely many values.

This formula can be extended on the case where one of the quantifies is a random vector to define the *joint entropy of a random vector V and discrete quantity F* as

$$H(V, F) := h(V|F) + H(F). \tag{6.7.11}$$

The joint conditional entropy $H_{G=g}(V, F) = H_g(V, F) =: \mathfrak{H}(g)$ is the joint entropy of V and F with respect to the probability given $G = g$, and $H(V, F|G) := \boldsymbol{E}\mathfrak{H}(G)$ is the averaged joint conditional entropy. The conditional mutual information $I_{G=g}(V, F)$ and the averaged conditional mutual information $I(V, F|G)$ are defined likewise.

Some Properties of the Entropy and Mutual Information

Now we list general facts concerning the entropy and information that are required to prove the necessity part of Theorem 6.4.1. In doing so, we suppose that the symbol F (possibly, with indices) stands for random quantities assuming finitely many values, either $\widehat{V} := V$ or $\widehat{V} := (V, F_1)$, $\mathcal{V}$ is a deterministic function, and $h(V) \in \mathbb{R}$.

$$-\infty < H(\widehat{V}|F) \leq H(\widehat{V}) < +\infty; \tag{6.7.12}$$

$$I(V, F) = h(V) - h(V|F) \in \mathbb{R}; \quad I[V, (F, F_1)] = I[V, F] + I[V, F_1|F]; \tag{6.7.13}$$

$$H(V, F_1|F) = h(V|F_1, F) + H(F_1|F) \geq h(V|F); \tag{6.7.14}$$

$$h(V) \leq \frac{s}{2} \log_2 \left(2\pi e \boldsymbol{E}\|V\|^2\right); \tag{6.7.15}$$

$$h(V|F) = h\big[V - \mathcal{V}(F)|F\big]. \tag{6.7.16}$$

Remark 6.7.10. Inequality (6.7.15) expresses the maximizing property of the Gaussian distribution: among all random vectors with zero mean and given variance, those with symmetric Gaussian distribution have the maximum differential entropy [40].

Whenever the random variables V, F_1, F_2 form a Markov chain (i.e., V and F_2 are independent given F_1), the following *data processing inequality* holds:

$$I(V, F_2) \leq I(F_1, F_2). \tag{6.7.17}$$

Justification of the listed properties can be found in many textbooks and monographs on information theory, see, e.g., [38, 40, 50, 60, 65, 68, 152, 188, 190]. At the same time, though the joint entropy (6.7.11) is defined in terms of the classic entropy of a discrete random quantity and the differential entropy of a random vector, it is not as conventional and well studied a tool as the last two kinds of entropy. In view of this, the formal justification of the properties (6.7.12) and (6.7.14) concerning the joint entropy is offered in Appendix D.

6.7.4 Probability of Large Estimation Errors

So far as asymptotic tracking does not concern the stable modes, it seems more or less clear that the proof can be confined to systems with only unstable ones. This fact will be formally justified in Subsect. 6.7.6 (starting on p. 156). It will also be shown that without the loss of generality, one can assume that the initial state is a.s. bounded and has a finite differential entropy. So from this point and until otherwise stated, we adopt one more assumption.

Assumption 6.7.11. *The system* (6.2.1) *has no stable* $|\lambda| < 1$ *eigenvalues* λ. *The initial state* x_0 *has a finite differential entropy. A (deterministic) constant* $b_0 \in (0, \infty)$ *exists such that* $\|x_0\| \leq b_0$ *a.s.*

In this subsection, we show that whenever the capacity (6.3.3) of the channel $\mathfrak{c}$ is less $\mathfrak{c} < H(A)$ than the topological entropy $H(A) = \log_2 |\det A|$ of the system (6.2.1), arbitrarily large estimation errors unavoidably occur with the probability $\approx 1 - \frac{\mathfrak{c}}{H(A)}$ (for large t). Specifically, the following claim holds.

Proposition 6.7.12. *Let* $\mathfrak{c} < H(A)$. *Then for any coder* (6.2.3) *and decoder* (6.2.2)*,*

$$\boldsymbol{P}\Big[\|x(t) - \widehat{x}(t)\| > b\Big] \geq 1 - \frac{\mathfrak{c}}{H(A)} - \frac{1}{t} \times \frac{1 - h(x_0) + \mathfrak{c} + \frac{n}{2}\log_2\left(2\pi e \max\{b^2, b_0^2\}\right)}{H(A)} \quad \forall b > 0, t \geq 1, \tag{6.7.18}$$

where b_0 *is the constant from Assumption* 6.7.11 *and* $n = \dim(x)$.

For any sequence $\{b(t) > 0\}_{t=0}^{\infty}$ *such that* $\frac{\log_2 b(t)}{t} \to 0$ *as* $t \to \infty$*, we have*

$$\varliminf_{t\to\infty} \boldsymbol{P}\Big[\|x(t) - \widehat{x}(t)\| > b(t)\Big] \geq 1 - \frac{\mathfrak{c}}{H(A)}. \tag{6.7.19}$$

Remark 6.7.13. In particular, (6.7.19) holds whenever $b(t)$ is a constant or polynomial in t.

Remark 6.7.14. It is easy to see that (6.7.19) is a simple corollary of (6.7.18).

Remark 6.7.15. Relation (6.7.19) implies that $\mathfrak{c} \geq \eta(A)$ whenever coder and decoder exist that keep the mathematical expectation of the error (or at least the time average $\frac{1}{t}\sum_{\theta=0}^{t-1} \boldsymbol{E}\|x(\theta) - \widehat{x}(\theta)\|$) bounded.

Comment 6.7.16. An inequality similar to (6.7.18) can be obtained from Lemma 3.2 [201].

By means of Lemma 6.7.4, Proposition 6.7.12 can easily be extended on the stabilization problem. This results in the following.

Corollary 6.7.17. *Let* $\mathfrak{c} < H(A)$. *Then for any controller consisting of a coder* (6.2.3) *and decoder* (6.5.2)*, the statements of Proposition* 6.7.12 *remain true, provided that in* (6.7.18) *and* (6.7.19)*,*

$$\boldsymbol{P}\Big[\|x(t)\| > b\Big] \quad \text{and} \quad \boldsymbol{P}\Big[\|x(t)\| > b(t)\Big]$$

are substituted in place of, respectively,

$$\boldsymbol{P}\Big[\|x(t) - \widehat{x}(t)\| > b\Big] \quad \text{and} \quad \boldsymbol{P}\Big[\|x(t) - \widehat{x}(t)\| > b(t)\Big].$$

The remainder of the subsection is devoted to the proof of Proposition 6.7.12. We start with a technical fact. To state it, we denote by $E_0^t := \{e(\theta)\}_{\theta=0}^t$ and $S_0^t := \{s(\theta)\}_{\theta=0}^t$ the total data injected into and received over the channel, respectively, up to time t.

Lemma 6.7.18. *For any coder–decoder pair with a feedback* (6.2.2)*,* (6.2.3)*, the conditional differential entropy* $h\big[x(t)|S_0^t\big]$ *is finite and*

$$h\big[x(t)|S_0^t\big] \geq h\big[x_0\big] + t\big[H(A) - \mathfrak{c}\big] - \mathfrak{c}. \tag{6.7.20}$$

Proof. Due to (6.2.1), the probability density $p_t(\cdot|\theta)$ of $x(t)$ given S_0^θ evolves as follows:

$$p_t(x|\theta) = |\det A|^{-t} \times p_0(A^{-t}x|\theta).$$

By (6.7.10), this implies

$$h\big[x(t)\big|S_0^\theta\big] = h\big[x_0\big|S_0^\theta\big] + t\log_2|\det A|, \tag{6.7.21}$$

where $\log_2|\det A| = H(A)$ due to Assumption 6.7.11.

Since, thanks to Assumption 6.3.4, the random quantities $x_0, e(t), s(t)$ form a Markov chain given S_0^{t-1}, they satisfy (6.7.17):

$$I\big[x_0, s(t)|S_0^{t-1}\big] \leq I\big[e(t), s(t)|S_0^{t-1}\big].$$

Here $I\big[e(t), s(t)|S_0^{t-1}\big] \leq \mathfrak{c}$ by (6.3.3) (on p. 138). So with regard to the second relation from (6.7.13), we get

$$I[x_0, S_0^t] = I\big[x_0, (s(t), S_0^{t-1})\big] = I[x_0, S_0^{t-1}] + I\big[x_0, s(t)|S_0^{t-1}\big] \\ \leq I[x_0, S_0^{t-1}] + \mathfrak{c}.$$

Iterating the obtained inequality yields that

$$I[x_0, S_0^t] \leq \mathfrak{c}(t+1).$$

To complete the proof, we note that

$$I\big[x_0, S_0^t\big] = h[x_0] - h\big[x_0\big|S_0^t\big]$$

by (6.7.13), and employ (6.7.21) with $\theta := t$. □

Proof of Proposition 6.7.12. Pick t and denote by Ω the random event

$$\Omega \equiv \Big\{ \|x(t) - \widehat{x}(t)\| \leq b \Big\},$$

by p its probability, and by $\mathfrak{I}$ its indicator: $\mathfrak{I} = 1$ if Ω holds and $\mathfrak{I} = 0$ otherwise. Then

$$H\left[x(t), \mathfrak{I} \middle| S_0^t\right] \overset{(6.7.14)}{\geq} h\left[x(t) \middle| S_0^t\right] \overset{(6.7.20)}{\geq} h[x_0] - \mathfrak{c} + t\big[H(A) - \mathfrak{c}\big]. \qquad (6.7.22)$$

The random variable $\mathfrak{I}$ takes only two values. So its entropy (given any event) does not exceed 1. Hence

$$H\left[x(t), \mathfrak{I} \middle| S_0^t\right] \overset{(6.7.14)}{=\!=} h\left[x(t) \middle| \mathfrak{I}, S_0^t\right] + H\left[\mathfrak{I} \middle| S_0^t\right] \leq 1 + \sum_{\sigma=0,1} \boldsymbol{P}(\mathfrak{I} = \sigma) h_{\mathfrak{I}=\sigma}\left[x(t) \middle| S_0^t\right].$$

Repeating the arguments underlying (6.7.21) shows that

$$h_{\mathfrak{I}=0}\left[x(t) \middle| S_0^t\right] = h_{\mathfrak{I}=0}\left[x_0 \middle| S_0^t\right] + tH(A).$$

Hence

$$\begin{aligned} H\left[x(t), \mathfrak{I} \middle| S_0^t\right] &\leq 1 + (1-p) h_{\mathfrak{I}=0}\left[x_0 \middle| S_0^t\right] + (1-p)tH(A) + p h_{\mathfrak{I}=1}\left[x(t) \middle| S_0^t\right] \\ &\overset{(6.7.12)}{\leq} 1 + (1-p) h_{\mathfrak{I}=0}\left[x_0\right] + (1-p)tH(A) + p h_{\mathfrak{I}=1}\left[x(t) \middle| S_0^t\right]. \end{aligned}$$

Here $\|x_0\| \leq b_0$ a.s. by Assumption 6.7.11. So

$$h_{\mathfrak{I}=0}\left[x_0\right] \overset{(6.7.15)}{\leq} \frac{n}{2} \log_2 \left[2\pi e \boldsymbol{E}\left(|x_0|^2 \middle| \mathfrak{J} = 0\right)\right] \leq \frac{n}{2} \log_2 \left[2\pi e b_0^2\right].$$

Furthermore,

$$\begin{aligned} h_{\mathfrak{I}=1}\left[x(t) \middle| S_0^t\right] &\overset{(6.2.2),(6.7.16)}{=\!=\!=\!=} h_{\mathfrak{I}=1}\left[x(t) - \widehat{x}(t) \middle| S_0^t\right] \overset{(6.7.12)}{\leq} h_{\mathfrak{I}=1}\left[x(t) - \widehat{x}(t)\right] \\ &\overset{(6.7.15)}{\leq} \frac{n}{2} \log_2 \Big[2\pi e \boldsymbol{E}\big(\underbrace{\|x(t) - \widehat{x}(t)\|^2}_{\leq b^2 \text{ whenever } \Omega \text{ holds}} \big| \Omega\big)\Big] \leq \frac{n}{2} \log_2 \left[2\pi e b^2\right]. \end{aligned}$$

Thus we see that

$$\begin{aligned} H\left[x(t), \mathfrak{I} \middle| S_0^t\right] &\leq 1 + (1-p)tH(A) + \frac{n}{2}\left[(1-p)\log_2\left(2\pi e b_0^2\right) + p \log_2\left(2\pi e b^2\right)\right] \\ &\leq 1 + (1-p)tH(A) + \frac{n}{2} \log_2 \left(2\pi e \max\{b_0^2, b^2\}\right). \end{aligned}$$

By combining this with (6.7.22), we get the following formula:

$$t\Big\{\big[1 - (1-p)\big] H(A) - \mathfrak{c}\Big\} \leq 1 + \frac{n}{2} \log_2 \left(2\pi e \max\{b_0^2, b^2\}\right) - h(x_0) + \mathfrak{c}\ .$$

It clearly implies (6.7.18). Remark 6.7.14 completes the proof. $\square$

6.7.5 Proof of Proposition 6.7.1 under Assumption 6.7.11

By Lemma 6.7.6, it suffices to prove the statement (i). Consideration can evidently be confined to the system with full observation: $y = x, C = I$ in (6.2.1). Suppose to the contrary to (i) that a coder–decoder pair exists that keeps the estimation error bounded with a positive probability. By sacrificing a small probability, the error can be made uniformly bounded: a constant $b > 0$ exists such that

$$\boldsymbol{P}\Big[\|x(t) - \widehat{x}(t)\| \leq b \; \forall t\Big] > 0. \tag{6.7.23}$$

Since $H(A) > \mathfrak{c}$ by the hypotheses of Proposition 6.7.1, it follows from Proposition 6.7.12 that for any $1 > \rho > \frac{\mathfrak{c}}{H(A)}$, a non-random time $\tau_1 > 0$ exists such that

$$\boldsymbol{P}\Big[\|x(t) - \widehat{x}(t)\| \leq b\Big] \leq \rho \qquad \forall t \geq \tau_1.$$

Now we consider the tail of the process

$$x(t), \widehat{x}(t), e(t), s(t), \qquad t \geq \tau_1 + 1$$

in the conditional probability space given that $\|x(\tau_1) - \widehat{x}(\tau_1)\| \leq b$ and $S|_0^{\tau_1} = \boldsymbol{S}$. Here we employ an $\boldsymbol{S} \in \mathfrak{S}^{\tau_1+1}$ such that

$$\boldsymbol{P}\big[\Omega^1_{\boldsymbol{S}}\big] > 0, \qquad \text{where} \quad \Omega^1_{\boldsymbol{S}} := \big\{\|x(\tau_1) - \widehat{x}(\tau_1)\| \leq b \wedge S|_0^{\tau_1} = \boldsymbol{S}\big\}.$$

The initial state $x(\tau_1 + 1) = A^{\tau_1+1}x_0$ of this tail is a.s. bounded and

$$h(x_0) \in \mathbb{R} \xrightarrow{(6.7.12),(6.7.21)} h\big[x(\tau_1 + 1)|\Omega^1_{\boldsymbol{S}}\big] \in \mathbb{R}.$$

At the same time, the above conditioning does not alter the channel (considered for $t > \tau_1$) due to Assumptions 6.3.1 and 6.3.4 (on pp. 136 and 137). The signals

$$\widehat{x}(t), e(t), s(t), \qquad t \geq \tau_1 + 1$$

are still generated by (6.2.2) and (6.2.3), where $\boldsymbol{S}$ and

$$A^{-\tau_1-1}x(\tau_1 + 1), A^{-\tau_1}x(\tau_1 + 1), \ldots, A^{-1}x(\tau_1 + 1)$$

are substituted for

$$\big[s(0), \ldots, s(\tau_1)\big] \quad \text{and} \quad \big[y(0), \ldots, y(\tau_1)\big],$$

respectively. Thus Proposition 6.7.12 can be applied once more. It follows that

$$\boldsymbol{P}\Big[\|x(t) - \widehat{x}(t)\| \leq b|\Omega^1_{\boldsymbol{S}}\Big] \leq \rho \qquad \forall t \geq \tau_2(\boldsymbol{S}).$$

For $\tau_2 := \max_{\boldsymbol{S}} \tau_2(\boldsymbol{S})$, we have

$$\begin{aligned} &\boldsymbol{P}\Big[\big\|x(\tau_2)-\widehat{x}(\tau_2)\big\| \le b \Big| \|x(\tau_1)-\widehat{x}(\tau_1)\| \le b\Big] \\ &\quad = \sum_{\boldsymbol{S}} \boldsymbol{P}\Big[S|_0^{\tau_1} = \boldsymbol{S}\Big| \|x(\tau_1)-\widehat{x}(\tau_1)\| \le b\Big] \boldsymbol{P}\Big[\big\|x(\tau_2)-\widehat{x}(\tau_2)\big\| \le b \Big| \Omega^1_{\boldsymbol{S}}\Big] \\ &\qquad\qquad \le \rho \sum_{\boldsymbol{S}} \boldsymbol{P}\Big[S|_0^{\tau_1} = \boldsymbol{S}\Big| \|x(\tau_1)-\widehat{x}(\tau_1)\| \le b\Big] = \rho. \end{aligned}$$

Now we repeat the above arguments with respect to the tail on $t > \tau_2$ and conditioning given that

$$\|x(\tau_1)-\widehat{x}(\tau_1)\| \le b, \quad \|x(\tau_2)-\widehat{x}(\tau_2)\| \le b, \quad S|_0^{\tau_2} = \boldsymbol{S}.$$

By continuing likewise, we get a sequence $0 < \tau_1 < \tau_2 < \ldots$ such that

$$\begin{aligned} &p_{i+1|1,\ldots,i} := \\ &:= \boldsymbol{P}\Big[\big\|x(\tau_{i+1}) - \widehat{x}(\tau_{i+1})\big\| \le b \Big| \|x(\tau_1) - \widehat{x}(\tau_1)\| \le b, \ldots, \|x(\tau_i) - \widehat{x}(\tau_i)\| \le b\Big] \le \rho \end{aligned}$$

for all i. Hence

$$\begin{aligned} &\boldsymbol{P}\Big[\big\|x(t)-\widehat{x}(t)\big\| \le b \quad \forall t\Big] \le \boldsymbol{P}\Big[\big\|x(\tau_i)-\widehat{x}(\tau_i)\big\| \le b \quad \forall i\Big] \\ &\qquad = \lim_{k\to\infty} \boldsymbol{P}\Big[\big\|x(\tau_i)-\widehat{x}(\tau_i)\big\| \le b \quad \forall i = 1,\ldots,k\Big] \\ &= \lim_{k\to\infty} \boldsymbol{P}\Big[\big\|x(\tau_1)-\widehat{x}(\tau_1)\big\| \le b\Big] \times \prod_{i=2}^{k} p_{i|1,\ldots,i-1} \le \lim_{k\to\infty} \prod_{i=1}^{k} \rho \overset{\rho<1}{=\!=} 0, \end{aligned}$$

in violation of (6.7.23). The contradiction obtained proves (6.7.1). □

6.7.6 Completion of the Proofs of Propositions 6.7.1 and 6.7.2: Dropping Assumption 6.7.11

We do not suppose any longer that Assumption 6.7.11 (on p. 152) holds.

Extension of Proposition 6.7.1 (on p. 147) on systems with both unstable and stable modes is based on the following proposition. To state it, we introduce the invariant subspace M_{unst} of the matrix A related to the unstable part $\sigma^+ := \{\lambda \in \sigma(A) : |\lambda| \ge 1\}$ of its spectrum, and the restriction A_+ of A on M_{unst} viewed as an operator in M_{unst}.

Proposition 6.7.19. *Suppose that some coder* (6.2.3) *and decoder* (6.2.2) *keep the estimation error bounded*[8] *with the probability better than* p *for the primal system* (6.2.1)*. Then such coder and decoder can be constructed for the following system:*

$$x_+(t+1) = A_+x_+(t), \; x_+(t) \in M_{\text{unst}}, \; x_+(0) = x_0^+, \quad y_+(t) = Cx_+(t) \quad (6.7.24)$$

with some initial random vector $x_0^+ \in M_{\text{unst}}$ *that satisfies Assumptions* 6.3.4 *and* 6.3.5 *(on p.* 137*) and is a.s. bounded and has a finite differential entropy.*

[8] See Definition 6.2.1 on p. 135.

Remark 6.7.20. Generally speaking, the system (6.7.24) is considered on a new underlying probability space. However, Assumptions 6.3.1 and 6.3.4–6.3.6 are still true and the channel parameters $W(s|e)$ remain unchanged.

Explanation 6.7.21. Equations (6.7.24) describe the processes in the primal system (6.2.1) starting at $x(0) = x_0^+ \in M_{\text{unst}}$. A certain technical nontriviality of Proposition 6.7.19 comes from the fact that due to Assumption 6.3.5, the probability to start at $x(0) \in M_{\text{unst}}$ is zero (if $M_{\text{unst}} \neq \mathbb{R}^n$). At the same time, the assumptions of the lemma allow the initial coder–decoder pair to produce asymptotically infinite estimation errors with not only zero but also a positive probability. To keep the estimation error bounded for the processes in the system (6.7.24), this pair will be modified.

Remark 6.7.22. Proposition 6.7.1 follows from Lemma 6.7.6 (on p. 149), Proposition 6.7.19, and the fact established in Subsect. 6.7.5.

Indeed, it suffices to prove (i) of Proposition 6.7.1 by Lemma 6.7.6. Suppose to the contrary that (i) is violated: A coder–decoder pair exists that keeps the estimation error bounded with a nonzero probability for the original system (6.2.1). Then by Proposition 6.7.19, such a pair also exists for the auxiliary system (6.7.24). However this contradicts the fact established in Subsect. 6.7.5 since $H(A) = H(A_+)$. The contradiction obtained proves that (i) does hold.

Remark 6.7.23. Proposition 6.7.2 (on p. 148) follows from Proposition 6.7.1 (on p. 147) by Corollary 6.7.5 (on p. 149).

Thus it remains to prove Proposition 6.7.19. We start with a simple computation.

Lemma 6.7.24. *Suppose that Assumptions* 6.3.1 *and* 6.3.4 *(on pp.* 136 *and* 137*) hold and that a decoder* (6.2.2) *and a feedback coder* (6.2.3) *are taken. Then the joint distribution of the variables* $x_0, E_0^t = (e_0, \dots, e_t)$, *and* $S_0^t = (s_0, \dots, s_t)$ *is given by*

$$\boldsymbol{P}[dx, dS_0^t, dE_0^t] = \prod_{j=0}^{t} W(s_j|e_j)\delta\Big[e_j, \mathcal{E}_x\big(j, x, S_0^{j-1}\big)\Big] ds_j de_j \boldsymbol{P}_0(dx), \quad (6.7.25)$$

where $\mathcal{E}_x(\cdot)$ *is obtained from the right-hand side of the coder equation* (6.2.3)*:*

$$\mathcal{E}_x[t, x_0, S_0^{t-1}] := \mathcal{E}[t, Cx_0, \dots, CA^t x_0, S_0^{t-1}] \ \left(\overset{(6.2.1),(6.2.3)}{=\!=\!=} e(t)\right). \quad (6.7.26)$$

Furthermore, $\boldsymbol{P}_0(dx)$ *is the probability distribution of* x_0, *and* $\delta(e, e') := 1$ *if* $e = e'$ *and* $\delta(e, e') := 0$ *otherwise.*

Proof. The proof will be by induction on t. For $t = 0$,

$$\boldsymbol{P}[dx, dS_0^t, dE_0^t] = \boldsymbol{P}[dx, ds_0, de_0] = \boldsymbol{P}[dx, ds_0|e_0]\boldsymbol{P}(de_0) \overset{\text{Assumptions 6.3.1, 6.3.4}}{=\!=\!=\!=} \boldsymbol{P}[dx|e_0]\boldsymbol{P}[ds_0|e_0]\boldsymbol{P}(de_0) \overset{\text{Assumption 6.3.1}}{=\!=\!=\!=}$$

$$= W(s_0|e_0)\boldsymbol{P}[dx, de_0]ds_0 \overset{(6.7.26)}{=\!=\!=} W(s_0|e_0)\delta\big[e_0, \mathcal{E}_x(0, x)\big] ds_0 de_0 \boldsymbol{P}_0(dx);$$

i.e., (6.7.25) does hold for $t = 0$. Suppose that it holds for some $t = 0, 1, \ldots$. Then

$$\begin{aligned}\boldsymbol{P}\left[dx, dS_0^{t+1}, dE_0^{t+1}\right] &= \boldsymbol{P}\left[dx, dS_0^t, ds_{t+1}, dE_0^t, de_{t+1}\right] \\ &= \boldsymbol{P}\left[dx, dS_0^t, ds_{t+1}, dE_0^t \middle| e_{t+1}\right] \boldsymbol{P}\left[de_{t+1}\right] \\ &\overset{\text{Assumptions 6.3.1, 6.3.4}}{=\!=\!=} \boldsymbol{P}\left[dx, dS_0^t, dE_0^t \middle| e_{t+1}\right] \boldsymbol{P}\left[ds_{t+1} \middle| e_{t+1}\right] \boldsymbol{P}\left[de_{t+1}\right] \\ &= W(s_{t+1}|e_{t+1}) \boldsymbol{P}\left[dx, dS_0^t, dE_0^t, de_{t+1}\right] ds_{t+1} \\ &\overset{(6.7.26)}{=\!=} W(s_{t+1}|e_{t+1}) \delta\left[e_{t+1}, \mathcal{E}_x(t+1, x, S_0^t)\right] \boldsymbol{P}\left[dx, dS_0^t, dE_0^t, \right] ds_{t+1} de_{t+1}.\end{aligned}$$

This and the induction hypothesis show that (6.7.25) does hold for $t = t + 1$. □

Corollary 6.7.25. *Given a coder and a decoder-estimator, we denote by Ω the random event of keeping the error bounded (see the first formula in* (6.2.6) *on p.* 135*). The conditional probability of this event given $x(0) = x$ can be chosen so that it does not depend on the distribution of the initial state $x(0)$ provided Assumption* 6.3.4 *holds. This is true irrespective of whether this distribution has a probability density.*

Indeed thanks to Lemma 6.7.24, the conditional distribution

$$\begin{aligned}\boldsymbol{P}\left[dS|_0^t \middle| x(0) = x\right] &= \sum_{E|_0^t} \prod_{j=0}^t ds_j W(s_j|e_j) \delta\left[e_j, \mathcal{E}_x(j, x, S|_0^{j-1})\right] \\ &= \prod_{j=0}^t ds_j W\left[s_j | \mathcal{E}_x(j, x, S|_0^{j-1})\right]\end{aligned}$$

does not depend on the distribution of the initial state. The proof is completed by observing that

$$\boldsymbol{P}(\Omega|x) = \lim_{k\to\infty} \lim_{l\to\infty} \int_{\left\{S_0^l : |A^t x - \mathcal{X}(t, S|_0^t)| < k \, \forall t = 0, \ldots, l\right\}} \boldsymbol{P}\left[dS|_0^l \middle| x(0) = x\right],$$

where $\mathcal{X}(\cdot)$ is the function from (6.2.2) (on p. 134).

We proceed with the following simple observation.

Lemma 6.7.26. *Any random vector $V \in \mathbb{R}^s$ with a bounded probability density $p_V(v) \le p_\infty$ and a finite variance has a finite differential entropy $h(V) \in \mathbb{R}$.*

Proof. The inequality $h(V) < \infty$ follows from from (6.7.15). At the same time,

$$h(V) = -\boldsymbol{E} \log_2 p_V(V) \ge -\log_2 p_\infty > -\infty. \qquad \square$$

Proof of Proposition 6.7.19. Consider a coder (6.2.3) and a decoder (6.2.2) that keep the estimation error bounded with the probability better than p for the primal system (6.2.1). By invoking the notation $p_0(\cdot)$ from Assumption 6.3.5 (on p. 137) and Ω from Corollary 6.7.25 and putting

$$Q_{>p} := \{x \in \mathbb{R}^n : \boldsymbol{P}(\Omega|x) > p\},$$

we get

$$p < \boldsymbol{P}(\Omega) = \int_{\mathbb{R}^n} \boldsymbol{P}(\Omega|x) p_0(x)\, dx \Rightarrow \int_{Q_{>p}} p_0(x)\, dx = \boldsymbol{P}(x_0 \in Q_{>p}) > 0 \Rightarrow$$
$$\Rightarrow \exists c > 0 : \boldsymbol{P}(x_0 \in Q) > 0, \quad \text{where} \quad Q := \{x \in Q_{>p} : p_0(x) \leq c\}.$$

Then a compact subset $\overset{\circ}{Q} \subset Q$ exists such that [57, Sec.134Fb]

$$\boldsymbol{P}[x_0 \in \overset{\circ}{Q}] = \int_{\overset{\circ}{Q}} p_0(x)\, dx > 0.$$

Now we pass to the probability space related to the probability given $x_0 \in \overset{\circ}{Q}$. This evidently keeps Assumptions 6.3.1 and 6.3.4–6.3.6 true and the channel parameters $W(s|e)$ unchanged. We assume that all random variables inherit their initial notations. Note also that in the new probability space, the initial vector x_0 is a.s. bounded and has a bounded density.

Now we denote by M_{st} the invariant subspace of A related to the stable part $\sigma^- := \{\lambda \in \sigma(A) : |\lambda| < 1\}$ of its spectrum $\sigma(A)$. We also introduce the projector π_+ onto M_{unst} parallel to M_{st} and the compact set $Q_+ := \pi_+ \overset{\circ}{Q} \subset M_{\text{unst}}$, and we define the initial vector in (6.7.24) to be $x_0^+ := \pi_+ x_0$. This vector evidently has a bounded probability density,

$$x_0^+ \in Q_+ \quad \text{almost surely,} \tag{6.7.27}$$

and so the second moment of x_0 is finite. Then Lemma 6.7.26 yields $h(x_0^+) \in \mathbb{R}$.

The multivalued function

$$x_+ \in Q_+ \mapsto \mathfrak{V}(x_+) := \{x_- \in L_- : x_+ + x_- \in \overset{\circ}{Q}\}$$

has a closed graph $\overset{\circ}{Q}$ and so is upper-hemicontinuous. Thus there exists a single-valued measurable selector [217, Sec.I.7]

$$x_+ \in Q_+ \mapsto \chi_-(x_+) \in \mathfrak{V}(x_+).$$

By extending this selector as a measurable function on M_{unst} and putting

$$\chi(x_+) := x_+ + \chi_-(x_+),$$

we get

$$x_+ \in Q_+ \Rightarrow \chi(x_+) \in \overset{\circ}{Q} \subset Q \Rightarrow \boldsymbol{P}\big[\Omega|\chi(x_+)\big] > p. \tag{6.7.28}$$

Now we are in a position to transform the original coder–decoder pair (6.2.3), (6.2.2) serving the primal system into that keeping the estimation error bounded for the auxiliary one (6.7.24). We note first that the system (6.7.24) is observable thanks to Assumption 6.3.6 (on p. 137). So for any $t \geq n-1$, there exists a *deadbeat observer*, i.e., a linear transformation

$$[y_+(0), \ldots, y_+(t)] \overset{\mathcal{S}_t}{\mapsto} x_+(0),$$

where $y_+(i)$ are taken from (6.7.24). We define a new coder and decoder as follows. For $t = 0, \ldots, n-1$, they in fact do nothing. However, for the sake of definiteness, we pick $e_* \in \mathfrak{E}$ and put

$$\mathcal{E}_+[t, y(0), \ldots, y(t), S_0^{t-1}] := e_*, \qquad \mathcal{X}_+[t, S_0^t] := 0.$$

For $t \geq n$, the new coder and decoder act as follows:

$$\omega = \left[y_+(0), \ldots, y_+(t), S_0^{t-1}\right] \overset{\mathcal{S}_t}{\mapsto} \left[x_+(0), S_0^{t-1}\right] \mapsto \mathcal{E}_+[t, \omega]$$
$$:= \mathcal{E}\left\{t-n, C\chi[x_+(0)], \ldots, CA^{t-n}\chi[x_+(0)], S_n^{t-1}\right\},$$
$$\widehat{x}_+(t) := \mathcal{X}_+\left[t, S_0^t\right] := \pi_+ A^n \mathcal{X}\left[t-n, S_n^t\right].$$

Now consider the process

$$\xi(t) = [x(t), y(t), e(t), s(t), \widehat{x}(t)], \qquad t = 0, 1, \ldots$$

generated by the original coder–decoder pair in the primal system (6.2.1) when started with the initial random state $\chi[x_0^+]$. It is easy to see that

$$\pi_+ x(t), \quad y_+(t) = C\pi_+ x(t), \quad e(t-n), \quad s(t-n), \quad \pi_+ A^n \widehat{x}(t-n)$$

is a process generated by the new coder and decoder in the auxiliary system (6.7.24). Here $\widehat{x}(t) := 0, e(t) := e_*$ for $t < 0$, and $s(-n), \ldots, s(-1)$ are mutually independent and independent of $\xi(t), t = 0, 1, \ldots$ random quantities each with the distribution $W(s|e_*)$. Hence for $t \geq n$,

$$\|x_+(t) - \widehat{x}_+(t)\| = \|\pi_+ x(t) - \pi_+ A^n \widehat{x}(t-n)\| \leq \|\pi_+\| \|x(t) - A^n \widehat{x}(t-n)\|$$
$$= \|\pi_+\| \|A^n x(t-n) - A^n \widehat{x}(t-n)\| \leq \|\pi_+\| \|A^n\| \|x(t-n) - \widehat{x}(t-n)\|.$$

So for the new coder–decoder pair and the system (6.7.24), the probability of keeping the estimation error bounded is no less than that for the process $\xi(t), t = 0, 1, \ldots$. The proof is completed by noting that the latter is given by

$$\boldsymbol{P}(\Omega) = \boldsymbol{EP}[\Omega|\chi(x_0^+)] \xrightarrow{(6.7.27),(6.7.28)} \boldsymbol{P}(\Omega) > p. \qquad \square$$

6.8 Tracking with as Large a Probability as Desired: Proof of the $\mathfrak{c} > H(A) \Rightarrow$ b) Part of Theorem 6.4.1

In this section, we suppose that the assumptions of Theorem 6.4.1 (on p. 140) hold and $\mathfrak{c} > H(A)$. The objective is to construct coder–decoder pairs that ensure b) of Theorem 6.4.1. In doing so, one more assumption is adopted until otherwise stated.

Assumption 6.8.1. *The system* (6.2.1) *has no stable* $|\lambda| < 1$ *eigenvalues* λ.

In the general case, a tracking observer will be constructed in Sect. 6.10 by applying that presented below to the unstable part of the system, like in Subsect. 3.8.11 (starting on p. 88).

Observation 6.8.2. *Due to Assumption* 6.8.1*,* $H(A) = \log_2 |\det A|$ *and so the conidition* $\mathfrak{c} > H(A)$ *takes the form*

$$\log_2 |\det A| < \mathfrak{c}. \tag{6.8.1}$$

We start with preliminaries concerning some well known facts and constructions from the classic information theory (see, e.g., [37, 38, 50, 59, 190]).

6.8.1 Error Exponents for Discrete Memoryless Channels

Block Code

A *block code* with *block length* r is a finite number N of the channel input code words, each of length r,

$$E^1, \dots, E^N, \qquad E^i = (e_0^i, \dots, e_{r-1}^i), \qquad e_j^i \in \mathfrak{E}. \tag{6.8.2}$$

This code is used to notify the recipient which choice of N possibilities, labeled by i, is taken by the informant by sending the corresponding word E^i across the channel. The average number of bits per channel use that can be communicated in this way

$$R := \frac{\log_2 N}{r} \tag{6.8.3}$$

is called the *rate* of the code. The *decoding rule* is a method to associate a unique i with any output word of length r, that is, a map $\mathcal{D}_r : \mathfrak{S}^r \to [1 : N]$, where

$$\mathfrak{S}^r = \left\{ S = (s_0, \dots, s_{r-1}) : s_j \in \mathfrak{S} \ \forall j \right\}$$

is the set of all such output words. The probability of incorrect decoding given that the word E^i is sent over the channel is as follows:

$$\mathrm{err}_i := \boldsymbol{P}\left[\mathcal{D}_r(S) \neq i \middle| E^i \right] = \sum_{S : \mathcal{D}_r(S) \neq i} \prod_{j=0}^{r-1} W(s_j | e_j^i), \tag{6.8.4}$$

where $W(s|e)$ are the channel transition probabilities from Assumption 6.3.1 (on p. 136). The maximum probability of error is given by

$$\mathrm{ERR} = \mathrm{ERR}\left[E^1, \dots, E^N, \mathcal{D}_r(\cdot), r\right] := \max_{i=1,\dots,N} \mathrm{err}_i. \tag{6.8.5}$$

Error Exponent

One of the basic results of the Shannon theory [188, 190, 214] is the following.

Theorem 6.8.3. *For any probability value* $p \in (0,1)$ *and block length* r*, let* $R(r,p) := \max R$ *denote the maximum rate* (6.8.3) *achievable over block codes with block length* r *and the error probability* ERR $\leq p$*. Then*

$$\lim_{r\to\infty} R(r,p) = \mathfrak{c} \qquad \forall p \in (0,1),$$

where $\mathfrak{c}$ *is the capacity* (6.3.3) *of the channel.*

This yields that it is possible to send information at any rate $R < \mathfrak{c}$ through the channel with as small a probability of error as desired by means of a proper block code. This claim is not true if $R > \mathfrak{c}$.

In the first case $R < \mathfrak{c}$, the error probability decreases exponentially as a function of the block length. For a rigorous statement, we employ the symbols $\lesssim$ and $\approx$ introduced in Notation 3.8.32 (on p. 78). The following result is straightforward from, e.g., Lemma IV.1 and Theorem IV.1 [37] (see also [50, 52, 59]).

Theorem 6.8.4. *For any* $0 < R < \mathfrak{c}$ *and* $r = 1, 2, \ldots,$ *there exist* $N \approx 2^{rR}$ *input code words* (6.8.2) *and a decoding rule* $\mathcal{D}_r(\cdot)$ *such that the maximum probability of error* (6.8.5) *obeys the bound*

$$\text{ERR} \lesssim 2^{-rF(R,W)}, \qquad F(R,W) > 0. \tag{6.8.6}$$

Here $F(R,W)$ *is independent of* r *but depends on the rate* R *and the channel* W.

6.8.2 Coder–Decoder Pair without a Communication Feedback

Now we introduce a coder–decoder pair (6.2.4), (6.2.2) that underlies the $\mathfrak{c} > H(A) \Rightarrow$ b) part of Theorem 6.4.1 (on p. 140). This pair resembles that from Subsect. 3.8.4. In particular, it employs a contracted quantizer.[9] The major difference is that now the operation epochs $[\tau_i : \tau_{i+1})$ of the observer are of increasing duration: $\tau_{i+1} - \tau_i = r_i \to \infty$ as $i \to \infty$, whereas this duration was constant in Subsect. 3.8.4.

Remark 6.8.5. This difference will be discarded in the case where a communication feedback is available (see Subsect. 6.9.1 starting on p. 168).

To construct an observer, we pick

1) two numbers η and R such that

$$\log_2 |\det A| = H(A) < \eta < R < \mathfrak{c};$$

2) a parameter $\gamma > \|A\|$;
   and then for any $r = 1, 2, \ldots$, we choose

[9] See Definition 3.8.7 on p. 70, where $A_{ss} := A$.

3) a code book $\mathfrak{C}_r$ with $N = N'_r \approx 2^{rR}$ input code words (6.8.2) each of length r and a decoding rule $\mathcal{D}_r(\cdot)$ with the properties described in Theorem 6.8.4; and
4) an r-contracted quantizer $\mathfrak{Q}_r$ from Lemma 3.8.36 (on p. 80) applied to $A_{ss} := A$.[10].

Explanation 6.8.6. Inequality (6.8.1) makes 1) possible.

Explanation 6.8.7. In 2), $\|A\| = \max_{x:\|x\|=1} \|Ax\|$ is the matrix norm associated with the vector norm $\|\cdot\|$ from Definition 3.8.6 (on p. 69).

Observation 6.8.8. *Whenever r is large enough $r \geq r_*$, the quantizer outputs including the alarm signal ✠ can be encoded by code words from the code book $\mathfrak{C}_r$.*

Indeed, this holds whenever $N''_r + 1 \leq N'_r$, where N''_r denotes the number of the quantizer levels. By Lemma 3.8.36, $N''_r \lesssim 2^{r\eta}$, whereas $N'_r \approx 2^{rR}$ and $\eta < R$. It follows that the inequality $N''_r + 1 \leq N'_r$ does hold for $r \approx \infty$.

Finally, we introduce the sequence of integers

$$r_i := i + r_0 \qquad i = 0, 1, \ldots, \tag{6.8.7}$$

where

$$r_0 \geq \max\{n, r_*\} \tag{6.8.8}$$

is an integer parameter of the observer and r_* is taken from Observation 6.8.8.

Description of the Coder and Decoder

Both coder and decoder compute their own estimates $\widehat{x}_c(t), \widehat{x}_d(t)$ and bounds for the estimate exactness $\delta_c(t), \delta_d(t)$, respectively. Initially, they are given common and arbitrarily chosen values of

$$\widehat{x}_c(0) = \widehat{x}_d(0) = \widehat{x}_0 \quad \text{and} \quad \delta_c(0) = \delta_d(0) = \delta_0 > 0.$$

Remark 6.8.9. The inequality $\delta_0 \geq \|\widehat{x}_0 - x(0)\|$ may be violated.

At any time t, both coder and decoder compute the next estimates and the bounds via the recursions

$$\widehat{x}_c(t+1) := A\widehat{x}_c(t), \quad \widehat{x}_d(t+1) := A\widehat{x}_d(t),$$
$$\delta_c(t+1) := \delta_c(t), \quad \delta_d(t+1) := \delta_d(t). \tag{6.8.9}$$

However, at the times $t = \tau_i$, where

$$\tau_i := r_0 + \cdots + r_{i-1} = i \cdot r_0 + \frac{i(i-1)}{2} \tag{6.8.10}$$

and r_i is given by (6.8.7), they preface (6.8.9) by the following operations.

The coder (at the times $t = \tau_i, i = 1, 2, \ldots$)

[10]We also invoke Remark 3.8.37 on p. 80 here.

c.1) Proceeding from the previous measurements, calculates the current state $x(\tau_i)$;
c.2) Employs the quantizer $\mathfrak{Q}_{r_i}$ and computes the quantized value $q_c(\tau_i)$ of the current scaled estimation error

$$\varepsilon(\tau_i) := \left[\delta_c(\tau_i)\right]^{-1}\left[x(\tau_i) - \widehat{x}_c(\tau_i)\right]; \tag{6.8.11}$$

c.3) Encodes the quantized value $q_c(\tau_i)$ by means of the code book $\mathfrak{E}_{r_i}$. The obtained code word of length r_i is transmitted across the channel during the next operation epoch $[\tau_i : \tau_{i+1})$;
c.4) Finally, corrects the estimate and then the exactness bound

$$\widehat{x}_c(\tau_i) := \widehat{x}_c(\tau_i) + \delta_c(\tau_i)\stackrel{\star}{q}_c(\tau_i), \qquad \delta_c(\tau_i) := \delta_c(\tau_i) \times \left(\langle q_c(\tau_i)\rangle_{\varkappa,\gamma}\right)^{r_i},$$

$$\text{where} \quad \stackrel{\star}{q} := \begin{cases} q & \text{if } q \neq \maltese \\ 0 & \text{otherwise} \end{cases}, \qquad \langle q\rangle_{\varkappa,\gamma} := \begin{cases} \varkappa & \text{if } q \neq \maltese \\ \gamma & \text{otherwise} \end{cases}, \tag{6.8.12}$$

and $\varkappa \in (0,1)$ is the parameter from Lemma 3.8.36 (on p. 80).

Only after this, the coder performs the computations in accordance with (6.8.9).

Explanation 6.8.10. The step **c.1)** is possible since the system (6.2.1), which has no stable modes by Assumption 6.8.1, is observable thanks to Assumption 6.3.6 (on p. 137).

Remark 6.8.11. We recall that the quantized value of any vector outside the unit ball is the alarm symbol $\maltese$.

Observation 6.8.12. *The above coder does not employ the communication feedback, i.e., has the form* (6.2.4) *(on p.* 135*).*

The decoder (at times $t = \tau_i, i = 2, 3, \ldots$)

d.1) Applies the decoding rule $\mathcal{D}_{r_{i-1}}$ to the data received within the previous operation epoch $[\tau_{i-1} : \tau_i)$ and thus computes the decoded value $q_d(\tau_i)$ of the quantized and scaled estimation error $q_c(\tau_{i-1})$;
d.2) Then corrects successively the estimate and the exactness bound

$$\widehat{x}_d(\tau_i) := \widehat{x}_d(\tau_i) + \delta_d(\tau_i)A^{r_{i-1}}\stackrel{\star}{q}_d(\tau_i),$$
$$\delta_d(\tau_i) := \delta_d(\tau_i) \times \left(\langle q_d(\tau_i)\rangle_{\varkappa,\gamma}\right)^{r_{i-1}}. \tag{6.8.13}$$

Only after this does it perform the computations from (6.8.9).

Remark 6.8.13. The quantized valued $q_d(\tau_i)$ determined by the decoder may be incorrect $q_d(\tau_i) \neq q_c(\tau_{i-1})$ because of communication errors.

Remark 6.8.14. Instead of multiplying $\delta_c(\tau_i)$ by $\varkappa^{r_i}$ or γ^{r_i} at the time τ_i with keeping $\delta_c(t)$ constant during the next operation epoch $[\tau_i : \tau_{i+1})$, the coder can constantly multiply $\delta_c(t)$ by $\varkappa$ or γ at each step. Computing the large power $A^{r_{i-1}}$ employed in (6.8.13) can be distributed over the epoch $[\tau_{i-1} : \tau_i)$ in the same way.

Remark 6.8.15. The hint from the previous remark cannot be directly extended on the decoder and quantity δ_d since the decoder becomes aware of the multiplier ($\varkappa$ or γ) only at the end of the current epoch $[\tau_{i-1} : \tau_i)$. However, the decoder can perform both computations and at time τ_i choose between $\delta_d(\tau_{i-1})\varkappa^{r_{i-1}}$ and $\delta_d(\tau_{i-1})\gamma^{r_{i-1}}$.

Remark 6.8.16. To communicate information across the channel, the proposed coder–decoder pair employs block codes with increasing block lengths r_i. At the same time, the estimate $\widehat{x}(t)$ of the current state $x(t)$ is produced at the current time t.

6.8.3 Tracking with as Large a Probability as Desired without a Communication Feedback

Now we show that the above coder–decoder pair fits to track the state with as large a probability as desired.

Proposition 6.8.17. *For the arbitrary probability value $p \in (0,1)$, the coder–decoder pair described in Subsect.* 6.8.2 *tracks* (6.2.5) *the state with probability p or better provided that the parameter r_0 from* (6.8.7) *is large enough $r_0 \geq \overline{r}_0(p)$.*

Corollary 6.8.18. *Suppose that Assumption* 6.8.1 *holds and that $H(A) < \mathfrak{c}$. Then the statement* b) *from Theorem* 6.4.1 *(on p.* 140*) is true.*

The remainder of the subsection is devoted to the proof of Proposition 6.8.17.

The values of the quantities $\widehat{x}_d, \widehat{x}_c, \delta_c, \delta_d$ before and after the update at the time τ_i are marked by $^-$ and $^+$, respectively . We start with the following key fact.

Lemma 6.8.19. *At any event where the decoder always decodes the data correctly*

$$q_d(\tau_i) = q_c(\tau_{i-1}) \qquad \forall i \geq 2,$$

the coder–decoder pair ensures asymptotic tracking (6.2.5) *(on p.* 135*). Furthermore*

$$\left\| x(\tau_i) - \widehat{x}_c^-(\tau_i) \right\| \leq K\varkappa^{\tau_i}, \qquad i = 1, 2, \ldots, \tag{6.8.14}$$

where the constant K does not depend on i (but may depend on the event).

Proof. We start with showing that an index $i = 1, 2, \ldots$ exists for which

$$\|\boldsymbol{\varepsilon}(\tau_i)\| \leq 1, \tag{6.8.15}$$

where the scaled error $\boldsymbol{\varepsilon}(\tau_i)$ is defined in (6.8.11). Indeed otherwise,

$$q_c(\tau_i) = \maltese, \quad \delta_c^-(\tau_{i+1}) = \delta_c^-(\tau_i)\gamma^{r_i} \quad \forall i \geq 1; \quad \widehat{x}_c(t+1) = A\widehat{x}_c(t) \quad \forall t.$$

So for $i \geq 2$, we have

$$\begin{aligned}\|\boldsymbol{\varepsilon}(\tau_i)\| = \left[\gamma^{\sum_{j=1}^{i-1} r_j}\delta_0\right]^{-1} \left\| A^{\tau_i}\left[x_0 - \widehat{x}_0\right] \right\| \overset{(6.8.10)}{=\!=} \frac{\gamma^{r_0}}{\delta_0}\gamma^{-\tau_i} \left\| A^{\tau_i}\left[x_0 - \widehat{x}_0\right] \right\| \\ \leq \left(\frac{\|A\|}{\gamma}\right)^{\tau_i} \gamma^{r_0}\frac{\|x_0 - \widehat{x}_0\|}{\delta_0} \xrightarrow{\gamma > \|A\|} 0 \qquad \text{as} \quad i \to \infty\end{aligned}$$

in violation of the hypothesis $\|\varepsilon(\tau_i)\| > 1 \, \forall i$. Thus (6.8.15) does hold for some i.

Now consider an index i such that (6.8.15) holds. Then (6.8.15) is still true for $i := i+1$. Indeed

$$\begin{aligned} \|\varepsilon(\tau_{i+1})\| &\overset{(6.8.11)}{=\!=} \left[\delta_c^-(\tau_{i+1})\right]^{-1} \left\|x(\tau_{i+1}) - \widehat{x}_c^-(\tau_{i+1})\right\| \\ &\overset{(6.2.1),(6.8.10),(6.8.12)}{=\!=\!=\!=} \varkappa^{-r_i}\left[\delta_c(\tau_i)\right]^{-1} \times \left\| A^{r_i} x(\tau_i) - A^{r_i}\left[\widehat{x}_c^-(\tau_i) + \delta_c^-(\tau_i) q_c(\tau_i)\right]\right\| \\ &= \varkappa^{-r_i}\left\| A^{r_i}\Big\{ \underbrace{\delta_c^-(\tau_i)]^{-1}\left[x(\tau_i) - \widehat{x}_c^-(\tau_i)\right]}_{v} - q_c(\tau_i)\Big\}\right\|. \end{aligned}$$

Here $q_c(\tau_i)$ is the quantized value of the vector v. So far as the quantizer is taken from Lemma 3.8.36 (on p. 80), it is r_i-contracted with the contraction rate $\varkappa^{2r_i}$. So (3.8.9) (on p. 70) yields

$$\|\varepsilon(\tau_{i+1})\| \le \varkappa^{r_i} < 1; \tag{6.8.16}$$

i.e., (6.8.15) does hold for $i := i+1$.

It follows that (6.8.15) is true for all $i \geq \bar{\imath}$, where $\bar{\imath}$ is large enough. By (6.8.12),

$$\delta_c^-(\tau_i) = \delta_c^-(\tau_{\bar{\imath}})\varkappa^{\sum_{j=\bar{\imath}}^{i-1} r_j}.$$

We proceed by taking into account (6.8.11) and (6.8.15)

$$\left\|x(\tau_i) - \widehat{x}_c^-(\tau_i)\right\| \le \delta_c^-(\tau_i) = \overline{\delta}\varkappa^{\sum_{j=0}^{i-1} r_j} \overset{(6.8.10)}{=\!=} \overline{\delta}\varkappa^{\tau_i},$$

$$\text{where} \quad \overline{\delta} := \delta_c(\tau_{\bar{\imath}})\varkappa^{-\sum_{j=0}^{\bar{\imath}-1} r_j}.$$

This evidently implies (6.8.14) and shows that the coder tracks the state.

As for the decoder, observe that

$$\widehat{x}_d^+(\tau_i) = \widehat{x}_c^-(\tau_i), \qquad i = 1, 2, \ldots. \tag{6.8.17}$$

Indeed for $i = 1$, this relation is evident. Suppose that this relation is true for some $i \geq 1$. Due to the absence of transmission errors,

$$\delta_d^{\pm}(\tau_j) = \delta_c^{\pm}(\tau_{j-1}), \qquad j = 2, 3, \ldots.$$

So

$$\begin{aligned} \widehat{x}_d^+(\tau_{i+1}) &\overset{(6.8.13)}{=\!=} \widehat{x}_d^-(\tau_{i+1}) + \delta_d^-(\tau_{i+1}) A^{r_i} \overset{\star}{q}_d(\tau_{i+1}) \\ &\overset{(6.8.9)}{=\!=} A^{r_i}\widehat{x}_d^+(\tau_i) + \delta_c^-(\tau_i) A^{r_i} \overset{\star}{q}_c(\tau_i) \overset{(6.8.17)}{=\!=} A^{r_i}\left[\widehat{x}_c^-(\tau_i) + \delta_c^-(\tau_i) \overset{\star}{q}_c(\tau_i)\right] \\ &\overset{(6.8.12)}{=\!=} A^{r_i}\widehat{x}_c^+(\tau_i) \overset{(6.8.9)}{=\!=} \widehat{x}_c^-(\tau_{i+1}); \end{aligned}$$

i.e., (6.8.17) holds for $i := i+1$. Thus this relation is true for all $i \geq 1$.

Whenever $\tau_i < t \leq \tau_{i+1}$, we have by (6.8.9)

$$\left\|x(t) - \widehat{x}_d(t)\right\| = \left\|A^{t-\tau_i}\left[x(\tau_i) - \widehat{x}_d^+(\tau_i)\right]\right\| \overset{(6.8.17)}{\leq} \|A\|^{t-\tau_i}\left\|x(\tau_i) - \widehat{x}_c^-(\tau_i)\right\|,$$

$$\max_{\tau_i < t \leq \tau_{i+1}} \left\|x(t) - \widehat{x}_d(t)\right\| \overset{(6.8.14)}{\leq} K\|A\|^{r_i} \varkappa^{\tau_i} = K 2^{r_i \log_2 \|A\| + \tau_i \log_2 \varkappa}$$

$$\overset{(6.8.7),(6.8.10)}{=\!=\!=} K 2^{(i+r_0)\log_2 \|A\| + \left[i \cdot r_0 + \frac{i(i-1)}{2}\right] \log_2 \varkappa}. \quad (6.8.18)$$

So far as $\log_2 \varkappa < 0$, this maximum converges to 0 as $i \to \infty$; i.e., (6.2.5) (on p. 135) does hold with $\widehat{x}(t) := \widehat{x}_d(t)$. □

Now we show that the assumption of Lemma 6.8.19 holds with high probability provided the parameter r_0 in (6.8.7) is large enough.

Lemma 6.8.20. *The probability $\boldsymbol{p}_{\text{err}}$ that the decoder decodes at least one message incorrectly does not exceed*

$$\boldsymbol{p}_{\text{err}} \leq K_{R,W,F} 2^{-r_0 F}.$$

Here the constant $K_{R,W,F}$ does not depend on r_0 and the inequality holds with any $F \in (0, F(R,W))$, where $F(R,W)$ is taken from (6.8.6)*.*

Proof. Denote by E_i and S_i the messages of length r_{i-1} formed by the coder at time τ_{i-1} and received by the decoder at time τ_i, respectively. For simplicity of notations, we assume that the map $\mathcal{D}_r(\cdot)$ from Theorem 6.8.4 takes values directly in the input code book. The symbol $\boldsymbol{p}_{\text{err}}(i)$ stands for the probability that decoding of S_i is wrong: $\boldsymbol{p}_{\text{err}}(i) = \boldsymbol{P}\left\{\mathcal{D}_{r_i}[S_i] \neq E_i\right\}$. Since the estimate (6.8.6), along with (6.8.4) and (6.8.5), implies that

$$\max_{E \in \mathfrak{E}_{r_{i-1}}} \boldsymbol{P}\left\{\mathcal{D}_{r_{i-1}}[S_i] \neq E_i \middle| E_i = E\right\} \leq c_{R,W,F} 2^{-r_{i-1} F},$$

we have

$$\boldsymbol{p}_{\text{err}}(i) = \sum_{E \in \mathfrak{E}_{r_{i-1}}} \boldsymbol{P}\left[E_i = E\right] \boldsymbol{P}\left\{\mathcal{D}_{r_{i-1}}[S_i] \neq E_i \middle| E_i = E\right\}$$

$$\leq c_{R,W,F} \sum_{E \in \mathfrak{E}_{r_{i-1}}} \boldsymbol{P}\left[E_i = E\right] 2^{-r_{i-1} F} \overset{(6.8.7)}{=\!=} c_{R,W,F} 2^{-(i-1+r_0)F};$$

$$\boldsymbol{p}_{\text{err}} \leq \sum_{i=1}^{\infty} \boldsymbol{p}_{\text{err}}(i+1) \leq c_{R,W,F} \sum_{i=1}^{\infty} 2^{-(i+r_0)F} = \frac{c_{R,W,F}}{2^F - 1} 2^{-r_0 F}. \qquad \square$$

Comment 6.8.21. As was shown in [164], the probability of error cannot be made small when stationary fixed-length block coding–decoding schemes are employed.

Proof of Proposition 6.8.17. It results from Lemmas 6.8.19 and 6.8.20. □

6.9 Tracking Almost Surely by Means of Fixed-Length Code Words: Proof of the $\mathfrak{c} > H(A) \Rightarrow$ a) part of Theorem 6.4.1

The observer from the previous section employs code words whose lengths increase without limits as the estimation process progresses. So the complexities of the coder and decoder should increase accordingly.[11] In this section, we show that whenever a communication feedback is available, asymptotic state tracking can be ensured by a coder and decoder that perform a limited number of operations per step and communicate information by means of fixed-length code words. Moreover, the state can be tracked almost surely, whereas a weaker tracking with as large a probability as desired was ensured in the previous section.

In doing so, we still consider the system (6.2.1) with no stable modes, i.e., adopt Assumption 6.8.1, until otherwise stated. Extensions on systems with both stable and unstable modes will be given in Sect. 6.10. We also suppose that the assumptions of Theorem 6.4.1 (on p. 140) hold and that $\mathfrak{c} > H(A)$.

6.9.1 Coder–Decoder Pair with a Communication Feedback

To ensure almost sure state tracking, the coder–decoder pair from Subsect. 6.8.2 is modified as follows:

i) The operation epochs are chosen to be of equal and fixed duration r_0; i.e., (6.8.7) and (6.8.10) are replaced by, respectively,

$$r_i := r_0 \qquad \text{and} \qquad \tau_i := ir_0;$$

ii) Instead of forming its own sequences of state estimates $\{\widehat{x}_c(t)\}$ and exactness bounds $\{\delta_c(t)\}$, the coder duplicates those generated by the decoder.

Explanation 6.9.1. To accomplish ii), the coder should be aware about the results $\boldsymbol{s}(t)$ of transmission across the channel. This becomes possible thanks to the communication feedback.

Specifically, now the coder prefaces (6.8.9) by the following actions at times $t = \tau_i, i = 1, 2, \ldots$:

- It carries out the step c.1) (see p. 164) of the previous coder;
- Then it duplicates the steps d.1) and d.2) of the decoder;
- After this, the steps c.2) and c.3) of the previous coder are carried out.

Explanation 6.9.2. Now step c.4) of the previous coder is in fact accomplished by carrying out step d.2) of the decoder.

For the convenience of the reader, now we describe the operation of the new coder in a more systematic way.

The coder (at the times $t = \tau_i, i = 1, 2, \ldots$)

[11] The same feature is characteristic for anytime coding–decoding schemes considered in [164, 166].

cc.1) Proceeding from the measurements obtained during the previous operation epoch, calculates the current state $x(\tau_i)$;

cc.2) Applies the decoding rule $\mathcal{D}_{r_0}$ to the data received within the previous epoch $[\tau_{i-1} : \tau_i]$ via the feedback communication channel and thus gets aware of the decoded value $q_d(\tau_i)$ produced by the decoder at time τ_i;

cc.3) Corrects successively the estimate and the exactness bound by duplicating the actions of the decoder:

$$\widehat{x}_c(\tau_i) := \widehat{x}_c(\tau_i) + \delta_c(\tau_i) A^{r_0} \overset{\star}{q}_d(\tau_i),$$
$$\delta_c(\tau_i) := \delta_c(\tau_i) \times \Big(\langle q_d(\tau_i) \rangle_{\varkappa,\gamma} \Big)^{r_0}; \quad (6.9.1)$$

cc.4) Employs the quantizer $\mathfrak{Q}_{r_0}$ and computes the quantized value $q_c(\tau_i)$ of the current scaled estimation error

$$\varepsilon(\tau_i) := \big[\delta_c(\tau_i)\big]^{-1} \big[x(\tau_i) - \widehat{x}_c(\tau_i)\big]; \quad (6.9.2)$$

cc.5) Encodes the quantized value $q_c(\tau_i)$ by means of the code book $\mathfrak{E}_{r_0}$. The obtained code word of the fixed length r_0 is transmitted across the channel during the next operation epoch $[\tau_i : \tau_{i+1})$.

Only after this does the coder perform the computations in accordance with (6.8.9).

The decoder is not altered. In other words, it still operates as follows.

The decoder (at times $t = \tau_i, i = 2, 3, \ldots$)

d.1) Applies the decoding rule $\mathcal{D}_{r_0}$ to the data received within the previous operation epoch $[\tau_{i-1} : \tau_i)$ and thus computes the decoded value $q_d(\tau_i)$ of the quantized and scaled estimation error $q_c(\tau_{i-1})$;

d.2) Corrects successively the estimate and the exactness bound in accordance with (6.9.1), where $\widehat{x}_d$ and δ_d are substituted in place of $\widehat{x}_c$ and δ_c, respectively.

Only after this does the decoder perform the computations from (6.8.9).

Explanation 6.9.3. For technical convenience, we put $q_c(\tau_0) := q_d(\tau_1) := \maltese$ and suppose that at times $t = \tau_0, \tau_1$ the coder and decoder act accordingly.

Remark 6.9.4. As follows from the foregoing, the coder and decoder generate common state estimates and their upper bounds:[12]

$$\widehat{x}_c^{\pm}(\tau_i) = \widehat{x}_d^{\pm}(\tau_i) \quad \text{and} \quad \delta_c^{\pm}(\tau_i) = \delta_d^{\pm}(\tau_i). \quad (6.9.3)$$

Remark 6.9.5. Step cc.1) is possible by the arguments from Explanation 6.8.10 since the duration of the operation epoch $r_0 \geq n$ by (6.8.8). Moreover, the current state can be obtained as a linear function of n previous measurements.

[12]We recall that their values before and after the update are marked by $^-$ and $^+$, respectively.

Observation 6.9.6. *The coder and decoder introduced in this section are semirational r_0-step recursive.*[13]

This is straightforward from the description of the coder and decoder with regard to the fact that the employed quantizer is taken from Lemma 3.8.36 (on p. 80) and so is polyhedral.[14]

Explanation 6.9.7. In (6.3.4) (on p. 139) and (6.3.6) (on p. 140), the states of the coder and decoder can be defined as $z_c := [\widehat{x}_c^+, \delta_c^+]$ and $z_d := [\widehat{x}_d^+, \delta_d^+]$, respectively.

6.9.2 Tracking Almost Surely by Means of Fixed-Length Code Words

The main result of the section is as follows.

Proposition 6.9.8. *The coder–decoder pair introduced in Subsect.* 6.9.1 *detects the state (i.e.,* (6.2.5) *on p.* 135 *holds) almost surely, provided that the duration of the operation epoch is large enough* $r_0 \geq \overline{r}(A, \varkappa, W, \gamma, R)$.

Here $\varkappa$ is taken from Lemma 3.8.36 (on p. 80), $W(\cdot|\cdot)$ is the matrix of the channel transition probabilities, and R, γ are the observer parameters from 1) and 2) (on p. 162).

The value of $\overline{r}(A, \varkappa, W, \gamma, R)$ will be specified in Lemma 6.9.20 (on p. 177).

By taking into account Observation 6.9.6, we arrive at the following.

Corollary 6.9.9. *Suppose that Assumption* 6.8.1 *(on p.* 161*) holds and that* $H(A) < \mathfrak{c}$. *Then the statement* a) *from Theorem* 6.4.1 *(on p.* 140*) is true.*

The remainder of the section is devoted to the proof of Proposition 6.9.8.

Ideas and Facts Underlying the Proof of Proposition 6.9.8

We start with an informal discussion. The operation epoch $[\tau_{i-1} : \tau_i)$ is said to be *regular* if during it a message different from the alarm one is correctly transmitted from the coder to the decoder

$$q_d(\tau_i) = q_c(\tau_{i-1}) \neq \maltese$$

and $\delta_d^+(\tau_{i-1})$ is a true bound for the estimation error:

$$\delta_d^+(\tau_{i-1}) \geq \|x(\tau_{i-1}) - \widehat{x}_d^+(\tau_{i-1})\|.$$

Then the update (6.8.13) at time $t = \tau_i$ improves the error bound via multiplying by $\varkappa^{r_0} < 1$, while keeping it correct for the updated estimate, which can be proved like (6.8.16).

Unfortunately, not each epoch is regular. First, the initial bound δ_0 may be incorrect. This reason is weak since the algorithm would make the bound correct for a

[13] See Definitions 6.3.9 and 6.3.10 on pp. 139 and 140.

[14] See Definition 3.8.8 on p. 70.

finite time in the absence of decoding errors (see the proof of Lemma 6.8.19). Second, the epoch may be irregular due to such errors. Any of them may make not only the current epoch irregular but also launch a whole "tail" of irregular epochs even if the messages transmitted across the channel during the subsequent epochs were decoded correctly. This holds if the transmission error makes the upper bound δ incorrect. During this tail, the error bound would increase via multiplying by $\gamma^{r_0} > 1$ in order to become correct once more. So any error has an after-effect, which evidently remains true in the real circumstances where the subsequent epochs are not necessarily "errorless." A priori, it is not clear even that the chain of irregular epochs will be broken and that a regular one will occur.

The proof is based on the fact that the probability of the decoding errors can be made as small as desired by properly picking the duration of the operation epoch r_0. By the strong law of large numbers, this entails that the average frequency of the decoding errors is small almost surely. In other words, the errors are rarely encountered. The next step is to evaluate the duration of the after-effect of each error and to show that the average frequency ω_{irr} of the irregular epochs does not exceed the average frequency of the above errors multiplied by a fixed factor. So ω_{irr} is also small. Hence not only regular epochs do follow any irregular one but also the average frequency of regular epochs $\omega_{\text{reg}} >> \omega_{\text{irr}}$. By taking into account that at any irregular epoch, the bound δ_d is increased at most by multiplying by γ^{r_0}, we conclude that (approximately)

$$\delta_d^-(\tau_i) \leq \delta_0 \varkappa^{ir_0\omega_{\text{reg}}} \gamma^{ir_0\omega_{\text{irr}}} \to 0 \quad \text{as} \quad i \to \infty.$$

This convergence is extended on the estimation error on the grounds that $\delta_d^-(\tau_i)$ is the correct bound for this error for most of the i's.

Strong Law of Large Numbers

To carry out the first step of this plan, we shall use the following variant of the strong law of large numbers [95, §32, p.53] (see also [145, 161]).

Theorem 6.9.10. *Suppose that $\mathfrak{F}_i$ is a flow of nondecreasing σ-algebras in a probability space, the random variable J_i is $\mathfrak{F}_i$-measurable, and $b_i \uparrow \infty, b_i > 0, i = 1, 2, \ldots$. Suppose also that $\boldsymbol{E}|J_i| < \infty$ and*

$$\sum_{i=1}^{\infty} \frac{1}{b_i^2} \boldsymbol{E}\Big\{ \big[J_i - \boldsymbol{E}(J_i|\mathfrak{F}_{i-1})\big]^2 \Big\} < \infty. \tag{6.9.4}$$

Then with probability 1,

$$\frac{1}{b_k} \sum_{i=1}^{k} \big[J_i - \boldsymbol{E}(J_i|\mathfrak{F}_{i-1})\big] \to 0 \quad \text{as} \quad k \to \infty. \tag{6.9.5}$$

6.9.3 Proof of Proposition 6.9.8

We start with estimating the average frequency with which the decoding errors occur.

Frequency of Decoding Errors

We consider the stochastic process generated by the coder–decoder pair in connection with the system (6.2.1). The symbols E_i and S_i stand for the code words formed by the coder at time τ_{i-1} and received by the decoder at time τ_i, respectively. We also introduce the error indicator function:

$$I^{\text{err}}(i) := \begin{cases} 1 & \text{if} \quad \mathcal{D}_{r_0}[S_i] \neq E_i, \; i \geq 2, \\ 0 & \text{otherwise} \end{cases}. \tag{6.9.6}$$

Lemma 6.9.11. *We pick* $0 < F < F(R, W)$, *where* $F(R, W)$ *is taken from* (6.8.6). *Then the following relation holds almost surely, provided that the duration* r_0 *of the operation epoch is sufficiently large:*

$$\overline{\lim_{k \to \infty}} \frac{1}{k} \sum_{i=1}^{k} I^{\text{err}}(i) \leq 2^{-F r_0}. \tag{6.9.7}$$

Proof. We are going to apply Theorem 6.9.10 to $J_i := I^{\text{err}}(i)$ and $b(i) := i$. The σ-algebra $\mathfrak{F}_i$ is taken to be that generated by the random quantities

$$x_0, S_0, \ldots, S_i.$$

Due to the construction of the coder,

$$E_i = \mathcal{E}_*[i, x_0, S_0, \ldots, S_{i-1}],$$

where $\mathcal{E}_*(\cdot)$ is a deterministic function. It follows that $I^{\text{err}}(i)$ is $\mathfrak{F}_i$-measurable. Furthermore,

$$0 \leq I^{\text{err}}(i) \leq 1 \Rightarrow 0 \leq \boldsymbol{E}[I^{\text{err}}(i) | \mathfrak{F}_{i-1}] \leq 1 \quad \text{a.s.,}$$

which implies (6.9.4). So by Theorem 6.9.10,

$$\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \Big\{ I^{\text{err}}(i) - \boldsymbol{E}[I^{\text{err}}(i) | \mathfrak{F}_{i-1}] \Big\} = 0 \quad \text{a.s.} \tag{6.9.8}$$

Now we are going to estimate $\boldsymbol{E}[I^{\text{err}}(i) | \mathfrak{F}_{i-1}]$. Let I_Ω denote the indicator of the random event Ω. By invoking Assumptions 6.3.1 and 6.3.4 (on pp. 136 and 137), we get

$$\begin{aligned} \boldsymbol{E}[I^{\text{err}}(i) | \mathfrak{F}_{i-1}] &= \boldsymbol{E}\Big\{ I^{\text{err}}(i) \Big| x_0, S_0, \ldots, S_{i-1}, E_i = \mathcal{E}_*[i, x_0, S_0, \ldots, S_{i-1}] \Big\} \\ &= \sum_{\boldsymbol{E}} \boldsymbol{P}\big[\mathcal{D}_{r_0}[S_i] \neq \boldsymbol{E} \big| x_0, S_0, \ldots, S_{i-1}, E_i = \boldsymbol{E}\big] I_{E_i = \boldsymbol{E}} \\ &= \sum_{\boldsymbol{E}} \boldsymbol{P}\big[\mathcal{D}_{r_0}[S_i] \neq \boldsymbol{E} \big| E_i = \boldsymbol{E}\big] I_{E_i = \boldsymbol{E}} \overset{(6.8.6)}{\lesssim} 2^{-r_0 F(R,W)}. \end{aligned}$$

Since $F < F(R, W)$, this implies

$$\boldsymbol{E}[I^{\text{err}}(i)|\mathfrak{F}_{i-1}] \leq 2^{-Fr_0} \qquad \forall i \qquad \text{whenever} \quad r_0 \approx \infty. \tag{6.9.9}$$

So by invoking (6.9.8), we see that almost surely

$$\overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=1}^{k} I^{\text{err}}(i)$$
$$= \overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{E}[I^{\text{err}}(i)|\mathfrak{F}_{i-1}] + \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} \Big\{ I^{\text{err}}(i) - \boldsymbol{E}[I^{\text{err}}(i)|\mathfrak{F}_{i-1}] \Big\} \leq 2^{-Fr_0}. \ \square$$

Influence of the Channel Errors on the Estimation Process

Our next goal is to discover how the channel noise affects the estimation errors. To this end, we introduce the indicator functions of the following events:

$$\begin{array}{rcl} I_0(i) & \longleftrightarrow & q_d(\tau_i) = q_c(\tau_{i-1}) \neq \maltese; \\ I_{\maltese}(i) & \longleftrightarrow & q_d(\tau_i) = q_c(\tau_{i-1}) = \maltese; \\ I^{\text{err}}_{\text{c}\maltese}(i) & \longleftrightarrow & q_d(\tau_i) \neq q_c(\tau_{i-1}) = \maltese; \\ I^{\text{err}}_{\text{d}\maltese}(i) & \longleftrightarrow & \maltese = q_d(\tau_i) \neq q_c(\tau_{i-1}); \\ I^{\text{err}}_{0}(i) & \longleftrightarrow & \maltese \neq q_d(\tau_i) \neq q_c(\tau_{i-1}) \neq \maltese. \end{array} \tag{6.9.10}$$

Remark 6.9.12. It is easy to see that

$$I^{\text{err}}_{\text{c}\maltese}(i) + I^{\text{err}}_{\text{d}\maltese}(i) + I^{\text{err}}_{0}(i) = I^{\text{err}}(i) \quad \text{and} \quad I_0(i) + I_{\maltese}(i) = 1 - I^{\text{err}}(i). \tag{6.9.11}$$

We first examine the evolution of the following quantities:

$$\delta_i := \delta_c^+(\tau_i) \quad \text{and} \quad z_i := \|\widehat{x}_c^+(\tau_i) - x(\tau_i)\|. \tag{6.9.12}$$

Lemma 6.9.13. *The following relations hold for any $i \geq 1$:*

$$\delta_i = \delta_{i-1} \Big\{ \varkappa^{r_0} \big[I_0(i) + I^{\text{err}}_0(i) + I^{\text{err}}_{c\maltese}(i) \big] + \gamma^{r_0} \big[I_{\maltese}(i) + I^{\text{err}}_{d\maltese}(i) \big] \Big\}; \tag{6.9.13}$$

$$z_i \leq z_{i-1} \|A\|^{r_0} \big[I_{\maltese}(i) + I^{\text{err}}_{d\maltese}(i) \big] + \delta_{i-1} \varkappa^{2r_0} I_0(i)$$
$$+ \|A\|^{r_0} (z_{i-1} + \delta_{i-1}) \big[I^{\text{err}}_0(i) + I^{\text{err}}_{c\maltese}(i) \big]. \tag{6.9.14}$$

Here $\varkappa \in (0,1)$ is taken from Lemma 3.8.36 *(on p.* 80*) and $\gamma > \|A\|$ is the parameter of the estimator from* 2) *on p.* 162.

Proof. To prove (6.9.13), we note that

$$\delta_i \overset{(6.9.12)}{=\!=\!=} \delta_c^+(\tau_i) \overset{(6.9.1)}{=\!=\!=} \delta_c^-(\tau_i) \Big(\langle q_d(\tau_i) \rangle_{\varkappa,\gamma} \Big)^{r_0} \overset{(6.9.12)}{=\!=\!=} \delta_{i-1} \Big(\langle q_d(\tau_i) \rangle_{\varkappa,\gamma} \Big)^{r_0}.$$

So (6.9.13) is immediate from (6.9.10) and the definition of $\langle\cdot\rangle_{\varkappa,\gamma}$ from (6.8.12). To justify (6.9.14), we observe that

$$z_i \overset{(6.9.12)}{=\!=} \|\widehat{x}_c^+(\tau_i) - x(\tau_i)\| \overset{(6.9.1)}{=\!=} \|\widehat{x}_c^-(\tau_i) + \delta_c^-(\tau_i) A^{r_0} \overset{\star}{q}_d(\tau_i) - x(\tau_i)\|$$
$$\overset{(6.2.1),(6.8.9)}{=\!=} \left\| A^{r_0}\left[\widehat{x}_c^+(\tau_{i-1}) + \delta_c^+(\tau_{i-1}) \overset{\star}{q}_d(\tau_i) - x(\tau_{i-1})\right]\right\|$$
$$\overset{(6.9.12)}{=\!=} \left\| A^{r_0}\left[\widehat{x}_c^+(\tau_{i-1}) - x(\tau_{i-1}) + \delta_{i-1} \overset{\star}{q}_d(\tau_i)\right]\right\|.$$

If $I_{\maltese}(i) + I^{\text{err}}_{\text{d}\maltese}(i) = 1$, then $q_d(\tau_i) = \maltese$ and $\overset{\star}{q}_d(\tau_i) = 0$ by (6.8.12). So

$$z_i \leq \|A\|^{r_0} \|\widehat{x}_c^+(\tau_{i-1}) - x(\tau_{i-1})\| \overset{(6.9.12)}{=\!=} \|A\|^{r_0} z_{i-1}.$$

If $I_0(i) = 1$, then $\maltese \neq q_c(\tau_{i-1}) = q_d(\tau_i) = \overset{\star}{q}_d(\tau_i)$. So

$$z_i = \delta_{i-1} \|A^{r_0}[\varepsilon(\tau_{i-1}) - q_c(\tau_{i-1})]\|$$

due to (6.8.11), where $q_c(\tau_{i-1})$ is the quantized value of $\varepsilon(\tau_{i-1})$. Hence by invoking (3.8.9) (on p. 70), where $\rho_{\mathfrak{Q}} = \varkappa^{2r_0}$ thanks to Lemma 3.8.36 (on p. 80), we get $z_i \leq \varkappa^{2r_0}\delta_{i-1}$. Finally, let $I_0^{\text{err}}(i) + I^{\text{err}}_{\text{c}\maltese}(i) = 1$. Then $\|\overset{\star}{q}_d(\tau_i)\| \leq 1$ and so

$$z_i \leq \|A\|^{r_0}\left(\|\widehat{x}_c^+(\tau_{i-1}) - x(\tau_{i-1})\| + \delta_{i-1}\right) \overset{(6.9.12)}{=\!=} \|A\|^{r_0}(z_{i-1} + \delta_{i-1}).$$

Summarizing, we arrive at (6.9.14). □

Lemma 6.9.13 entails an important conclusion about the evolution of the ratio $\xi_i := z_i/\delta_i$, which determines whether $\maltese$ is sent over the channel:

$$q_c(\tau_i) = \maltese \Leftrightarrow \xi_i = z_i/\delta_i > 1. \tag{6.9.15}$$

Corollary 6.9.14. *For $i \geq 1$, the following inequality holds:*

$$\xi_i \leq \left\{\begin{array}{ll} \rho\xi_{i-1} & \text{if } \xi_{i-1} > 1 \\ \varkappa^{r_0} & \text{if } \xi_{i-1} \leq 1 \end{array}\right\}\left[1 - I^{\text{err}}(i)\right] + b/2\left[\xi_{i-1} + 1\right] I^{\text{err}}(i), \tag{6.9.16}$$

where $I^{\text{err}}(i)$ is the error indicator function (6.9.6)*, and*

$$\rho := \left(\frac{\|A\|}{\gamma}\right)^{r_0}, \quad b := 2\left(\frac{\|A\|}{\varkappa}\right)^{r_0}. \tag{6.9.17}$$

The proof is by merely checking (6.9.16) on the basis of (6.9.13) and (6.9.14).

How Often the Bound δ_c Generated by the Coder Is Incorrect?

The corresponding event can be also equivalently described in each of the following two ways:

- The alarm symbol ✠ is emitted into the channel;
- The inequality $\xi_i > 1$ holds.

We are interested in the average frequency of this event.

The following lemma reveals an important relationship between this event and the channel errors.

Lemma 6.9.15. *Whenever* $\xi_i > 1$ *for* $i = \overline{i}+1, \ldots \overline{i}+k$*, the number* l *of the channel errors within the interval* $[\overline{i}+1 : \overline{i}+k]$ *obeys the lower bound*

$$l := \left|\{j = \overline{i}+1, \ldots \overline{i}+k : I^{\text{err}}(j) = 1\}\right| \geq k\frac{\log_2[\rho^{-1}]}{\log_2 b + \log_2[\rho^{-1}]} - \frac{\log_2 \max\left\{\xi_{\overline{i}}, \frac{\xi_{\overline{i}}+1}{2}\right\}}{\log_2 b + \log_2[\rho^{-1}]}. \quad (6.9.18)$$

Proof. If $\xi_{\overline{i}} \leq 1$ and $I^{\text{err}}(\overline{i}+1) = 0$, then (6.9.16) implies $\xi_{\overline{i}+1} = \varkappa^{r_0} < 1$ in violation of the hypotheses of the lemma. Thus

$$\xi_{\overline{i}} \leq 1 \Rightarrow I^{\text{err}}(\overline{i}+1) = 1.$$

By invoking (6.9.16) once more, we get for $i = \overline{i}+1, \ldots, \overline{i}+k$,

$$\xi_i \leq \rho\xi_{i-1}[1 - I^{\text{err}}(i)] + b\xi_{i-1}I^{\text{err}}(i) + \frac{b}{2}[1-\xi_{i-1}]I^{\text{err}}(i)$$
$$\leq \xi_{i-1}\left\{\rho[1 - I^{\text{err}}(i)] + bI^{\text{err}}(i)\right\} + \frac{b}{2}\max\{1-\xi_{i-1}, 0\}I^{\text{err}}(i).$$

The last summand may not vanish only if $i = \overline{i}+1$ and $I^{\text{err}}(\overline{i}+1) = 1$. Hence

$$1 < \xi_{\overline{i}+k} \leq \xi_{\overline{i}}\rho^{k-l}b^l + \frac{b}{2}\max\{1-\xi_{\overline{i}}, 0\}\rho^{k-l}b^{l-1} = \rho^{k-l}b^l \max\left\{\xi_{\overline{i}}, \frac{\xi_{\overline{i}}+1}{2}\right\},$$
$$0 < (l-k)\log_2[\rho^{-1}] + l\log_2 b + \log_2 \max\left\{\xi_{\overline{i}}, \frac{\xi_{\overline{i}}+1}{2}\right\} \;\Big|\; \Rightarrow (6.9.18). \qquad \square$$

Now we are in a position to give an answer to the question posed.

Corollary 6.9.16. *For the indicator function* $I_{\xi>1}(i) \longleftrightarrow \xi_i > 1$*, the following relation holds almost surely:*

$$\overline{\lim_{k\to\infty}} \frac{1}{k}\sum_{i=1}^{k} I_{\xi>1}(i-1) \leq \beta := \mu^{-1} \overline{\lim_{k\to\infty}} \frac{1}{k}\sum_{i=1}^{k} I^{\text{err}}(i), \quad (6.9.19)$$

where

$$\mu := \frac{\log_2[\rho^{-1}]}{\log_2 b + \log_2[\rho^{-1}]}.$$

Remark 6.9.17. The $\overline{\lim}$ in the right-hand side of (6.9.19) is estimated by (6.9.7).

Proof of Corollary 6.9.16. If $\beta \geq 1$, the claim is obvious. Suppose that $\beta < 1$. Then $\xi_i \leq 1$ for some $i = i^*$. Indeed otherwise, Lemma 6.9.15 with $\overline{i} := 0$ and arbitrary k yields

$$\overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=1}^{k} I^{\text{err}}(i) \geq \mu + \overline{\lim_{k\to\infty}} -\frac{1}{k} \frac{\log_2 \max\{\xi_0, \frac{\xi_0+1}{2}\}}{\log_2 b + \log_2[\rho^{-1}]} = \mu,$$

which implies $\beta \geq 1$ in violation of the hypothesis. For $k > i^*$, the set

$$\{i^* \leq i \leq k : I_{\xi>1}(i) = 1\}$$

disintegrates into several intervals of respective sizes $k_1, \ldots, k_s$, which do not contain i^* and are separated by intervals where $\xi_i \leq 1$. Now we apply Lemma 6.9.15 to the jth interval, picking $\overline{i}$ to be the integer preceding its left end. Then $\xi_{\overline{i}} \leq 1$, the second ratio in (6.9.18) is nonpositive, and so the number l_j of errors contained by the interval at hand is no less than $k_j\mu$. Hence

$$\overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=1}^{k} I^{\text{err}}(i) \geq \overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=j}^{s} l_j \geq \mu \overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=j}^{s} k_j$$
$$= \mu \overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=i^*}^{k} I_{\xi>1}(i) = \mu \overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=1}^{k} I_{\xi>1}(i-1) \Bigg| \Rightarrow \text{(6.9.19)}. \qquad \square$$

Corollary 6.9.18. *The indicator function* $I(i) \longleftrightarrow I^{\text{err}}(i) = 1 \vee I_{\xi>1}(i-1) = 1$ *a.s. obeys the inequality*

$$\overline{\lim_{r\to\infty}} \frac{1}{k} \sum_{i=1}^{k} I(i) \leq \left\{2 + \frac{\log_2 b}{\log_2[\rho^{-1}]}\right\} \times \overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=1}^{k} I^{\text{err}}(i) \overset{(6.9.7)}{\leq} \overline{p}, \qquad (6.9.20)$$

where

$$\overline{p} := 2^{-Fr_0} \left\{2 + \frac{\log_2 b}{\log_2[\rho^{-1}]}\right\}. \qquad (6.9.21)$$

Indeed, this is immediate from Corollary 6.9.16, (6.9.7), and the apparent inequality

$$I(i) \leq I^{\text{err}}(i) + I_{\xi>1}(i-1).$$

Observation 6.9.19. *It is easy to see that the first inequality in* (6.9.20) *is a direct consequence of* (6.9.16)*. In other words, it holds for any nonnegative solution* ξ_i *of the recursive inequalities* (6.9.16) *with* $i = 1, 2, \ldots$*, where* $\{I^{\text{err}}(i)\}$ *is an arbitrary sequence of reals* $I^{\text{err}}(i) = 0, 1$*, and* $\rho, \varkappa \in (0,1), b > 1$ *are arbitrary numbers.*

Sufficient Conditions for Almost Sure Tracking

Now we bring the pieces together. We start with conditions, which can be used to pick the duration r_0 of the operation epoch in order to ensure almost sure state tracking.

Lemma 6.9.20. *The coder–decoder pair introduced in Subsect.* 6.9.1 *tracks the state a.s. whenever*

$$\omega := \log_2[\varkappa^{-1}] - \overline{p}\big\{ \log_2 \gamma + \log_2[\varkappa^{-1}]\big\} > 0 \quad \text{and} \\ \chi := \omega(1-\overline{p}) - \overline{p}\log_2 \|A\| > 0. \qquad (6.9.22)$$

Here $\varkappa \in (0,1)$ *is the constant from Lemma* 3.8.36 *(on p.* 80*),* γ *is the parameter from* 2) *on p.* 162*, and* $\overline{p}$ *is given by* (6.9.17) *and* (6.9.21)*.*

Proof. The symbol c (with a possible index) will be used to denote random constants independent of i and r_0. For any $\alpha > 0$, (6.9.20) implies

$$\mathfrak{I}(k) := \sum_{i=1}^{k} I(i) \leq k(\overline{p}+\alpha) \quad \text{for} \quad k \approx \infty.$$

By 2) on p. 162, $\gamma > \|A\|$, where $\|A\| \geq 1$ due to Assumption 6.8.1 (on p. 161). Thus $\varkappa < 1 < \gamma$. So (6.9.10) and (6.9.13) yield

$$\delta_i \leq \delta_{i-1}\big\{\varkappa^{r_0}[1-I(i)] + \gamma^{r_0} I(i)\big\} \;\forall i \geq 1 \;\Rightarrow\; \delta_k \leq \delta_0 \varkappa^{kr_0} \prod_{i=1}^{k} \left(\frac{\gamma}{\varkappa}\right)^{r_0 I(i)} \\ = \delta_0 \varkappa^{kr_0} \left(\frac{\gamma}{\varkappa}\right)^{r_0 \mathfrak{I}(k)} \overset{k\approx\infty}{\leq} \delta_0 \varkappa^{kr_0} \left(\frac{\gamma}{\varkappa}\right)^{kr_0(\overline{p}+\alpha)} = \delta_0 2^{-kr_0\omega_\alpha},$$

where

$$\omega_\alpha := \log_2[\varkappa^{-1}] - [\overline{p}+\alpha]\big\{ \log_2 \gamma + \log_2[\varkappa^{-1}]\big\} \xrightarrow{\alpha\to 0} \omega > 0.$$

Thus for $\alpha \approx 0$, we have $\omega_\alpha > 0$ and

$$\delta_i \leq c' 2^{-ir_0\omega_\alpha} \to 0 \quad \text{as} \quad i \to \infty. \qquad (6.9.23)$$

Now we note that due to (6.9.14),

$$z_i \leq \delta_{i-1}\varkappa^{2r_0}\big[1-I(i)\big] + \|A\|^{r_0}(z_{i-1}+\delta_{i-1})I(i) \leq \|A\|^{r_0} z_{i-1} I(i) + c2^{-ir_0\omega_\alpha}, \\ z_k \leq z_0 \prod_{i=1}^{k}\Big[\|A\|^{r_0} I(i)\Big] + c\sum_{i=1}^{k} 2^{-ir_0\omega_\alpha} \prod_{j=i+1}^{k}\Big[\|A\|^{r_0} I(j)\Big].$$

The first relation from (6.9.22) implies $\overline{p} < 1$. So

$$\{i \geq 1 : I(i) = 1\} \neq \{i = 1, 2, \ldots\}$$

due to (6.9.20). It follows that for $k \approx \infty$, the first summand vanishes and

$$z_k \leq c \sum_{i=k-l}^{k} 2^{-ir_0\omega_\alpha} \|A\|^{r_0(k-i)},$$

where $[k-l+1:k]$ is the largest subinterval of the set

$$\Omega_k := \{1 \le i \le k : I(i) = 1\}$$

containing k. (If $k \notin \Omega_k$, then $l := 0$.) We proceed by taking into account the inequality $l \le \sum_{i=1}^{k} I(i) = \mathfrak{I}(k) \le k(\overline{p}+\alpha)\ \forall k \approx \infty$:

$$\begin{aligned} z_k &\le c 2^{-kr_0\omega_\alpha} \sum_{i=k-l}^{k} 2^{(k-i)r_0\omega_\alpha} \|A\|^{r_0(k-i)} = c2^{-kr_0\omega_\alpha} \sum_{i=0}^{l} \left(2^{\omega_\alpha}\|A\|\right)^{r_0 i} \\ &\le c\frac{2^{-kr_0\omega_\alpha}\left(2^{\omega_\alpha}\|A\|\right)^{r_0 l}}{1-\left(2^{\omega_\alpha}\|A\|\right)^{-r_0}} \overset{k\approx\infty}{\le} c\frac{2^{-kr_0\omega_\alpha}\left(2^{\omega_\alpha}\|A\|\right)^{r_0 k(\overline{p}+\alpha)}}{1-\left(2^{\omega_\alpha}\|A\|\right)^{-r_0}} \\ &= \frac{c}{1-\left(2^{\omega_\alpha}\|A\|\right)^{-r_0}} 2^{-kr_0\chi_\alpha}, \end{aligned}$$

where

$$\chi_\alpha := \omega_\alpha[1-(\overline{p}+\alpha)] - (\overline{p}+\alpha)\log_2\|A\| \xrightarrow[\alpha\to 0]{(6.9.22)} \chi > 0.$$

Thus $\chi_\alpha > 0$ for $\alpha \approx 0$. So

$$z_k \overset{(6.9.12)}{=\!=} \|\widehat{x}_c^+(\tau_k) - x(\tau_k)\| \to 0 \quad \text{as} \quad k \to \infty.$$

Here $\widehat{x}_c^+(\tau_k) = \widehat{x}_d^+(\tau_k)$ thanks to (6.9.3). To complete the proof, we note that for $\tau_k < t \le \tau_{k+1} = \tau_k + r_0$, we have

$$\begin{aligned} \|\widehat{x}_d(t) - x(t)\| &\overset{(6.2.1),(6.8.9)}{=\!=\!=} \left\|A^{t-\tau_k}[\widehat{x}_d^+(\tau_k) - x(\tau_k)]\right\| \\ &\le \|A\|^{r_0}\|\widehat{x}_d^+(\tau_k) - x(\tau_k)\| \xrightarrow{t\to\infty} 0, \end{aligned}$$

since $k \to \infty$ as $t \to \infty$. □

Completion of the Proof of Proposition 6.9.8

By Lemma 6.9.20, it suffices to show that (6.9.22) does hold whenever r_0 is large enough. In its turn, this is true if $\overline{p} \to 0$ as $r_0 \to \infty$. The required property is established as follows:

$$\begin{aligned} \overline{p} &\overset{(6.9.20)}{=\!=} 2^{-Fr_0}\left\{2 + \frac{\log_2 b}{\log_2[\rho^{-1}]}\right\} \\ &\overset{(6.9.17)}{=\!=} 2^{-Fr_0}\left\{2 + \frac{1 + r_0[\log_2\|A\| + \log_2 \varkappa^{-1}]}{r_0\left(\log_2\gamma - \log_2\|A\|\right)}\right\} \to 0 \quad \text{as} \quad r_0 \to \infty. \end{aligned}$$ □

6.10 Completion of the Proof of Theorem 6.4.1 (on p. 140): Dropping Assumption 6.8.1 (on p. 161)

The implication $\mathfrak{c} > H(A) \Rightarrow$ a) $\wedge$ b) has been already justified for systems with no stable modes. Now we consider the general case. Suppose that $\mathfrak{c} > H(A)$, and consider the matrix A_+ from (6.7.24) (on p. 156), i.e., the "unstable part" of A. Since $H(A) = H(A_+)$, the claims a) and b) are true for the system (6.7.24) with $x_0^+ := \pi_+ x_0$ by Corollaries 6.8.18 (on p. 165) and 6.9.9 (on p. 170).

Now we apply the corresponding coder–decoder pair to the primal system (6.2.1). In doing so, we also alter the coder's step c.1) (see p. 164) or cc.1) (see p. 169), where it determines the current state $x_+(\tau_i)$ of (6.7.24). Formerly this step was done on the basis of the past measurements from (6.7.24). Now we employ the observations from (6.2.1). Then thanks to Assumption 6.3.6 (on p. 137), it is possible to compute $\pi_+ x(\tau_i)$ as a linear function of the measurements received during the previous operation epoch, provided that $r_0 \geq n$. Since evidently $x_+(t) := \pi_+ x(t)$, this does not alter the operation of the observer. To complete the proof of the implication $\mathfrak{c} > H(A) \Rightarrow$ a) $\wedge$ b), we note that so far as $x_-(t) := x(t) - x_+(t) \to 0$ as $t \to \infty$, this observer tracks that state of (6.2.1)

$$\|x(t) - \widehat{x}_+(t)\| = \|x_-(t) + x_+(t) - \widehat{x}_+(t)\| \leq \|x_-(t)\| + \|x_+(t) - \widehat{x}_+(t)\| \xrightarrow{t\to\infty} 0$$

whenever it detects the state of (6.7.24): $\|x_+(t) - \widehat{x}_+(t)\| \to 0$ as $t \to \infty$.

The implications a) $\vee$ b) $\Rightarrow$ c) $\Rightarrow$ d) are apparent, whereas d) $\Rightarrow \mathfrak{c} \geq H(A)$ was justified in Sect. 6.7. □

6.11 Stabilizing Controller and the Proof of the Sufficient Conditions for Stabilizability

In this section, we prove the $\mathfrak{c} > H(A) \Rightarrow$ a) part of Theorem 6.6.1 (on p. 145). So we suppose that the assumptions of Theorem 6.6.1 hold and that $\mathfrak{c} > H(A)$. Like in Sect. 6.8, we first assume that the plant (6.5.1) (on p. 143) has no stable eigenvalues; i.e., we adopt Assumption 6.8.1 (on p. 161) until otherwise stated. In the general case of the plant with both unstable and stable modes, stabilization will be achieved by applying the controller presented below to the unstable part of the system, like in Subsect. 3.8.11 (starting on p. 88) and Sect. 6.10.

Remark 6.11.1. Assumptions 6.3.6, 6.5.5, and 6.8.1 (on pp. 137, 144, and 161) imply that the pairs (A, B) and (A, C) are controllable and observable, respectively.

For the problem of state estimation via a noisy channel, two algorithms were proposed in Sects. 6.8 and 6.9, respectively. The first of them works in the absence of the feedback communication link, employs code words whose lengths increase as the estimation process progresses, and ensures observability with as large a probability as desired. The second algorithm uses fixed-length code words, has a limited complexity, and guarantees almost sure observability, but it relies on the complete communication feedback. This feedback makes the coder aware about the result of the transmission across the channel by the time of the next transmission.

In this section, we show that an almost sure stabilizing controller with the features like those of the second observer can be constructed even if no special feedback communication link is available. This will be accomplished via two steps.

Step 1. We show that such a controller can be designed whenever a feedback link is yet available but has an arbitrarily small capacity.

Step 2. We show that in fact even such a low-capacity special link is not needed since the required low-rate feedback information flow can be arranged by means of proper control actions.

Remark 6.11.2. The feedback link concerned at step 1 should transmit one bit per operation epoch by notifying the coder whether the signal received by the decoder at the end of the previous epoch was the alarm one ✠. The average rate of this communication can be made arbitrarily small by taking the duration of the operation epoch large enough.

Explanation 6.11.3. Communication of information can be arranged by means of control thanks to the fact that the controller influences the plant motion, whereas the sensor observes this motion. So the controller can encode a message by imparting the motion of a certain specific feature. In its turn, the sensor can receive the message by observing the motion and detecting this feature.

Remark 6.11.4. In the context of this section, only one bit per operation epoch should be communicated by means of control. At the same time, communication of as much information as desired can be arranged in such a way without violating the main objective of stabilization [114].

6.11.1 Almost Sure Stabilization by Means of Fixed-Length Code Words and Low-Rate Feedback Communication

In this subsection, we do step 1 and introduce the corresponding stabilizing controller.

Components Used to Assemble the Coder and Decoder

The first four components are basically those chosen in Subsect. 6.8.2 to construct an observer.

Specifically, we pick

1) two numbers η and R such that

$$\log_2 |\det A| < \eta < R < \mathfrak{c};$$

2) a parameter $\gamma > \|A\|$ and a duration $r = 1, 2, \ldots$ of the operation epoch;

3) a code book $\mathfrak{E}_r$ with $N = N'_r \approx 2^{rR}$ input code words (6.8.2) (on p. 161) each of length r and a decoding rule $\mathcal{D}_r(\cdot)$ with the properties described in Theorem 6.8.4 (on p. 162);
4) the r-contracted quantizer $\mathfrak{Q}_r$ from Lemma 3.8.36 [15] applied to $A_{ss} := A$;
5) A deadbeat stabilizer,[16] i.e., a linear transformation of an initial state

$$x(0) = x \xrightarrow{\mathcal{N}} \boldsymbol{U} = [u(0), u(1), \ldots, u(n-1), 0, 0, \ldots] \tag{6.11.1}$$

into a sequence of controls driving the state to zero $x(n) = 0$;
6) An *alarm control sequence*

$$\boldsymbol{U}_{\maltese} = [u_0, \ldots, u_{s-1}, 0, 0, \ldots],$$

which drives the system from $x(0) = 0$ to $x(s) = 0$.

Explanations 6.8.6 and 6.8.7 (on p. 163) concerning 1) and 2), respectively, remain active.

Remark 6.11.5. Due to Remark 6.11.1, a deadbeat stabilizer does exist [10, p. 253].

Explanation 6.11.6. The control sequence is extended by zeros from $[0 : n-1]$ onto $[0 : \infty)$ in (6.11.1) for technical convenience.

Explanation 6.11.7. The alarm control sequence and its *active length* s will be specified further. The role of this sequence will be elucidated in the next subsection.

Notation 6.11.8. *We extend the deadbeat stabilizer on the alarm symbol* $\maltese$ *by putting*

$$\mathcal{N}(\maltese) := \boldsymbol{U}_{\maltese}. \tag{6.11.2}$$

Definition 6.11.9. *The number* $L(\mathcal{N}) := \max\{n, s\}$ *is called the* length *of the deadbeat stabilizer.*

Explanation 6.11.10. We pick the duration r from 2) so large $r \geq r_*$ that

$$r > n + L(\mathcal{N}) \tag{6.11.3}$$

and the quantizer outputs including the alarm symbol $\maltese$ can be encoded by code words from the code book $\mathfrak{E}_r$. The latter is possible by Remark 6.8.8 (on p. 163).

Operation Epochs and an Intermediate Temporary Assumption

Like in the case of the observer from Subsect. 6.9.1, the controller operation is organized into epochs

$$[\tau_i : \tau_{i+1}), \qquad \tau_i := ir \tag{6.11.4}$$

[15] See p. 80; we also invoke Remark 3.8.37 from p. 80 here.
[16] See p. 72 for the definition.

of equal duration r. A fixed and independent of epoch sequence of operations is executed within any of them. In particular, at the beginning of any epoch, the coder converts a quantizer output (which may equal the alarm symbol ✠) into a code word of length r. This word is then transmitted across the channel during the operation epoch. At the end of the epoch, the decoder decodes the sequence of messages received within the epoch, thus trying to determine the original quantizer output.

Until the next subsection, we suppose that a low-rate feedback communication is available by adopting the following.

Assumption 6.11.11. *By the end τ_{i+1} of the current operation epoch, the coder almost surely gets aware of whether the message received by the decoder as a result of decoding at the beginning τ_i of this epoch was the alarm one ✠.*

Remark 6.11.12. This feedback communication has the size of a one-bit-per-operation-epoch. By increasing the epoch duration r, the average bit rate of this communication can be made arbitrarily small.

Remark 6.11.13. Assumption 6.11.11 may be true due to not only the presence of a special feedback communication link but also the fact that the alarm signal is transmitted over an especially reliable feedforward subchannel.

Remark 6.11.14. As will be shown in Subsect. 6.11.2, Assumption 6.11.11 can always be ensured by means of control via a proper choice of the alarm control sequence and epoch duration.

Coder–Decoder Pair Consuming a Low-Rate Communication Feedback

Both coder and decoder compute controls $u_c(t)$, $u_d(t)$ and upper bounds for the state norm $\delta_c(t), \delta_d(t)$, respectively. Actually acting upon the plant is the control $u_d(t)$. The initial bound is common:

$$\delta_c(0) = \delta_d(0) = \delta_0 > 0.$$

Remark 6.11.15. The bounds $\delta_c(t)$ and $\delta_d(t)$ may be incorrect. In particular, the inequality $\delta_0 \geq \|x(0)\|$ may be violated.

Within any operation epoch $[\tau_i : \tau_{i+1})$, the coder consecutively emits into the channel the symbols of the code word of the length r formed at time τ_i, and the decoder carries out the control program

$$\boldsymbol{U}_i^d := \textbf{col}\left[u_d(\tau_i), u_d(\tau_i+1), \ldots, u_d(\tau_{i+1}-1)\right] \tag{6.11.5}$$

generated at time τ_i. These actions are prefaced at the times $t = \tau_i$ by the following operations.

The coder (at the times $t = \tau_i, i = 1, 2, \ldots$)

cs.1) Proceeding from the previous measurements, calculates the current state $x(\tau_i)$;

cs.2) Computes the prognosis of the state at the time $t = \tau_{i+1}$:

$$\widehat{x}_c(t) := A^r x(\tau_i) + \sum_{j=\tau_i}^{t-1} A^{t-1-j} B u_c(j); \tag{6.11.6}$$

cs.3) If $i \geq 3$, corrects the state norm upper bound:

$$\begin{aligned} \delta_c(\tau_i) &:= \delta_c(\tau_i) \frac{\langle q_d(\tau_{i-1}) \rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{i-2}) \rangle_{\varkappa,\gamma}^r} \\ &= \delta_c(\tau_i) \times \begin{cases} \left(\frac{\gamma}{\varkappa}\right)^r & \text{if } q_d(\tau_{i-1}) = \maltese \;\&\; q_c(\tau_{i-2}) \neq \maltese, \\ \left(\frac{\varkappa}{\gamma}\right)^r & \text{if } q_d(\tau_{i-1}) \neq \maltese \;\&\; q_c(\tau_{i-2}) = \maltese, \\ 1 & \text{if } \begin{array}{c} q_d(\tau_{i-1}) = \maltese \;\&\; q_c(\tau_{i-2}) = \maltese, \\ \text{or} \\ q_d(\tau_{i-1}) \neq \maltese \;\&\; q_c(\tau_{i-2}) \neq \maltese, \end{array} \end{cases} . \end{aligned} \tag{6.11.7}$$

Here γ is the parameter of the controller (from 2) on p. 180), $\varkappa \in (0, 1)$ is the constant from Lemma 3.8.36 (on p. 80), which determines the contraction rate of the quantizer at hand, and $\langle q \rangle_{\varkappa,\gamma}$ is defined in (6.8.12); i.e.,

$$\langle q \rangle_{\varkappa,\gamma} = \begin{cases} \varkappa & \text{if } q \neq \maltese \\ \gamma & \text{otherwise} \end{cases} . \tag{6.11.8}$$

cs.4) Employs the quantizer $\mathfrak{Q}_r$ and computes the quantized value $q_c(\tau_i)$ of the scaled state prognosis at the time τ_{i+1}:

$$\varepsilon(\tau_i) := \left[\delta_c(\tau_i)\right]^{-1} \widehat{x}_c(\tau_{i+1}), \qquad q_c(\tau_i) := \mathfrak{Q}_r\left[\varepsilon(\tau_i)\right]; \tag{6.11.9}$$

cs.5) Encodes this quantized value $q_c(\tau_i)$ by means of the code book $\mathfrak{E}_r$ and thus obtains the code word to be transmitted over the channel during the next operation epoch $[\tau_i : \tau_{i+1})$;

cs.6) Finally, the coder computes the control program

$$\boldsymbol{U}_{i+1}^c = \textbf{col}\left[u_c(\tau_{i+1}), \ldots, u_c(\tau_{i+2} - 1)\right]$$

for the operation epoch $[\tau_{i+1} : \tau_{i+2})$ following the next one $[\tau_i : \tau_{i+1})$ and then corrects the state upper bound:

$$\boldsymbol{U}_{i+1}^c := \delta_c(\tau_i) \mathcal{N}\left[q_c(\tau_i)\right], \qquad \delta_c(\tau_i) := \delta_c(\tau_i) \times \langle q_c(\tau_i) \rangle_{\varkappa,\gamma}^r, \tag{6.11.10}$$

where $\langle q \rangle_{\varkappa,\gamma}$ is given by (6.11.8) and $\mathcal{N}$ is the deadbeat stabilizer.

The decoder (at the times $t = \tau_i, i = 2, 3, \ldots$)

ds.1) Applies the decoding rule $\mathcal{D}_r(\cdot)$ to the data received within the previous operation epoch $[\tau_{i-1} : \tau_i)$ and thus acquires the decoded value $q_d(\tau_i)$ of $q_c(\tau_{i-1})$;

ds.2) Computes the control program (6.11.5) for the next operation epoch $[\tau_i : \tau_{i+1})$ and corrects the state upper bound

$$\boldsymbol{U}_i^d := \delta_d(\tau_i)\mathcal{N}\big[q_d(\tau_i)\big], \qquad \delta_d(\tau_i) := \delta_d(\tau_i) \times \langle q_d(\tau_i)\rangle_{\varkappa,\gamma}^r. \tag{6.11.11}$$

Remark 6.11.16. For uniformity of subsequent formulas, we assume that

$$q_c(\tau_k) := q_d(\tau_{1+k}) := \maltese \qquad \forall k \leq 0,$$

and that the coder at time $t = \tau_0$ performs cs.6) and the decoder at time $t = \tau_1$ accomplishes ds.2) accordingly. Then (6.11.7) is in fact active for $i = 1, 2$ and $\boldsymbol{U}_1^c = \boldsymbol{U}_1^d = \delta_0 \boldsymbol{U}_{\maltese}$. For consistency, we also put $\boldsymbol{U}_0^c := \boldsymbol{U}_0^d := \delta_0 \boldsymbol{U}_{\maltese}$.

Explanation 6.11.17. Although the coder is unaware of the entire sequence of controls $u_d(t)$ actually acting upon the plant, the operation cs.1) is possible. Moreover, the current state $x(\tau_i)$ can be found as a linear function of n previous observations.

This holds since the dynamics of the closed-loop system (6.5.1) (on p. 143) is free $u(t) = 0$ for at least n time steps before τ_i thanks to Definition 6.11.9 and Remark 6.11.1, along with (6.11.1), (6.11.3), (6.11.4), (6.11.11), and Definition 6.11.9.

Remark 6.11.18. The decoded value $q_d(\tau_i)$ from ds.1) may be incorrect $q_d(\tau_i) \neq q_c(\tau_{i-1})$ due to the channel errors.

Explanation 6.11.19. The operation cs.3) is possible thanks to Assumption 6.11.11.

Remark 6.11.20. The operation cs.3) makes the bounds δ_c and δ_d identical whenever the transmission across the channel is errorless.

This claim is more rigorously specified by the lemma to follow. To state it, we mark the values of δ_c and δ_d after and just before the updates in accordance with (6.11.10) and (6.11.11) with the $^+$ and $^-$ indices, respectively. So the value $\delta_c^-(\tau_i)$ is taken after the correction (6.11.7).

Lemma 6.11.21. *The step* cs.3) *ensures that whenever the current transmission is errorless, the next state norm upper bounds produced by the coder and decoder, respectively, are identical:*

$$q_c(\tau_{i-1}) = q_d(\tau_i) \quad \Longrightarrow \quad \delta_c^-(\tau_i) = \delta_d^-(\tau_{i+1}), \qquad i = 1, 2, \dots. \tag{6.11.12}$$

Proof. It suffices to show that for $i = 1, 2, \dots$

$$\delta_c^-(\tau_i) = \delta_d^-(\tau_{i+1}) \left[\frac{\langle q_c(\tau_{i-1})\rangle_{\varkappa,\gamma}}{\langle q_d(\tau_i)\rangle_{\varkappa,\gamma}} \right]^r. \tag{6.11.13}$$

The proof will be by induction on i. For $i = 1$, the claim is evident. Suppose that (6.11.13) holds for some $i \geq 1$. Then

$$\delta_c^-(\tau_{i+1}) \overset{(6.11.7)}{=\!=} \delta_c^+(\tau_i)\frac{\langle q_d(\tau_i)\rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{i-1})\rangle_{\varkappa,\gamma}^r} \overset{(6.11.10)}{=\!=} \delta_c^-(\tau_i)\,\langle q_c(\tau_i)\rangle_{\varkappa,\gamma}^r \frac{\langle q_d(\tau_i)\rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{i-1})\rangle_{\varkappa,\gamma}^r}$$
$$\overset{(6.11.13)}{=\!=} \delta_d^-(\tau_{i+1})\,\langle q_c(\tau_i)\rangle_{\varkappa,\gamma}^r \overset{(6.11.11)}{=\!=} \delta_d^-(\tau_{i+2})\frac{\langle q_c(\tau_i)\rangle_{\varkappa,\gamma}^r}{\langle q_d(\tau_{i+1})\rangle_{\varkappa,\gamma}^r};$$

i.e., (6.11.13) with $i := i+1$ does hold. □

Stabilization by the Coder–Decoder Pair Consuming a Low-Rate Communication Feedback

The main property of the above coder–decoder pair is given by the following proposition.

Proposition 6.11.22. *Suppose that Assumption* 6.11.11 *holds. The coder–decoder pair introduced in this subsection stabilizes the system almost surely, provided that the duration r of the operation epoch is large enough: $r \geq \overline{r}(A, B, \varkappa, W, \gamma, R)$.*

Explanation 6.11.23. Here $\varkappa$ is taken from Lemma 3.8.36 (on p. 80), $W(\cdot|\cdot)$ is the matrix of the channel transition probabilities, and R, γ are the controller parameters from 1), 2) (on p. 180).

The proof of Proposition 6.11.22 will be given in Subsect. 6.11.3. The bound $\overline{r}(A, B, \varkappa, W, \gamma, R)$ can be specified from this proof (see Lemma 6.11.40 on p. 193).

Remark 6.11.24. Proposition 6.11.22 holds for any choice of the alarm control sequence $\boldsymbol{U}_{\maltese}$.

6.11.2 Almost Sure Stabilization by Means of Fixed-Length Code Words in the Absence of a Special Feedback Communication Link

Now we show that Assumption 6.11.11 can be always ensured, even if there is no special feedback communication link. This means that almost sure stabilization by means of fixed-length code words and a controller with a limited computational power can be ensured in the absence of a special feedback communication link.

Communication Feedback by Means of Control

We start with an informal discussion. Assumption 6.11.11 to be ensured means that at time τ_{i+1}, the coder can recognize whether the message determined by the decoder as a result of decoding at time τ_i was the alarm one $\maltese$.

The idea is roughly as follows. In the first case $q_d(\tau_i) = \maltese$, the plant is affected by a scaled (i.e., multiplied by the scalar positive factor) alarm control sequence $\boldsymbol{U}_{\maltese}$ during the epoch $[\tau_i : \tau_{i+1})$ by (6.11.2) and (6.11.11). In the second case, it is driven by a control program $\boldsymbol{U}_i$ produced by the deadbeat stabilizer $\boldsymbol{U}_i = \mathcal{N}(q)$ from a vector $q := \delta_d(\tau_i) q_d(\tau_i) \in \mathbb{R}^n$. So it suffices that the coder be able to recognize at time τ_{i+1} whether the control program $\boldsymbol{U}$ that acted upon the plant during the epoch was the scaled alarm one or of the kind $\mathcal{N}(q), q \in \mathbb{R}^n$. In doing so, the coder should proceed from the sequence of observations received within this epoch:

$$Y_i := \mathbf{col}\,[y(\tau_i), \ldots, y(\tau_{i+1}-1)]. \tag{6.11.14}$$

Now we show that the coder can correctly accomplish this if the alarm control sequence and the duration r of the operation epoch are chosen properly. Indeed, let us first observe that the tuple of observations (6.11.14) is a linear function of the control sequence and n-dimensional state $x(\tau_i)$. In the case where $\boldsymbol{U}_i = \mathcal{N}(q), q \in \mathbb{R}^n$, this sequence lies in an n-dimensional space $\mathbf{Im}\,\mathcal{N}$ as well, since the operator $\mathcal{N}$ is linear. It follows that whenever $q_d(\tau_i) \neq \maltese$, the tuple (6.11.14) lies in a certain specific $2n$-dimensional linear subspace L of the space of all possible observation sequences. However, this space may be of larger dimension for large r. This makes it possible to pick the alarm control sequence so that it generates the sequence of observations not in L. Then the coder may recognize the event $q_d(\tau_i) = \maltese$ by checking the relation

$$Y_i \notin L.$$

To be specific, now we offer a particular example of this scheme.

Picking the Alarm Control Sequence

We first pick

- A control u_* such that $Bu_* \neq 0$ and
- A control sequence $\boldsymbol{U}_- := \mathbf{col}\,[u_0^-, \ldots, u_{n-1}^-]$ that drives the system from the state $x(0) = A^n Bu_*$ to $x(n) = 0$.

The alarm control sequence is defined as follows:

$$\boldsymbol{U}_{\maltese} := \mathbf{col}\,\Big[\underbrace{0, \ldots, 0}_{2n}, u_*, \underbrace{0, \ldots, 0}_{n}, u_0^-, \ldots, u_{n-1}^-, 0, \ldots\Big], \quad s := 4n+1. \tag{6.11.15}$$

Remark 6.11.25. It is easy to see that this sequence drives the system from $x(0) = 0$ to $x(s) = 0$, as is required.

As will be shown (see Remark 6.11.27), this alarm control sequence does ensure correct recognition of the event $q_d(\tau_i) = \maltese$.

Coder–Decoder Pair with No Communication Feedback

This is exactly the coder–decoder pair introduced in Subsect. 6.11.1 in which the coder acquires the knowledge about $q_d(\tau_{i-1})$ to perform the step cs.3) (see p. 183) in such a way that no special feedback communication link is employed.

So the description of the decoder (see p. 183) is not altered, whereas two more steps are inserted into the description of the coder between steps cs.2) and cs.3).

Specifically, the coder first carries out steps cs.1) and cs.2) (see p. 183), then prefaces step cs.3) with the following two ones to recognize the event $q_d(\tau_{i-1}) = \maltese$, and proceeds by executing steps cs.3)–cs.6) (see pp. 183 and 183), as before.

Additional Intermediate Steps

cs.2-3$_1$) Proceeding from the previous measurements, the coder computes the states $x(\tau_{i-1} + 2n)$ and $x(\tau_{i-1} + 3n + 1)$;

cs.2-3$_2$) The coder decides that $q_d(\tau_{i-1}) = \maltese$ if and only if

$$x(\tau_{i-1} + 3n + 1) \neq A^{n+1} x(\tau_{i-1} + 2n).$$

Remark 6.11.26. The step cs.2-3$_1$) is possible. Moreover, the required states can be determined as linear functions of n previous measurements.

This is basically justified by the arguments underlying Explanation 6.11.17 since the dynamics of the system is free $u(t) = 0$ for at least n time steps before the times $\tau_{i-1} + 2n$ and $\tau_{i-1} + 3n + 1$ due to (6.11.1) and (6.11.15).

Remark 6.11.27. The steps cs.2-3$_1$) and cs.2-3$_2$) ensure correct recognition of the event $q_d(\tau_{i-1}) = \maltese$.

Indeed, it suffices to note that due to (6.11.1) and (6.11.15),

$$x(\tau_{i-1}+3n+1) - A^{n+1} x(\tau_{i-1}+2n) = \begin{cases} \delta_d(\tau_{i-1}) A^n B u_* \neq 0 & \text{if} \quad q_d(\tau_{i-1}) = \maltese \\ 0 & \text{otherwise} \end{cases}.$$

Since the coder and decoder at hand are in fact those from the previous subsection, Proposition 6.11.22 remains true for them with Assumption 6.11.11 dropped. This gives rise to the following.

Proposition 6.11.28. *The coder–decoder pair introduced in this subsection stabilizes the system almost surely, provided that the duration r of the operation epoch is large enough: $r \geq \overline{r}(A, B, \varkappa, W, \gamma, R)$.*

Explanation 6.11.29. Explanation 6.11.23 equally concerns the case under current consideration.

Observation 6.11.30. *The coder and decoder considered in this subsection are semi-rational r-step recursive.*[17]

This is straightforward from the constructions of the coder and decoder with taking into account that the employed quantizer is taken from Lemma 3.8.36 (on p. 80) and so is polyhedral,[18] and cs.2-3$_2$) consists of checking a system of linear inequalities.

Remark 6.11.31. The inner dynamic variables z_c and z_d of the coder and decoder, respectively, from (6.3.4) (on p. 139) and (6.5.5) (on p. 145) can be chosen as

$$z_c(ir) := \Big[\delta_c^-(\tau_i), \delta_c^-(\tau_{i-1}), \widehat{x}_c(\tau_{i+1}), \widehat{x}_c(\tau_i), \widehat{x}_c(\tau_{i-1})\Big], \quad z_d(ir) := \delta_d^-(\tau_i).$$

[17] See Definitions 6.3.9 and 6.5.7 on pp. 139 and 145.

[18] See Definition 3.8.8 on p. 70.

As will be shown (see Remark 6.11.41 on p. 195), these inner variables a.s. converge to 0 as $t \to \infty$ under the assumptions of Proposition 6.11.28.

The following claim is immediate from Proposition 6.11.28 and Observation 6.11.30.

Corollary 6.11.32. *Suppose that Assumption* 6.8.1 *(on p.* 161*) holds and that* $H(A) < \mathfrak{c}$. *Then the statement* a) *from Theorem* 6.6.1 *(on p.* 145*) is true.*

The remainder of the section is devoted to the proof of Proposition 6.11.22 and to the completion of the proof of Theorem 6.6.1.

6.11.3 Proof of Proposition 6.11.22

The proof resembles that of Proposition 6.9.8.[19] However there are important differences. They basically proceed from the fact that now the coder and decoder are not completely synchronized via the communication feedback, which is contrary to Proposition 6.9.8. More precisely, the coder and decoder considered in Proposition 6.9.8 produce common error upper bounds δ_c, δ_d and state estimates $\widehat{x}_c, \widehat{x}_d$. Now only the bounds δ_c, δ_d are synchronized in a weaker sense: They are not always common but only when the previous transmission across the channel is errorless (see Lemma 6.11.21). At the same time, the controls produced by the coder are not put in harmony with those generated by the decoder. In fact, the goal of the proof is to demonstrate that being properly adjusted to the current circumstances, the arguments from Subsect. 6.9.3 are not destroyed by these differences.

To start with, we rewrite the state prognosis (6.11.6) in a form more convenient for further analysis.

Another Formula for the State Prognosis Generated by the Coder

Lemma 6.11.33. *The state prognosis* (6.11.6) *obeys the following equation:*

$$\widehat{x}_c(\tau_{i+1}) = \delta_c^-(\tau_{i-1}) A^r \left[\varepsilon(\tau_{i-1}) - \overset{\star}{q}_c(\tau_{i-1})\right] + A^r \sum_{j=\tau_{i-1}}^{\tau_i - 1} A^{\tau_i - 1 - j} B\left[u_d(j) - u_c(j)\right], \qquad i \geq 2. \tag{6.11.16}$$

Here $\overset{\star}{q}$ *is given by* (6.8.12) *(on p.* 164*) and* $\varepsilon(\tau_{i-1})$ *is defined by* (6.11.9)*, where* $i := i-1$ *and* $\delta_c(\tau_{i-1}) = \delta_c^-(\tau_{i-1})$.

Proof. Suppose first that $q_c(\tau_{i-1}) \neq$ ✠. Due to the first formula from (6.11.10) with $i := i-1$ and the definition of the deadbeat stabilizer, the sequence of controls

$$u_c(\tau_i), \dots, u_c(\tau_{i+1} - 1) \tag{6.11.17}$$

[19] See Subsect. 6.9.3 starting at p. 171.

drives the system from the state $\delta_c^-(\tau_{i-1})q_c(\tau_{i-1})$ at time τ_i to the state 0 at time $\tau_i + n$. Since $u_c(t) = 0$ for $t = \tau_i + n, \ldots, \tau_{i+1} - 1$, the state 0 is kept unchanged until the time $\tau_{i+1} = \tau_i + r_0$. Hence

$$\delta_c^-(\tau_{i-1})A^r \overset{\star}{q}_c (\tau_{i-1}) + \sum_{j=\tau_i}^{\tau_{i+1}-1} A^{\tau_{i+1}-1-j}Bu_c(j) = 0. \tag{6.11.18}$$

This is still true if $q_c(\tau_{i-1}) = \maltese$. Indeed, then $\overset{\star}{q}_c (\tau_{i-1}) = 0$, whereas (6.11.17) is the alarm control sequence up to a scalar multiplier and so it drives the system from the state $x(\tau_i) = 0$ to $x(\tau_{i+1}) = 0$.

Subtraction (6.11.6) and (6.11.18) yields

$$\begin{aligned}\widehat{x}_c(\tau_{i+1}) &= A^r\left[x(\tau_i) - \delta_c^-(\tau_{i-1}) \overset{\star}{q}_c (\tau_{i-1})\right] \\ &= \delta_c^-(\tau_{i-1})A^r\left[\delta_c^-(\tau_{i-1})^{-1}\widehat{x}_c(\tau_i) - \overset{\star}{q}_c (\tau_{i-1})\right] + A^r\left[x(\tau_i) - \widehat{x}_c(\tau_i)\right].\end{aligned}$$

Here by (6.11.6) with $i := i-1$ and (6.5.1) (on p. 143),

$$\widehat{x}_c(\tau_i) = A^r x(\tau_{i-1}) + \sum_{j=\tau_{i-1}}^{\tau_i - 1} A^{\tau_i-1-j}Bu_c(j),$$

$$x(\tau_i) = A^r x(\tau_{i-1}) + \sum_{j=\tau_{i-1}}^{\tau_i - 1} A^{\tau_i-1-j}Bu_d(j).$$

As a result, we arrive at (6.11.16) by taking into account (6.11.9). □

Frequency of Decoding Errors

Now we consider the stochastic process generated by the coder and decoder in connection with the system (6.5.1) (on p. 143). The symbols E_i and S_i stand for the code words formed by the coder at time τ_{i-1} and received by the decoder at time τ_i, respectively. We also invoke the indicator function (6.9.6) (on p. 172) of the decoding error $\mathfrak{D}_r[S_i] \neq E_i$, and we pick $0 < F < F(R, W)$, where $F(R, W)$ is taken from (6.8.6) (on p. 162).

Observation 6.11.34. *Lemma* 6.9.11 *(on p.* 172*) is still true with* $r_0 := r$.

This can be easily seen by retracing the arguments from the proof of this lemma.

Corollary 6.11.35. *For the indicator function*

$$\widehat{I}^{\text{err}}(i) \leftrightarrow I^{\text{err}}(i) = 1 \vee I^{\text{err}}(i-1) = 1 \vee I^{\text{err}}(i-2) = 1, \tag{6.11.19}$$

the following inequality holds a.s.:

$$\overline{\lim_{k\to\infty}} \frac{1}{k} \sum_{i=1}^{k} \widehat{I}^{\text{err}}(i) \leq 3 \cdot 2^{-Fr}. \tag{6.11.20}$$

Indeed, this is immediate from (6.9.7) (on p. 172) and the inequality

$$\widehat{I}^{\text{err}}(i) \leq I^{\text{err}}(i) + I^{\text{err}}(i-1) + I^{\text{err}}(i-2).$$

Influence of the Channel Errors on the Evolution of the Closed-Loop System

To analyze this influence, we first introduce the following linear operators by employing the deadbeat stabilizer (6.11.1) and its length $L(\mathcal{N})$[20]:

$$\mathfrak{C}[\,\mathbf{col}\,(u_0,\ldots,u_{L(\mathcal{N})-1})] := \sum_{j=0}^{L(\mathcal{N})-1} A^{-1-j}Bu_j;$$

$$x \in \mathbb{R}^n \xrightarrow{\mathcal{N}} \big[u(0),\ldots,u(n-1),0,0,\ldots\big] \xrightarrow{\mathfrak{C}} \mathfrak{B}(x). \quad (6.11.21)$$

We also put $\tau_{-1} := -1$ and invoke Remark 6.11.16 (on p. 184) and the indicator functions (6.9.10).

We first study the evolution of

$$\delta_i := \delta_c^-(\tau_i) \quad \text{and} \quad z_i := \|\widehat{x}_c(\tau_{i+1})\|. \quad (6.11.22)$$

Lemma 6.11.36. *The following relations hold for $j \geq 1$ and $i \geq 2$:*

$$\delta_j = \delta_{j-1}\Big\{\varkappa^r\big[I_0(j) + I_0^{\text{err}}(j) + I_{d\maltese}^{\text{err}}(j)\big] + \gamma^r\big[I_{\maltese}(j) + I_{c\maltese}^{\text{err}}(j)\big]\Big\}\times$$
$$\times\left\{\left(\frac{\gamma}{\varkappa}\right)^r I_{d\maltese}^{\text{err}}(j-1) + \left(\frac{\varkappa}{\gamma}\right)^r I_{c\maltese}^{\text{err}}(j-1) + [1 - I_{d\maltese}^{\text{err}}(j-1) - I_{c\maltese}^{\text{err}}(j-1)]\right\}; \quad (6.11.23)$$

$$z_i \leq z_{i-1}\|A\|^r\big[I_{\maltese}(i) + I_{c\maltese}^{\text{err}}(i)\big] + \delta_{i-1}\varkappa^{2r}\big[I_0(i) + I_0^{\text{err}}(i) + I_{d\maltese}^{\text{err}}(i)\big] + \delta_{i-2}d(r)\widehat{I}^{\text{err}}(i). \quad (6.11.24)$$

Here $\varkappa \in (0,1)$ and $\widehat{I}^{\text{err}}(i)$ are taken from Lemma 3.8.36 (on p. 80) and Corollary 6.11.35, respectively; $\gamma > \|A\|$ is the parameter of the controller from 2) on p. 180; $I_0, I_{\maltese}, I_0^{\text{err}}, I_{d\maltese}^{\text{err}}, I_{c\maltese}^{\text{err}}$ are defined by (6.9.10) (on p. 173); and

$$d(r) := \|A\|^{2r}\left[1 + \left(\frac{\gamma}{\varkappa}\right)^r\right]\max\big\{\|\mathfrak{B}\|, \|\mathfrak{C}\boldsymbol{U}_{\maltese}\|\big\}, \quad (6.11.25)$$

where $\boldsymbol{U}_{\maltese}$ is the alarm control sequence and $\mathfrak{B}, \mathfrak{C}$ are given by (6.11.21).

Proof. We start with proving (6.11.23):

$$\delta_j \overset{(6.11.22)}{=\!=} \delta_c^-(\tau_j) \overset{(6.11.7)}{=\!=} \delta_c^+(\tau_{j-1})\frac{\langle q_d(\tau_{j-1})\rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{j-2})\rangle_{\varkappa,\gamma}^r}$$
$$\overset{(6.11.10),(6.11.22)}{=\!=\!=} \delta_{j-1}\,\langle q_c(\tau_{j-1})\rangle_{\varkappa,\gamma}^r\,\frac{\langle q_d(\tau_{j-1})\rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{j-2})\rangle_{\varkappa,\gamma}^r}.$$

[20]See Definition 6.11.9 on p. 181.

Due to (6.9.10) (on p. 173) and (6.11.8), here the second multiplier and the ratio in the last expression equal the first and second expressions in the curly brackets { } from (6.11.23), respectively.

To justify (6.11.24), we denote by s' and s'' the first and second summands from (6.11.16), respectively. Since in (6.11.16), $q_c(\tau_{i-1})$ is the quantized value of $\varepsilon(\tau_{i-1})$ by means of the r-contracted quantizer $\mathfrak{Q}_r$ with the contraction rate $\rho_{\mathfrak{Q}_r} = \varkappa^{2r}$, relation (3.8.9) (on p. 70) yields

$$\|s'\| \leq \delta_{i-1} \times \begin{Bmatrix} \varkappa^{2r} & \text{if } q_c(\tau_{i-1}) \neq \maltese \\ \|A\|^r \|\varepsilon(\tau_{i-1})\| & \text{if } q_c(\tau_{i-1}) = \maltese \end{Bmatrix}$$

$$\overset{(6.9.10)}{=\!=} \delta_{i-1}\varkappa^{2r}\big[I_0(i)+I_0^{\text{err}}(i)+I_{\text{d}\maltese}^{\text{err}}(i)\big]+\delta_c^-(\tau_{i-1})\|\varepsilon(\tau_{i-1})\|\|A\|^r\big[I_\maltese(i)+I_{\text{c}\maltese}^{\text{err}}(i)\big].$$

Here

$$\delta_c^-(\tau_{i-1})\|\varepsilon(\tau_{i-1})\| = \|\widehat{x}_c(\tau_i)\| = z_{i-1}$$

by (6.11.9) and (6.11.22). As a result, we see that $\|s'\|$ does not exceed the sum of the first two summands from (6.11.24).

The second summand s'' from (6.11.16) can be rewritten in the following form due to (6.11.10), (6.11.11), and (6.11.21):

$$s'' = A^{2r}\Big\{\delta_d^-(\tau_{i-1})\beta\big[q_d(\tau_{i-1})\big] - \delta_c^-(\tau_{i-2})\beta\big[q_c(\tau_{i-2})\big]\Big\},$$

$$\text{where} \quad \beta(q) := \begin{cases} \mathfrak{B}(q) \text{ if } q \neq \maltese \\ \mathfrak{C}\boldsymbol{U}_\maltese \text{ otherwise} \end{cases};$$

$$s'' \overset{(6.11.13),(6.11.22)}{=\!=\!=\!=} \delta_{i-2}A^{2r}\left\{\frac{\langle q_d(\tau_{i-2})\rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{i-3})\rangle_{\varkappa,\gamma}^r}\beta\big[q_d(\tau_{i-1})\big] - \beta\big[q_c(\tau_{i-2})\big]\right\}. \tag{6.11.26}$$

Whenever $\widehat{I}^{\text{err}}(i) = 0$, we have by (6.9.6) (on p. 172) and (6.11.19),

$$q_d(\tau_{i-1}) = q_c(\tau_{i-2}), \quad q_d(\tau_{i-2}) = q_c(\tau_{i-3}),$$

and so the expression embraced by the last curly brackets {} in (6.11.26) vanishes. In any case,

$$\|\beta(q)\| \leq \max\{\|\mathfrak{B}\|, \|\mathfrak{C}\boldsymbol{U}_\maltese\|\} \quad \text{for} \quad q := q_d(\tau_{i-1}), q_c(\tau_{i-2}),$$

since

$$q \neq \maltese \Rightarrow \|q\| \leq 1 \Rightarrow \|\beta(q)\| \leq \|\mathfrak{B}\|.$$

At the same time,

$$\frac{\langle q_d(\tau_{i-2})\rangle_{\varkappa,\gamma}}{\langle q_c(\tau_{i-3})\rangle_{\varkappa,\gamma}} \leq \gamma/\varkappa$$

due to (6.11.8). As a result, we see that $\|s''\|$ does not exceed the last summand from (6.11.24), which completes the proof. □

Now we focus on the evolution of the ratio $\xi_i := z_i/\delta_i$ determining whether the alarm message $\maltese$ is dispatched over the channel:

$$q_c(\tau_i) = \maltese \Leftrightarrow \xi_i = z_i/\delta_i > 1. \tag{6.11.27}$$

Lemma 6.11.37. *For $i \geq 2$, the following inequality holds:*

$$\xi_i \leq \begin{Bmatrix} \rho\xi_{i-1} & \text{if } \xi_{i-1} > 1 \\ \varkappa^r & \text{if } \xi_{i-1} \leq 1 \end{Bmatrix} \left[1 - \widehat{I}^{\text{err}}(i)\right] + b/2\left[\xi_{i-1} + 1\right]\widehat{I}^{\text{err}}(i), \tag{6.11.28}$$

where the indicator function $\widehat{I}^{\text{err}}(i)$ was introduced by (6.11.19) *and*

$$\rho := \left(\frac{\|A\|}{\gamma}\right)^r, \quad b := 2\left(\frac{\gamma}{\varkappa^2}\right)^{2r}\left[1 + d(r)\right]. \tag{6.11.29}$$

Proof. Thanks to (6.11.23), (6.11.24), and (6.11.29)

$$\xi_i \leq \Big\{\xi_{i-1}\rho\big[I_{\maltese}(i) + I^{\text{err}}_{\text{c}\maltese}(i)\big] + \varkappa^r\left[I_0(i) + I^{\text{err}}_0(i) + I^{\text{err}}_{\text{d}\maltese}(i)\right]\Big\}\times$$
$$\times \underbrace{\left\{\left(\frac{\varkappa}{\gamma}\right)^r I^{\text{err}}_{\text{d}\maltese}(i-1) + \left(\frac{\gamma}{\varkappa}\right)^r I^{\text{err}}_{\text{c}\maltese}(i-1) + \left[1 - I^{\text{err}}_{\text{d}\maltese}(i-1) - I^{\text{err}}_{\text{c}\maltese}(i-1)\right]\right\}}_{F}$$
$$+ \frac{\delta_{i-2}}{\delta_i} d(r)\widehat{I}^{\text{err}}(i).$$

By (6.11.23),

$$\frac{\delta_{j-1}}{\delta_j} \leq \gamma^r\varkappa^{-2r} \Rightarrow \frac{\delta_{i-2}}{\delta_i} \leq \gamma^{2r}\varkappa^{-4r}.$$

Due to (6.9.6) (on p. 172), (6.9.10) (on p. 173), and (6.11.19),

$$F \leq \left(\frac{\gamma}{\varkappa}\right)^r \widehat{I}^{\text{err}}(i) + 1 - \widehat{I}^{\text{err}}(i), \qquad I(i)[1 - \widehat{I}^{\text{err}}(i)] = 0,$$

for $I := I^{\text{err}}_0, I^{\text{err}}_{\text{d}\maltese} I^{\text{err}}_{\text{c}\maltese}$. Hence

$$\xi_i \leq \Big[\xi_{i-1}\rho I_{\maltese}(i) + \varkappa^r I_0(i)\Big]\Big[1 - \widehat{I}^{\text{err}}(i)\Big]$$
$$+ \Big[\xi_{i-1}\rho\big\{I_{\maltese}(i) + I^{\text{err}}_{\text{c}\maltese}(i)\big\} + \varkappa^r\left\{I^{\text{err}}_0(i) + I^{\text{err}}_{\text{d}\maltese}(i) + I_0(i)\right\}\Big] \times \left(\frac{\gamma}{\varkappa}\right)^r \widehat{I}^{\text{err}}(i)$$
$$+ \gamma^{2r}\varkappa^{-4r} d(r)\widehat{I}^{\text{err}}(i) \leq \Big[\xi_{i-1}\rho I_{\maltese}(i) + \varkappa^r I_0(i)\Big]\Big[1 - \widehat{I}^{\text{err}}(i)\Big]$$
$$+ \left\{\left(\frac{\gamma}{\varkappa}\right)^r\left[\xi_{i-1}\rho + \varkappa^r\right] + \gamma^{2r}\varkappa^{-4r} d(r)\right\}\widehat{I}^{\text{err}}(i).$$

Here $\rho < 1$ owing to (6.11.29) and 2) on p. 180, and $\varkappa < 1$. So the factor multiplying $\widehat{I}^{\text{err}}(i)$ in the last summand does not exceed

$$(\xi_{i-1}+1)\left(\frac{\gamma}{\varkappa}\right)^r + \left(\frac{\gamma}{\varkappa^2}\right)^{2r} d(r) \leq \left(\frac{\gamma}{\varkappa^2}\right)^{2r}[1+d(r)](\xi_{i-1}+1) \overset{(6.11.29)}{=\!=} b/2(\xi_{i-1}+1).$$

Summarizing, we arrive at (6.11.28). □

Remark 6.11.38. By (6.11.22) and (6.11.27), $\xi_i = \left[\delta_c^-(\tau_i)\right]^{-1}\|\widehat{x}_c(\tau_{i+1})\|$. Since the state prognosis $\widehat{x}_c(\tau_{i+1})$ is computed by the coder only for $i \geq 1$, the quantity ξ_i is defined only for such i's. Nevertheless, we pick $\xi_0 \in (1, \infty)$ for technical convenience: This makes the equivalence from (6.11.27) true for all i due to Remark 6.11.16 (on p. 184).

Corollary 6.11.39. *The indicator function*

$$I(i) \longleftrightarrow \widehat{I}^{\text{err}}(i) = 1 \vee \xi_{i-1} > 1 \tag{6.11.30}$$

almost surely obeys the inequality

$$\overline{\lim_{k\to\infty}} \frac{1}{k}\sum_{i=1}^{k} I(i) \leq \overline{p} := 3 \cdot 2^{-Fr}\left\{2 + \frac{\log_2 b}{\log_2[\rho^{-1}]}\right\}, \tag{6.11.31}$$

where F is taken from (6.11.20)*.*

Indeed, thanks to Lemma 6.11.37 and Observation 6.9.19 (on p. 176),

$$\overline{\lim_{k\to\infty}} \frac{1}{k}\sum_{i=1}^{k} I(i) \leq \left\{2 + \frac{\log b}{\log[\rho^{-1}]}\right\} \times \overline{\lim_{k\to\infty}} \frac{1}{k}\sum_{i=1}^{k} \widehat{I}^{\text{err}}(i) \xrightarrow{(6.11.20)} (6.11.31).$$

Sufficient Conditions for Almost Sure Stability

Now we bring the pieces together. We start with conditions, which can be used to pick the duration r of the operation epoch in order to ensure almost sure stability of the closed-loop system.

Lemma 6.11.40. *The coder and decoder at hand stabilize the system a.s. whenever*

$$\omega := \log_2[\varkappa^{-1}] - 2\overline{p}\{\log_2 \gamma + \log_2[\varkappa^{-1}]\} > 0 \quad \text{and} \quad \chi := \omega(1-\overline{p}) - \overline{p}\log_2\|A\| > 0. \tag{6.11.32}$$

Here $\varkappa \in (0,1)$ is the constant from Lemma 3.8.36 *(on p.* 80*), γ is the parameter of the controller from* 2) *on p.* 180*, and $\overline{p}$ is given by* (6.11.31)*, along with* (6.11.25) *and* (6.11.29)*.*

Proof. The symbol c (with a possible index) will be used to denote random constants independent of i and r. For any $\alpha > 0$, (6.11.31) implies

$$\mathfrak{I}(k) := \sum_{i=1}^{k} I(i) \leq k(\overline{p} + \alpha)$$

for $k \approx \infty$. Since $\varkappa < 1 < \gamma$, (6.9.10) (on p. 173), (6.11.19), (6.11.23), and (6.11.30) yield

$$\delta_i \leq \delta_{i-1} \left\{ \varkappa^r [1 - I(i)] + \left(\frac{\gamma^2}{\varkappa} \right)^r I(i) \right\} \quad \forall i \geq 1 \Rightarrow \delta_k \leq \delta_0 \varkappa^{rk} \prod_{i=1}^{k} \left(\frac{\gamma}{\varkappa} \right)^{2rI(i)}$$
$$= \delta_0 \varkappa^{rk} \left(\frac{\gamma}{\varkappa} \right)^{2r\Im(k)} \overset{k \approx \infty}{\leq} \delta_0 \varkappa^{rk} \left(\frac{\gamma}{\varkappa} \right)^{2rk(\overline{p}+\alpha)} = \delta_0 2^{-rk\omega_\alpha},$$
$$\text{where} \quad \omega_\alpha := \log_2[\varkappa^{-1}] - 2[\overline{p} + \alpha] \big\{ \log_2 \gamma + \log_2[\varkappa^{-1}] \big\} \xrightarrow{\alpha \to 0} \omega > 0.$$

Thus for $\alpha \approx 0$, we have $\omega_\alpha > 0$ and

$$\delta_i \leq c' 2^{-ir\omega_\alpha} \to 0 \quad \text{as} \quad i \to \infty. \tag{6.11.33}$$

This along with (6.11.10), (6.11.11), (6.11.13), and (6.11.22) imply that

$$u_c(t) \to 0, \quad u(t) = u_d(t) \to 0 \qquad \text{as} \quad t \to \infty. \tag{6.11.34}$$

In particular, the second relation from (6.5.3) (on p. 144) holds.

To prove the first one, we note that due to (6.11.24) and (6.11.33)

$$z_i \leq z_{i-1} \|A\|^r I(i) + c''[\delta_{i-1} + \delta_{i-2}] \leq z_{i-1} \|A\|^r I(i) + c 2^{-ir\omega_\alpha} \ \forall i \geq 2;$$
$$z_k \leq z_1 \prod_{i=2}^{k} \Big[\|A\|^r I(i) \Big] + c \sum_{i=2}^{k} 2^{-ir\omega_\alpha} \prod_{j=i+1}^{k} \Big[\|A\|^r I(j) \Big].$$

(We recall that $\prod_{j=\alpha}^{\beta} \ldots := 1$ whenever $\beta < \alpha$.) The first relation from (6.11.32) implies $\overline{p} < 1$. So an index $i \geq 2$ exists such that $I(i) = 0$ due to (6.11.31). It follows that for $k \approx \infty$, the first summand vanishes and

$$z_k \leq c \sum_{i=k-l}^{k} 2^{-ir\omega_\alpha} \|A\|^{r(k-i)},$$

where $\{k - l + 1, \ldots, k\}$ is the largest subinterval of the set

$$\Omega_k := \{2 \leq i \leq k : I(i) = 1\}$$

containing k. (If $k \notin \Omega_k$, then $l := 0$.) Now we take into account that

$$l \leq \sum_{i=1}^{k} I(i) = \Im(k) \leq k(\overline{p} + \alpha) \qquad \forall k \approx \infty$$

and proceed as follows:

$$z_k \le c2^{-kr\omega_\alpha} \sum_{i=k-l}^{k} 2^{(k-i)r\omega_\alpha} \|A\|^{r(k-i)} = c2^{-kr\omega_\alpha} \sum_{i=0}^{l} \left(2^{\omega_\alpha}\|A\|\right)^{ri}$$
$$\le c\frac{2^{-kr\omega_\alpha}\left(2^{\omega_\alpha}\|A\|\right)^{rl}}{1-\left(2^{\omega_\alpha}\|A\|\right)^{-r}} \overset{k\approx\infty}{\le} c\frac{2^{-kr\omega_\alpha}\left(2^{\omega_\alpha}\|A\|\right)^{rk(\overline{p}+\alpha)}}{1-\left(2^{\omega_\alpha}\|A\|\right)^{-r}}$$
$$= \frac{c}{1-\left(2^{\omega_\alpha}\|A\|\right)^{-r}} 2^{-kr\chi_\alpha},$$
$$\text{where} \quad \chi_\alpha := \omega_\alpha[1-(\overline{p}+\alpha)] - (\overline{p}+\alpha)\log_2\|A\| \xrightarrow[\alpha\to 0]{(6.11.32)} \chi > 0.$$

Thus $\chi_\alpha > 0$ for $\alpha \approx 0$. So

$$z_k \overset{(6.11.22)}{=\!=} \|\widehat{x}_c(\tau_{k+1})\| \to 0 \quad \text{as} \quad k \to \infty. \tag{6.11.35}$$

This and (6.11.6), (6.11.34) yield

$$A^r x(\tau_i) \to 0 \quad \text{as} \quad i \to \infty.$$

Since the matrix A has no stable modes by Assumption 6.8.1 (on p. 161), the matrix A^{-r} is well defined and so $x(\tau_i) \to 0$ as $i \to \infty$. To obtain the first relation from (6.5.3) (on p. 144), we note that for $\tau_i \le t < \tau_{i+1} = \tau_i + r$,

$$\|x(t)\| = \left\| A^{t-\tau_i} x(\tau_i) + \sum_{j=\tau_i}^{t-1} A^{t-1-j} B u_d(j) \right\|$$
$$\le \|A\|^r \left(\|x(\tau_i)\| + \|B\| \sum_{j=\tau_i}^{\tau_{i+1}-1} \|u_d(j)\| \right)$$

and invoke (6.11.34). □

Remark 6.11.41. As follows from Remark 6.11.31, (6.11.13), and (6.11.33)–(6.11.35), the inner dynamical variables of the coder and decoder converge to 0 almost surely.

Proof of Proposition 6.11.22

By Lemma 6.11.40, it suffices to show that (6.11.32) does hold whenever r is large enough. Owing to (6.11.25) and (6.11.29),

$$\frac{1}{r}\log_2[1+d(r)] = \frac{1}{r}\Big\{\log_2 d(r)+\log_2\Big[1+\frac{1}{d(r)}\Big]\Big\} = 2\log_2\|A\|+\log_2\gamma+\log_2\frac{1}{\varkappa}$$
$$+\frac{1}{r}\log_2\left[1+\left(\frac{\varkappa}{\gamma}\right)^r\right]+\frac{1}{r}\log_2\max\{\|\mathfrak{B}\|,\|\mathfrak{C}\boldsymbol{U}_{\maltese}\|\}+\frac{1}{r}\log_2[1+d(r)^{-1}]$$
$$\xrightarrow{r\to\infty}\Delta_\infty := 2\log_2\|A\|+\log_2\gamma+\log_2\varkappa^{-1},$$
$$\frac{\log_2 b}{\log_2[\rho^{-1}]} = \frac{1+2r[\log_2\gamma+2\log_2\varkappa^{-1}]+\log_2[1+d(r)]}{r[\log_2\gamma-\log_2\|A\|]}$$
$$\xrightarrow{r\to\infty}\frac{2[\log_2\gamma+2\log_2\varkappa^{-1}]+\Delta_\infty}{\log_2\gamma-\log_2\|A\|}.$$

This and (6.11.31) yield $\overline{p}\to 0$ as $r\to\infty$, and we see that (6.11.32) does hold for $r\approx\infty$. □

6.11.4 Completion of the Proof of Theorem 6.6.1

By Corollary 6.11.32, the implication $\mathfrak{c} > H(A) \Rightarrow$ a) is true for systems with no stable modes. Now we consider the general case of systems with both unstable and stable modes.

Suppose that $\mathfrak{c} > H(A)$, consider the invariant subspaces M_{unst} and M_{st} of the matrix A related to the unstable $\{\lambda : |\lambda| \geq 1\}$ and stable $\{\lambda : |\lambda| < 1\}$ parts of its spectrum, respectively, and denote by π_+ and π_- the projector onto M_{unst} parallel to M_{st} and vice versa, respectively, and by A_+ and A_- the operator A acting in M_{unst} and M_{st}, respectively. The claim a) is true for the system

$$x_+(t+1) = A_+x_+(t)+\pi_+Bu(t),\; x_+(0) := \pi_+x_0,\; y_+(t) = Cx_+(t) \quad (6.11.36)$$

by Corollary 6.11.32 since $H(A) = H(A_+)$.

Let us consider a coder and decoder stabilizing this system. While constructing them, let us employ the state dimension n of the original system in (6.11.3) and the alarm control sequence (6.11.15). Now we apply this coder and decoder to the primal system (6.5.1). In doing so, we also alter the coder's steps cs.1) (on p. 182) and cs.2-3_1) (on p. 187), where it determines the state $x_+(\tau)$ for $\tau = \tau_i, \tau_i+2n, \tau_i+3n+1$. Formerly this was done on the basis of the past measurements from (6.11.36). Now we employ the observations from (6.5.1). Then thanks to Assumption 6.3.6 (on p. 137), it is possible to compute $\pi_+x(\tau_i) = x_+(\tau_i)$ because the dynamics of the system (6.5.1) is free $u(t) = 0$ at least n time steps before τ_i. Hence the coder and decoder at hand can be applied to the primal system indeed.

As a result, we see that for the corresponding closed-loop system,

$$\|\pi_+x(t)\|\to 0 \quad \text{and} \quad \|u(t)\|\to 0 \qquad \text{as} \quad t\to\infty \qquad \text{a.s.} \qquad (6.11.37)$$

So to complete the proof, it suffices to show that $x_-(t) := \pi_-x(t)\to 0$ whenever (6.11.37) holds. To this end, we note that

$$x_-(t+1) = A_-x_-(t)+\pi_-Bu(t), \qquad \|A_-^m\|\leq c\mu^m,\; m = 0,1,2,\ldots$$

for some $\mu \in (0,1)$. Hence for any given t_* and $t > t_*$, we have

$$\|x_-(t)\| = \left\| A_-^t x_-(0) + \sum_{j=0}^{t-1} A_-^{t-1-j} \pi_- B u(j) \right\| \leq c\mu^t \|x_-(0)\|$$

$$+ c\|B\|\|\pi_-\| \left[\sum_{j=0}^{t_*} \mu^{t-1-j} \|u(j)\| + \sum_{j=t_*}^{t-1} \mu^{t-1-j} \|u(j)\| \right],$$

$$\overline{\lim_{t\to\infty}} \|x_-(t)\| = c\|B\|\|\pi_-\| \overline{\lim_{t\to\infty}} \sum_{j=t_*}^{t-1} \mu^{t-1-j} \|u(j)\|$$

$$\leq \frac{c\|B\|\|\pi_-\|}{1-\mu} \sup_{t \geq t_*} \|u(t)\| \to 0 \quad \text{as} \quad t_* \to \infty,$$

where the last relation follows from (6.11.37). Thus a) does hold.

The implications a) $\Rightarrow$ b) $\Rightarrow$ c) $\Rightarrow$ d) are apparent, whereas d) $\Rightarrow$ $\mathfrak{c} \geq H(A)$ holds by Remark 6.7.3 (on p. 148). $\square$

7

An Analog of Shannon Information Theory: State Estimation and Stabilization of Linear Noisy Plants via Noisy Discrete Channels

7.1 Introduction

In this chapter, we continue to address state estimation and stabilization over noisy channels for discrete-time linear partially observed systems. As compared with the previous chapter, the critical feature of this one is the account for the plant disturbances and sensor noises. The major points concern the case where these disturbances are uniformly and arbitrarily small (at any sample and time). We demonstrate that in the face of both channel and system noises, the strong objective of trajectory-wise (i.e., almost sure) stability or observability cannot be achieved by any means and under any circumstances for many discrete memoryless channels of practical interest, although it may yet be achieved for rather special channels.

We offer an exhaustive description of these two classes of channels. This description results from showing that the capability of the noisy channel to ensure almost sure stability/observability of the plant is identical to its capability to transmit information with the zero probability of error. The latter capability is studied in the zero error information theory.[1] This theory reveals noisy channels capable of errorless transmission of information and offers the corresponding coding–decoding schemes. The zero error capacity $\mathfrak{c}_0$ [189] is the standard parameter characterizing the maximum rate at which data can be transmitted over the channel with no error. The results of this chapter state that the boundary of the almost sure stabilizability/observability domain is given by the channel zero error capacity.

We also show that if this boundary is trespassed, an unstable linear system can never be stabilized or observed: The error is almost surely unbounded for all nonanticipating time-varying algorithms of stabilization/observation (with infinite memories). It should be stressed once more that this holds under uniformly small plant disturbances. So this phenomenon has nothing in common with, e.g., the well-known fact that for the stable linear plant affected by the Gaussian white noise, the stabilization error is yet a.s. unbounded. Indeed, the latter unboundedness ultimately results from the fact that the sample sequences of the white noise are a.s. unbounded. On the

[1] We refer the reader to [84] for an excellent survey of this area.

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_7,

contrary, we show that in the face of channel errors, external disturbances obeying a common and arbitrarily small deterministic bound at any sample and time unavoidably accumulate and cause, sooner or later, arbitrarily large stabilization/estimation errors.

Although noisy channels with a positive zero error capacity exist, the zero error capacity of many communication channels of practical interest is equal to zero [84, 214]. For example, this holds for the erasure channel with the probability of erasures $p > 0$, as well as for the binary symmetric channel with the crossover probability $0 < p < 1$. The above negative result implies that an asymptotically unstable linear plant can be neither stabilized nor observed with a bounded error over a discrete memoryless channel whose zero error capacity is zero.

A negative fact similar in spirit was established in [164] for a simple scalar stochastic process (random walk) and binary channel. This fact concerns a small and special class of estimation/stabilization schemes that employ block encoders with a fixed block length and static decoders producing only a finite number of outputs. This strong simplification played a crucial role in the proof from [164]. On the contrary, the results of this chapter deal with all time-varying nonanticipating deterministic algorithms of stabilization/estimation. They enjoy unlimited memories and produce outputs (estimates/controls) whose number may go to infinity as time progresses. Furthermore, we examine general linear plants and discrete memoryless channels.

The works [102, 103] deal with the channel that transmits binary code words with dropping a random number of concluding bits. This particular case of a discrete memoryless channel generalizes the erasure channel and is motivated by certain wireless communication applications in [101]. It was shown in [102, 103] that for a noise-free LTI plant to be uniformly stabilizable (i.e., with an error uniformly bounded over the initial states from the unit ball), it is necessary that a certain number $r_{\min}$ of bits is not lost under any circumstances (i.e., with probability 1), and this number $r_{\min}$ exceeds the topological entropy $H(A)$ of the open-loop system. Since $r_{\min}$ equals the zero error capacity of the channel at hand, this claim is in harmony with the results of this chapter. However there is a difference. It concerns not only the fact that this chapter examines general discrete memoryless channels with data distortion more complicated than dropout of several bits. As compared with the current chapter, the works [102, 103] justify the necessity of the zero error capacity bound for stability in a stronger sense. To illustrate this, we note that the lack of uniform stability does not exclude that the worst case, where the error is unbounded, holds only with a very small (up to negligible) probability (given by the probability density of the initial state). Moreover, the results of Chap. 6 state that exactly this situation holds whenever the necessary condition $H(A) < r_{\min}$ [102, 103] for the uniform stability is violated but $H(A)$ is less than the Shannon ordinary capacity of the channel at hand. Then the stabilization error can be made almost surely converging to zero, as was shown in Chap. 6. This implies that the plant can be uniformly stabilized with as large a probability as desired,[2] even if the necessary

[2] In other words, the error is bounded uniformly over all initial states except for those from a set with an arbitrarily small probability measure.

condition from [102, 103] is violated. In contradistinction to [102, 103], the results of this chapter show that the violation of the zero error capacity bound implies that the stabilization error is unbounded with probability 1. It should be also stressed that the above results from [102, 103] concern noise-free plants, whereas this chapter is devoted to the study of phenomena related to the interplay of an additive noise in the plant with the channel errors.

Mean-square (and more generally mth moment) stabilizability/observability under communication capacity constraints and exhogenous additive stochastic disturbances was studied in, e.g., [77, 102, 103, 138, 164–166, 192, 193] for scalar linear plants and noisy channels [77, 102, 103, 164–166, 192, 193] and multidimensional linear plants and noiseless channels [138]. The stronger trajectory-wise stabilizability/observability was examined in, e.g., [202, 204, 215] for noiseless channels in a deterministic setting. The observers/controllers proposed in [77, 138, 164, 165, 192, 193, 202, 204, 215] use some parameters of the system noise.

Unlike these works, the observer/controller proposed in this chapter does not rely on the knowledge of parameters of the plant noises. In other words, the same observer/controller ensures bounded estimation/stabilization error under arbitrary bounded noise, although the accuracy of estimation/stabilization depends on the noise level. This is of interest, e.g., in the cases where statistical knowledge of the noise is not available or the noise properties may change during the system operation.

Traditionally, the zero error capacity is considered in both cases where the (feedforward) communication channel is and is not equipped with a (perfect) feedback communication link, respectively. Via such a link, the informant gets aware about the results of the previous transmissions across the erroneous feedforward channel. The zero error capacity may in some cases, be greater with feedback than without [189]. The relationship between these two kinds of the zero error capacity and the observability domain is rather straightforward: This domain is given by the zero error capacity with or without a feedback, depending on whether this feedback is in fact available. On the contrary, the domain of stabilizability is given by the channel zero error capacity with a feedback link, irrespective of whether such a link is in fact available.

The reason for this was basically illuminated in Subsect. 6.11.2 (starting on p. 185): No special feedback communication link is needed since the required feedback information flow from the controller to the sensor site can be arranged by means of control actions upon the plant. However the algorithm proposed in Subsect. 6.11.2 ensures feedback data transmission at only a small rate (one bit per potentially long operation epoch). This is not enough to fabricate a controller of a (feedforward) data transmission scheme relaying on a complete communication feedback: To make such a scheme work properly, the entire previous transmission result should be sent back to the sensor site at each time step. In view of this, we offer another algorithm of data transmission by means of control in this chapter. This algorithm ensures feedback data transmission at an arbitrarily large rate without violating the main objective of keeping the stabilization error a.s. bounded.

The main results of the chapter were originally published in [112, 113, 116, 117, 121].

The remainder of the chapter is organized as follows. Sections 7.2 and 7.4 offer the statement of the state estimation problem and the corresponding main results, respectively. These results employ the concept of the zero error capacity of the noisy channel, which is discussed in Sect. 7.3. In Sect. 7.5, a stabilization problem is posed and the corresponding results are stated. Sections 7.6 and 7.7 are devoted to the proofs of, respectively, the necessary and sufficient conditions (for observability and stabilizability).

7.2 Problem of State Estimation in the Face of System Noises

In Chap. 6, we studied noiseless linear plants. Now we pass to the case where the plant is affected by additive exogenous disturbances $\xi(t)$ and the sensor is noisy:

$$x(t+1) = Ax(t) + \xi(t), \qquad x(0) = x_0, \qquad y(t) = Cx(t) + \chi(t). \tag{7.2.1}$$

Here $x(t) \in \mathbb{R}^n$ is the state; $\xi(t) \in \mathbb{R}^n$ is the exogenous disturbance; $y(t) \in \mathbb{R}^{n_y}$ is the measured output; $\chi(t) \in \mathbb{R}^{n_y}$ is the sensor noise, and $t = 0, 1, \ldots$. The noises $\chi(t)$ and $\xi(t)$ are deterministic and bounded:

$$\|\xi(t)\| \le D, \quad \|\chi(t)\| \le D_\chi \qquad \forall t. \tag{7.2.2}$$

The initial state x_0 is a random vector. Like in Chap. 6, the plant (7.2.1) is unstable, and Assumptions 6.3.1 and 6.3.4–6.3.6 (on pp. 136, 137) are supposed to hold. The objective is to estimate the current state on the basis of the prior measurements.

This estimate is still required at a remote location, where data from the sensor can be communicated only via a given noisy discrete memoryless channel with inputs e and outputs s. So like in Chap. 6, the observer consists of a decoder

$$\widehat{x}(t) = \mathcal{X}\left[t, s(0), s(1), \ldots, s(t)\right] \in \mathbb{R}^n \tag{7.2.3}$$

and coder. We still consider two classes of coders depending on whether a communication feedback is available. We recall that the coders with and without a feedback are described by equations of the following forms, respectively:

$$e(t) = \mathcal{E}[t, y(0), \ldots, y(t), s(0), \ldots, s(t-1)]; \tag{7.2.4}$$

$$e(t) = \mathcal{E}[t, y(0), \ldots, y(t)]. \tag{7.2.5}$$

Is it possible to design a coder–decoder pair keeping the estimation error bounded?

Explanation 7.2.1. Because of the noise, the estimation error cannot be made decaying to zero (6.2.5) (on p. 135), like in Chap. 6. At the same time, this error can be made bounded whenever the decoder has access to the entire observation $y(t)$, and the pair (A, C) is detectable [8, 12].

To specify the above question, we introduce two concepts of observability: the weak and strong ones.

Definition 7.2.2. *The coder–decoder pair is said to* keep the estimation error bounded *under* (D, D_χ)*-bounded noises if*

$$\overline{\lim_{t\to\infty}} \|x(t) - \widehat{x}(t)\| < \infty \tag{7.2.6}$$

for any noises $\{\xi(t)\}$ *and* $\{\chi(t)\}$ *satisfying* (7.2.2) *with the given bounds* D *and* D_χ.

Remark 7.2.3. Such a coder–decoder pair may depend on the bounds D and D_χ and fail to ensure (7.2.6) for bounded noises tresspassing these bounds.

Explanation 7.2.4. Since the initial state x_0 is random, the event (7.2.6) of keeping the estimation error bounded is also random.

Definition 7.2.5. *The coder–decoder pair is said to* uniformly keep the estimation error bounded *under bounded noises if* (7.2.6) *holds uniformly over all noises satisfying* (7.2.2) *with some* D *and* D_χ *and irrespective of the values of* D *and* D_χ*:*

$$\overline{\lim_{t\to\infty}} \sup_{\{\xi(t)\},\{\chi(t)\}} \|x(t) - \widehat{x}(t)\| < \infty \qquad \forall D \geq 0, D_\chi \geq 0, \tag{7.2.7}$$

where sup *is over all noises that obey* (7.2.2) *with the bounds* D *and* D_χ *considered.*

Remark 7.2.6. Such a pair does not depend on the noise bounds and ensures observability under all bounded noises.

A practical difference between the coder–decoder pairs considered in Definitions 7.2.2 and 7.2.5, respectively, is that unlike the second pair, the first one requires a knowledge of an estimate of the noises levels.

Remark 7.2.7. In (7.2.6) and (7.2.7), $\overline{\lim}_{t\to\infty}$ can be evidently replaced by $\sup_{t\geq 0}$.

In this chapter, we examine the border between the cases where the state of the noisy plant can and, respectively, cannot be observed in the sense introduced by either Definition 7.2.2 or 7.2.5. We show that the border is common for these two definitions. Unlike Chap. 6, it is constituted by not the ordinary capacity (6.3.3) (on p. 138) of the channel. Instead, it is given by another fundamental characteristic of the noisy channel, introduced by Shannon [189] and discussed in the next section.

7.3 Zero Error Capacity of the Channel

By Theorem 6.8.3 (on p. 162), the capacity (6.3.3) (on p. 138) employed in Chap. 6 is the least upper bound of rates at which it is possible to transmit information across the channel with as small a probability of error as desired. The zero error capacity is the least upper bound of rates at which it is possible to transmit information with zero probability of error. Unlike the former, the latter may depend on whether the communication feedback is available [189].

Zero Error Capacity of Channels without a Feedback Link

Consider a block code (6.8.2) (on p. 161) with block length r

$$E^1, \dots, E^N, \qquad E^i = (e_0^i, \dots, e_{r-1}^i), \qquad e_j^i \in \mathfrak{E} \tag{7.3.1}$$

endowed with a decoding rule $\mathcal{D}(\cdot) : \mathfrak{S}^r \to [1 : N]$. Here $\mathfrak{S}^r$ is the set of all output r-words $S = (s_0, \dots, s_{r-1})$. This rule is *errorless* if $\mathcal{D}(S) = i$ for any i and any output word S that can be received with a positive probability

$$\boldsymbol{P}(S|E^i) = \prod_{\nu=0}^{N-1} W(s_\nu|e_\nu^i) > 0$$

given that E^i is sent.[3] The *zero error capacity*

$$\mathfrak{c}_0 := \sup R, \qquad R := \frac{\log_2 N}{r},$$

where R is the rate of the code and sup is over all block codes of arbitrary lengths r for which errorless decoding is possible. If no such code exists, $\mathfrak{c}_0 := 0$.

Definition 7.3.1. *Two input code words E' and E'' are* distinguishable *at the receiving end of the channel if they cannot both result in a common output word S with a positive probability. In other words, the following two sets are disjoint:*

$$\{S : \boldsymbol{P}(S|E') > 0\} \qquad \text{and} \qquad \{S : \boldsymbol{P}(S|E'') > 0\}.$$

Remark 7.3.2. It is easy to see that a block code (7.3.1) can be endowed with an errorless decoding rule if and only if it consists of mutually distinguishable words.

It follows that the zero error capacity of the channel

$$\mathfrak{c}_0 = \sup_{r=1,2,\dots} \frac{1}{r} \log_2 N_{\max}(r), \tag{7.3.2}$$

where $N_{\max}(r)$ is the maximal number of mutually distinguishable code words of length r.

Remark 7.3.3. In (7.3.2), $\sup_r$ can be replaced by $\lim_{r\to\infty}$, where the limit does exist [84].

To pave the way to channels with feedback, we start with the following.

Remark 7.3.4. For channels with no feedback, encoding by block codes is identical to encoding i via *block functions*; i.e., recursive rules of the form $i \mapsto E^i = [e(0), \dots, e(r-1)]$, where

$$e(0) := \mathcal{E}_*[0, i], \qquad e(t) := \mathcal{E}_*[t, e(0), \dots, e(t-1), i], \quad t = 1, \dots, r-1. \tag{7.3.3}$$

Indeed, such a rule does nothing but associates every i with an r-word E^i. Conversely, any block code (7.3.1) can be generated in such a way by putting

$$\mathcal{E}_*[t, \cdot, i] := e_t^i.$$

[3] Here $W(s|e)$ are the channel transition probabilities from Assumption 6.3.1 on p. 136.

Zero Error Capacity of Channels with Complete and Perfect Feedback

In this case, the block function (7.3.3) with block length r takes the form [189]

$$e(t) = \mathcal{E}_*[t, e(0), \dots, e(t-1), s(0), \dots, s(t-1), i], \quad t = 1, \dots, r-1;$$
$$e(0) = \mathcal{E}_*[0, i] \quad (7.3.4)$$

and is still used to encode messages labeled by $i = 1, \dots, N$ for transmission over the channel. The other particulars in the definition of the zero error capacity remain unchanged; the corresponding capacity is denoted by $\mathfrak{c}_{0F}$.

Remark 7.3.5. The zero error capacity may, in some cases, be greater with feedback than without $\mathfrak{c}_{0F} > \mathfrak{c}_0$ [189].

Remark 7.3.6. Similarly to Remark 7.3.3, $\mathfrak{c}_{0F} = \lim_{r\to\infty} R_{\max}(r)$, where $R_{\max}(r)$ is the maximal rate of errorless block functions with block length r [84].

The general formula for $\mathfrak{c}_0$ is still missed [84], whereas for $\mathfrak{c}_{0F}$, it is well known [189]

$$2^{-\mathfrak{c}_{0F}} = \min \max_{s\in\mathfrak{S}} \sum_{e\in\mathfrak{E}_s} \boldsymbol{P}(e).$$

Here $\min$ is over all probability distributions $\{\boldsymbol{P}(e)\}$ on the input channel alphabet $\mathfrak{E}$, and $\mathfrak{E}_s$ is the set of all input symbols e that cause the output symbol s with a positive probability $W(s|e) > 0$. The above formula for $2^{-\mathfrak{c}_{0F}}$ is true if there is a pair e', e'' of distinguishable input symbols: $W(s|e')W(s|e'') = 0 \ \forall s$. Otherwise, $\mathfrak{c}_{0F} = \mathfrak{c}_0 = 0$. Moreover, each of the inequalities $\mathfrak{c}_{0F} > 0$ and $\mathfrak{c}_0 > 0$ holds if and only if there is a pair of distinguishable input symbols [189].

Remark 7.3.7. The ordinary capacity (6.3.3) (on p. 138) is typically nonzero, whereas

$$\mathfrak{c}_{0F} = \mathfrak{c}_0 = 0$$

for many discrete memoryless channels of practical interest [214].

The simplest examples are as follows.

- *Erasure channel* with an arbitrary alphabet $\mathfrak{E}$ of size $K = 2, 3, \dots$. Such a channel transmits a message $e \in \mathfrak{E}$ correctly with probability $1-p$ and loses it with probability $p \in (0, 1]$. The output alphabet $\mathfrak{S} = \mathfrak{E} \cup \{\circledast\}$ (where $s(t) = \circledast \Leftrightarrow$ the message $e(t)$ is lost). Since $W(\circledast|e) = p > 0 \, \forall e$, any two input letters are indistinguishable and so $\mathfrak{c}_{0F} = \mathfrak{c}_0 = 0$: There is no way to transmit data without an error. However, data can be transmitted across this channel with as small a probability of error as desired at any rate $R < \mathfrak{c} = (1-p)\log_2 K > 0$ [190].
- *Binary symmetric channel with crossover probability* $0 < p < 1$. This channel has binary input and output alphabets $\mathfrak{E} = \mathfrak{S} = \{0, 1\}$, flips bits with probability $p \in (0, 1)$, and transmits them correctly with probability $1 - p$:

$$W(i|j) = \begin{cases} 1-p & \text{if} \quad i=j \\ p & \text{otherwise} \end{cases}.$$

Since $W(1|0) > 0$ and $W(1|1) > 0$, the input letters 0 and 1 are indistinguishable and so $\mathfrak{c}_{0F} = \mathfrak{c}_0 = 0$. The ordinary capacity (6.3.3) (on p. 138) is given by [190]

$$\mathfrak{c} = 1 + p\log_2 p + (1-p)\log_2(1-p) \quad (> 0 \text{ if } p \neq 1/2).$$

- *General discrete memoryless channel with positive transition probabilities* $W(s|e) > 0\, \forall s, e$ has 0 zero error capacity by the same argument. Moreover

$$\boxed{\exists s : W(s|e) > 0\ \forall e} \Rightarrow \mathfrak{c}_{0F} = \mathfrak{c}_0 = 0.$$

An opposite example of a noisy channel with a positive zero error capacity is any discrete memoryless channel with the pentagon *characteristic graph*. Its vertices are associated with the symbols from the input channel alphabet $\mathfrak{E}$. Two vertices (in fact, input symbols) are linked with a (nonoriented) edge if and only if these symbols are indistinguishible. The zero error capacity is completely determined by the characteristic graph [189]. The pentagon graph is depicted in Fig. 7.1 and related to channels with five input symbols. The zero error capacity of any channel with pentagon characteristic graph equals $\mathfrak{c}_0 = \frac{1}{2}\log_2 5 > 0$ [96].

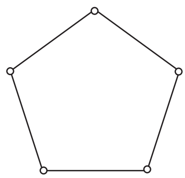


Fig. 7.1. Pentagon graph.

Remark 7.3.8. In [189], the zero error capacity was calculated for all discrete memoryless channels with input alphabets containing no more than four elements.

This capacity is long studied in the so-called zero error information theory. We refer the reader to [84] for an excellent survey of this area.

7.4 Conditions for Almost Sure Observability of Noisy Plants

We start with the necessary conditions for observability.

Theorem 7.4.1. *Suppose that Assumptions* 6.3.1 *and* 6.3.4 *(on pp.* 136 *and* 137*) hold and that the noise does occur in the plant:* $D > 0$ *and* $D_\chi \geq 0$ *in* (7.2.2) *(on* *p.* 202*). Consider the zero error capacities* $\mathfrak{c}_{0F}$ *and* $\mathfrak{c}_0$ *of the channel with and without a communication feedback, respectively, and the topological entropy* $H(A)$ *of the open-loop plant with the noise removed:*

$$H(A) = \sum_{\lambda_j : |\lambda_j| \geq 1} \log_2 |\lambda_j|. \tag{7.4.1}$$

Here $\lambda_1, \ldots, \lambda_n$ *are the eigenvalues of the system* (7.2.1) *(on p.* 202*) repeating in accordance with their algebraic multiplicities.*

Let some coder-decoder pair without (with) a feedback produce estimates of the states of the noisy plant (7.2.1)*. If this pair keeps the estimation error bounded under* (D, D_χ)*-bounded noises*[4] *with a nonzero probability, then* $\mathfrak{c}_0 \geq H(A)$ *in the case of the pair without a feedback and* $\mathfrak{c}_{0F} \geq H(A)$ *in the case of the pair with a feedback.*

The proof of this theorem will be given in Sect. 7.6.

Corollary 7.4.2. *Asymptotically unstable plants* $H(A) > 0$ *can never be observed with a bounded error over channels whose zero error capacity is equal to zero: The estimation error is almost surely unbounded, irrespective of which estimation scheme is employed.*

This in particular holds for the binary symmetric channel with crossover probability $0 < p < 1$ and erasure channel with arbitrary alphabet (of size ≥ 2) and positive erasure probability.

Explanation 7.4.3. In Corollary 7.4.2, the phrase "is almost surely unbounded" means that with probability 1, noise sequences $\{\xi(\cdot)\}, \{\chi(\cdot)\}$ obeying the prescribed bound (7.2.2) (on p. 202) exist for which the estimation error is unbounded. As will be specified by Proposition 7.6.2 (on p. 211), this is true with the zero sensor noise.

Remark 7.4.4. The level $D > 0$ of the exogenous disturbance is immaterial for the conclusions of Theorem 7.4.1 and Corollary 7.4.2 to hold. In particular, this level may be arbitrarily small.

In view of this, Corollary 7.4.2 means that arbitrarily and uniformly small plant disturbances unavoidably accumulate and cause arbitrarily large estimation errors.

The next theorem demonstrates that the necessary conditions given by Theorem 7.4.1 are "almost" sufficient.

Theorem 7.4.5. *Suppose that Assumptions* 6.3.1, 6.3.4*, and* 6.3.6 *(on pp.* 136 *and* 137*) hold. If* $\mathfrak{c}_0 > H(A)$ *(or* $\mathfrak{c}_{0F} > H(A)$*), then a semirational finite-step recursive*[5] *coder–decoder pair without (respectively, with) a feedback exists that with probability* 1 *uniformly keeps the estimation error bounded under bounded noises.*[6]

The proof of this theorem will be given in Sect. 7.7. An explicit description of the observer will be offered in Subsect. 7.7.1 (starting on p. 223).

[4] See Definition 7.2.2 on p. 203.

[5] See Definitions 6.3.9 and 6.3.10 on pp. 139 and 140.

[6] See Definition 7.2.5 on p. 203.

Remark 7.4.6. The coder–decoder pair from Theorem 7.4.5 does not depend on the noise bounds from (7.2.2) (on p. 202) and ensures state tracking with uniformly bounded error (7.2.7) (on p. 203) irrespective of these bounds.

Remark 7.4.7. In Theorem 7.4.5, the coder–decoder pair can be chosen so that whenever the initial state x_0 is bounded $\|x_0\| \leq D_x$ a.s., this pair ensures that the estimation error is bounded uniformly over almost all samples assumed by the initial state: The $\overline{\lim}$ in (7.2.7) is a.s. upper bounded by a constant, which does not depend on the elementary random event and depends only on D_x, D, D_χ.

If $\det A \neq 0$, this claim is straightforward from Theorem 7.4.5. Indeed, it suffices to consider the process started at $t = -1$ at $x(-1) := 0$ under $\xi(-1) := A^{-1}x_0, \chi(-1) := 0$. The required minor comments concerning the case of singular matrix A will be given in the footnote on p. 231.

Comment 7.4.8. For noiseless discrete channels, the statement of Theorem 7.4.5 was established in [204] (see also [215]) in the case, where the observation is noiseless and complete $y(t) = x(t)$ and an upper bound of the initial state is known, with respect to a weaker form of observability: (7.2.7) was ensured only for given noise bounds from (7.2.2). In particular, the observers proposed in [204, 215] depend on the noises bounds D, D_χ with no guarantee of convergence in the case where the bounds employed in the observer design are trespassed.

Remark 7.4.9. By properly increasing the sample period in the vein of Subsect. 3.8.4 (starting on p. 68), the case of a noisy channel with the zero error capacity $> H(A)$ can be reduced to the case of a noiseless discrete channel with capacity $> H(A)$ considered in [204, 215].

7.5 Almost Sure Stabilization in the Face of System Noises

In this section, we show that Theorems 7.4.1 and 7.4.5 basically remain true for the stabilization problem. However, in this case, there is a difference as compared with the observation problem, which concerns the communication feedback.

In this section, we consider a controlled version of the plant (7.2.1) (on p. 202):

$$x(t+1) = Ax(t) + Bu(t) + \xi(t), \qquad x(0) = x_0, \qquad y(t) = Cx(t) + \chi(t), \quad (7.5.1)$$

where $u(t) \in \mathbb{R}^{n_u}$ is the control. This plant is unstable, and the objective is to stabilize it.

We examine a remote control setup: The controls are generated at a remote location, where data from the sensor can be communicated only via a given noisy discrete memoryless channel with inputs e and outputs s. So the controller is constituted by a coder and decoder(-controller):

$$u(t) = \mathfrak{U}[t, s(0), s(1), \dots, s(t)] \in \mathbb{R}^{n_u}. \quad (7.5.2)$$

We still consider two classes of coders: those with (7.2.4) and without (7.2.5) a communication feedback, respectively.

Similarly to Definitions 7.2.2 and 7.2.5 (on p. 203), we introduce two concepts of stabilizability in the face of system noises.

Definition 7.5.1. *The coder–decoder pair is said to* stabilize *the plant under* (D, D_χ)-*bounded noises if*

$$\overline{\lim_{t\to\infty}} \|x(t)\| < \infty \tag{7.5.3}$$

for any noises $\{\xi(t)\}$ *and* $\{\chi(t)\}$ *satisfying* (7.2.2) *with the given bounds* D *and* D_χ.

Definition 7.5.2. *The coder–decoder pair is said to* uniformly stabilize *the plant under bounded noises if* (7.5.3) *holds uniformly over all noises satisfying* (7.2.2) *and irrespective of the values of* D *and* D_χ*:*

$$\overline{\lim_{t\to\infty}} \sup_{\{\xi(t)\},\{\chi(t)\}} \|x(t)\| < \infty \qquad \forall D \geq 0, D_\chi \geq 0, \tag{7.5.4}$$

where sup *is over all noises that obey* (7.2.2) *with the bounds* D *and* D_χ *considered.*

Remarks 7.2.3, 7.2.6, and 7.2.7 and Explanation 7.2.4 (on p. 203) remain true with respect to Definitions 7.5.1 and 7.5.2.

Now we are in a position to state the main result of the section.

Theorem 7.5.3. *Suppose that Assumptions* 6.3.1, 6.3.4–6.3.6*, and* 6.5.5 *(on pp.* 136*,* 137*, and* 144*) hold. Consider the zero error capacity* $\mathfrak{c}_{0F}$ *of the channel with a communication feedback and the topological entropy* $H(A)$ *of the uncontrolled open-loop plant with the noise removed* (7.4.1) *(on p.* 207*). Then the following two claims are true:*

i) If the plant disturbance does occur $D > 0$ *and a coder–decoder pair exists that with a nonzero probability, stabilizes the noisy plant* (7.5.1) *under* (D, D_χ)-*bounded noises, then* $\mathfrak{c}_{0F} \geq H(A)$*;*

ii) Conversely, if $\mathfrak{c}_{0F} > H(A)$*, a semirational finite-step recursive coder–decoder pair*[7] *without a feedback exists that almost surely uniformly stabilizes the plant under bounded noises.*

The proof of this theorem will be given in Sects. 7.6 and 7.8. A coder–decoder pair ensuring ii) will be described in Subsect. 7.8.3 (starting on p. 236).

Explanation 7.5.4. Claim i) holds for coder–decoder pairs both with (7.2.4) and without (7.2.5) a feedback.

Remark 7.5.5. By ii), the condition $\mathfrak{c}_{0F} > H(A)$ on the channel zero error capacity **with** a feedback ensures existence of a stabilizing coder–decoder pair **without** a feedback.

[7] See Definitions 6.3.9 and 6.5.7 on pp. 139 and 145, respectively.

This is in contrast with the state estimation problem, where the conditions on the channel zero error capacities with $\mathfrak{c}_{0F}$ and without $\mathfrak{c}_0$ a feedback are sufficient and almost necessary for observability with and without a feedback link, respectively.

Explanation 7.5.6. The phenomenon concerned in Remark 7.5.5 is underlyed by the fact discussed in Subsect. 6.11.1 (starting on p. 185): In control systems, data can be communicated by means of control, and so a special feedback communication link is not required.

As is known, $\mathfrak{c}_{0F} \geq \mathfrak{c}_0$ and for some channels, $\mathfrak{c}_{0F} > \mathfrak{c}_0$ [189].

Observation 7.5.7. *In the absence of a feedback communication link, noisy plants* (7.5.1) *with*

$$\mathfrak{c}_{0F} > H(A) > \mathfrak{c}_0$$

are a.s. stabilizible but are not observable in the open loop ($u(t) \equiv 0$*).*

This is in contrast with the results of Chap. 6, where the conditions for stabilizability and observability are common.

Comment 7.5.8. Theorem 7.5.3 remains true for stochastic plant disturbances [121] .

Specifically, they remain true in the case where the disturbances are identically distributed according to some probability density with a bounded support,[8] mutually independent, and independent of the channel and initial state.

The important Corollary 7.4.2 (along with Remark 7.4.4 on p. 207) evidently remains true for the stabilization problem: Arbitrarily and uniformly small plant disturbances unavoidably accumulate and cause arbitrarily large stabilization errors. More presicely, the following claim holds.

Corollary 7.5.9. *Asymptotically unstable plants* $H(A) > 0$ *can never be stabilized with a bounded error over channels whose zero error capacity equals zero: The stabilization error is almost surely unbounded, no matter what stabilization scheme is employed.*

Explanation 7.4.3 (on p. 207) extends on this corollary.

Remark 7.5.10. The phenomenon addressed by Corollary 7.5.9 has nothing in common with the well-known fact that for the stable linear plant affected by the Gaussian white noise, the stabilization error is yet a.s. unbounded. Indeed, this unboundedness ultimately results from the facts that first, the sample sequences of the white noise are a.s. unbounded and second, an unpredictable external disturbance cannot be compensated. On the contrary, we show that in the face of channel errors, external disturbances obeying a common and arbitrarily small deterministic bound at any sample and time unavoidably accumulate and cause, sooner or later, arbitrarily large stabilization errors.

Remark 7.4.6 and Comment 7.4.8 (on p. 208) with “ii) of Theorem 7.5.3” substituted in place of “Theorem 7.4.5” remain valid for the stabilization problem.

[8] The bounded support assumption is required only for part ii) of Theorem 7.5.3.

7.6 Necessary Conditions for Observability and Stabilizability: Proofs of Theorem 7.4.1 and i) of Theorem 7.5.3

In this section, Assumptions 6.3.1 and 6.3.4 (on pp. 136 and 137) of these theorems are supposed to hold. So far as necessary conditions are concerned, we can confine ourselves to consideration of certain specific noises satisfying (7.2.2) (on p. 202). Specifically, we shall consider the zero sensor noise throughout the section:

$$\chi(t) \equiv 0.$$

We also consider a certain specific disturbance bound $D > 0$ in (7.2.2).

Definition 7.6.1. *A disturbance* $\{\xi(t)\}$ *satisfying* (7.2.2) *is said to be* admissible.

In the remainder of the section, we focus on proving the following proposition, which evidently implies Theorem 7.4.1 and i) of Theorem 7.5.3.

Proposition 7.6.2. *The following two claims hold:*

i) Suppose that $H(A) > \mathfrak{c}_0$ *(or* $H(A) > \mathfrak{c}_{0F}$*) and for the uncontrolled plant* (7.2.1) *(on p.* 202*), consider arbitrary coder and decoder-estimator without (respectively, with) a feedback. Then with probability* 1*, an admissible disturbance exists for which the estimation error is unbounded:*

$$\overline{\lim_{t\to\infty}} \|x(t) - \widehat{x}(t)\| = \infty. \tag{7.6.1}$$

ii) Suppose that $H(A) > \mathfrak{c}_{0F}$ *and for the controlled plant* (7.5.1)*, consider arbitrary coder and decoder-controller. Then with probability* 1*, an admissible disturbance exists for which the stabilization error is unbounded:*

$$\overline{\lim_{t\to\infty}} \|x(t)\| = \infty. \tag{7.6.2}$$

Remark 7.6.3. In the particular case where $H(A) > 0$ and the zero error capacity of the channel is zero, this proposition comes to Corollaries 7.4.2 and 7.5.9 (on pp. 207 and 210).

Remark 7.6.4. The statement ii) follows from i).

This holds since Lemma 6.7.4 (on p. 148) clearly remains true for the noisy plant (7.5.1), provided that the noises are introduced in the equations of the system (6.7.7).

Remark 7.6.4 permits us to focus on proving i) of Proposition 7.6.2.

7.6.1 Proof of i) in Proposition 7.6.2 for Erasure Channels

We start with a particular case where the proof is especially simple. Specifically, we suppose that communication from the coder to decoder is over an erasure channel with arbitrary finite alphabet $\mathfrak{E}$ of size ≥ 2. We recall that this channel transmits a message $e \in \mathfrak{E}$ correctly with probability $1-p$ and loses it with probability $p \in (0, 1]$.

Thus, the only form of data distorion is dropout.[9] This channel has 0 zero error capacity. So i) of Proposition 7.6.2 means that (7.6.1) holds almost surely for any coder and decoder-estimator, whenever the plant is asymptotically unstable.

In this subsection, we prove the following claim.

Proposition 7.6.5. *Let the matrix A from* (7.2.1) *(on p.* 202*) have an unstable eigenvalue* $|\lambda| \geq 1$*, the erasure probability* $p > 0$*, and the plant disturbance does occur* $D > 0$*. Then for arbitrary coder and decoder-estimator, the conclusion of* i) *in Proposition* 7.6.2 *holds.*

Remark 7.6.6. In fact, this claim is a bit stronger than i) of Proposition 7.6.2 projected on the case of the erasure channel.

Indeed, i) concerns matrices A with $H(A) > 0$, which implies that A has an eigenvalue λ with $|\lambda| > 1$.

The proof of Proposition 7.6.5 is prefaced by two simple technical facts.

Lemma 7.6.7. *Let* λ *be an unstable* $|\lambda| \geq 1$ *eigenvalue of* A*. Given a time* $\tau = 1, 2, \ldots$ *and an initial state* $\overline{x}_0$*, an admissible disturbance exists such that*

$$\|x_\tau - x_\tau^0\| \geq D\nu_{|\lambda|}(\tau), \quad \text{where} \quad \nu_a(\tau) := \begin{cases} \frac{a^\tau - 1}{a-1} & \text{if} \quad a > 1 \\ \tau & \text{if} \quad a = 1 \end{cases}. \tag{7.6.3}$$

Here x_τ *and* x_τ^0 *stand for the states to which the system* (7.2.1) *(on p.* 202*) is driven from the initial state* $\overline{x}_0$ *under the above and zero disturbances, respectively.*

Proof. Suppose first that λ is real. The required admissible disturbance is given by

$$\xi(t) = D(\operatorname{sgn}\lambda)^t \zeta_\lambda,$$

where ζ_λ is a normalized $\|\zeta_\lambda\| = 1$ eigenvector $A\zeta_\lambda = \lambda\zeta_\lambda$. Indeed,

$$x_\tau = \underbrace{A^\tau \overline{x}_0}_{x_\tau^0} + q, \quad \text{where} \quad q := \sum_{t=0}^{\tau-1} A^{\tau-1-t}\xi(t) = D\sum_{t=0}^{\tau-1}(\operatorname{sgn}\lambda)^t A^{\tau-1-t}\zeta_\lambda$$
$$= D\sum_{t=0}^{\tau-1}(\operatorname{sgn}\lambda)^t \lambda^{\tau-1-t}\zeta_\lambda = (\operatorname{sgn}\lambda)^{\tau-1} D\nu_{|\lambda|}(\tau)\zeta_\lambda,$$

which implies (7.6.3).

If the eigenvalue is not real

$$\lambda = |\lambda|(\cos\varphi + \imath\sin\varphi), \qquad \varphi \neq 0, \pm\pi, \pm 2\pi, \ldots,$$

we consider disturbances

[9] This may be due to the activity of a filter at the receiving end of the channel, which blocks any message suspicious as damaged.

$$\xi(t) = D|\lambda|^{-t} A^t \zeta(\theta)$$

depending on the parameter $\theta \in \mathbb{R}$. Here

$$\zeta(\theta) := \zeta_\lambda^1 \cos\theta - \zeta_\lambda^2 \sin\theta = \text{ Re } \left(e^{\imath\theta}\zeta_\lambda\right),$$

where $\zeta_\lambda \in \mathbb{C}^n$ is a λ-eigenvector $A\zeta_\lambda = \lambda\zeta_\lambda$ and $\zeta_\lambda^1, \zeta_\lambda^2 \in \mathbb{R}^n$ are its real and imaginary parts $\zeta_\lambda = \zeta_\lambda^1 + \imath\zeta_\lambda^2$. The vector ζ_λ is normalized so that

$$\max_\theta \|\zeta(\theta)\| = 1.$$

It is easy to see that then

$$q = D\nu_{|\lambda|}(\tau)\zeta[\theta + (\tau - 1)\varphi].$$

After picking θ such that $\|\zeta[\theta + (\tau - 1)\varphi]\| = 1$, the proof is completed just as in the case of the real eigenvalue. $\square$

Notation 7.6.8. *Denote by I_t the indicator function of the erasure:*

$$I_t = \begin{cases} 1 & \text{if the message} \quad e(t) \quad \text{is lost} \\ 0 & \text{otherwise} \end{cases}.$$

Lemma 7.6.9. *Let $k = 1, 2, \ldots$ be given. Then $k + 1$ successive erasures occur during some time interval $I_{t-k} = I_{t-k+1} = \cdots = I_t = 1$ of duration k almost surely. Moreover, this happens infinitely many times with the average frequency*

$$\frac{1}{r}\sum_{i=1}^{r} J_i \to p^{k+1} > 0 \quad \text{as} \quad r \to \infty \quad \text{a.s.}.$$

Here

$$J_i := \begin{cases} 1 & \text{if } I_{t-k} = I_{t-k+1} = \cdots = I_t = 1 \text{ for } t := (k+1)i \\ 0 & \text{otherwise} \end{cases}.$$

Proof. Due to Assumption 6.3.1 (on p. 136), the random quantities I_t are independent and identically distributed with $\boldsymbol{P}(I_t = 1) = p > 0$, hence, so are the quantities J_i, and $\boldsymbol{P}(J_i = 1) = p^{k+1}$. The proof is completed by the strong law of large numbers (see Theorem 6.9.10 on p. 171). $\square$

Proof of Proposition 7.6.5. By Lemma 7.6.9, a.s. $k+1$ successive erasures $I_{t-k} = \cdots = I_t = 1$ are encountered infinitely many times for any given k. It follows that a sequence of random times $0 = \tau_0 < \tau_1 < \tau_2 < \ldots$ exists such that

$$\tau_i > \tau_{i-1} + i \quad \text{and} \quad I_{\tau_i - i} = I_{\tau_i - i+1} = \cdots = I_{\tau_i} = 1 \quad \text{for all} \quad i = 1, 2, \ldots \quad \text{a.s.}$$

The admissible disturbance ensuring (7.6.1) will be constructed by consecutive extension of the sequence $\{\xi(t)\}$ from $[0 : \tau_i]$ on $[0 : \tau_{i+1}]$ via induction on i. The induction is started at $i = 0$ by putting $\xi(0) := 0$. Suppose that the sequence $\{\xi(t)\}$ has been already constructed for $t = 0, \dots, \tau_{i-1}$. For $\tau_{i-1} < t < \tau_i - i$, we put $\xi(t) := 0$. Since all messages carried across the channel at times $t = \tau_i - i, \dots, \tau_i$ are lost, the decoder is unable to notice any difference between the cases where different continuations of the chosen disturbance sequence are applied on the interval $[\tau_i - i : \tau_i]$. Thus the decoder produces a common sequence of estimates

$$\widehat{x}(\tau_i - i), \dots, \widehat{x}(\tau_i)$$

for all such continuations.

Now we apply Lemma 7.6.7 to $\tau := i$ and $\overline{x}_0 := x(\tau_i - i)$. Then we prolong the sequence of disturbances from $[0 : \tau_{i-1} - i)$ onto $[0 : \tau_i]$ in two ways: first, by the disturbace taken from Lemma 7.6.7 and second, by zeros. Due to (7.6.3), these two extensions drive the system to the states $x(\tau_i)^+$ and $x(\tau_i)^0$, respectively, such that

$$\|x(\tau_i)^+ - x(\tau_i)^0\| \geq D\nu_{|\lambda|}(i).$$

It follows that $\|x - \widehat{x}(\tau_i)\| \geq \frac{1}{2}D\nu_{|\lambda|}(i)$ for either $x := x(\tau_i)^+$ or $x := x(\tau_i)^0$. In other words,

$$\|x(\tau_i) - \widehat{x}(\tau_i)\| \geq \frac{1}{2}D\nu_{|\lambda|}(i) \tag{7.6.4}$$

for one of these extensions. It is this extension that is put in use.

Relation (7.6.1) is immediate from (7.6.4) since $\nu_{|\lambda|}(i) \to \infty$ as $i \to \infty$ due to the last formula from (7.6.3). □

Remark 7.6.10. For general discrete memoryless channels, some forms of data distortion different from mere dropout may occur, whereas the pure dropout may not hold.

In this case, the proof of i) from Proposition 7.6.5 requires more sophisticated arguments. This proof is offered in the remainder of the section.

7.6.2 Preliminaries

Three technical facts are established in this subsection. First, we state a variant of the strong law of large numbers required for the proof of Proposition 7.6.2. Second, we show that in this proof, the attention can be switched from the weak to uniform observability. Third, we demonstrate that exogenous additive disturbances make the system state more and more distributed over the space as time progresses.

Preliminaries from the Probability Theory

Let $\{\mathfrak{F}_k\}_{k=0}^{\infty}$ be a nondecreasing $\mathfrak{F}_k \subset \mathfrak{F}_{k+1}$ flow of σ-algebras in a probability space.

Definition 7.6.11. *A random variable* $\tau \in \{0, 1, \ldots, \infty\}$ *is called a* Markov time *(moment) with respect to this flow if*

$$\{\tau = k\} \in \mathfrak{F}_k \qquad \forall k < \infty.$$

Notation 7.6.12. *For two Markov times* τ_1 *and* τ_2*, we write*

$$\tau_1 \prec \tau_2$$

if and only if with probability 1*,*

$$\text{either} \quad \tau_1 = \tau_2 = \infty \quad \text{or} \quad \tau_1 < \tau_2.$$

To prove Proposition 7.6.2, we need the following fact.

Proposition 7.6.13. *Suppose that* $\tau_1 \prec \tau_2 \prec \ldots$ *is an infinite sequence of Markov times with respect to the flow* $\{\mathfrak{F}_k\}$*, and* $V_0, V_1, \ldots$ *are random variables with values in a measurable space* $[\mathfrak{V}, \Sigma]$*, identically distributed according to a probability distribution* $\boldsymbol{P}(dv)$*. Suppose also that* V_k *is* $\mathfrak{F}_{k+1}$*-measurable but independent of* $\mathfrak{F}_k$ *for all* $k = 0, 1, \ldots$*. Then for any* $\mathcal{V} \in \Sigma$*,*

$$\frac{1}{k}\sum_{\nu=1}^{k} I_{V_{\tau_\nu} \in \mathcal{V} \wedge \tau_\nu < \infty} - \boldsymbol{P}(\mathcal{V})\frac{1}{k}\sum_{\nu=1}^{k} I_{\tau_\nu < \infty} \to 0 \quad \text{as} \quad k \to \infty \quad \text{a.s.} \tag{7.6.5}$$

Explanation 7.6.14. Here $I_{\mathcal{R}}$ is the indicator of the random event $\mathcal{R}$:

$$I_{\mathcal{R}} = \begin{cases} 1 & \text{if} \quad \mathcal{R} \quad \text{holds} \\ 0 & \text{otherwise} \end{cases}.$$

The proof of Proposition 7.6.13 is given in Appendix A.

Corollary 7.6.15. *If* $\boldsymbol{P}(\mathcal{V}) > 0$*, then the sequence* $V_{\tau_0}, V_{\tau_1}, V_{\tau_2}, \ldots$ *a.s. visits the set* $\mathcal{V}$ *infinitely many times with the average frequency* $\boldsymbol{P}(\mathcal{V})$*, given that* $\tau_k < \infty \, \forall k$*.*

Switching Attention to Uniform Observability in the Proof of Proposition 7.6.2

A possibility to do so is justified by the following lemma.

Lemma 7.6.16. *Suppose that for any initial state distribution satisfying Assumption* 6.3.4 *(on p.* 137*), arbitrary coder and decoder-estimator without (with) a feedback almost surely do not uniformly keep the estimation error bounded:*

$$\overline{\lim_{t\to\infty}} \sup_{\{\xi(t)\}:|\xi(t)| \leq D \, \forall t} \|x(t) - \widehat{x}(t)\| = \infty \qquad \text{a.s.} \tag{7.6.6}$$

Then the conclusion of i) in Proposition 7.6.2 *holds.*

Proof. Consider arbitrary coder and decoder-estimator without (with) a feedback, and pick a sequence

$$\{b_i > 0\}_{i=1}^{\infty}, \qquad b_i \uparrow \infty.$$

Thanks to (7.6.6), there exists a random time τ_1 and random admissible disturbance

$$\xi(t), \qquad t = 0, \dots, \tau_1 - 1$$

such that

$$\|x(\tau_1) - \widehat{x}(\tau_1)\| \geq b_1 \qquad \text{a.s.}$$

Let us take the least such a time τ_1. Then τ_1 is the Markov time with respect to the flow of σ-algebras generated by $x_0, s(0), \dots, s(k)$ for $k = 0, 1, \dots$. The above disturbance extended $\xi(t) := 0$ on $t = \tau_1$ is a function of these random variables and τ_1. Now we consider the tail $t \geq \tau_1 + 1$ of the process

$$x(t), \quad y(t), \quad \widehat{x}(t), \quad e(t), \quad s(t)$$

in the (regular [94]) probability space obtained by conditioning over

$$\tau_1 = k, \quad x_0 = \boldsymbol{x}, \quad s(0) = \boldsymbol{s}_0, \dots, s(k) = \boldsymbol{s}_k.$$

Employed is consistent data $\boldsymbol{d} = [k, \boldsymbol{x}, \boldsymbol{s}_0, \dots, \boldsymbol{s}_k]$. It uniquely determines the state $x(\tau_1 + 1) = \boldsymbol{x}_{k+1}$ and the observations $y(0) = \boldsymbol{y}_0, \dots, y(k) = \boldsymbol{y}_k$. Hence the tail starts at $t = k + 1$ at $x(k+1) = \boldsymbol{x}_{k+1}$ and is governed by the coder and decoder that are obtained from putting

$$y(0) := \boldsymbol{y}_0, \dots, y(k) := \boldsymbol{y}_k \quad \text{and} \quad s(0) := \boldsymbol{s}_0, \dots, s(k) := \boldsymbol{s}_k$$

into (7.2.4) or (7.2.5) and (7.2.3) (on p. 202), respectively. Since the above conditioning does not alter the transition probabilities of the channel by Assumption 6.3.4 (on p. 137), (7.6.6) still holds for the process at hand. So by repeating the starting arguments of the proof, we see that there exist a random time $\Delta\tau_2^{\boldsymbol{d}}$ and disturbance

$$\xi^{\boldsymbol{d}}(t), \qquad t = k + 1, \dots, k + \Delta\tau_2^{\boldsymbol{d}} - 1$$

for which

$$\|x(k + \Delta\tau_2^{\boldsymbol{d}}) - \widehat{x}(k + \Delta\tau_2^{\boldsymbol{d}})\| \geq b_2.$$

Now we put

$$\tau_2 := \tau_1 + \Delta\tau_2^{\tau_1, x_0, s(0), \dots, s(\tau_1)}$$

and continue the disturbance

$$\xi(t) := \xi^{\tau_1, x_0, s(0), \dots, s(\tau_1)}(t) \quad \text{on} \quad t = \tau_1 + 1, \dots, \tau_2 - 1.$$

After this, we have

$$\|x(\tau_2) - \widehat{x}(\tau_2)\| \geq b_2 \qquad \text{a.s.}$$

By continuing likewise, we construct a sequence $\{\tau_i\}$ of random times and a random admissible disturbance such that

$$\|x(\tau_i) - \widehat{x}(\tau_i)\| \geq b_i \qquad \text{a.s.} \qquad \text{for all} \quad i,$$

which clearly implies (7.6.1). □

Estimate of the Rate of the State Uncertainty Caused by Disturbances

Note first that by the assumptions of Proposition 7.6.2, $H(A) = \sum_{j:|\lambda_j|\geq 1} \log_2 |\lambda_j| > 0$. It follows that the strictly unstable part

$$\sigma_\oplus := \{\lambda \in \sigma(A) : |\lambda| > 1\}$$

of the spectrum $\sigma(A)$ of A is not empty. Let $M_{\text{s-unst}}$ denote the invariant subspace of A related to $\sigma_\oplus$, and $A_\oplus$ the operator A acting in $M_{\text{s-unst}}$. It is easy to see that

$$H(A) = \log_2 |\det A_\oplus|.$$

The following definition extends Definition 2.3.7 (on p. 16) on the case where the set is not necessarily that constituted by trajectories of a dynamical system.

Definition 7.6.17. *A set $V \subset \mathbb{R}^n$ is called b-*separated *if $\|v_1 - v_2\| \geq b$ for any two elements $v_1 \neq v_2 \in V$.*

The maximal number of elements in a b-separated set can be viewed as a lower estimate of uncertainty about the vector v if the available knowledge about it is expressed by the inclusion $v \in V$.

Lemma 7.6.18. *For any $b > 0$ and time interval $[0 : r-2]$ of arbitrary duration $r-1$, there exist*

$$N \geq \left(\frac{D}{b}\right)^{\dim M_{\text{s-unst}}} 2^{(r-2)H(A)} \tag{7.6.7}$$

admissible disturbances $\Xi_1, \ldots, \Xi_N$ defined on this interval

$$\Xi_i = [\xi_i(0), \ldots, \xi_i(r-2)]$$

that drive the system (7.2.1) *(on p. 202) from the zero initial state to the states $x_i(r-1) \overset{\Xi_i}{\longleftarrow} 0$ forming a b-separated set:*

$$\|x_{i'}(r-1) - x_{i''}(r-1)\| \geq b \qquad \text{whenever} \quad i' \neq i''.$$

Proof. We shall look for the required disturbances among those of the form

$$\xi_i(0) \in M_{\text{s-unst}}, \quad \xi_i(1) = \cdots = \xi_i(r-2) = 0.$$

For them,

$$x_i(r-1) = A_\oplus^{r-2} \xi_i(0).$$

Among the subsets of the closed ball $\overline{B_0^D} \subset M_{\text{s-unst}}$ transformed by $A_\oplus^{r-2}$ into b-separated sets, we pick one $Z = \{\xi_1(0), \ldots, \xi_N(0)\}$ with the maximal cardinality N. It is clear that

$$A_\oplus^{r-2} \overline{B_0^D} \subset \bigcup_{i=1}^N \overline{B_{x_i(r-1)}^b},$$

where all balls are in $M_{\mathrm{s-unst}}$: Otherwise, one more point can be put in Z. Hence

$$|\det A_\oplus|^{r-2}\boldsymbol{V}\left(B_0^D\right) = \boldsymbol{V}\left[A_\oplus^{r-2}\overline{B_0^D}\right] \leq \sum_{i=1}^{N} \boldsymbol{V}\left[\overline{B_{x_i(r-2)}^b}\right] = N\boldsymbol{V}\left(B_0^b\right),$$

$$N \geq |\det A_\oplus|^{r-2}\frac{\boldsymbol{V}\left(B_0^D\right)}{\boldsymbol{V}\left(B_0^b\right)} = 2^{(r-2)H(A)}\left(\frac{D}{b}\right)^{\dim M_{\mathrm{s-unst}}}. \qquad \square$$

7.6.3 Block Codes Fabricated from the Observer

Now we in fact start the direct proof of i) from Proposition 7.6.2. So we suppose that

$$H(A) > \mathfrak{c}_0 \qquad (\text{or } H(A) > \mathfrak{c}_{0F}) \tag{7.6.8}$$

and consider a coder and decoder-estimator without (respectively, with) a feedback. In view of Lemma 7.6.16, it should be demonstrated that (7.6.6) holds.

The plan of the proof is as follows.

1. We argue by contradiction; i.e., suppose that (7.6.6) fails to be true;
2. From the above coder and decoder, we fabricate infinitely many block codes transmitting information at rates $R \approx H(A)$;
3. We show that there is an errorless code among the above infinitely many ones;
4. By the definition of the zero error capacity, this contradicts (7.6.8). This contradiction proves that (7.6.6) is yet correct.

Comment 7.6.19. In [164, 166], a stabilizing coder–decoder pair was transformed into a data communication device, which at any time produces updated estimates of all messages dispatched over the channel since the initial until the current time. This was accomplished by converting the incoming bit stream into a “simulated” plant disturbance. As was shown, the resultant device ensures that the probability of error in decoding the early messages decays to zero fast enough as the time elapsed since the message dispatch becomes large. In this section, it is required to establish the stronger fact: A (finite-length) block code exists that produces no error at all.

In this subsection, we introduce the infinite variety of block codes mentioned in 2. In the next subsection, we prove that this variety contains an errorless code.

From now and until otherwise stated, we adopt the following temporary assumption matching 1 of the above plan.

Assumption 7.6.20. *Relation* (7.6.6) *fails to be true.*

Explanation 7.6.21. This means that with a nonzero probability, the estimation error is bounded; i.e., (7.2.7) (on p. 203) holds for given D and $D_\chi := 0$.

Remark 7.6.22. By sacrificing a small probability, the error can be made uniformly bounded over elementary random events in the sense that

$$\sup_{\{\xi(t)\}:\|\xi(t)\|\leq D} \|x(t) - \widehat{x}(t)\| < b_* \qquad \text{for all} \quad t \geq t_* \tag{7.6.9}$$

with a positive probability for some nonrandom t_* and $b_* > 0$.

Construction of Block Codes

To create a block code with block length $r \geq 2$, we do the following.

- Consider the constants t_* and b_* from (7.6.9) and pick $b > 2b_*$;
- Take the disturbances $\Xi_1, \dots, \Xi_N$ from Lemma 7.6.18;
- Introduce the time instants $t_*^k := t_* + kr$ enumerated by $k = 1, 2, \dots$;
- For any $i = 1, \dots, N$ and $k = 1, 2, \dots$, consider the process

$$x_i^k(t), \quad y_i^k(t), \quad \widehat{x}_i^k(t), \quad e_i^k(t), \quad s_i^k(t), \qquad t = 0, 1, \dots, t_*^{k+1} - 1 \quad (7.6.10)$$

in the system (7.2.1) (on p. 202) endowed by the observer.

Explanation 7.6.23. This process is generated by the disturbance that is zero until $t = t_*^k - 1$ and identical to Ξ_i (up to the shift in time) on the interval $[t_*^k : t_*^{k+1} - 2]$.

Comment 7.6.24. The choice of the disturbance on the interval $[0 : t_*^k - 1]$ is immaterial. However, the disturbance should be common on this interval for all i and given k.

Remark 7.6.25. The process (7.6.10) does not depend on i, k on the interval $[0 : t_*^{k_*}]$, provided that $k_* \leq k$.

Notation 7.6.26. *In* (7.6.10)*, the indices $\substack{k \\ i}$ are further dropped whenever $t \leq t_*^k$.*

Remark 7.6.27. The process (7.6.10) is stochastic since the initial state and the channel noise are random.

The codes to be introduced are enumerated by $k = 1, 2, \dots$.

The kth block code (7.3.3) or (7.3.4) encodes $i = 1, \dots, N$ by acting just as the coder (7.2.4), respectively, (7.2.5) (on p. 202) does under the disturbance Ξ_i applied on $[t_*^k : t_*^{k+1} - 2]$:

$$\mathcal{E}_*^k \Big[t, e_0, \dots, e_{t-1}, \boxed{s_0, \dots, s_{t-1},} \ i \Big] = \mathcal{E}_*^k \Big[t, \boxed{s_0, \dots, s_{t-1},} \ i \Big]$$
$$:= \mathcal{E} \Big[t + t_*^k, y(0), \dots, y(t_*^k), y_i^k(t_*^k + 1), y_i^k(t_*^k + 2), \dots,$$
$$\dots, y_i^k(t^k + t - 1), y_i^k(t^k + t), \boxed{s(0), \dots, s(t_*^i - 1), s_0, \dots, s_{t-1}} \Big] \quad (7.6.11)$$

for $t = 0, \dots, r - 1$.

Explanation 7.6.28. For $t = 0$, all arguments of the forms $s_\theta, y_i^k(\theta)$ are dropped on the right. The dashed expressions are omitted in the case where no communication feedback is available.

Remark 7.6.29. By (7.6.7), the rate R of this block code obeys the lower bound:

$$R = \frac{\log_2 N}{r} \geq \frac{\dim M_{\text{s-unst}}}{r} [\log_2 D - \log_2 b] + H(A) \left[1 - \frac{2}{r}\right]. \tag{7.6.12}$$

The acompanying kth decoding rule $\mathfrak{D}^k : \mathfrak{S}^r \to [1 : N]$ is fabricated from the decoder (7.2.3) (on p. 202):

$$\mathfrak{D}^k[s_0, \ldots, s_{r-1}] :=$$
$$:= \begin{cases} i & \begin{cases} \text{if the ball in } \mathbb{R}^n \text{ with the radius } b_* \text{ centered at} \\ \mathfrak{X}[t_*^{k+1} - 1, s(0), \ldots, s(t_*^k - 1), s_0, \ldots, s_{r-1}] \\ \text{contains } x_i^k(t_*^{k+1} - 1) \\ \text{and does not contain } x_{i'}^k(t_*^{k+1} - 1) \text{ with any } i' \neq i \end{cases} \\ 1 & \text{otherwise} \end{cases}. \tag{7.6.13}$$

Explanation 7.6.30. Here 1 is taken for the definiteness; in the "otherwise" case, no reasonable decision can be made.

Remark 7.6.31. The kth coding–decoding pair

$$\mathfrak{P}_{\text{cd}}^k := [\mathcal{E}_*^k(\cdot), \mathfrak{D}^k(\cdot)]$$

is random due to Remark 7.6.27.

We consider the variety of all deterministic samples assumed by these random block codes as k runs over $k = 1, 2, \ldots$. In doing so, the block length r is fixed.

7.6.4 Errorless Block Code Hidden within a Tracking Observer

Now we show that the above variety contains a deterministic erorrless block code.

To this end, it will be convenient to think about the channel as a sequence of independent and identically distributed random maps G_t from the input $\mathfrak{E}$ into the output $\mathfrak{S}$ channel alphabet such that

$$\boldsymbol{P}[G_t(e) = s] = W[s|e],$$

where $W(\cdot|\cdot)$ are the channel transition probabilities from Assumption 6.3.1 (on p. 136). Then

$$s(t) = G_t[e(t)],$$

and the coding–decoding pair $\mathfrak{P}_{\text{cd}}^k$ is clearly determined by x_0 and $G_0, \ldots, G_{t_*^k - 1}$.

We also introduce the channel block maps G_k^r. They act componentwise on the code words of length r:

$$G_k^r(E) := \left[G_{t_*^k}(e_0), G_{t_*^k+1}(e_1) \ldots, G_{t_*^{k+1}-1}(e_{r-1})\right], \quad E = [e_0, \ldots, e_{r-1}].$$

Notation 7.6.32. *When a block coding–decoding pair* $\mathfrak{P}_{\text{cd}}$ *with block length* r *is used to transmit a message* i *during the time interval* $[t_*^k : t_*^{k+1} - 1]$*, the result* i_{tr} *depends on not only this pair but also* G_k^r*. With a slight abuse of notation, we write this as follows:*

$$i_{tr} = \mathfrak{P}_{\text{cd}}[i, G_k^r].$$

Now we are in a position to state the key property of the coding–decoding pair $\mathfrak{P}_{\text{cd}}^k$ introduced in the previous subsection.

Lemma 7.6.33. *Whenever* (7.6.9) *holds, decoding is errorless*

$$i = \mathfrak{P}_{\text{cd}}^k[i, G_k^r] \qquad \forall i = 1, \ldots, N \quad \text{and} \quad k = 1, 2, \ldots. \tag{7.6.14}$$

Proof. By construction, the rule $\mathfrak{E}_*^k$ encodes i into

$$E_i^k := [e_i^k(t_*^k), \ldots, e_i^k(t_*^k + r - 1)].$$

Furthermore,

$$\begin{aligned} G_k^r[E_i^k] = [s_i^k(t_*^k), \ldots, s_i^k(t_*^k + r - 1)], \quad \widehat{x}_i^k(t_*^{k+1} - 1) \\ = \mathfrak{X}[t_*^{k+1} - 1, s(0), \ldots, s(t_*^k - 1), s_i^k(t_*^k), \ldots, s_i^k(t_*^{k+1} - 1)]. \end{aligned} \tag{7.6.15}$$

Since the state at $t = t_*^k$ is common for all i by Remark 7.6.25, the set

$$\left\{x_1^k(t_*^{k+1} - 1), \ldots, x_N^k(t_*^{k+1} - 1)\right\}$$

is a displacement of the set

$$\left\{x_1(r - 1), \ldots, x_N(r - 1)\right\}$$

from Lemma 7.6.18. Since the latter is b-separated, so is the former. It follows that any ball of radius $b_* < b/2$ contains no more than one point of the form

$$x_{i'}^k(t_*^{k+1} - 1).$$

At the same time, (7.6.9) ensures that the ball centered at $\widehat{x}_i^k(t_*^{k+1} - 1)$ contains $x_i^k(t_*^{k+1} - 1)$. This and (7.6.13), (7.6.15) imply (7.6.14). □

Remark 7.6.34. Lemma 7.6.33 does not mean that (7.6.13) is an errorless decoding rule for the block code (7.6.11).

Indeed, this lemma states that this rule does not make an error for only a particular sample assumed by the random channel block map G^r. At the same time, the rule is errorless if it does not make errors for any sample that is assumed with a nonzero probability.

Now we are going to show that the variety of samples assumed by the random coding–decoding pairs $\mathfrak{P}_{\text{cd}}^k, k = 1, 2, \ldots$ contains an errorless deterministic pair.

We start with the following simple fact.

Observation 7.6.35. *In any sample sequence of the stochastic process* $\mathfrak{P}_{\mathrm{cd}}^k, k = 1, 2, \ldots$*, some particular coding–decoding pair is encountered infinitely many times.*

This holds since there are only finitely many such pairs with a given block length r and the number N of messages.

With this argument in mind, let us observe all sample sequences that correspond to elementary events for which (7.6.9) holds. Then we arrive at the following.

Observation 7.6.36. *A particular (deterministic) pair* $\mathfrak{P}_{\mathrm{cd}}$ *exists such that with a positive probability, both* (7.6.9) *is true and* $\mathfrak{P}_{\mathrm{cd}}$ *is encountered in the sample sequence* $\{\mathfrak{P}_{\mathrm{cd}}^k\}_{k=1}^{\infty}$ *infinitely many times.*

In other words, a random sequence $1 \leq \tau_1 < \tau_2 < \ldots$ exists such that with a nonzero probability,

$$\mathfrak{P}_{\mathrm{cd}}^{\tau_k} = \mathfrak{P}_{\mathrm{cd}} \quad \forall k \quad \text{and} \quad \text{(7.6.9) holds.} \tag{7.6.16}$$

The following fact plays a key role in the proof of Proposition 7.6.2.

Lemma 7.6.37. *The coding–decoding pair* $\mathfrak{P}_{\mathrm{cd}}$ *is errorless.*

Proof. It is convenient to pick τ_k as a Markov time[10] with respect to the flow $\{\mathfrak{F}_k\}$, where $\mathfrak{F}_k$ is the σ-algebra generated by

$$x_0, G_0, \ldots, G_{t_*^k - 1}. \tag{7.6.17}$$

To this end, we put $\tau_0 := 0$, and given that τ_j has already been chosen, we define τ_{j+1} as the least index $\tau > \tau_j$ for which $\mathfrak{P}_{\mathrm{cd}}^{\tau} = \mathfrak{P}_{\mathrm{cd}}$ if such a τ exists; otherwise, $\tau_{j+1} := \infty$. Clearly, the event $\tau_j = k$ is recognizable from the knowledge of $\mathfrak{P}_{\mathrm{cd}}^1, \ldots, \mathfrak{P}_{\mathrm{cd}}^k$, which are the functions of (7.6.17). Hence $\{\tau_j = k\} \in \mathfrak{F}_k$; i.e., τ_j is a Markov time. Furthermore, $\tau_1 \prec \tau_2 \prec \ldots$ by Notation 7.6.12 (on p. 215).

Putting $k := \tau_j$ into (7.6.14) yields

$$i = \mathfrak{P}_{\mathrm{cd}}[i, G_{\tau_j}^r] \quad \forall i, j \qquad \text{whenever (7.6.16) holds.}$$

To complete the proof, it suffices to show that here $G_{\tau_j}^r, j = 1, 2, \ldots$ a.s.[11] runs over all samples assumed by the random block map G^r with nonzero probabilities.

To this end, we pick such a sample $g(\cdot)$ and apply Corollary 7.6.15 (on p. 215) to $V_k := G_k^r$ and $\mathcal{V} := \{g(\cdot)\}$. This is possible since G_k^r is clearly $\mathfrak{F}_{k+1}$-measurable and independent of $\mathfrak{F}_k$. By this corollary, $G_{\tau_j}^r, j = 1, 2, \ldots$ does run through $g(\cdot)$ infinitely many times as j runs over $j = 1, 2, \ldots$, which completes the proof. □

[10] See Definition 7.6.11 on p. 215.

[11] Given that (7.6.16) holds.

7.6.5 Completion of the Proofs of Proposition 7.6.2, Theorem 7.4.1, and i) of Theorem 7.5.3

Now we bring the pieces together. By Remark 7.6.4 (on p. 211) and Lemma 7.6.16 (on p. 215), it suffices to justify (7.6.6) (on p. 215). We supposed to the contrary that (7.6.6) fails to be true for some coder–decoder pair without (with) a feedback by adopting Assumption 7.6.20 (on p. 218). Lemma 7.6.37 shows that then this pair hides an errorless block code without (with) a feedback with the rate R satisfying (7.6.12) (on p. 220). The block length r can be chosen arbitrariry. By letting $r \to \infty$ in (7.6.12) and invoking the definition of the zero error capacity, we get $\mathfrak{c}_0 \geq H(A)$ or $\mathfrak{c}_{0F} \geq H(A)$, in violation of the assumptions of Proposition 7.6.2. This demonstrates that (7.6.6) does hold, which completes the proof. □

Proof of Theorem 7.4.1 (on p. 206*) and i) of Theorem 7.5.3 (on p.* 209*).* As was remarked, these claims are immediate from Proposition 7.6.2. □

7.7 Almost Sure State Estimation in the Face of System Noises: Proof of Theorem 7.4.5

In this section, we suppose that the assumptions of Theorem 7.4.5 (on p. 207) hold and the channel zero error capacity exceeds the topological entropy of the system:

$$\mathfrak{c}_0 > H(A) \qquad (\text{or } \mathfrak{c}_{0F} > H(A)). \tag{7.7.1}$$

The objective is to construct a coder–decoder pair without (respectively, with) a feedback that (with probability 1) uniformly keeps the estimation error bounded under bounded noises.[12] As in Chaps. 3 and 6, such a pair will be first constructed under the following additional assumption and then extended on the general case of a plant with both unstable and stable modes.

Assumption 7.7.1. *The system* (7.2.1) *has no stable* $|\lambda| < 1$ *eigenvalues* λ.

7.7.1 Construction of the Coder and Decoder-Estimator

The coder and decoder to be considered are basically built on the ideas from Subsect. 3.8.4 (starting on p. 68) subjected to the following two modifications.

m.1) At the beginning τ_i of each operation epoch, the coder determines not the exact state $x(\tau_i)$ but its estimate $\widetilde{x}(\tau_i)$ proceeding from the previous measurements;
m.2) The coder and decoder update the upper bound δ for the estimation error in such a way that this bound does not approach zero as time progresses.

Explanation 7.7.2. The coder is unable to determine the exact state due to the plant and sensor noises. At the same time, it is able to estimate it with an accuracy, which is determined by the noise level and so is bounded in time.

[12] See Definition 7.2.5 on p. 203.

Explanation 7.7.3. Since the estimation error cannot be made decaying to zero because of the system noises, it is not natural to employ, like in Subsects. 3.8.4, 6.8.2, and 6.9.1, an algorithm making the upper bound δ of this error converging to zero.

We recall that previously the basic rule to update δ was of the form:

$$\delta(\tau_i) := \delta_c(\tau_i) \times \mu_i,$$

where $\mu_i := \gamma > 1$ if the alarm message was received at time τ_i and $\mu_i := \rho \in (0, 1)$ otherwise. Here γ and ρ are the parameters of the observer. If the alarm is not frequent, this rule makes δ converging to zero.

Now the rule to update δ will be modified as follows:

$$\delta(\tau_i) := \delta_c(\tau_i) \times \mu_i + \delta_*, \tag{7.7.2}$$

where $\delta_* > 0$ is one more parameter. This prevents δ from approaching zero.

For the convenience of the reader, we offer a systematic description of the coder and decoder. As discussed, they are reminiscent of those previously considered.

Components of the Observer

To construct an observer, we pick

1) two numbers η and R_* such that

$$H(A) < \eta < R_* < \mathfrak{c}_0 \quad (\text{or } \mathfrak{c}_{0F}); \tag{7.7.3}$$

2) two parameters

$$\gamma > \|A\| \qquad \text{and} \qquad \delta_* > 0; \tag{7.7.4}$$

3) an r-contracted quantizer $\mathfrak{Q}_r$ from Lemma 3.8.36 (on p. 80) applied to $A_{ss} := A$ (we also invoke Remark 3.8.37 on p. 80 here) for any $r = 1, 2, \ldots$;
4) a block code (7.3.1) (or block function (7.3.4)) with block length r, rate $R \geq R_*$, and an errorless decoding rule $\mathcal{D}_r(\cdot)$ for all sufficiently large r;
5) a particular sufficiently large r such that all outputs of the quantizer (including the alarm symbol) can be encoded by means of this block-code (or block function);
6) a deadbeat observer; i.e., a device producing an estimate $\widetilde{x}(t)$ of the current state $x(t)$ with a bounded error

$$\big\|x(t) - \widetilde{x}(t)\big\| \leq \overline{D} < \infty \qquad \forall t \tag{7.7.5}$$

in the form of a sliding average of $n - 1$ past observations

$$\widetilde{x}(t) = \sum_{\theta = t-n+1}^{t} F_{t-\theta} y(\theta).$$

Explanation 7.7.4. • 1) is possible thanks to (7.7.1).

- In 2), $\|A\| = \max_{x:\|x\|=1} \|Ax\|$ is the matrix norm associated with the vector norm $\|\cdot\|$ from Definition 3.8.6 (on p. 69).
- 4) is possible due to Remarks 7.3.3 and 7.3.6 (on pp. 204 and 205).
- 5) is possible since the number of quantizer outputs (including the alarm symbol) $N' \lesssim 2^{r\eta} + 1$ by Lemma 3.8.36 (on p. 80), whereas the block code is capable of encoding $N'' = 2^{Rr} \geq 2^{R_* r}$ messages for errorless transmission, and $N'' > N'$ for all $r \approx \infty$ by (7.7.3).
- In 6), a deadbeat observer does exist since the pair (A, C) is observable by Assumptions 6.3.6 and 7.7.1 (on pp. 137, 223) [8]. It can be chosen so that the estimate accuracy linearly depends $\overline{D} = k_{A,C}(D + D_\chi)$ on the noise bounds from (7.2.2) (on p. 202) [8].

Finally, we introduce the operation epochs of the common duration r:

$$[\tau_i : \tau_{i+1}), \qquad \tau_i := ir.$$

Description of the Coder and Decoder

The decoder computes not only a state estimate $\widehat{x}(t)$ but also an upper (and maybe incorrect) bound $\delta(t)$ of its exactness. To do so, it employs the recursion

$$\widehat{x}(t+1) = A\widehat{x}(t), \qquad \delta(t+1) = \delta(t) \tag{7.7.6}$$

at any time. However at times $t = \tau_i$, this recursion is prefaced by correcting $\widehat{x}(t)$ and $\delta(t)$[13] on the basis of the message received from the coder over the channel during the previous operation epoch. The errorless block code (function) is employed to serve this transmission. So the coder is aware of the message driving the decoder and is thus able to duplicate its operations and thus to compute $\widehat{x}(t)$ and $\delta(t)$ by itself.

Finally, both coder and decoder carry out the recursion (7.7.6) at any time and preface it by the following operations at times $t = \tau_i$.

The coder (at times $t = \tau_i, i = 1, 2, \ldots$)

c.1) Proceeding from the previous measurements, calculates the estimate $\widetilde{x}(\tau_i)$ of the current state by means of the deadbeat observer;

c.2) Computes the corrected values $\widehat{x}^+(\tau_i)$ and $\delta^+(\tau_i)$ by duplicating the operations of the decoder;

c.3) Employs the quantizer $\mathfrak{Q}_r$ and computes the quantized value $q(\tau_i)$ of the current scaled discrepancy between the estimates produced by the coder–decoder pair and the deadbeat observer, respectively:

$$\varepsilon(\tau_i) := \left[\delta^+(\tau_i)\right]^{-1}\left[\widetilde{x}(\tau_i) - \widehat{x}^+(\tau_i)\right]; \tag{7.7.7}$$

c.4) Encodes this value by means of the block code (function) and transmits the obtained code word of length r across the channel during the next operation epoch $[\tau_i : \tau_{i+1})$.

[13]The values before and after correction are marked by the indices $^-$ and $^+$, respectively.

Only after this, does the coder perform the current step (7.7.6).

The decoder (at times $t = \tau_i, i = 2, 3, \ldots$)

d.1) Applies the errorless decoding rule $\mathcal{D}_r(\cdot)$ to the data received within the previous operation epoch $[\tau_{i-1} : \tau_i)$ and thus computes the quantized and scaled discrepancy $q(\tau_{i-1})$;

d.2) Corrects successively the estimate and the exactness bound:

$$\widehat{x}^+(\tau_i) := \widehat{x}^-(\tau_i) + \delta^-(\tau_i) A^r \overset{\star}{q}(\tau_{i-1}),$$
$$\delta^+(\tau_i) := \delta^-(\tau_i) \times \left[\langle q(\tau_{i-1}) \rangle_{\varkappa,\gamma} \right]^r + \delta_*. \quad (7.7.8)$$

Here $\varkappa \in (0, 1)$ is the parameter from Lemma 3.8.36 (on p. 80) and

$$\overset{\star}{q} := \begin{cases} q & \text{if } q \neq \maltese \\ 0 & \text{otherwise} \end{cases}, \qquad \langle q \rangle_{\varkappa,\gamma} := \begin{cases} \varkappa & \text{if } q \neq \maltese \\ \gamma & \text{otherwise} \end{cases}. \quad (7.7.9)$$

Only after this does the decoder perform the step (7.7.6).

Remark 7.7.5. Initially, the coder and decoder are given common and arbitrarily chosen values

$$\widehat{x}(0) = \widehat{x}_0 \quad \text{and} \quad \delta(0) = \delta_0 \geq \delta_*. \quad (7.7.10)$$

The bound δ_0 may be incorrect; i.e., the inequality $\delta_0 \geq \|x_0 - \widehat{x}_0\|$ may be violated.

Remark 7.7.6. For technical convenience, we put $q(\tau_0) := \maltese$ and suppose that the coder and decoder act accordingly.

Explanation 7.7.7. The step c.2) of the coder comes to carrying out d.2).

Observation 7.7.8. *The coder and decoder do not employ the noise bounds D, D_χ from* (7.2.2) *(on p.* 202*).*

Remark 7.7.9. Although the coder and decoder compute $\widehat{x}(t)$ and $\delta(t)$ independently, they generate common sequences $\widehat{x}(t)$ and $\delta(t), t = 0, 1, \ldots$.

Observation 7.7.10. *The coder and decoder introduced in this Subsection are semi-rational finite-step recursive.*[14]

This is straightforward from the description of the coder and decoder with regard to the fact that in c.3) the quantizer is taken from Lemma 3.8.36 (on p. 80) and so is polyhedral.[15] In (6.3.4) (on p. 139) and (6.3.6) (on p. 140), the states of the coder and decoder can be defined as $z_c := [\widehat{x}^+, \widetilde{x}, \delta^+]$ and $z_d := [\widehat{x}^+, \delta^+]$, respectively.

Observation 7.7.11. *Due to* (7.7.6)*,* (7.7.8)*, and* (7.7.10)*,*

$$\delta(t) \geq \delta_* > 0 \qquad \forall t \geq 0. \quad (7.7.11)$$

[14] See Definitions 6.3.9 and 6.3.10 on pp. 139 and 140, respectively.

[15] See Definition 3.8.8 on p. 70.

7.7.2 Almost Sure State Estimation in the Face of System Noises

The main result of the subsection is as follows.

Proposition 7.7.12. *Suppose that Assumption* 7.7.1 *(on p.* 223*) holds. The coder–decoder pair introduced in Subsect.* 7.7.1 *uniformly keeps the estimation error bounded under bounded noises*[16] *with probability 1.*

The temporary additional Assumption 7.7.1 can be dropped by applying the coder–decoder pair at hand to the unstable part of the plant, like in Chaps. 3 and 6. This gives rise to the following.

Remark 7.7.13. Modulo the arguments from Sect. 6.10 (starting on p. 179), Proposition 7.7.12 implies Theorem 7.4.5 (on p. 207).[17]

The remainder of the subsection is devoted to the proof of Proposition 7.7.12.

Estimating the Quantities Generated by the Observer

Now we study the process in the system equipped by the coder and decoder, with focusing attention on

$$\delta_i := \delta^+(\tau_i), \quad z_i := \left\| \widehat{x}^+(\tau_i) - \widetilde{x}(\tau_i) \right\|, \quad \varsigma_i := \frac{z_i}{\delta_i}. \tag{7.7.12}$$

Here δ_i is the exactness bound generated by the coder–decoder pair, z_i is the discrepancy between the estimates produced by this pair and the deadbeat observer, and ς_i determines whether the alarm symbol ✠ is sent over the channel:

$$\varsigma_i > 1 \Leftrightarrow q(\tau_i) = \maltese.$$

Explanation 7.7.14. To make z_i and ς_i defined for $i := 0$, we choose $\widehat{x}^+(0)$ to be equal to the initial state estimate $\widehat{x}_0$ and $\widetilde{x}(0) \neq \widehat{x}_0$ so that (7.7.5) be true for $t = 0$, and we pick $\varsigma_0 := 1 + z_0/\delta_0$.

We stress that these conventions do not concern the operation of the coder and decoder. They establish notation that makes subsequent formulas more homogeneous.

We are going to show that the following quantities:

$$W_i := \delta_i^\sigma \max\{\varsigma_i; \varkappa^r\} \tag{7.7.13}$$

satisfy recursive inequalities of the form

$$W_i \leq \mu_r W_{i-1} + g_r, \qquad i = 1, 2, \dots. \tag{7.7.14}$$

Here $\sigma \in (0,1)$ is an arbitrarily chosen parameter, r is the duration of the operation epoch, and $\varkappa \in (0,1)$ is taken from Lemma 3.8.36 (on p. 80). To this end, we put

[16] See Definition 7.2.5 on p. 203.

[17] See Subsect. 7.8.5 starting on p. 244 for details.

$$\rho := \frac{\|A\|}{\gamma} \overset{(7.7.4)}{<} 1, \qquad D_r := \overline{D}(1 + \|A\|^r) + D \sum_{j=0}^{r-1} \|A\|^j; \tag{7.7.15}$$

$$\mu_r := \max\left\{\varkappa^{\sigma r}; (\gamma^r + 1)^\sigma \max\{\rho^r, \varkappa^r\}\right\},$$
$$g_r := \varkappa^r(\varkappa^r + 1)^\sigma \delta_*^\sigma + \frac{D_r}{\delta_*^{1-\sigma}}, \tag{7.7.16}$$

where γ and δ_* are the parameters of the observer from 2) (on p. 224), $\overline{D}$ is the exactness of the deadbeat observer from (7.7.5), and D is the plant noise upper bound from (7.2.2) (on p. 202).

Proposition 7.7.15. *For any $i \geq 1$ and $\sigma \in (0, 1)$, inequalities* (7.7.14) *hold.*

Corollary 7.7.16. *If $\sigma \in (0, 1)$ is chosen so that*

$$\sigma < -r\frac{\max\{\log_2 \rho; \log_2 \varkappa\}}{\log_2(\gamma^r + 1)} = \frac{-\max\{\log_2 \rho; \log_2 \varkappa\}}{\log_2 \gamma \log_2(1 + \gamma^{-r})}, \tag{7.7.17}$$

then $\mu_r < 1$ and so (7.7.14) *means that $W(\delta, \varsigma) := \delta^\sigma \max\{\varsigma; \varkappa^r\}$ is a Lyapunov function.*

By the standard arguments, this implies that the quantities (7.7.13) are bounded as i runs over $i = 1, 2, \ldots$. As will be shown, it follows from this that δ_i, z_i and hence the estimation error $x(t) - \widehat{x}(t)$ are also bounded, which completes the proof of Proposition 7.7.12.

Remark 7.7.17. The right-hand side of (7.7.17) is positive since $\rho < 1$ by (7.7.15) and $\varkappa < 1$ by Lemma 3.8.36 (on p. 80). Hence $\sigma \in (0, 1)$ can be chosen to satisfy (7.7.17).

The proof of Proposition 7.7.15 is prefaced by a technical lemma. To state it, we introduce the indicator functions of the following events:

$$I_0(i) \longleftrightarrow \varsigma_{i-1} \leq 1, \qquad I_{\maltese}(i) \longleftrightarrow \varsigma_{i-1} > 1. \tag{7.7.18}$$

Lemma 7.7.18. *The following relations hold for any $i \geq 1$:*

$$\delta_i = \delta_{i-1}\Big[\varkappa^r I_0(i) + \gamma^r I_{\maltese}(i)\Big] + \delta_*; \tag{7.7.19}$$

$$z_i \leq z_{i-1}\|A\|^r I_{\maltese}(i) + \delta_{i-1}\varkappa^{2r} I_0(i) + D_r. \tag{7.7.20}$$

Proof. Equation (7.7.19) is immediate from (7.7.6), (7.7.8), (7.7.9), and (7.7.18).

To justify (7.7.20), we observe that

$$z_i \overset{(7.7.12)}{=\!=} \|\widehat{x}^+(\tau_i) - \widetilde{x}(\tau_i)\| \overset{(7.7.8)}{=\!=} \|\widehat{x}^-(\tau_i) + \delta^-(\tau_i)A^r \overset{\star}{q}(\tau_{i-1}) - \widetilde{x}(\tau_i)\|$$
$$\overset{(7.7.6)}{=\!=} \Big\|A^r\Big\{\big[\widehat{x}^+(\tau_{i-1}) - \widetilde{x}(\tau_{i-1})\big] + \delta^+(\tau_{i-1})\overset{\star}{q}(\tau_{i-1})\Big\} + A^r\widetilde{x}(\tau_{i-1}) - \widetilde{x}(\tau_i)\Big\|$$
$$\leq a_1 + a_2,$$

where

$$a_1 := \left\| A^r \left\{ \left[\widetilde{x}(\tau_{i-1}) - \widehat{x}^+(\tau_{i-1}) \right] - \delta^+(\tau_{i-1}) \overset{\star}{q} (\tau_{i-1}) \right\} \right\|;$$

$$a_2 := \left\| A^r \widetilde{x}(\tau_{i-1}) - \widetilde{x}(\tau_i) \right\| = \left\| A^r \left[\widetilde{x}(\tau_{i-1}) - x(\tau_{i-1}) \right] + \left[x(\tau_i) - \widetilde{x}(\tau_i) \right] \right.$$
$$\left. + \left[A^r x(\tau_{i-1}) - x(\tau_i) \right] \right\| \overset{(7.2.1),(7.7.5)}{\leq} \overline{D} + \|A\|^r \overline{D} + \left\| \sum_{\theta=\tau_{i-1}}^{\tau_i - 1} A^{\tau_i - 1 - \theta} \xi(\theta) \right\|$$
$$\overset{(7.2.2)}{\leq} \overline{D}\left(1 + \|A\|^r\right) + D \sum_{j=0}^{r-1} \|A\|^j \overset{(7.7.15)}{=\!=} D_r.$$

If $I_{\maltese}(i) = 1$, then

$$q(\tau_{i-1}) = \maltese \xrightarrow{(7.7.9)} \overset{\star}{q} (\tau_{i-1}) = 0$$

and so

$$a_1 \leq \|A\|^r \|\widehat{x}^+(\tau_{i-1}) - \widetilde{x}(\tau_{i-1})\| \overset{(7.7.12)}{=\!=} \|A\|^r z_{i-1}.$$

If $I_0(i) = 1$, then

$$\maltese \neq q(\tau_{i-1}) \overset{(7.7.9)}{=\!=} \overset{\star}{q} (\tau_{i-1}),$$

where $q(\tau_{i-1})$ is the quantized value of $\varepsilon(\tau_{i-1})$ given by (7.7.7). So (3.8.9) (on p. 70) (where $\rho_{\mathfrak{Q}} = \varkappa^{2m}$ by Lemma 3.8.36 on p. 80) implies

$$a_1 = \delta^+(\tau_{i-1}) \left\| A^r \left[\varepsilon(\tau_{i-1}) - q(\tau_{i-1}) \right] \right\| \overset{(7.7.12)}{\leq} \delta_{i-1} \varkappa^{2r}.$$

Summarizing, we arrive at (7.7.20). □

With regard to the last and first formulas from (7.7.12) and (7.7.15), respectively, and (7.7.18), Lemma 7.7.18 yields the following.

Corollary 7.7.19. *For* $i \geq 1$, *the following inequality holds:*

$$\varsigma_i \leq \left\{ \begin{array}{ll} \rho^r \varsigma_{i-1} + \frac{D_r}{\delta_i} & \text{if } \varsigma_{i-1} > 1 \\ \varkappa^r + \frac{D_r}{\delta_i} & \text{if } \varsigma_{i-1} \leq 1 \end{array} \right\}. \tag{7.7.21}$$

Proof of Proposition 7.7.15. If $\varsigma_{i-1} \leq 1$, then

$$W_i \overset{(7.7.13)}{=\!=} \delta_i^\sigma \max\left\{\varsigma_i; \varkappa^r\right\} \overset{(7.7.21)}{\leq} \delta_i^\sigma \max\left\{\varkappa^r + \frac{D_r}{\delta_i}; \varkappa^r\right\} = \delta_i^\sigma \varkappa^r + \frac{D_r}{\delta_i^{1-\sigma}}$$
$$\overset{(7.7.11)}{\leq} \delta_i^\sigma \varkappa^r + \frac{D_r}{\delta_*^{1-\sigma}} \overset{(7.7.19)}{=\!=} \varkappa^r \left(\varkappa^r \delta_{i-1} + \delta_*\right)^\sigma + \frac{D_r}{\delta_*^{1-\sigma}}.$$

It is easy to check that the function

$$\varphi(\delta) := (\varkappa^r \delta + \delta_*)^\sigma - \varkappa^{\sigma r} \delta^\sigma$$

decreases for $\delta > 0$. So $\varphi(\delta) \leq \varphi(\delta_*)$ for $\delta \geq \delta_*$, which implies

$$(\varkappa^r \delta + \delta_*)^\sigma \leq \varkappa^{\sigma r} \delta^\sigma + \varphi(\delta_*) \leq \varkappa^{\sigma r} \delta^\sigma + (\varkappa^r + 1)^\sigma \delta_*^\sigma.$$

Hence

$$W_i \leq \varkappa^r \varkappa^{\sigma r} \delta_{i-1}^\sigma + \varkappa^r (\varkappa^r + 1)^\sigma \delta_*^\sigma + \frac{D_r}{\delta_*^{1-\sigma}}$$
$$\overset{(7.7.13)}{\leq} \varkappa^{\sigma r} W_{i-1} + \varkappa^r (\varkappa^r + 1)^\sigma \delta_*^\sigma + \frac{D_r}{\delta_*^{1-\sigma}} \overset{(7.7.16)}{\leq} \mu_r W_{i-1} + g_r;$$

i.e., (7.7.14) is valid.

Now consider the case where $\varsigma_{i-1} > 1$. Then

$$W_i \overset{(7.7.13)}{=\!=} \delta_i^\sigma \max\left\{\varsigma_i; \varkappa^r\right\} \overset{(7.7.21)}{\leq} \delta_i^\sigma \max\left\{\rho^r \varsigma_{i-1} + \frac{D_r}{\delta_i}; \varkappa^r\right\} \overset{\varsigma_{i-1}>1}{\leq}$$
$$\delta_i^\sigma \max\left\{\rho^r \varsigma_{i-1} + \frac{D_r}{\delta_i}; \varkappa^r \varsigma_{i-1}\right\} \leq \delta_i^\sigma \left[\max\left\{\rho^r; \varkappa^r\right\}\varsigma_{i-1} + \frac{D_r}{\delta_i}\right]$$
$$\leq \delta_i^\sigma \max\left\{\rho^r; \varkappa^r\right\} \max\left\{\varsigma_{i-1}; \varkappa^r\right\} + \frac{D_r}{\delta_i^{1-\sigma}}$$
$$\overset{(7.7.19)}{=\!=} (\gamma^r \delta_{i-1} + \delta_*)^\sigma \max\left\{\rho^r; \varkappa^r\right\} \max\left\{\varsigma_{i-1}; \varkappa^r\right\} + \frac{D_r}{\delta_i^{1-\sigma}}$$
$$\overset{(7.7.11)}{\leq} (\gamma^r + 1)^\sigma \delta_{i-1}^\sigma \max\left\{\rho^r; \varkappa^r\right\} \max\left\{\varsigma_{i-1}; \varkappa^r\right\} + \frac{D_r}{\delta_*^{1-\sigma}}$$
$$\overset{(7.7.13),\,(7.7.16)}{\leq} \mu_r W_{i-1} + g_r;$$

i.e., (7.7.14) is true. □

Lemma 7.7.20. *The quantities* $\varsigma_i, z_i, \delta_i$ *are bounded as* i *runs over* $0, 1, \ldots$*:*

$$\varsigma_i \leq \overline{\varsigma}, \quad \delta_i \leq \overline{\delta}, \quad z_i \leq \overline{z}. \tag{7.7.22}$$

Here all upper bounds are uniquely determined by

$$\delta_0,\ \delta_*,\ r,\ D_r,\ \varkappa,\ \|A\|,\ \gamma \tag{7.7.23}$$

and an arbitrary upper bound on z_0*.*

Proof. By invoking Corollary 7.7.16 and Remark 7.7.17 (on p. 228), we pick $\sigma \in (0, 1)$ so that $\mu_r < 1$ in (7.7.14). By induction on i, it is easy to check that

$$W_i \leq \mu_r^i W_0 + g_r \frac{1 - \mu_r^i}{1 - \mu_r} \overset{(7.7.12),\,(7.7.13)}{\leq} \mu_r^i \delta_0^\sigma \max\left\{\frac{z_0}{\delta_0} + 1; \varkappa^r\right\} + \frac{g_r}{1 - \mu_r}.$$

Here the last inequality holds thanks to Explanation 7.7.14. It follows that

$$W_i \le \overline{W} := \delta_0^\sigma \max\left\{\frac{z_0}{\delta_0}+1; \varkappa^r\right\} + \frac{g_r}{1-\mu_r}.$$

Due to (7.7.11) and (7.7.13),

$$W_i \ge \delta_*^\sigma \varsigma_i, \qquad W_i \ge \delta_i^\sigma \varkappa^r.$$

Hence

$$\varsigma_i \le \overline{\varsigma} := \delta_*^{-\sigma}\overline{W}, \quad \delta_i \le \overline{\delta} := \left(\varkappa^{-r}\overline{W}\right)^{\frac{1}{\sigma}}, \quad z_i \overset{(7.7.12)}{=\!=} \delta_i\varsigma_i \le \overline{z} := \overline{\delta}\overline{\varsigma}. \qquad \square$$

Observation 7.7.21. *As follows from the proof,* (7.7.22) *is a consequence of* (7.7.18)–(7.7.20)*; the last and first formulas from* (7.7.12) *and* (7.7.15)*, respectively; and the inequality* $\delta_i \ge \delta_*$.

Proof of Proposition 7.7.12: Concluding Part

Thanks to Explanation 7.7.14,

$$z_0 = \|\widehat{x}_0 - \widetilde{x}(0)\| \le \|\widehat{x}_0\| + \|x_0\| + \overline{D}.$$

So (7.7.22) holds with the upper bounds determined by $\|\widehat{x}_0\|, \|x_0\|, \overline{D}$, and the parameters from (7.7.23).[18] Hence the estimation error at time $t = \tau_i$ is also bounded:

$$\text{err}_i := \left\|x(\tau_i) - \widehat{x}^+(\tau_i)\right\| \le \|x(\tau_i) - \widetilde{x}(\tau_i)\| + \left\|\widetilde{x}(\tau_i) - \widehat{x}^+(\tau_i)\right\| \overset{(7.7.5),\,(7.7.12)}{\le} \overline{D} + z_i \le \overline{D} + \overline{z}.$$

It remains to extend this conclusion on all $t = 0, 1, \ldots$. For any $t \ne \tau_j \ \forall j$, an index i exists such that $\tau_i < t < \tau_{i+1}$. Then

$$\|x(t) - \widehat{x}(t)\| \overset{(7.2.1),(7.7.6)}{=\!=\!=} \left\| A^{t-\tau_i}x(\tau_i) + \sum_{\theta=\tau_i}^{t-1} A^{t-1-\theta}\xi(\theta) - A^{t-\tau_i}\widehat{x}^+(\tau_i)\right\|$$
$$\le \|A\|^{t-\tau_i}\left\|x(\tau_i) - \widehat{x}^+(\tau_i)\right\| + \sum_{\theta=\tau_i}^{t-1}\|A\|^{t-1-\theta}\|\xi(\theta)\|$$
$$\overset{(7.2.2),\|A\|\ge 1,\tau_{i+1}-\tau_i=r}{\le} \|A\|^r\left(\overline{D}+\overline{z}\right) + D\sum_{\theta=0}^{r-1}\|A\|^\theta,$$

where the inequality $\|A\| \ge 1$ holds by Assumption 7.7.1 (on p. 223). Definition 7.2.5 (on p. 203) completes the proof. $\square$

[18] If the initial state is bounded $\|x_0\| \le D_x$ a.s., then $\|x_0\|$ can be replaced by D_x here. Modulo the arguments to follow, this proves Remark 7.4.7 (on p. 208).

7.7.3 Completion of the Proof of Theorem 7.4.5

For systems with no stable modes (i.e., under Assumption 7.7.1 on p. 223), this theorem is justified by Proposition 7.7.12. Thus completion of the proof comes to dropping Assumption 7.7.1. This is accomplished just like in Subsects. 3.8.11 and 6.7.6 (starting on pp. 88 and 156, respectively) by estimating only the "unstable" part of the state.

Thus we do not suppose any longer that Assumption 7.7.1 holds. At the same time, we assume that (7.7.1) is true and, temporarily, that the initial state is bounded $\|x_0\| \leq D_x$ a.s. We also focus on the unstable part of the system (7.2.1) (on p. 202); i.e., the noisy analog of the system (6.7.24) (on p. 156)

$$x_+(t+1) = A_+x_+(t) + \pi_+\xi(t),\ x_+(t) := \pi_+x(t) \in M_{\text{unst}};$$
$$x_+(0) = x_0^+ := \pi_+x_0, \quad y(t) = Cx_+(t) + \chi^+(t). \quad (7.7.24)$$

We recall that M_{unst} and M_{st} are the invariant subspaces of A related to the unstable and stable parts of its spectrum, respectively, and π_+, π_- are the projections onto M_{unst} parallel to M_{st} and vice versa. Furthermore,

$$\chi^+(t) := \chi(t) + Cx_-(t), \quad \text{where} \quad x_-(t) := \pi_-x(t).$$

Our immediate goal is to show that the noise $\chi^+(t)$ is bounded. To this end, we denote by A_- the operator A acting on its stable invariant subspace M_{st} and note that A_- is asymptotically stable: $\|A_-^t\| \leq cv^t\ \forall t$, where $c > 0$ and $v \in (0,1)$. Hence

$$\|x_-(t)\| = \|\pi_-x(t)\| \overset{(7.2.1)}{=\!=} \left\| A_-^t\pi_-x_0 + \sum_{\theta=0}^{t-1} A_-^{t-1-\theta}\pi_-\xi(\theta) \right\|$$
$$\leq \|\pi_-\| \left[\|A_-^t\|\|x_0\| + \sum_{\theta=0}^{t-1} \|A_-^{t-1-\theta}\|\|\xi(\theta)\| \right] \overset{(7.2.2)}{\leq} c\|\pi_-\| \left[v^tD_x + D\sum_{\theta=0}^{t-1} v^{t-1-\theta} \right]$$
$$\leq \overline{c}_- := c\|\pi_-\| \left[D_x + \frac{D}{1-v} \right], \qquad \|\chi^+(t)\| \overset{(7.2.2)}{\leq} D_\chi + \overline{c}_-\|C\|. \quad (7.7.25)$$

Thus the sensor noise in (7.7.24) is bounded.

As was shown in Subsect. 6.7.6 (starting on p. 156), Assumptions 6.3.1, 6.3.4, 6.3.6 (on pp. 136 and 137), and inequality (7.7.1) (on p. 223) remain true for the reduced plant (7.7.24), which also satisfies Assumption 7.7.1 (on p. 223). Thus Proposition 7.7.12 is true for the plant (7.7.24). Now we apply the corresponding observer to the primal system (7.2.1) (on p. 202) and interpret its output $\widehat{x}_+(t)$ as an estimate $\widehat{x}(t) := \widehat{x}_+(t)$ of the state $x(t)$ of the primal plant. By invoking (7.7.25), we see that

$$\|x(t) - \widehat{x}(t)\| = \|x_-(t) + x_+(t) - \widehat{x}_+(t)\| \leq \|x_-(t)\| + \|x_+(t) - \widehat{x}_+(t)\|$$
$$\leq \|x_+(t) - \widehat{x}_+(t)\| + \overline{c}_-.$$

It follows that relation (7.2.7) (on p. 203) is true for the primal plant (7.2.1) since it holds for $x(t) := x_+(t)$ and $\widehat{x}(t) := \widehat{x}_+(t)$ by Proposition 7.7.12.

To complete the proof, the assumption that x_0 is bounded should be dropped. This is accomplished by considering the process in the conditional probability space given that $\|x_0\| \leq D_x$, along with letting $D_x \to \infty$. □

7.8 Almost Sure Stabilization in the Face of System Noises: Proof of (ii) from Theorem 7.5.3

In this section, we suppose that the assumptions of Theorem 7.5.3 (on p. 209) hold and that the channel zero error capacity exceeds the topological entropy of the system:

$$\mathfrak{c}_{0F} > H(A). \tag{7.8.1}$$

The objective is to construct a coder–decoder pair without a feedback that (with probability 1) uniformly keeps the stabilization error bounded under bounded noises.[19] As in the previous section, such a pair will be first constructed for systems with no stable modes and then extended on the general case, where both unstable and stable modes may occur. So we adopt Assumption 7.7.1 (on p. 223), until otherwise stated. Modulo the assumptions of Theorem 7.5.3, this implies that the pairs (A, B) and (A, C) are controllable and observable, respectively.

The coder and decoder to be considered are basically those from Subsect. 3.8.4 (starting on p. 68) subjected to the modifications m.1), m.2) (on p. 223). The stability of the closed-loop system will be proved along the lines from the previous section. The major new issue to be addressed now proceeds from the fact that in (7.8.1), the zero error capacity with a complete feedback is concerned, whereas no feedback communication link is in fact available. It should be shown that nevertheless, the controller is able to make the sensor aware about the previous result of transmission across the erroneous feedforward communication channel at each time step. We show that this may be accomplished by means of control actions upon the plant.

7.8.1 Feedback Information Transmission by Means of Control

In this subsection, we show that as much information as desired can be transmitted in such a way during a time step. Specifically, we bear in mind that the controller makes a choice from a set of N possibilities (with N arbitrary) and has to make the sensor aware about the choice, provided that the set is known at the sensor site.

Remark 7.8.1. Feedback data communication scheme from Subsect. 6.11.2 (starting on p. 185) does not achieve this objective since it ensures transmission of only a small average amount of data per unit time (one bit per potentially long operation epoch).

We start with a well-known technical fact [8].

[19] See Definition 7.5.2 on p. 209.

Lemma 7.8.2. *Given* $r = 0, 1, \ldots$, *the controls* $u(t-n+1), \ldots, u(t+r-1)$ *and observations* $y(t-n+1), \ldots, y(t)$, *the prognosis*[20] $\widetilde{x}(t+r|t)$ *of the state* $x(t+r)$ *can be constructed as a linear function of the available data*

$$\widetilde{x}(t+r|t) = \sum_{\theta=t-n+1}^{t} F_{t-\theta}^{[r]} y(\theta) + \sum_{\theta=t-n+1}^{t+r-1} G_{t-\theta}^{[r]} u(\theta) \tag{7.8.2}$$

so that the error does not exceed a bound $D_x^{\langle r \rangle} < \infty$, *which does not depend on time* t, *controls, and observations.*

Remark 7.8.3. The estimation accuracy linearly depends $D_x^{\langle r \rangle} = k_{A,C,B,r} \cdot (D + D_\chi)$ on the noise bounds from (7.2.2) (on p. 202).

The next lemma displays the key fact that enables one to establish the feedback communication by means of control. To state it, we denote

$$d := \min\{j = 1, 2, \ldots : CA^{j-1}B \neq 0\}. \tag{7.8.3}$$

Since the pairs (A, B) and (A, C) are controllable and observable, respectively, the integer d is well defined.

Lemma 7.8.4. *For any* $N \geq 1$, *there exists an ordered set of* N *different controls*

$$\mathfrak{U} = \{u_1, \ldots, u_N\} \subset \mathbb{R}^{n_u} \tag{7.8.4}$$

with the following property. Given

$$t \geq n + d, \quad y(t-n-d+1), \ldots, y(t), \quad u(t-n-d+1), \ldots, u(t-1-d)$$

and that $u(t-d) = u + u_\nu$, *where* u *is known but* ν *is not known, there is a way to determine* ν *from these data.*

Remark 7.8.5. At time t, the observations $y(t-n-d+1), \ldots, y(t)$ are known at the sensor site. If the controls $u(t-n-d+1), \ldots, u(t-1-d)$, u and the ordered set (7.8.4) are also known there, the value of ν can be determined at this site. Thus by applying the control $u + u_\nu$, the controller can communicate ν to the sensor site.

Definition 7.8.6. *The set* (7.8.4) *is called the* feedback control alphabet*.*

Proof of Lemma 7.8.4. We first observe that

$$y(t) = CA^d x(t-d) + \sum_{\theta=t-d}^{t-1} CA^{t-1-\theta}\big[Bu(\theta) + \xi(\theta)\big]$$

$$\overset{(7.8.3)}{=\!=} CA^d x(t-d) + CA^{d-1}Bu(t-d) + \sum_{\theta=t-d}^{t-1} CA^{t-1-\theta}\xi(\theta).$$

[20] Estimate for $r = 0$.

By employing the state estimate $\widetilde{x}(t-d|t-d)$ given by Lemma 7.8.2, we get

$$\left\|y(t) - CA^d\widetilde{x}(t-d|t-d) - CA^{d-1}Bu(t-d)\right\|$$
$$= \left\|CA^d[x(t-d) - \widetilde{x}(t-d|t-d)] + \sum_{\theta=t-d}^{t-1} CA^{t-1-\theta}\xi(\theta)\right\| \leq D_x^{\langle 0\rangle} \left\|CA^d\right\|$$
$$+ D\sum_{\theta=t-d}^{t-1} \left\|CA^{t-1-\theta}\right\| \xlongequal{j:=t-1-\theta} b_* := D_x^{\langle 0\rangle} \left\|CA^d\right\| + D\sum_{j=0}^{d-1} \left\|CA^j\right\|. \quad (7.8.5)$$

We proceed by invoking that $u(t-d) = u + u_\nu$

$$\left\|z(t) - CA^{d-1}Bu_\nu\right\| < b_* + 1, \quad (7.8.6)$$

where

$$z(t) := y(t) - CA^d\widetilde{x}(t-d|t-d) - CA^{d-1}Bu$$

is computable from the given data. Let us pick the set $\{u_\nu\}$ so that the points $y_\nu := CA^{d-1}Bu_\nu$ be $(2b_* + 3)$-separated: $\|y_{\nu'} - y_{\nu''}\| \geq 2b_* + 3$ if $\nu' \neq \nu''$. This is possible since $CA^{d-1}B \neq 0$ by (7.8.3). Then inequality (7.8.6) holds for the unique ν and so can serve as an errorless decision rule to determine ν. □

Remark 7.8.7. By employing the Euclidean norm $\|\cdot\|$ and taking the second power of the left- and right-hand sides of (7.8.6), we see that ν can be determined by checking finitely many inequalities with rational (in available data) terms. In other words, the decision map $y(t-n-d+1), \dots, y(t), u(t-n-d+1), \dots, u(t-1-d), u \mapsto \nu$ is semialgebraic.[21]

7.8.2 Zero Error Capacity with Delayed Feedback

By Lemma 7.8.4, the feedback communication by means of control suffers from the delay given by (7.8.3). In this subsection, we show that this delay does not influence the zero error capacity of the channel.

According to (7.3.4) (on p. 205), the notion of the zero error capacity with a feedback given in Sect. 7.3 assumes that the feedback link provides the unit delay. Now suppose that in this link, the messages incur the delay of $d \geq 1$ units of time. Then the definition of the zero error capacity should be modified by reducing the set of "feedforward" messages consumed by the block function (7.3.4):

$$\boxed{s(0), \dots, s(t-1)} \mapsto \boxed{s(0), \dots, s(t-d)}\,.$$

By keeping the other particulars in the definition unchanged, we arrive at the notion of the *zero error capacity with d-delayed feedback* $\mathfrak{c}_{0F}[d]$.

More formally, suppose that the informant takes a choice from N prespecified possible messages labeled by ν and notifies the recipient about this choice by emitting a code word $[e(0), \dots, e(r-1)]$ with a given length r into the channel. Let

[21] See Definition 3.4.9 (on p. 46).

this word be produced by a block function (code) employing a d-delayed feedback communication link ($d \geq 1$); i.e., a recursive rule of the form:

$$e(t) = \mathcal{E}_*[t, e(0), \ldots, e(t-1), s(0), \ldots, s(t-d), \nu], \\ t = d, \ldots, r-1, \qquad e(t) = \mathfrak{E}_*[t, \nu], \; t \leq d-1. \quad (7.8.7)$$

As before, the ratio $R := \frac{\log_2 N}{r}$ is called the rate of the code. To recognize the choice of the informant, a decoding rule, i.e., a map $\mathcal{D} : \mathfrak{S}^r \to [1 : N]$, is applied to the received code word $[s(0), \ldots, s(r-1)], s(t) \in \mathfrak{S}$. Such a rule is errorless if $\mathcal{D}(S) = \nu$ for any ν and any output word S that may be received with a positive probability given that ν is encoded and sent over the channel. The *zero error capacity with a d-delayed feedback* $\mathfrak{c}_{0F}[d] := \sup R$, where sup is over all block codes (7.8.7) of arbitrary lengths r for which errorless decoding is possible.

Observation 7.8.8. *The zero error capacity $\mathfrak{c}_{0F}$ introduced on p.* 205 *and employed in the main Theorems* 7.4.5, 7.5.3 *is that with the unit delay:* $\mathfrak{c}_{0F} = \mathfrak{c}_{0F}[1]$.

Now we are in a position to present the main result of the subsection.

Lemma 7.8.9. *The zero error capacity does not depend on the delay in the feedback link:* $\mathfrak{c}_{0F}[d] = \mathfrak{c}_{0F} \; \forall d = 1, 2, \ldots$.

Proof. The inequality $\mathfrak{c}_{0F}[d] \leq \mathfrak{c}_{0F}$ is evident. To establish the converse, we pick $R < \mathfrak{c}_{0F}$. Then a block code (7.3.4) (on p. 205) exists with block length r, the input code book of a size N, and the unit delay in the feedback link for which $\frac{\log_2 N}{r} \geq R$ and errorless decoding by a rule $\mathcal{D}$ is possible. Now let this code operate (i.e., form messages $e(t)$ and send them over the channel) at times $t = 0, d, 2d, \ldots, (r-1)d$. This is possible by means of the d-delayed feedback link since it makes the coder aware about the result of the previous transmission by the time of the next one, as is required. By applying $\mathcal{D}$ to $s[0], s[d], \ldots, s[(r-1)d]$, one acquires the way to transmit one message μ_1 from the set of N ones without errors. Another copy of this code operating at times $t = 1, d+1, 2d+1, \ldots, (r-1)d+1$ makes it possible to transmit one more μ_2 message chosen independently of the previous one. By continuing likewise, we see that there is a way to transmit without errors any of N^d messages $(\mu_1, \mu_2, \ldots, \mu_d)$ during the time interval $[0 : rd-1]$ of duration rd using the d-delayed feedback. Thus $\mathfrak{c}_{0F}[d] \geq \frac{\log_2 N^d}{rd} = \frac{\log_2 N}{r} \geq R$. Letting $R \to \mathfrak{c}_{0F} - 0$ completes the proof. □

7.8.3 Construction of the Stabilizing Coder and Decoder-Controller

The coder and decoder inherit many features of those described in Subsect. 3.8.4 and 7.7.1 (starting on pp. 68 and 223). The major novelty concerns a machinery for feedback data transmission from the controller to the sensor.

From now on, we assume that the condition $\mathfrak{c}_{0F} > H(A)$ from ii) of Theorem 7.5.3 (on p. 209) holds.

Components of the Controller

We pick

1) two numbers η and R_* such that

$$H(A) < \eta < R_* < \mathfrak{c}_{0F}; \tag{7.8.8}$$

2) two parameters

$$\gamma > \|A\| \qquad \text{and} \qquad \delta_* > 0; \tag{7.8.9}$$

3) an r-contracted quantizer $\mathfrak{Q}_r$ from Lemma 3.8.36 (on p. 80) applied to $A_{ss} := A$ (we also invoke Remark 3.8.37 on p. 80 here) for any $r = 1, 2, \ldots$;
4) for all large r, a d-delayed, where d is given by (7.8.3), block function (7.8.7) with block length $r - n, n := \dim(x)$, rate $R \geq R_*$, and an errorless decoding rule $\mathcal{D}_r(\cdot)$, which is capable of encoding all outputs of the quantizer $\mathfrak{Q}_r$;
5) a feedback control alphabet $\{{}^a\boldsymbol{u}_s\}_{s\in\mathfrak{S}}$ (see Lemma 7.8.4 and Definition 7.8.6) of size equal to that of the output alphabet $\mathfrak{S}$ of the channel; the elements of the control alphabet are labeled by the channel output symbols;
6) the predictor (7.8.2);
7) a deadbeat stabilizer

$$x(0) \xrightarrow{\mathcal{N}} \boldsymbol{U} = \big[u(0), u(1), \ldots, u(n-1), 0, \ldots\big]. \tag{7.8.10}$$

Explanation 7.8.10. • In 2), $\|A\| = \max_{x:\|x\|=1} \|Ax\|$ is the matrix norm associated with the vector norm $\|\cdot\|$ from Definition 3.8.6 (on p. 69).

- In 4), the block length is n time units less than that r of the operation epoch. This permits us to withdraw the block function from service and so do not apply "communication" controls ensuring the feedback transmission of data from the controller to the sensor during concluding n time steps of any epoch. These steps are used to cancel the detrimental influence of the previous "communication" controls on the plant by producing a proper correcting control sequence.
- We recall that the deadbeat stabilizer (7.8.10) is a linear transformation producing a control sequence that drives the unperturbed ($\xi(t) \equiv 0$) plant (7.5.1) (on p. 208) from the initial state $x(0)$ to 0 at time $t = n$. For technical convenience, we define

$$\mathcal{N}(\maltese) := [0, 0, 0, 0, \ldots]. \tag{7.8.11}$$

- To show that 4) is possible, we invoke that the number of the quantizer levels (including $\maltese$) $N'_r \lesssim 2^{r\eta} + 1$ by Lemma 3.8.36 (on p. 80). Thanks to (7.8.8) and Lemma 7.8.9, a d-delayed feedback block function with a block length r_* exists for which errorless decoding is possible, and $\frac{\log_2 N}{r_*} \geq R_0 > R_*$, where N is the number of encoded messages. By applying this coding–decoding rule $i := \lfloor (r-n)/r^* \rfloor$ times (where $r > r_* + n$ is given) on the consecutive time intervals $[0 : r_*), [r_* : 2r_*), \ldots, [(i-1)r_* : ir_*)$ and doing nothing on $[ir_* : r-n)$, we get a block code with block length $r - n$, which makes it possible to transmit without an error any message from a specific set of $N''_r = N^i \geq 2^{ir_* R_0} \geq 2^{(r-n-r_*)R_0}$

ones. The rate of this code $R = \frac{(r-n-r_*)R_0}{r} \approx R_0 > R_*$ for $r \approx \infty$. It remains to pick the number r so large that $R > R_*$ and $N_r'' \geq N_r' + 1$, which is possible by (7.8.8).

Remark 7.8.11. The block function from 4) will be used to communicate outputs q of the quantizer from 3). So from now on, we assume that $\nu \equiv q$ in (7.8.7).

To make the subsequent formulas more homogeneous, we assume that $r > n+d$.

Description of the Coder and Decoder

As before, their operation is organized into epochs $[\tau_i = ir : \tau_{i+1})$ of duration r chosen to meet the requirement from 4) on p. 237. Both coder and decoder compute controls $u_c(t)$, $u_d(t)$ and upper bounds for the state norm $\delta_c(t), \delta_d(t)$, respectively. Acting upon the plant is the control $u_d(t)$. The initial bound is common $\delta_c(0) = \delta_d(0) = \delta_0 \geq \delta_*$ and not necessarily correct.

The control $u_d(t)$ is produced as the sum $u_d(t) = u_d^b(t) + u_d^{\text{com}}(t)$ of two parts. The *basic control* $u_d^b(t)$ aims to stabilize the plant, whereas the *communication control* $u_d^{\text{com}}(t)$ serves the feedback communication of $s(t)$ from the decoder to the coder. The basic controls are generated at times τ_i in the form of a control program for the entire epoch $[\tau_i : \tau_{i+1})$. The current communication control is generated at the current time t and encodes the message $s(t)$ currently received over the channel: $u_d^{\text{com}}(t) := {}^a\boldsymbol{u}_{s(t)}$. This ensures d-delayed communication of $s(t)$ to the coder, as is required by the block code at hand.

The coder employs this code to transmit the quantized value of the state prognosis. This value is determined at the beginning of the operation epoch τ_i, the prognosis concerns the state at its end τ_{i+1}, and transmission is during the epoch $[\tau_i : \tau_{i+1})$. However since the length $r - n$ of the block code is less than the epoch duration r, the transmission will be completed at time $\tau_{i+1} - n - 1$. During the remainder $[\tau_{i+1} - n : \tau_{i+1})$, the coder sends nothing over the channel. Hence for $t \in [\tau_{i+1} - n - d : \tau_{i+1} - 1]$, there is no need to communicate $s(t)$ from the decoder to the coder and thus employ the rule $u_d^{\text{com}}(t) := {}^a\boldsymbol{u}_{s(t)}$. The decoder uses this to cancel the influence of the previously generated sequence of communication controls $u_d^{\text{com}}(\tau_i), \ldots, u_d^{\text{com}}(\tau_{i+1} - n - d - 1)$ on the plant. To this end, it puts $u_d^{\text{com}}(\tau_{i+1} - d) := \cdots := u_d^{\text{com}}(\tau_{i+1} - 1) := 0$ and picks

$$u_d^{\text{com}}(\tau_{i+1} - n - d), \ldots, u_d^{\text{com}}(\tau_{i+1} - d - 1) \tag{7.8.12}$$

so that the entire sequence $\{u_d^{\text{com}}(t)\}_{t=\tau_i}^{\tau_{i+1}-1}$ drives the noiseless plant from 0 at time τ_i to 0 at time τ_{i+1}:

$$x(\tau_i) = 0 \xrightarrow{u_d^{\text{com}}(\tau_i), \ldots, u_d^{\text{com}}(\tau_{i+1}-1)} x(\tau_{i+1}) = 0 \tag{7.8.13}$$

provided that $u(t) := u_d^{\text{com}}(t)$ and $\xi(t) \equiv 0$ in (7.5.1) (on p. 208).

Specifically, the decoder computes the state

$$x \xleftarrow{u_d^{\text{com}}(\tau_i),\ldots,u_d^{\text{com}}(\tau_{i+1}-n-d-1)} 0$$

(assuming that $\xi(t) \equiv 0$) and then applies the deadbeat stabilizer (7.8.10) to x. Then given $x(\tau_i)$, only the basic controls $u_d^b(t)$ determine the state $x(\tau_{i+1})$ (provided $\xi(t) \equiv 0$). So at time $t = \tau_i$, the coder is able to compute a reliable prognosis of the state $x(\tau_{i+1})$ from the available measurements and knowledge of the basic controls $u_d^b(\theta)$ at times $\theta \in [\tau_i : \tau_{i+1})$, along with the entire real controls $u_d(\theta)$ at times $\theta = t - n + 1, \ldots, t$.

The coder generates the controls $u_c(t)$ so that they be replicas of $u_d(t)$. To this end, it calculates the basic controls by itself with overtaking the decoder by one epoch. For $t \in [\tau_i : \tau_{i+1} - n - d - 1]$, it gets aware of $s(t)$ and thus $u_d^{\text{com}}(t)$ at time $t + d$. So at time $\tau_{i+1} - n$, the coder knows all communication controls that do communicate information; i.e., it knows

$$\boldsymbol{U}_i^{\text{com}} := \textbf{col}\left[u_d^{\text{com}}(\tau_i), \ldots, u_d^{\text{com}}(\tau_{i+1} - n - d - 1)\right].$$

Hence at time $\tau_{i+1} - n$, the coder is able to compute the "canceling tail" (7.8.12), which is uniquely determined by $\boldsymbol{U}_i^{\text{com}}$. It follows that the coder acquires the controls u_d acting upon the plant with delay d at any time t, and with no delay at $t = \tau_i$.

Whereas the foregoing described the ideas, now we come to formal details. The actions of the coder and decoder are described first for times $t = \tau_i$ and then for times within the operation epochs.

The coder (at the times $t = \tau_i, i = 1, 2, \ldots$)

cs.1) Calculates the state prognosis $\widetilde{x}(\tau_{i+1}|\tau_i)$ by means of (7.8.2);

cs.2) Computes the quantized value $q_c(\tau_i)$ of the scaled state prognosis:

$$\varepsilon(\tau_i) := \frac{\widetilde{x}(\tau_{i+1}|\tau_i)}{\delta_c(\tau_i)}, \quad q_c(\tau_i) := \mathfrak{Q}_r\left[\varepsilon(\tau_i)\right]; \tag{7.8.14}$$

This value is to be transmitted to the decoder within the next epoch $[\tau_i : \tau_{i+1})$;

cs.3) Computes the basic control program

$$\boldsymbol{U}_{i+1}^{c,b} = \textbf{col}\left[u_c^b(\tau_{i+1}), \ldots, u_c^b(\tau_{i+2} - 1)\right]$$

for the operation epoch $[\tau_{i+1} : \tau_{i+2})$ following the forthcoming one $[\tau_i : \tau_{i+1})$ and then corrects the state upper bound:

$$\boldsymbol{U}_{i+1}^{c,b} := \delta_c(\tau_i)\mathcal{N}\left[q_c(\tau_i)\right],$$
$$\delta_c(\tau_i) := \delta_c(\tau_i) \times \left(\langle q_c(\tau_i)\rangle_{\varkappa,\gamma}\right)^r + \delta_* \,. \tag{7.8.15}$$

Here $\mathcal{N}$ is the deadbeat stabilizer (7.8.10), $\langle q\rangle_{\varkappa,\gamma} := \varkappa$ if $q \neq \maltese$ and $\langle q\rangle_{\varkappa,\gamma} := \gamma$ otherwise, where γ, δ_* are the parameters from (7.8.9) and $\varkappa \in (0, 1)$ is taken from Lemma 3.8.36 (on p. 80).

The decoder (at the times $t = \tau_i, i = 2, 3, \ldots$)

ds.1) Applies the decoding rule from 4) on p. 237 to the code word $[s(\tau_{i-1}),\dots,s(\tau_i-n-1)]$ and thus acquires the decoded value $q_d(\tau_i)$ of $q_c(\tau_{i-1})$;

ds.2) Computes the basic control program

$$U_i^{d,b} = \textbf{col}\left[u_d^b(\tau_i),\dots,u_d^b(\tau_{i+1}-1)\right]$$

for the next operation epoch $[\tau_i : \tau_{i+1})$ and corrects the bound δ_d:

$$U_i^{d,b} := \delta_d(\tau_i)\mathcal{N}\big[q_d(\tau_i)\big],$$
$$\delta_d(\tau_i) := \delta_d(\tau_i)\left(\langle q_d(\tau_i)\rangle_{\varkappa,\gamma}\right)^r + \delta_*. \quad (7.8.16)$$

Remark 7.8.12. For the technical convenience, the initial basic programs $U_0^{c,b}, U_0^{d,b}$ are taken to be zero. We also suppose that $q_c(\tau_0) := q_d(\tau_1) :=$ ✠ and the coder (at $t=\tau_0$) and decoder (at $t=\tau_1$) act accordingly. This implies that $U_1^{c,b} = U_1^{d,b} = 0$.

Within any epoch $t \in [\tau_i : \tau_{i+1}), i \geq 1$, the **decoder**

ds.3) Generates the current communication control $u_d^{\text{com}}(t)$:

$$u_d^{\text{com}}(t) := \begin{cases} {}^a\boldsymbol{u}_{s(t)} & \text{if } \tau_i \leq t < \tau_{i+1}^- \\ \left.\begin{array}{l}\text{the } \left(t-\tau_{i+1}^-\right) \text{th term} \\ \text{of the sequence (7.8.10) given by} \\ U := \\ \mathcal{N}\left[\sum_{\theta=\tau_i}^{\tau_{i+1}^- -1} A^{\tau_{i+1}^- -1-\theta} B u_d^{\text{com}}(\theta)\right], \end{array}\right\} & \text{if } \tau_{i+1}^- \leq t < \tau_{i+1} \end{cases},$$

where $\tau_{i+1}^- := \tau_{i+1} - n - d$;

ds.4) applies the control $u_d(t) := u_d^b(t) + u_d^{\text{com}}(t)$ to the plant;

whereas the **coder**

cs.4) for $\tau_i \leq t \leq \tau_i + d - 1$, determines the message to the decoder $e(t) := \mathcal{E}_*[t-\tau_i, q_c(\tau_i)]$ by means of the code (7.8.7), and emits $e(t)$ into the channel;

cs.5) for $\tau_i + d \leq t < \tau_{i+1} - n$, employs the decision rule (7.8.6) with $u := u_c^b(t-d)$ to determine $s(t-d)$ and thus $u_d^{\text{com}}(t-d)$, puts $u_c(t-d) := u_c^b(t-d) + u_d^{\text{com}}(t-d)$, then employs the code (7.8.7) and dispatches the message

$$e(t) := \mathcal{E}_*[t-\tau_i, e(\tau_i),\dots,e(t-1), s(\tau_i),\dots,s(t-d), q_c(\tau_i)];$$

cs.6) at time $t = \tau_{i+1} - n$, calculates both the communication controls $u_d^{\text{com}}(\theta)$ and controls $u_c(\theta) := u_c^b(\theta) + u_d^{\text{com}}(\theta)$ (actually acting upon the plant) for all $\theta = \tau_{i+1} - n - d, \dots, \tau_{i+1} - 1$ by applying the formula from ds.3)

$$U := \mathcal{N}\left[\sum_{\theta=\tau_i}^{t-d-1} A^{t-1-d-\theta} B u_d^{\text{com}}(\theta)\right]$$

to determine $u_d^{\text{com}}(\theta)$;

For consistency, we put $u_c^{\text{com}}(\theta) := u_d^{\text{com}}(\theta) := 0 \; \forall \theta < \tau_1$.

Observation 7.8.13. *Thanks to* ds.3)*, the sequence of communication controls produced within a given epoch* $[\tau_i : \tau_{i+1})$ *drives the unperturbed plant (*$\xi(t) \equiv 0$*) from* 0 *at* $t = \tau_i$ *to* 0 *at* $t = \tau_{i+1}$*; i.e.,* (7.8.13) *does hold.*

By induction on i and $t \in [\tau_i : \tau_{i+1})$, the foregoing implies the following.

Lemma 7.8.14. *The above coder–decoder pair is well defined: The coder has access to all data required to compute the state prognosis at step* cs.1) *and to apply the decision rule* (7.8.6) *at step* cs.5)*. This rule does determine* $s(t-d)$*. The feedforward transmission across the channel is errorless* $q_d(\tau_i) = q_c(\tau_{i-1})$*. The coder and decoder generate common controls* $\boldsymbol{U}_i^{c,b} = \boldsymbol{U}_i^{d,b}, u_c(t) = u_d(t)$ *and state norm upper bounds* $\delta_d^+(\tau_i) = \delta_c^-(\tau_i)$*.*

As before, the indices $^-$ and $^+$ mark the bound before and after the update at time τ_i, respectively.

Observation 7.8.15. *The coder and decoder introduced in this subsection are semirational finite-step recursive.*[22]

This is straightforward from the description of the coder and decoder with regard to the fact that in cs.2) the quantizer is taken from Lemma 3.8.36 (on p. 80) and so is polyhedral.[23] In (6.3.4) (on p. 139) and (6.5.5) (on p. 145), the states of the coder and decoder can be defined as follows:

$$z_c(ir) := \Big[\delta_c^-(\tau_i), \delta_c^-(\tau_{i-1}), \widetilde{x}(\tau_{i+1}|\tau_i), \widetilde{x}(\tau_i|\tau_{i-1}), \boldsymbol{U}_{i-1}^c\Big], \quad z_d(ir) := \delta_d^-(\tau_i),$$

where

$$\boldsymbol{U}_{i-1}^c := \mathbf{col}\Big[u_c(\tau_{i-1}), \dots, u_c(\tau_i - 1)\Big].$$

7.8.4 Almost Sure Stabilization in the Face of Plant Noises

The main result of the subsection is as follows.

Proposition 7.8.16. *Suppose that Assumption* 7.7.1 *(on p.* 223*) holds and* $\mathfrak{c}_{0F} > H(A)$*. Then the coder–decoder pair described in the previous subsection on pp.* 239–240 *uniformly stabilizes the plant under bounded noises.*[24]

Like in Subsect. 7.7.3 (starting on p. 232), the temporary additional Assumption 7.7.1 will be dropped by applying the coder–decoder pair at hand to the unstable part of the plant. This gives rise to the following.

Remark 7.8.17. Modulo the arguments from Sect. 6.10 (starting on p. 179), Proposition 7.8.16 implies (ii) of Theorem 7.5.3 (on p. 209). (For details, see Subsect. 7.8.5 starting on p. 244.)

The remainder of the subsection is devoted to the proof of Proposition 7.8.16. Basically, the proof follows the lines of Subsect. 7.7.2 (starting on p. 227).

[22] See Definitions 6.3.9 and 6.5.7 on pp. 139 and 145, respectively.

[23] See Definition 3.8.8 on p. 70.

[24] See Definition 7.5.2 on p. 209.

Estimating the Quantities Generated by the Controller

As in that subsection, we study the process in the system equipped by the coder and decoder, with focusing attention on

$$\delta_i := \delta_c^-(\tau_i), \quad z_i := \|\widetilde{x}(\tau_{i+1}|\tau_i)\|, \quad \varsigma_i := \begin{cases} \delta_i^{-1} z_i & \text{if } i \geq 1 \\ \delta_0^{-1} z_0 + 1 & \text{for } i = 0 \end{cases}. \qquad (7.8.17)$$

By cs.2), ς_i determines whether the alarm symbol $\maltese$ is sent over the channel:

$$\varsigma_i > 1 \Leftrightarrow q_c(\tau_i) = \maltese. \qquad (7.8.18)$$

Explanation 7.8.18. To make this formula true for all i, the quantity ς_i is defined by an exclusive rule for $i = 0$ since always $q_c(\tau_0) = \maltese$ by Remark 7.8.12. Furthermore, $\widetilde{x}(\tau_1|\tau_0)$ cannot be defined by formula (7.8.2) since no observations has been received by $t = \tau_0 = 0$. In the subsequent considerations, we suppose that $\widetilde{x}(\tau_1|\tau_0) \neq 0$ is some vector satisfying the exactness bound $\|\widetilde{x}(\tau_1|\tau_0) - x(\tau_1)\| \leq D_x^{\langle r \rangle}$ from Lemma 7.8.2 (on p. 234).

Lemma 7.8.19. *For any $i \geq 1$, the quantities* (7.8.17) *satisfy relations* (7.7.19) *and* (7.7.20) *(on p.* 228*), where $I_0(i), I_{\maltese}(i)$ are given by* (7.7.18) *(on p.* 228*) and*

$$D_r := D_x^{\langle r \rangle} \left(1 + \|A\|^r\right) + D \sum_{\theta=0}^{r-1} \|A\|^\theta. \qquad (7.8.19)$$

Here D and $D_x^{\langle r \rangle}$ are taken from (7.2.2) *(on p.* 202*) and Lemma* 7.8.2 *(on p.* 234*), respectively.*

Proof. Equation (7.7.19) is straightforward from (7.7.18), (7.8.15), (7.8.17), and (7.8.18). To justify (7.7.20), we observe that

$$\begin{aligned}
z_i &\overset{(7.8.17)}{=\!=} \|\widetilde{x}(\tau_{i+1}|\tau_i)\| = \|\widetilde{x}(\tau_{i+1}|\tau_i) - x(\tau_{i+1})\| + \|x(\tau_{i+1})\| \\
&\overset{a)}{\leq} D_x^{\langle r \rangle} + \|x(\tau_{i+1})\| \overset{(7.5.1)}{=\!=} D_x^{\langle r \rangle} + \left\| A^r x(\tau_i) + \sum_{\theta=\tau_i}^{\tau_{i+1}-1} A^{\tau_{i+1}-1-\theta} \left[B u_d(\theta) + \xi(\theta)\right] \right\| \\
&\leq D_x^{\langle r \rangle} + \|A^r \left[x(\tau_i) - \widetilde{x}(\tau_i|\tau_{i-1})\right]\| \\
&\quad + \left\| A^r \widetilde{x}(\tau_i|\tau_{i-1}) + \sum_{\theta=\tau_i}^{\tau_{i+1}-1} A^{\tau_{i+1}-1-\theta} \left[B u_d(\theta) + \xi(\theta)\right] \right\| \\
&\overset{b)}{\leq} \left(1 + \|A\|^r\right) D_x^{\langle r \rangle} + D \sum_{\theta=\tau_i}^{\tau_{i+1}-1} \|A\|^{\tau_{i+1}-\theta-1} \\
&\quad + \left\| A^r \widetilde{x}(\tau_i|\tau_{i-1}) + \sum_{\theta=\tau_i}^{\tau_{i+1}-1} A^{\tau_{i+1}-1-\theta} B \left[u_d^b(\theta) + u_d^{com}(\theta)\right] \right\| \\
&\overset{(7.8.13),\,(7.8.19)}{=\!=\!=\!=} D_r + \left\| A^r \widetilde{x}(\tau_i|\tau_{i-1}) + \sum_{\theta=\tau_i}^{\tau_{i+1}-1} A^{\tau_{i+1}-1-\theta} B u_d^b(\theta) \right\| =: a.
\end{aligned}$$

Here a) holds by Lemma 7.8.2 (on p. 234); and b) holds by this lemma, (7.2.2) (on p. 202), and ds.4). By Lemma 7.8.14,

$$a = D_r + \left\| A^r \widetilde{x}(\tau_i|\tau_{i-1}) + \sum_{\theta=\tau_i}^{\tau_{i+1}-1} A^{\tau_{i+1}-1-\theta} B u_c^b(\theta) \right\| .$$

If $I_{\maltese}(i) = 1$, then $u_c^b(\theta) = 0$ for all θ concerned due to (7.7.18) (on p. 228), (7.8.11), (7.8.15), and (7.8.18). Hence $a \le D_r + \|A\|^r z_{i-1}$ owing to (7.8.17). Now suppose that $I_0(i) = 1$. Then by (7.7.18), (7.8.10), and (7.8.15), we have

$$\delta_c(\tau_{i-1}) A^r q_c(\tau_{i-1}) + \sum_{\theta=\tau_i}^{\tau_{i+1}-1} A^{\tau_{i+1}-1-\theta} B u_c^b(\theta) = 0.$$

So

$$a = D_r + \|A^r \widetilde{x}(\tau_i|\tau_{i-1}) - \delta_c(\tau_{i-1}) A^r q_c(\tau_{i-1})\| \overset{(7.8.14),\,(7.8.17)}{=\!=\!=} D_r + \delta_{i-1} \|A^r [\varepsilon(\tau_{i-1}) - q_c(\tau_{i-1})]\| .$$

We recall that $q_c(\tau_{i-1})$ is the quantized value of $\varepsilon(\tau_{i-1})$ by means of the $\varkappa^{2r}$-contracted quantizer $\mathfrak{Q}_r$. So (3.8.9) (on p. 70) (where $\rho_{\mathfrak{Q}} = \varkappa^{2m}$ by Lemma 3.8.36 on p. 80) implies that $a \le D_r + \varkappa^{2r} \delta_{i-1}$. Summarizing, we arrive at (7.7.20) (on p. 228). □

Now we observe that $\delta_i \ge \delta_*$ thanks to (7.8.15) and (7.8.17). So with regard to Observation 7.7.21 (on p. 231), we arrive at the following.

Corollary 7.8.20. *Lemma* 7.7.20 *(on p.* 230*) remains true for the quantities* (7.8.17)*.*

Proof of Proposition 7.8.16: Concluding Part

Thanks to (7.8.17) and Explanation 7.8.18,

$$\begin{aligned} z_0 = \|\widetilde{x}(\tau_1|\tau_0)\| &\le \|\widetilde{x}(\tau_1|\tau_0) - x(\tau_1)\| + \|x(\tau_1)\| \\ &\le D_x^{\langle r \rangle} + \|x(\tau_1)\| \overset{\text{Remark 7.8.12}}{=\!=\!=} D_x^{\langle r \rangle} + \left\| A^r x_0 + \sum_{\theta=0}^{r-1} A^{r-1-\theta} \xi(\theta) \right\| \\ &\overset{(7.2.2)}{\le} D_x^{\langle r \rangle} + \|A\|^r \|x_0\| + D \sum_{\theta=0}^{r-1} \|A\|^\theta . \end{aligned}$$

So (7.7.22) (on p. 230) holds with the upper bounds determined by $\|x_0\|$, $D_x^{\langle r \rangle}$ and the quantities from (7.7.23) (on p. 230). Hence the stabilization error at time $t = \tau_i$ is also bounded:

$$\text{err}_i := \|x(\tau_i)\| \leq \|x(\tau_i) - \widetilde{x}(\tau_i|\tau_{i-1})\| + \|\widetilde{x}(\tau_i|\tau_{i-1})\|$$
$$\overset{\text{Lemma 7.8.2}}{\leq} D_x^{\langle r \rangle} + \|\widetilde{x}(\tau_i|\tau_{i-1})\| \overset{(7.8.17)}{=\!=} D_x^{\langle r \rangle} + z_{i-1} \overset{(7.7.22)}{\leq} D_x^{\langle r \rangle} + \overline{z}.$$

Since the quantized value $q_c(\tau_i)$ lies in the unit ball whenever $q_c(\tau_i) \neq ✠$ and $\mathfrak{N}(✠) = 0$ by (7.8.11) (on p. 237), it follows from (7.8.15) that $\|\boldsymbol{U}_i^c\| \leq \delta_c^-(\tau_{i-1})\|\mathfrak{N}\|$. By taking into account the second inequality from (7.7.22) (on p. 230), we see that the sequence $\{\boldsymbol{U}_i^c\}_{i=0}^\infty$ is bounded. The communication controls $u_c^{\text{com}}(t)$ are taken from a finite (and so bounded) control alphabet. Hence ds.4) (on p. 240) and Lemma 7.8.14 (on p. 241) imply that the sequence of controls $\{u_d(t)\}_{t=0}^\infty$ produced by the decoder is also bounded

$$\|u_d(t)\| \leq \overline{u} \qquad \forall t.$$

To complete the proof, we consider arbitrary t and pick i such that $\tau_i < t \leq \tau_{i+1}$. Then

$$\|x(t)\| \overset{(7.5.1)}{=\!=} \left\| A^{t-\tau_i} x(\tau_i) + \sum_{\theta=\tau_i}^{t-1} A^{t-1-\theta} \left[B u_d(\theta) + \xi(\theta) \right] \right\|$$
$$\leq \|A\|^{t-\tau_i} \|x(\tau_i)\| + \sum_{\theta=\tau_i}^{t-1} \|A\|^{t-1-\theta} \Big[\|B\| \|u_d(\theta)\| + \|\xi(\theta)\| \Big]$$
$$\overset{(7.2.2), \|A\| \geq 1, \tau_{i+1} - \tau_i = r}{\leq} \|A\|^r \left(D_x^{\langle r \rangle} + \overline{z} \right) + (D + \|B\| \overline{u}) \sum_{\theta=0}^{r-1} \|A\|^\theta, \quad (7.8.20)$$

where the inequality $\|A\| \geq 1$ holds by Assumption 7.7.1 (on p. 223). Definition 7.5.2 (on p. 209) completes the proof. □

7.8.5 Completion of the Proof of (ii) from Theorem 7.5.3

For systems with no stable modes (i.e., under Assumption 7.7.1 on p. 223), the claim (ii) is justified by Proposition 7.8.16. Thus to complete the proof, one should drop Assumption 7.7.1. This will be accomplished just like in Subsects. 3.8.11 and 6.11.4 (starting on pp. 88 and 196) by stabilizing the "unstable" part of the system.

We do not suppose any longer that Assumption 7.7.1 holds. We also assume that $\mathfrak{c}_{0F} > H(A)$ and focus attention on the unstable part of the plant (7.5.1) (on p. 208):

$$x_+(t+1) = A_+ x_+(t) + \pi_+ B u(t) + \pi_+ \xi(t), \quad x_+(0) = x_0^+ := \pi_+ x_0. \quad (7.8.21)$$

The notations are as follows: $x_\pm(t) := \pi_\pm x(t)$; M_{unst} and M_{st} are the invariant subspaces of A related to the unstable and stable parts of its spectrum, respectively; π_+, π_- are the projections onto M_{unst} parallel to M_{st} and vice versa; and A_+, A_- denote the operator A acting in M_{unst} and M_{st}, respectively.

The plant (7.8.21) is controllable thanks to Assumption 6.5.5 (on p. 144), and $H(A_+) = H(A) > \mathfrak{c}_{0F}$. So by the foregoing, the system (7.8.21) can be uniformly stabilized under bounded noises by the coder and decoder constructed in

Subsect. 7.8.3 (starting on p. 236). This is true provided that the coder has access to proper observations. The idea is to feed the coder serving the artificial plant (7.8.21) by the real measurements $y(\theta)$ from (7.5.1) (on p. 208). This is possible and alters neither operation of the coder nor the conclusion of Proposition 7.8.16 applied to the plant (7.8.21) if the following two statements hold:

p.1) To perform step cs.1) (on p. 239), future states x_+ can be prognosticated with a bounded error on the basis of $y(\theta)$;

p.2) To perform step cs.5) (on p. 240), the message $s(t-d)$ can be correctly recognized by an analog of the decision rule (7.8.6) (on p. 235) applied to $y(\theta)$.

To prove these claims, we note that since the pairs (A, C) and (A, B) are detectable and stabilizable by Assumptions 6.3.6 (on p. 137) and 6.5.5 (on p. 144), respectively, the integer d is still well defined by (7.8.3) (on p. 234). Moreover, a linear prognosis (7.8.2) (on p. 234) of the state $x(t+r)$ can be constructed so that the error modulo the unobservable subspace of (A, C) is bounded by a constant $\overline{D}(D, D_\chi)$ determined by the noise levels D, D_χ from (7.2.2) (on p. 202) [8]. Due to Assumption 6.3.6, this subspace lies in M_{st}. It follows that $\pi_+\widetilde{x}(t+r|t)$ estimates $x_+(t+r)$ with a bounded error, and so p.1) does hold. Since the unobservable subspace lies in $\ker CA^j$ for any $j = 0, 1, \ldots$, the discrepancy $CA^d[x(t-d) - \widetilde{x}(t-d|t-d)]$ is bounded as $t \to \infty$. This keeps the arguments from (7.8.5) (on p. 235) and thus the entire Lemma 7.8.4 (on p. 234) true, which proves p.2).

Finally, the stabilizing coder–decoder pair for the plant (7.8.21) is constructed in correspondence with Subsect. 7.8.3 up to the above alterations concerning steps cs.1) and cs.5). The duration of the operation epoch is taken so that $r > \dim(x) + d$. By the foregoing, this pair uniformly stabilizes the unstable part of the primal plant (7.5.1) (on p. 208) under bounded noises:

$$\|x_+(t)\| + \|u(t)\| \leq D_+(D, D_\chi) \qquad \forall t = 0, 1, \ldots.$$

To extend this conclusion on the entire state $x(t) = x_+(t) + x_-(t)$, we note that the operator A_- is asymptotically stable: $\|A_-^t\| \leq cv^t \,\forall t$, where $c > 0$ and $v \in (0, 1)$. Hence

$$\begin{aligned}
\|x_-(t)\| = \|\pi_- x(t)\| &= \left\| A_-^t \pi_- x_0 + \sum_{\theta=0}^{t-1} A_-^{t-1-\theta} \pi_- \left[Bu(\theta) + \xi(\theta) \right] \right\| \\
&\leq \|\pi_-\| \left(\|A_-^t\| \|x_0\| + \sum_{\theta=0}^{t-1} \|A_-^{t-1-\theta}\| \Big[\|B\| \|u(\theta)\| + \|\xi(\theta)\| \Big] \right) \\
&\leq c\|\pi_-\| \left[v^t \|x_0\| + [D + D_+(D, D_\chi)] \sum_{\theta=0}^{t-1} v^{t-1-\theta} \right] \\
&\leq \overline{D}_- := c\|\pi_-\| \left[\|x_0\| + \frac{D + D_+(D, D_\chi)}{1 - v} \right], \\
\|x(t)\| = \|x_+(t)\| + \|x_-(t)\| &\leq D_+(D, D_\chi) + \overline{D}_-. \qquad (7.8.22)
\end{aligned}$$

The proof is completed by Definition 7.5.2 (on p. 209). □

8

An Analog of Shannon Information Theory: Stable in Probability Control and State Estimation of Linear Noisy Plants via Noisy Discrete Channels

8.1 Introduction

In this chapter, we proceed with study of problems of state estimation/stabilization involving communication errors and capacity constraints. Discrete-time partially observed linear systems perturbed by stochastic additive disturbances are still examined. The sensor signals are communicated to the estimator/controller over a limited capacity noisy link modeled as a stochastic discrete memoryless channel.

As was shown in the previous chapter, the strong objective of almost sure stability or observability cannot be achieved in such a context by any means and under any circumstances for many channels of practical interest. In other words, arbitrarily large estimation/stabilization errors are encountered systematically with probability 1. In this chapter, the focus is on the weaker form of stability/observability. The objective is to ensure only that the probability of large observation/stabilization errors be small. With the strong law of large numbers in mind, this gives an evidence that unavoidable large errors occur rarely.

It is shown that the capability of the noisy channel to ensure reliable in this sense state estimation/stabilization is identical to its capability to transmit information with as small a probability of error as desired. In other words, the classic Shannon capacity[1] $\mathfrak{c}$ of the channel constitutes the boundary of the observability/stabilizability domain. Since this capacity is typically nonzero [214], most communication channels fit to achieve this weaker form of observability/stability.

The necessity of the Shannon capacity bound is justified with respect to the widest class of observers/controllers that are causal and restricted by no further requirements. This result basically can be derived from Proposition 6.7.12 (on p. 152).[2]

The major results of the chapter concern the sufficiency of the Shannon capacity bound. They show that whenever this bound is met, a reliable observa-

[1] See Subsect. 6.3.3 starting on p. 138.

[2] By Comment 6.7.16 (on p. 152), it can also be derived from Lemma 3.2 [201]. A similar result that concerns a stronger uniform (with respect to bounded plant disturbances) stability and scalar plants is contained in [166].

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_8,

tion/stabilization can be ensured by means of an observer/controller consuming an uniformly bounded (as time progresses) computational complexity and memory per unit time. Moreover, traditional block encoding–decoding procedures can be employed for its implementation. The corresponding state estimator/controller is constructed explicitly. However, the scheme (coder and decoder) for transmission of information across the channel is not described in detail. The point is that the proposed observer/controller employs block codes transmitting data at a given rate below the channel capacity $\mathfrak{c}$ with a given probability of error. Classic information theory guarantees existence of such a code. Moreover, invention of such codes is the standard long-standing task in information sciences. It is supposed that a relevant solution should merely be employed.

For the state estimation problem, we assume a complete feedback communication link. Unlike Chap. 7, the noises are not supposed to be bounded. The proposed observer is universal in the sense that it does not depend on the noise parameters. In other words, the same observer ensures reliable state estimation under an arbitrary noise, although the accuracy of the estimation depends on the noise level. This is of interest if, e.g., the statistical knowledge of the noise is not available or the noise properties change during the system operation.

We consider two statements of the stabilization problem. The first setup supposes that the noises are bounded but does not assume any communication feedback. In this case, the proposed controller depends on the parameters (upper bounds) of the plant noises. The second problem setup deals with unbounded noises and assumes that a perfect communication feedback at an arbitrarily small data rate is available. In this case, the proposed controller is universal (i.e., does not depend on the noise parameters).

Stabilizability in probability and σth moment stabilizability of scalar linear plants with bounded additive disturbances over noisy limited capacity communication channels were examined in [166]. It is shown that the boundary of the domain of stabilizability in probability is given by the Shannon capacity of the channel. Unlike this chapter, the proposed controllers consume unlimited (as time progresses) memory and complexity per step and rely on the novel “anytime” approach to information transmission over a noisy channel proposed in [164], whereas the controllers from this chapter are based on the well-developed classic block coding procedures. [3]

Another advance was made in [101, 103], where robust rth moment stability of uncertain scalar linear plants with bounded additive disturbances was examined in

[3]Related to this approach is a novel concept of anytime capacity. It is argued in [164, 166] that this capacity is the correct figure of merit whenever the moment stability is concerned. Specifically, the moment stabilizability by means of arbitrary causal controllers (including ones consuming unlimited memories and complexities per step) is equivalent to possibility to transmit data across the channel at a sufficiently large rate (depending on the moment degree) by means of the so-called anytime encoder and decoder. The anytime decoder is defined as a device that keeps all previously decoded messages in memory and monitors and updates the entire stock, including the ancient messages, at any time when a new message arrives. On contrary to the Shannon capacity, a simple expression amenable to computation is not in general known for the anytime capacity [142].

the class of realistic controllers. By letting $r \to +0$, these works show that whenever the Shannon capacity bound is met, stability in probability can be ensured by means of explicitly-described controllers with uniformly limited complexity. At the same time, this result concerns only the particular case of discrete memoryless channel (DMC): The *truncation channel*, which transmits binary code words with dropping a random number of concluding bits. This example of DMC generalizes the classic erasure channel and is motivated by certain wireless communication applications in [101]. The proposed control algorithm and coding–decoding scheme rely on specific features of this channel. In this chapter, we examine general noisy DMC and multidimensional linear plants with additive stochastic disturbances that are not necessarily bounded.

Unlike the other chapters of this book, the proofs of the main results of this one are not presented. The reason is that during this book preparation, these results with complete proofs were under review as parts of articles submitted to international journals. A preliminary version of the main results of the chapter was originally published in the conference paper [119].

The body of the chapter is organized as follows. Section 8.2 presents the setup of the state estimation problem. Section 8.3 offers the related assumptions and a description of the observability domain. The explicit contruction of the reliable state estimator is given in Sect. 8.4. In Sects. 8.5 and 8.6, the stabilization problem is posed and the stabilizability domain is described, respectively. Stabilizing coder–decoder pairs are offered in Sect. 8.7. The concluding Sect. 8.8 contains the proofs of auxiliary illustrative facts.

8.2 Statement of the State Estimation Problem

We consider unstable discrete-time invariant linear systems of the form:

$$x(t+1) = Ax(t) + \xi(t), \quad x(0) = x_0, \quad y(t) = Cx(t) + \chi(t). \tag{8.2.1}$$

Here $x \in \mathbb{R}^n$ is the state; $y \in \mathbb{R}^{n_y}$ is the measured output; $\xi(t) \in \mathbb{R}^n$ is the exogenous disturbance; $\chi(t) \in \mathbb{R}^{n_y}$ is the sensor noise, and $t = 0, 1, \ldots$. The instability means that there is an eigenvalue λ of the matrix A with $|\lambda| \geq 1$. The initial state x_0 and the noises $\xi(t), \chi(t)$ are random vectors. The objective is to estimate the current state $x(t)$ on the basis of the prior measurements $y(0), \ldots, y(t)$.

This estimate is required at a remote location. The sensor signals are communicated to this location via a given discrete (digital) noisy channel. So to be transmitted, they should be first translated by a *coder* into a sequence of messages e from a finite *input alphabet* $\mathfrak{E}$ of the channel. During transmission, these messages are transformed $e \xrightarrow{\text{noise}} s$ by some sort of random disturbance into a sequence of channel's outputs s from a finite *output alphabet* $\mathfrak{S}$. Proceeding from the prior outputs s, the *decoder(-estimator)* generates an estimate $\widehat{x}$ of the current state x. In this situation illustrated in Fig. 8.1, the *observer* is constituted by the coder and decoder.

The decoder is defined by an equation of the form:

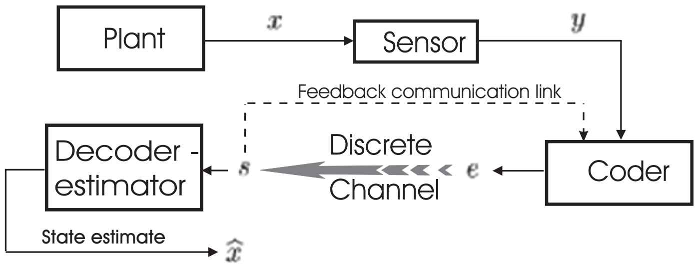


Fig. 8.1. Estimation via a limited capacity communication channel.

$$\widehat{x}(t) = \mathcal{X}\left[t, s(0), s(1), \ldots, s(t)\right]. \tag{8.2.2}$$

We also assume a feedback communication link [190, 219] from the decoder to the coder. Via this link, the result $s(t)$ of the current transmission across the "feedforward" channel becomes known at the coder site by the next time $t+1$. So the coder is described by an equation of the form:

$$e(t) = \mathcal{E}[t, y(0), \ldots, y(t), s(0), \ldots, s(t-1)] \in \mathfrak{E}. \tag{8.2.3}$$

In the face of the plant and sensor errors, a reasonable option is to accept that an observer succeeds if it keeps the estimation error bounded. We examine one of the weakest probabilistic specifications of this property. To introduce it, we start with the following analog of the standard definition of convergence in probability.

Definition 8.2.1. *The infinite sequence of random vectors $X_i \in \mathbb{R}^s, i = 1, 2, \ldots$ is said to be* bounded in probability *if for arbitrary probability value $p \approx 1, p < 1$, a bound $b = b_p \geq 0$ exists such that any vector X_i obeys this bound with probability p or better:*

$$\boldsymbol{P}\left[\|X_i\| \leq b\right] \geq p \approx 1 \qquad \forall i = 1, 2, \ldots.$$

Definition 8.2.2. *An observer is said to* track the state in probability *if the estimation error $x(t) - \widehat{x}(t)$ is bounded in probability.*

The following claim is evident.

Observation 8.2.3. *The observer tracks the state in probability whenever at least one of the following statements hold:*

(i) *The estimation error is almost surely bounded;*
(ii) $\sup_t \boldsymbol{E}\|x(t) - \widehat{x}(t)\|^r < \infty$ *for some $r \in (0, \infty)$;*
(iii) $\sup_t \boldsymbol{E}\psi[x(t) - \widehat{x}(t)] < \infty$, *where $\psi(\cdot) : \mathbb{R}^n \to [0, \infty)$ is a continuous function such that $\psi(x) \to \infty$ as $\|x\| \to \infty$.*

The objective is to discover the requirements to the noisy channel that are necessary and sufficient for existence of a coding–decoding pair tracking the state in probability via this channel. A reliable and realistic observer should be explicitly constructed in the case where these requirements are met.

Comments on the Problem Statement

Within the above problem setup, the issue of complexity is not addressed. In particular, coders and decoders with unlimited memories are allowed. It will be shown that whenever the observability conditions are met, tracking the state in probability can be ensured by means of a coder and decoder that consume limited (as $t \to \infty$) memory and perform a limited number of (algebraic) operations per unit time.

Comments on the communication feedback assumption were given on p. 136.

Tracking in probability is a weak performance criterion. However, it suffices to distinguish between reliable and unreliable observers in the classic problem setup.

To illustrate this, let us examine the time-invariant Luenberger-type observers with the Kalman filter structure. They generate the state estimate $\widehat{x}(t)$ via a recursion of the form

$$\widehat{x}(t+1) = A\widehat{x}(t) + K\big[y(t+1) - \widehat{y}(t+1|t)\big], \quad \text{where} \quad \widehat{y}(t+1|t) := CA\widehat{x}(t), \quad (8.2.4)$$

and K is the observer gain matrix. The initial estimate $\widehat{x}(-1)$ is a deterministic vector (e.g., $\boldsymbol{E}x_0$ if $\boldsymbol{E}x_0$ is known). As is well known, such an observer produces a reliable (in various reasonable senses) state estimate if and only if the eigenvalues of the matrix $A - KCA$ lie in the open unit disk of the complex plane. In its turn, this holds if and only if the observer tracks the state in probability, as is shown by the following lemma.

Lemma 8.2.4. *Let the system and sensor noise sequences be mutually independent and independent of the initial state, and let each of them be formed by mutually independent, identically distributed, and a.s. bounded random vectors. Suppose also that the system noises $\xi(t)$ have a probability density and that the initial state is a.s. bounded. Then the observer* (8.2.4) *tracks the state in probability if and only if the eigenvalues of the matrix $A - KCA$ lie in the open unit disk of the complex plane.*

The proof of this illustrating fact will be given in Sect. 8.8 (starting on p. 264).

8.3 Assumptions and Description of the Observability Domain

The assumptions to follows have much in common with those from Chaps. 6 and 7.

Assumption 8.3.1. *The communication channel is stationary, discrete, and memoryless. In other words, given a current input $e(t)$, the current output $s(t)$ is statistically independent of other inputs and outputs $e(j), s(j), j < t$, and the conditional probability $W(s|e) := \boldsymbol{P}[s(t) = s|e(t) = e]$ does not depend on time t.*

Assumption 8.3.2. *The system* (8.2.1) *does not affect the channel: Given an input $e(t)$, the output $s(t)$ is independent of the initial state x_0 and the plant $\xi(\theta)$ and sensor $\chi(\theta)$ noises (along with other channel inputs and outputs $e(j), s(j), j < t$).*

The next two assumptions are essential only for necessary conditions for observability.

Assumption 8.3.3. *The current plant disturbance $\xi(t)$ is independent of the past ones $\xi(0),\dots,\xi(t-1)$, both current $\chi(t)$ and past $\chi(0),\dots,\chi(t-1)$ sensor noises, and the initial state x_0.*

Assumption 8.3.4. *The initial state x_0 has a probability density $p_0(x)$ and finite differential entropy*[4] $h(x_0)\in\mathbb{R}$.

Conversely, the last assumption is important only for sufficient conditions. To state it, we put

$$\boldsymbol{m}_\sigma(X) := \begin{cases} [\boldsymbol{E}\|X\|^\sigma]^{\frac{1}{\sigma}} & \text{if } \sigma<\infty \\ \inf\{c : \|X\|\le c \text{ a.s.}\} & \text{otherwise} \end{cases} \tag{8.3.1}$$

for any random vector X. We also introduce the following definition.

Definition 8.3.5. *Let $\sigma\in(0,\infty]$. A random vector X and an infinite sequence $\{X_i\}$ of such vectors are said to be* σ-bounded *if the following relations hold, respectively:*

$$\boldsymbol{m}_\sigma(X)<\infty, \qquad \sup_i \boldsymbol{m}_\sigma(X_i)<\infty. \tag{8.3.2}$$

Assumption 8.3.6. *The initial state x_0 and the noises in both plant $\{\xi(t)\}_{t=0}^\infty$ and sensor $\{\chi(t)\}_{t=0}^\infty$ are σ-bounded.*

If $\sigma=\infty$, this means that the noises are a.s. uniformly bounded and that the initial state is distributed over a bounded set. If $\sigma<\infty$, relations (8.3.2) mean that this state and noises have finite σth moments, and the moments of the noises do not grow without limits as $t\to\infty$. It is clear that the larger the σ, the stronger Assumption 8.3.6.

As the following theorem shows, the conditions necessary for observability under Assumption 8.3.6 with $\sigma=\infty$ are "almost sufficient" for observability in the wider class of noises and initial states satisfying Assumption 8.3.6 with some $\sigma>0$.

Theorem 8.3.7. *Suppose that Assumptions* 8.3.1 *and* 8.3.2 *hold, and consider the ordinary Shannon capacity $\mathfrak{c}$ of the communication channel, which is given by* (6.3.3) *(on p.* 138*). Let $H(A)$ denote the topological entropy of the open-loop system* (8.2.1)*:*

$$H(A) := \sum_{\lambda_j:|\lambda_j|\ge 1} \log_2|\lambda_j|, \tag{8.3.3}$$

where $\lambda_1,\dots,\lambda_n$ are the eigenvalues of the system (8.2.1) *repeating in accordance with their algebraic multiplicities.*

The following two statements are true:

(i) If a coder (8.2.3) *and decoder* (8.2.2) *exist that track the state in probability under some random noises and initial state satisfying Assumptions* 8.3.3, 8.3.4, *and* 8.3.6 *with $\sigma:=\infty$, then the following inequality holds:*

$$H(A)\le\mathfrak{c}; \tag{8.3.4}$$

[4] See the definition on p. 150.

(ii) Conversely, let (8.3.4) *hold with the strict inequality sign* $H(A) < \mathfrak{c}$*, and let the pair* (A, C) *be detectable. Then there exists a semirational finite-step recursive*[5] *coder–decoder pair* (8.2.3)*,* (8.2.2) *that tracks the state in probability under any random noises and initial state satisfying Assumption* 8.3.6 *with some* $\sigma > 0$*.*

Thus inequality (8.3.4) is necessary and almost sufficient for observability in probability. The coder–decoder pair from (ii) will be explicitly described in the next section. This pair consumes a bounded (as time progresses) computational power and memory per step.

The proof of Theorem 8.3.7 is not presented in this text due to the reason explained on p. 249.

8.4 Coder–Decoder Pair Tracking the State in Probability

So far as asymptotic tracking does not concern the stable modes, it is clear that analysis can be focused on systems with only unstable ones. Extension on systems with both unstable and stable modes can be performed by applying the observer proposed in this section to the unstable part of the system, like in Sect. 6.10 and Subsect. 7.7.3 (starting on pp. 179 and 232, respectively). So the further consideration is focused on the case where the following assumption holds.

Assumption 8.4.1. *All eigenvalues of the matrix* A *lie outside the closed unit disk of the complex plane.*

We also suppose that the assumptions of (ii) from Theorem 8.3.7 hold. Then $H(A) < \mathfrak{c}$ and the pair (A, C) is observable in view of Assumption 8.4.1.

The observer to be proposed is in fact nothing but a combination of those from Subsects. 6.9.1 and 7.7.1 (described by cc.1)–cc.5), d.1), d.2) on pp. 169 and c.1)–c.4), d.1), d.2) on p. 225 and 226, respectively). In order to compensate for the channel errors, the coder will use the communication feedback to duplicate the actions of the decoder, like in Subsect. 6.9.1 (see cc.3) on p. 169). At the same time, the account for plant noises is like in Subsect. 7.7.1. In particular, the features described in m.1) and m.2) on p. 223 are inherited, and the rule to update the exactness upper bound δ will be of the form (7.7.2) (on p. 224) and so not forcing δ to decay to zero.

For the convenience of the reader, now we offer a systematic description of the coder–decoder pair. To construct it, we pick

1) two parameters

$$\delta_* > 0 \quad \text{and} \quad \gamma > \|A\|; \tag{8.4.1}$$

2) two numbers η and R such that

$$H(A) < \eta < R < \mathfrak{c}; \tag{8.4.2}$$

[5] See Definitions 6.3.9 and 6.3.10 on pp.139 and 140, respectively.

3) a code book $\{E^1, \ldots, E^N\}$ with $N = N_r' \approx 2^{rR}$ input code words E^i each of length r and a decoding rule $\mathcal{D}_r(\cdot)$ with the properties described in Theorem 6.8.4 (on p. 162);
4) an r-contracted quantizer $\mathfrak{Q}_r$ from Lemma 3.8.36 (on p. 80) applied to $A_{ss} := A$ (we also invoke Remark 3.8.37 on p. 80 here);
5) a particular sufficiently large r such that all outputs of the quantizer (including the alarm symbol) can be encoded by means of the code book from 3);
6) a deadbeat observer, i.e., a device producing an estimate $\widetilde{x}(t)$ of the current state $x(t)$ in the form of a sliding average of $n-1$ past observations

$$\widetilde{x}(t) = \sum_{\theta=t-n+1}^{t} F_{t-\theta} y(\theta) \tag{8.4.3}$$

so that this estimate is correct $\widetilde{x}(t) = x(t)$ for $t \geq n = \dim(x)$ in the absence of the noises $\xi(t) \equiv 0, \chi(t) \equiv 0$.

Explanation 8.4.2. • 2) is possible since $H(A) < \mathfrak{c}$.
• 5) is possible by Observation 6.8.8 (on p. 163).
• 6) is possible since the pair (A, C) is observable.

As before, the coder and decoder perform major operations only at times $\tau_i := ir, i = 0, 1, \ldots$. The decoder computes a state estimate $\widehat{x}(t)$ and an upper (and maybe incorrect) bound $\delta(t)$ of its exactness. Specifically, it executes the recursion

$$\widehat{x}(t+1) = A\widehat{x}(t), \qquad \delta(t+1) = \delta(t) \tag{8.4.4}$$

at any time. However at times $t = \tau_i$, this recursion is prefaced by correcting $\widehat{x}(t)$ and $\delta(t)$ on the basis of the code word received during the previous operation epoch $[\tau_{i-1} : \tau_i)$. (The values before and after correction are marked by the upper indices $^-$ and $^+$, respectively.) Thanks to the communication feedback, the coder becomes aware of this code word at time τ_i. This enables the coder to duplicate the operations of the decoder and thus to compute $\widehat{x}(t)$ and $\delta(t)$ by itself.

Specifically, the coder and decoder act as follows.

The coder (at times $t = \tau_i, i = 1, 2, \ldots$)

c.1) Proceeding from the previous measurements, calculates the estimate $\widetilde{x}(t)$ of the current state by (8.4.3);
c.2) Computes the corrected values $\widehat{x}^+(\tau_i)$ and $\delta^+(\tau_i)$ by duplicating the operations of the decoder;
c.3) Employs the quantizer $\mathfrak{Q}_r$ and computes the quantized value $q_c(\tau_i)$ of the current scaled discrepancy between the estimates produced by the coder–decoder pair and the deadbeat observer, respectively:

$$\varepsilon(\tau_i) := \left[\delta^+(\tau_i)\right]^{-1}\left[\widetilde{x}(\tau_i) - \widehat{x}^+(\tau_i)\right]; \tag{8.4.5}$$

c.4) Encodes this value by means of the code book $\mathfrak{E}_r$ and emits the obtained r-word into the channel during the next operation epoch $[\tau_i : \tau_{i+1})$.

Only after this, does the coder perform the current step (8.4.4).

The decoder (at times $t = \tau_i, i = 2, 3, \ldots$)

d.1) Applies the decoding rule $\mathcal{D}_r$ to the data received within the previous operation epoch $[\tau_{i-1} : \tau_i)$ and thus acquires the decoded value $q_d(\tau_i)$ of the quantized and scaled discrepancy $q_c(\tau_{i-1})$;

d.2) Corrects successively the estimate and the exactness bound:

$$\widehat{x}^+(\tau_i) := \widehat{x}^-(\tau_i) + \delta^-(\tau_i) A^r \overset{\star}{q}_d(\tau_i),$$
$$\delta^+(\tau_i) := \delta^-(\tau_i) \times \left(\langle q_d(\tau_i) \rangle_{\varkappa,\gamma} \right)^r + \delta_*. \quad (8.4.6)$$

Here

$$\overset{\star}{q} := \begin{cases} q & \text{if } q \neq \maltese \\ 0 & \text{otherwise} \end{cases}, \qquad \langle q \rangle_{\varkappa,\gamma} := \begin{cases} \varkappa & \text{if } q \neq \maltese \\ \gamma & \text{otherwise} \end{cases}, \quad (8.4.7)$$

$\varkappa$ is the constant from Lemma 3.8.36 (on p. 80), and γ, δ_* are taken from (8.4.1).

Only after this does the decoder perform the step (8.4.4).

Remark 8.4.3. 1. The decoded value $q_d(\tau_i)$ may differ from the true one $q_c(\tau_{i-1})$ due to transmission errors.

2. The step c.2) of the coder consists in carrying out d.1) and d.2). This is possible since at time τ_i, the coder knows the data received by the decoder within $[\tau_{i-1} : \tau_i)$ thanks to the communication feedback.
3. The coder and decoder do not employ any characteristics of the noises.
4. The complexity of the coder and decoder per step is basically determined by the parameter r of the algorithm and does not increase as time progresses. Specifically, the coder and decoder are semirational r-step recursive[6].
5. The coder and decoder compute $\widehat{x}(t)$ and $\delta(t)$ independently.

Initially, they are given common and arbitrarily chosen values $\widehat{x}(0) = \widehat{x}_0$ and $\delta(0) = \delta_0 \geq \delta_*$. (The bound δ_0 may be incorrect; i.e., the inequality $\delta_0 \geq \|x_0 - \widehat{x}_0\|$ may be violated.) For consistency, we put $q_c(\tau_0) := q_d(\tau_1) := \maltese$ and suppose that at times $t = \tau_0, \tau_1$ the coder and decoder act accordingly. It is easy to check that the coder and decoder do compute common sequences $\widehat{x}(t)$ and $\delta(t), t = 0, 1, \ldots$.

The following proposition in fact justifies (ii) of Theorem 8.3.7.

Proposition 8.4.4. *Suppose that the assumptions of* (ii) *from Theorem* 8.3.7 *and Assumption* 8.4.1 *hold. If the duration of the operation epoch is large enough* $r \geq \overline{r}(A, C, \varkappa, \delta_*, W, \gamma, R)$, *the above coder–decoder pair tracks the state in probability under any random noises and the initial state that are σ-bounded for some* $\sigma > 0$.

It should be stressed that the threshold duration $\overline{r}(A, C, \varkappa, \delta_*, W, \gamma, R)$ is not affected by the moment degree σ and the parameters of the noises.

The proof of Proposition 8.4.4 is not given due to the reason explained on p. 249.

[6] See Definitions 6.3.9 and 6.3.10 on pp.139 and 140.

8.5 Statement of the Stabilization Problem

Now we consider the controlled version of the plant (8.2.1) (on p. 249):

$$x(t+1) = Ax(t) + Bu(t) + \xi(t), \qquad x(0) = x_0;$$
$$y(t) = Cx(t) + \chi(t), \quad t = 0, 1, \dots, \tag{8.5.1}$$

where $u \in \mathbb{R}^{n_u}$ is the control. The plant is still unstable and should be stabilized.

The sensor signals are still communicated to the decision-maker (controller) via a given discrete noisy channel and prepared for transmission by a *coder*. Unlike Sect. 8.2, we do not assume a feedback communication link (see Fig. 8.2). The current control is generated by the *decoder(-controller)* on the basis of the prior channel outputs s. In this situation, the *controller* is constituted by the coder and decoder, which are given by equations of the following forms, respectively:

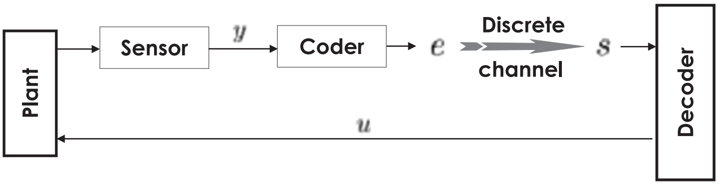


Fig. 8.2. Stabilization via a limited capacity communication channel.

$$e(t) = \mathcal{E}[t, y(0), \dots, y(t)], \tag{8.5.2}$$

$$u(t) = \mathcal{U}\left[t, s(0), s(1), \dots, s(t)\right]. \tag{8.5.3}$$

Explanation 8.5.1. These formulas describe the class of all nonanticipating (causal) controllers. Its study is instructive for necessary conditions for stabilizability. However, this class includes unrealistic controllers of unlimited complexity, which carry out an asymptotically infinite amount of computations per step. As will be shown, the conditions necessary for stabilization by means of such a controller are "almost" sufficient for stabilization by a realistic controller with a limited computational power, which is based on classic block encoding–decoding procedures of data transmission across the channel.

Like in Sect. 8.2, we assume that a coder–decoder pair succeeds if it keeps the (stabilization) error bounded in probability.

Definition 8.5.2. *A controller is said to* stabilize the plant in probability *if the time sequence of states* $x(t), t = 0, 1, \dots$ *is bounded in probability.*[7]

[7]See Definition 8.2.1 on p. 250.

Observation 8.2.3 naturally extends on the property introduced in this definition. In particular, almost sure stability ($\sup_t \|x(t)\| < \infty$ a.s.) clearly implies stability in probability. The last property also holds if $\sup_t \boldsymbol{E}\|x(t)\|^\sigma < \infty$ for some $\sigma > 0$.

Although stability in probability is a weak performance criterion, it suffices to distinguish between stable and unstable systems in the classic problem setup. To illustrate this, let us examine the static linear controller

$$u(t) = Ky(t). \tag{8.5.4}$$

As is known, it stabilizes the system (in various reasonable senses) if and only if the eigenvalues of the matrix $A + BKC$ lie in the open unit disk. In its turn, this holds if and only if the controller (8.5.4) stabilizes the plant in probability, as is shown by the following analog of Lemma 8.2.4 (on p. 251).

Lemma 8.5.3. *Suppose that the assumptions of Lemma* 8.2.4 *are true. Then the controller* (8.5.4) *stabilizes the plant in probability if and only if the eigenvalues of the matrix $A + BKC$ lie in the open unit disk of the complex plane.*

The proof of this illustrating fact will be given in Sect. 8.8 (starting on p. 264).

8.6 Stabilizability Domain

Theorem 8.6.1. *Let Assumptions* 8.3.1, 8.3.2*, and* 8.3.6 *with $\sigma = \infty$ (on pp.* 251 *and* 252*) be valid. Then the following claims hold:*

(i) If a coder (8.5.2) *and decoder* (8.5.3) *exist that stabilize the plant in probability under some random noises and initial state satisfying Assumptions* 8.3.3 *and* 8.3.4 *(on p.* 252*), then inequality* (8.3.4) *(on p.* 252*) holds;*
(ii) Conversely, suppose that (8.3.4) *holds with the strict inequality sign $H(A) < \mathfrak{c}$; the pairs (A, C) and (A, B) are detectable and stabilizable, respectively; and estimates D, D_χ of the noises in the system and sensor are known*

$$\|\xi(t)\| \leq D, \quad \|\chi(t)\| \leq D_\chi \quad \forall t \qquad \text{a.s.} \tag{8.6.1}$$

Then a semirational finite-step recursive[8] *coder–decoder pair* (8.5.2)*,* (8.5.3) *exists that stabilizes the plant in probability.*

Thus the inequality $H(A) \leq \mathfrak{c}$ is necessary and "almost sufficient" for stabilizability in probability. We recall that Theorem 8.6.1 concerns the case where there is no feedback communication link. It can be shown that if such a link is yet available and so the coder equation (8.5.2) can be taken in the form (8.2.3) (on p. 250), the above inequality remains necessary.

A stabilizing coder–decoder pair will be described in the next section. This pair depends on the noise parameters D, D_χ.

[8] See Definitions 6.3.9 and 6.3.10 on pp.139 and 140, respectively.

Remark 8.6.2. If a perfect communication feedback of an arbitrarily small rate is available, a universal stabilizing pair can be offered. This pair does not depend on the noise parameters (like the estimates D, D_χ) and stabilizes the plant under all noises that are σ-bounded[9] with some $\sigma > 0$.

The details will be specified, and the corresponding pair will be described in Subsect 8.7.2 (starting on p. 262). This pair consumes a bounded (as time progresses) computational power and memory per step.

The proof of Theorem 8.6.1 is not presented due to the reason explained on p. 249.

8.7 Stabilizing Coder–Decoder Pair

Like in Sect. 8.4 and by the reasons stated there, the focus is on systems with only unstable modes. In other words, we adopt Assumption 8.4.1 (on p. 253) in this section. We also suppose that the assumptions of (ii) from Theorem 8.6.1 hold. It follows that $H(A) < \mathfrak{c}$ and the pairs (A, C), (A, B) are observable and controllable, respectively.

8.7.1 Stabilizing Coder–Decoder Pair without a Communication Feedback

This pair can be viewed as a modification of that from Subsect. 6.11.2 (described by cs.1)–cs.6), ds.1), ds.2), cs.2-3_1), cs.2-3_1), on pp. 182, 183, and 187). Modification is caused by accounting for the plant noises and basically follows the lines of m.1) and m.2) on p. 223. In particular, the rule to update the bound δ will be of the form (7.7.2) (on p. 224) and so not forcing δ to decay to zero. The major point of novelty concerns the alarm control sequence (6.11.15) (on p. 186). We recall that by applying this sequence to the plant, the decoder notifies the coder about receiving the alarm signal ✠. The construction of this sequence should be altered in order to make it work in the presence of the plant noises. This will be the only point in the subsequent considerations where the knowledge of the noise bounds D, D_χ is used.

Components of the Coder and Decoder

To construct a stabilizing coder–decoder pair, we need a deadbeat stabilizer;[10] i.e., a linear transformation of an initial state

$$x(0) = x \xrightarrow{\mathcal{N}} \boldsymbol{U} = [u(0), u(1), \ldots, u(n-1), 0, 0, \ldots] \qquad (8.7.1)$$

into a sequence of controls driving the state to zero $x(n) = 0$ in the absence of the noise $\xi(t) \equiv 0$.

The alarm control sequence is introduced in the following lemma.

[9] See Definition 8.3.5 on p. 252.
[10] See p. 72 for the definition.

Lemma 8.7.1. *There exists a control sequence*

$$\boldsymbol{U}_{\maltese} = [u_0, \ldots, u_{4n}, 0, 0, \ldots] \tag{8.7.2}$$

with the following two properties:

(i) It drives the unperturbed ($\xi(t) \equiv 0$) system from $x(0) = 0$ to $x(4n+1) = 0$;
(ii) A constant $D_* > 0$ exists such that any control sequence produced by the deadbeat stabilizer (8.7.1) *is distinguishable from the sequence $\delta \boldsymbol{U}_{\maltese}$ with $\delta \geq D_*$ on the basis of observations $y(0), \ldots, y(4n+1)$ resulting from action of these sequences on the noisy plant* (8.5.1)*.*

The proof of this technical fact is given in Subsect. 8.7.3 (on p. 263).

As will be shown in the proof, the sequences from (ii) can be distinguished by means of a semialgebraic map[11] taking two values.

Remark 8.7.2. The constant D_* depends on D, D_χ from (8.6.1).

Remark 8.7.3. We extend the deadbeat stabilizer on $\maltese$ by putting $\mathcal{N}(\maltese) := \boldsymbol{U}_{\maltese}$, and we define its *length* to be $L(\mathcal{N}) := 4n + 1$.

Finally, we pick the components described in 1)–6) on pp. 253 and 254.

Remark 8.7.4. (i) In 5), the epoch duration r should be chosen so that $r > L(\mathcal{N}) + n$, like in Subsect. 6.11.2 (starting on p. 185).
(ii) In 1), the constant δ_* should be chosen large enough $\delta_* > D_*$, where D_* is the constant from (ii) of Lemma 8.7.1.
(iii) The conditions under which the estimate (8.4.3) (on p. 254) produced by the deadbeat observer from 6) is exact should be extended by the requirement that the controls are zero: $u(\theta) \equiv 0$ for $\theta \in (t - n : t)$ in (8.5.1).

Description of the Coder and Decoder

Like in Subsects. 6.11.1 and 6.11.2 (starting on pp. 180 and 185, respectively), both coder and decoder compute controls $u_c(t)$, $u_d(t)$ and upper bounds for the state norm $\delta_c(t), \delta_d(t)$, respectively. Acting upon the plant is $u_d(t)$. The initial bound is common: $\delta_c(0) = \delta_d(0) = \delta_0 \geq \delta_*$ and may be incorrect.

The coder (at the times $t = \tau_i := ir, i = 1, 2, \ldots$)

cs.1) Calculates the estimate $\widetilde{x}(t)$ of the current state by means of the deadbeat observer (8.4.3) (on p. 254);
cs.2) Computes the prognosis $\widehat{x}_c(\tau_{i+1})$ of the state at time $t = \tau_{i+1}$:

$$\widehat{x}_c(\tau_{i+1}) := A^r \widetilde{x}(\tau_i) + \sum_{\theta=\tau_i}^{\tau_{i+1}-1} A^{t-1-\theta} B u_c(\theta);$$

[11] See Definition 3.4.9 on p. 46.

cs.3) For $i = 3, 4, \ldots$, corrects the state norm upper bound $\delta_c(\tau_i)$ by the rule to be specified further on p. 261;

cs.4) Computes the quantized value $q_c(\tau_i)$ of the scaled state prognosis:

$$\varepsilon(\tau_i) := \left[\delta_c(\tau_i)\right]^{-1}\widehat{x}_c(\tau_{i+1}), \qquad q_c(\tau_i) := \mathfrak{Q}_r\left[\varepsilon(\tau_i)\right],$$

where $\mathfrak{Q}_r$ is the quantizer from 4) on p. 254;

cs.5) Employs the block code from 3) on p. 254 to convert $q_c(\tau_i)$ into a code word E_i to be sent over the channel during the next epoch $[\tau_i : \tau_{i+1})$;

cs.6) Computes the control program

$$\boldsymbol{U}_{i+1}^c = \mathbf{col}\left[u_c(\tau_{i+1}), \ldots, u_c(\tau_{i+2} - 1)\right]$$

for the operation epoch $[\tau_{i+1} : \tau_{i+2})$ following the next one $[\tau_i : \tau_{i+1})$ and then corrects δ_c:

$$\boldsymbol{U}_{i+1}^c := \delta_c(\tau_i)\mathcal{N}\left[q_c(\tau_i)\right], \quad \delta_c(\tau_i) := \delta_c(\tau_i) \cdot \langle q_c(\tau_i)\rangle_{\varkappa,\gamma}^r + \delta_*, \qquad (8.7.3)$$

where $\langle q\rangle_{\varkappa,\gamma}$ is defined by (8.4.7) (on p. 255), $\varkappa$ is the constant from Lemma 3.8.36 (on p. 80), γ, δ_* are taken from (8.4.1) (on p. 253), and $\mathcal{N}$ is the deadbeat stabilizer (8.7.1).

The decoder (at the times $t = \tau_i, i = 2, 3, \ldots$)

ds.1) Applies the decoding rule $\mathcal{D}_r$ to the data received within the previous epoch $[\tau_{i-1} : \tau_i)$ and thus acquires the decoded value $q_d(\tau_i)$ of $q_c(\tau_{i-1})$;

ds.2) Computes the control program

$$\boldsymbol{U}_i^d = \mathbf{col}\left[u_d(\tau_i), \ldots, u_d(\tau_{i+1} - 1)\right]$$

for the next epoch $[\tau_i : \tau_{i+1})$ and corrects the state upper bound

$$\boldsymbol{U}_i^d := \delta_d(\tau_i)\mathcal{N}\left[q_d(\tau_i)\right], \qquad \delta_d(\tau_i) := \delta_d(\tau_i) \cdot \langle q_d(\tau_i)\rangle_{\varkappa,\gamma}^r + \delta_*. \qquad (8.7.4)$$

Remark 8.7.5. The decoded value $q_d(\tau_i)$ may differ from the true one $q_c(\tau_{i-1})$ due to transmission errors.

Like in Remark 6.11.16 (on p. 184), we assume that $q_c(\tau_k) := q_d(\tau_{1+k}) := \maltese$ for $k \leq 0$ and that the coder at time $t = \tau_0$ performs cs.6) and the decoder at time $t = \tau_1$ accomplishes ds.2) accordingly. Then $\boldsymbol{U}_1^c = \boldsymbol{U}_1^d = \delta_0\boldsymbol{U}_{\maltese}$. For consistency, we also put $\boldsymbol{U}_0^c := \boldsymbol{U}_0^d := \delta_0\boldsymbol{U}_{\maltese}$.

To specify **cs.3)**, we need the following.

Observation 8.7.6. *Proceeding from the observations $y(t)$ during the operation epoch $[\tau_{i-1} : \tau_i]$, $i \geq 3$, the coder is able to recognize whether the code word received by the decoder at the beginning τ_{i-1} of this epoch was $\maltese$.*

Indeed, by (ii) of Remark 8.7.4 and the second formula from (8.7.4), $\delta_d(\tau_i) \geq \delta_* > D_*$ for any i. In view of this, the observation is immediate from the first formula in (8.7.4) , (ii) in Lemma 8.7.1, and (i) of Remark 8.7.4.

Observation 8.7.6 explains why the following specification of the step **cs.3)** is well defined.

cs.3) The correction is as follows:

$$\delta_c(\tau_i) := \left(\left[\frac{\delta_c(\tau_i) - \delta_*}{\langle q_c(\tau_{i-1}) \rangle_{\varkappa,\gamma}^r} - \delta_* \right] \lambda_i + \delta_* \right) \langle q_c(\tau_{i-1}) \rangle_{\varkappa,\gamma}^r + \delta_*, \quad (8.7.5)$$

where

$$\lambda_i := \left\{ \begin{array}{ll} \left(\frac{\gamma}{\varkappa}\right)^r & \text{if } q_d(\tau_{i-1}) = \maltese \ \& \ q_c(\tau_{i-2}) \neq \maltese \\ \left(\frac{\varkappa}{\gamma}\right)^r & \text{if } q_d(\tau_{i-1}) \neq \maltese \ \& \ q_c(\tau_{i-2}) = \maltese \\ 1 & \text{otherwise} \end{array} \right\} \overset{(8.4.7)}{=\!=} \frac{\langle q_d(\tau_{i-1}) \rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{i-2}) \rangle_{\varkappa,\gamma}^r}. \quad (8.7.6)$$

Remark 8.7.7. Like in Subsect. 6.11.1, the operation cs.3) makes the bounds δ_c and δ_d identical whenever the transmission across the channel is errorless.

This claim is rigorously specified by the lemma to follow. To state it, we mark the values of δ_c and δ_d after and just before the updates in accordance with (8.7.3) and (8.7.4) with the $^+$ and $^-$ indices, respectively. So the value $\delta_c^-(\tau_i)$ is taken after the correction (8.7.5). The following lemma is an analog of Lemma 6.11.21 (on p. 184).

Lemma 8.7.8. *The step* cs.3) *ensures that whenever the current transmission is errorless, the next state norm upper bounds produced by the coder and decoder, respectively, are identical:*

$$q_c(\tau_{i-1}) = q_d(\tau_i) \quad \Longrightarrow \quad \delta_c^-(\tau_i) = \delta_d^-(\tau_{i+1}), \qquad i = 1, 2, \dots. \quad (8.7.7)$$

Proof. It suffices to show that for $i \geq 1$, the following formula holds:

$$\delta_c^-(\tau_i) - \delta_* = \left[\delta_d^-(\tau_{i+1}) - \delta_*\right] \left[\frac{\langle q_c(\tau_{i-1}) \rangle_{\varkappa,\gamma}}{\langle q_d(\tau_i) \rangle_{\varkappa,\gamma}} \right]^r. \quad (8.7.8)$$

The proof will be by induction on i. For $i = 1$, the claim is evident. Suppose that (8.7.8) holds for some $i \geq 1$. Note that due to the conventions following Remark 8.7.5, formula (8.7.5) is true for not only $i \geq 3$ but also $i = 1, 2$. Then

$$\begin{aligned} \delta_c^-(\tau_{i+1}) - \delta_* &\overset{(8.7.5)}{=\!=} \left(\left[\frac{\delta_c^+(\tau_i) - \delta_*}{\langle q_c(\tau_i) \rangle_{\varkappa,\gamma}^r} - \delta_* \right] \lambda_{i+1} + \delta_* \right) \langle q_c(\tau_i) \rangle_{\varkappa,\gamma}^r \\ &\overset{(8.7.3)}{=\!=} \left(\left[\delta_c^-(\tau_i) - \delta_*\right] \lambda_{i+1} + \delta_* \right) \langle q_c(\tau_i) \rangle_{\varkappa,\gamma}^r \\ &\overset{(8.7.6)}{=\!=} \left(\left[\delta_c^-(\tau_i) - \delta_*\right] \frac{\langle q_d(\tau_i) \rangle_{\varkappa,\gamma}^r}{\langle q_c(\tau_{i-1}) \rangle_{\varkappa,\gamma}^r} + \delta_* \right) \langle q_c(\tau_i) \rangle_{\varkappa,\gamma}^r \\ &\overset{(8.7.8)}{=\!=} \delta_d^-(\tau_{i+1}) \langle q_c(\tau_i) \rangle_{\varkappa,\gamma}^r \overset{(8.7.4)}{=\!=} \left[\delta_d^-(\tau_{i+2}) - \delta_*\right] \frac{\langle q_c(\tau_i) \rangle_{\varkappa,\gamma}^r}{\langle q_d(\tau_{i+1}) \rangle_{\varkappa,\gamma}^r}; \end{aligned}$$

i.e., (8.7.8) with $i := i + 1$ does hold. □

Remark 8.7.9. (i) The proposed coder and decoder are semirational r-step recursive.[12] Their complexity per step is basically determined by the parameter r of the algorithm and does not increase as time progresses.

(ii) For $\delta_* := 0$, the above coder and decoder are identical to those from Subsect. 6.11.2 (starting on p. 185). So by Proposition 6.11.28 (on p. 187), they ensure a.s. stabilization $x(t) \to 0$ of the noiseless plant $\xi(t) \equiv 0, \chi(t) \equiv 0$ provided that the epoch duration r is large enough.

Stabilization in Probability in the Absence of a Feedback Communication Link

Proposition 8.7.10. *Suppose that the assumptions of (ii) from Theorem* 8.6.1 *(on p.* 257*) and Assumption* 8.4.1 *(on p.* 253*) hold. Then the above coder–decoder pair stabilizes the plant in probability, provided that the duration r of the operation epoch is large enough* $r \geq \overline{r}(A, B, C, \varkappa, \delta_*, W, \gamma, R)$.

The proof of this proposition is not given due to the reason explained on p. 249.

8.7.2 Universal Stabilizing Coder–Decoder Pair Consuming Feedback Communication of an Arbitrarily Low Rate

The controller introduced in the previous subsection deals with uniformly bounded noises (8.6.1) (on p. 257) and depends on their upper bounds D, D_χ. Specifically, these bounds determine the constant D_* by Remark 8.7.2 (on p. 259), whereas the parameter δ_* of the controller should exceed D_* by (ii) of Remark 8.7.4 (on p. 259).

Now we show that if an arbitrarily small perfect communication feedback is available (see Fig. 8.3), a universal stabilizing coder–decoder pair can be offered. This pair does not depend on the noise parameters (like D, D_χ) and stabilizes the plant under all noises that are σ-bounded[13] with some $\sigma > 0$.

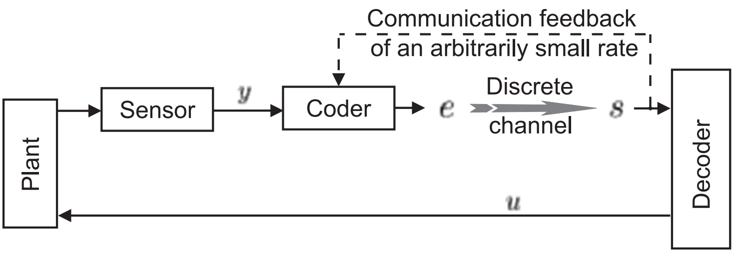


Fig. 8.3. Stabilization with arbitrarily small communication feedback.

One more assumption is adopted in this subsection.

[12] See Definitions 6.3.9 and 6.5.7 on pp. 139 and 145, respectively.

[13] See Definition 8.3.5 on p. 252.

Assumption 8.7.11. *There is a* feedback communication link *capable of transmitting one bit of information without an error from the decoder-controller to the sensor site during any time interval of sufficiently large duration.*

The average data rate of this link $R = 1\text{bit}/(\text{large interval duration}) \approx 0$.

Let the bit from Assumption 8.7.11 be transmitted during the operation epoch (which thus should be chosen to be large enough) and used to notify the coder whether the code word received by the decoder at the beginning τ_{i-1} of the epoch $[\tau_{i-1} : \tau_i)$ was ✠. The notification is received at $t = \tau_i$. Then Observation 8.7.6 and the alarm control sequence become useless. So is the inequality $\delta_* > D_*$ from (ii) of Remark 8.7.4, whose role is confined to serving Observation 8.7.6. This permits us to relax this inequality into $\delta_* > 0$. After this, no controller parameters related to the noise uniform upper bounds D, D_χ from (8.6.1) (on p. 257) remain in use. In other words, the controller becomes independent of the noises. At the same time, it stabilizes the plant, as is shown by the following.

Proposition 8.7.12. *Suppose that Assumptions* 8.3.1, 8.3.2, 8.4.1 *(on pp.* 251 *and* 253*), and* 8.7.11 *are valid,* $H(A) < \mathfrak{c}$*, and the pairs* (A, C)*,* (A, B) *are detectable and stabilizable, respectively. Let us employ the zero alarm control sequence* $\boldsymbol{U}_{✠} := 0$ *and an arbitrary positive parameter* $\delta_* > 0$ *in the controller* **cs.1)–cs.6), ds.1), ds.2)** *(see pp.* 259 *and* 261*) modified as was described in this subsection.*

If the duration r *of the epoch is large enough* $r \geq \overline{r}(A, B, C, \varkappa, \delta_*, W, \gamma, R)$*, this controller stabilizes the plant in probability under any noises that are* σ*-bounded with some* $\sigma > 0$*.*

The proof of this proposition is not given due to the reason explained on p. 249.

8.7.3 Proof of Lemma 8.7.1

We recall that the estimate produced by the deadbeat observer (8.4.3) (on p. 254) is exact provided that the noises are removed and the controls are zero. So elementary estimates based on (8.6.1) (on p. 257) give rise to the following.

Observation 8.7.13. *In the presence of the noises, the error produced by the observer* (8.4.3) *is bounded by a constant determined by the noise levels from* (8.6.1)*:*

$$\big\|x(t) - \widetilde{x}(t)\big\| \leq \overline{D} = \overline{D}(D, D_\chi) < \infty \qquad \text{whenever} \quad t \geq n, \quad \text{and } u(\theta) = 0 \ \forall \theta \in (t - n : t). \tag{8.7.9}$$

Now we pass to the immediate proof of Lemma 8.7.1. The alarm control sequence (8.7.2) is constructed in the form (6.11.15) (on p. 186):

$$\boldsymbol{U}_{✠} := \textbf{col}\,\big[\underbrace{0, \ldots, 0}_{2n}, u_*, \underbrace{0, \ldots, 0}_{n}, u_0^-, \ldots, u_{n-1}^-, 0, \ldots\big].$$

Here u_* is taken so that $Bu_* \neq 0$, and $u_0^-, \ldots, u_{n-1}^-$ is a control sequence that drives the unperturbed ($\xi(t) \equiv 0$) system from $x(0) = A^n B u_*$ to $x(n) = 0$. (It exists since the pair (A, B) is controllable.) As a result, the entire sequence $\boldsymbol{U}_{\maltese}$ drives the system from $x(0) = 0$ to $x(4n+1) = 0$, as is required by (i) from Lemma 8.7.1.

For both $\boldsymbol{U} := \delta \boldsymbol{U}_{\maltese}$ and any control sequence $\boldsymbol{U}$ generated by the deadbeat stabilizer (8.7.1) (on p. 258), the controls are zero within n time steps before both $\tau_{i-1} + 2n$ and $\tau_{i-1} + 3n + 1$. So the estimates produced by the deadbeat observer (8.4.3) (on p. 254) at times $t = \tau_{i-1} + 2n, \tau_{i-1} + 3n + 1$ in the presence of the plant noises obey the exactness bound (8.7.9). Now we put

$$w(\boldsymbol{U}) := \begin{cases} \delta A^n B u_* & \text{for} \quad \boldsymbol{U} = \delta \boldsymbol{U}_{\maltese} \\ 0 & \text{for } \boldsymbol{U} \text{ produced by the deadbeat stabilizer} \end{cases} \tag{8.7.10}$$

and note that

$$\begin{aligned} &\Big\| \underbrace{\widetilde{x}(\tau_{i-1} + 3n + 1) - A^{n+1}\widetilde{x}(\tau_{i-1} + 2n)}_{\varpi} - w(\boldsymbol{U}) \Big\| \\ &\leq \left\|\widetilde{x}(\tau_{i-1}+3n+1) - x(\tau_{i-1}+3n+1)\right\| + \left\|A^{n+1}\left[\widetilde{x}(\tau_{i-1}+2n) - x(\tau_{i-1}+2n)\right]\right\| \\ &\qquad + \left\|x(\tau_{i-1} + 3n + 1) - A^{n+1}x(\tau_{i-1} + 2n) - w(\boldsymbol{U})\right\| \\ &\overset{(8.7.9)}{\leq} \overline{D}\big(1 + \|A\|^{n+1}\big) + \left\|x(\tau_{i-1} + 3n + 1) - A^{n+1}x(\tau_{i-1} + 2n) - w(\boldsymbol{U})\right\| \\ &\qquad \overset{(8.5.1)}{=\!=} \overline{D}\big(1 + \|A\|^{n+1}\big) + \left\| \sum_{\theta=\tau_{i-1}+2n}^{\tau_{i-1}+3n} A^{\tau_{i-1}+3n-\theta}\xi(\theta) \right\| \\ &\qquad\qquad \overset{(8.6.1)}{\leq} \overline{D}\big(1 + \|A\|^{n+1}\big) + D\sum_{\theta=0}^{n} \|A\|^{\theta} =: D_0. \end{aligned}$$

Thus by computing ϖ, the coder gets to know $w(\boldsymbol{U})$ with the exactness D_0. Thanks to (8.7.10), this knowledge suffices to distinguish between $\boldsymbol{U}_{\maltese}$ and any control sequence produced by the deadbeat stabilizer if $\delta > D_* := 2D_0/(\|A^n B u_*\|)$. □

8.8 Proofs of Lemmas 8.2.4 and 8.5.3

It is easy to see that in the case of Lemma 8.2.4 (on p. 251), the estimation error $\Delta(t) := x(t) - \widehat{x}(t)$ evolves as follows:

$$\Delta(t+1) = M\Delta(t) + \zeta(t), \qquad \Delta(0) = \Delta_0, \tag{8.8.1}$$

where

$$M := A - KCA, \qquad \zeta(t) := (I - KC)\xi(t) - K\chi(t+1),$$
$$\Delta_0 := (I - KC)x_0 - M\widehat{x}(-1) - K\chi(0).$$

By substituting (8.5.4) (on p. 257) into (8.5.1) (on p. 256), we see that in the case of Lemma 8.5.3 (on p. 257), the evolution of $\Delta(t) := x(t)$ is given by (8.8.1) with

$$M := A + BKC, \qquad \zeta(t) := \xi(t) + BK\chi(t), \qquad \Delta_0 := x_0.$$

In both cases, the noise sequence $\{\zeta(t)\}$ is independent of the initial state Δ_0 and formed by mutually independent random vectors, which are identically distributed in accordance with a common probability density $p(\zeta)$. Furthermore, these vectors and the initial state are a.s. bounded:

$$\|\zeta(t)\| \le D_\zeta \quad \forall t, \qquad \|\Delta_0\| \le D_0 \quad \text{a.s.} \tag{8.8.2}$$

Thus both lemmas follow from the following.

Lemma 8.8.1. *The sequence* $\{\Delta(t)\}$ *generated by the recursion of the form* (8.8.1) *is bounded in probability if and only if the eigenvalues of the matrix* M *lie in the open unit disk.*

The remainder of the section is devoted to the proof of this lemma. We start with a technical claim. In fact, it states nothing but that in the absence of control, the state of the stochastically disturbed unstable plant grows without limits. Moreover, the state becomes more and more distributed over the space: The probability that it lies in a ball vanishes as $t \to \infty$ uniformly over all balls of a common and fixed radius.

Lemma 8.8.2. *Suppose that the system* (8.8.1) *is unstable: There is an eigenvalue* λ *of* M *with* $|\lambda| \ge 1$. *Then for any given* $b > 0$*, the following relation holds:*

$$\omega_t(b) := \sup_{\phi \in \mathbb{R}^n} \boldsymbol{P}\big\{\|\Delta_t^{-c} + \phi\| \le b\big\} \to 0 \quad \text{as} \quad t \to \infty, \tag{8.8.3}$$

where $\Delta_t^{-c} := \sum_{\theta=0}^{t-1} M^{t-1-\theta}\zeta(\theta)$ *is the state at time* t*, provided that the system starts at zero* $\Delta(0) = 0$.

Proof. We start with a simple observation. Let $M = \left(\begin{smallmatrix} M_{11} & M_{12} \\ 0 & M_{22}\end{smallmatrix}\right)$, where M_{11} and M_{22} are square matrices of respective dimensions n_1 and n_2, and the conclusion of the lemma be true for $M := M_{22}, n := n_2$, and $\zeta(t) := \zeta''(t)$. Here $\zeta = \begin{pmatrix} \zeta' \\ \zeta'' \end{pmatrix}$ is the partition of $\zeta \in \mathbb{R}^n$ matching the above partition of the matrix. Then this conclusion holds for the primal matrix M and disturbances $\zeta(t)$.

Indeed, it is easy to see that

$$\left[\Delta_t^{-c}\right]'' = \left[\sum_{\theta=0}^{t-1} M^{t-1-\theta}\zeta(\theta)\right]'' = \sum_{\theta=0}^{t-1} M_{22}^{t-1-\theta}\zeta''(\theta).$$

Furthermore a constant $c > 0$ exists such that $\|\Delta''\| \le c\|\Delta\|$ for any Δ. So $\|\Delta_t^{-c} + \phi\| < b \Rightarrow \|\left(\Delta_t^{-c}\right)'' + \phi''\| < cb$, and

$$\sup_{\phi \in \mathbb{R}^n} \boldsymbol{P}\big\{\|\Delta_t^{-c} + \phi\| < b\big\} \le \sup_{\phi'' \in \mathbb{R}^{n_2}} \boldsymbol{P}\big\{\|\left(\Delta_t^{-c}\right)'' + \phi''\| < cb\big\} \to 0$$

as $t \to \infty$. Thus (8.8.3) does hold for the primal matrix M.

By reducing the matrix M to the real Jordan form and employing the above observation, we see the the proof can be focused on the case where for some $\rho \geq 1$,

$$\text{either } n = 1 \text{ and } M = \pm\rho \quad \text{or } n = 2 \text{ and } M = \rho \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \varphi \neq 0, \pm\pi, \ldots.$$

We shall consider separately two cases.

1. Suppose first that $\rho > 1$. By denoting $\Delta_{1,t}^{-c} := \Delta_t^{-c} - M^{t-1}\zeta(0)$, we get

$$\begin{aligned}
\sup_\phi \boldsymbol{P}\left\{\|\Delta_t^{-c} + \phi\| < b\right\} &= \sup_\phi \int \boldsymbol{P}\left\{\|\Delta_t^{-c} + \phi\| < b \middle| \Delta_{1,t}^{-c} = \Delta\right\} \boldsymbol{P}_{\Delta_{1,t}^{-c}}(d\Delta) \\
&\leq \int \sup_\phi \boldsymbol{P}\left\{\|M^{t-1}\zeta(0) + \Delta + \phi\| < b \middle| \Delta_{1,t}^{-c} = \Delta\right\} \boldsymbol{P}_{\Delta_{1,t}^{-c}}(d\Delta) \\
&\overset{a)}{=} \int \sup_\psi \boldsymbol{P}\left\{\|M^{t-1}\zeta(0) + \psi\| < b \middle| \Delta_{1,t}^{-c} = \Delta\right\} \boldsymbol{P}_{\Delta_{1,t}^{-c}}(d\Delta) \\
&\overset{b)}{=} \int \sup_\psi \boldsymbol{P}\left\{\|M^{t-1}\zeta(0) + \psi\| < b\right\} \boldsymbol{P}_{\Delta_{1,t}^{-c}}(d\Delta) \\
&= \sup_\psi \boldsymbol{P}\left\{\|M^{t-1}[\zeta(0) + M^{1-t}\psi]\| < b\right\} \\
= \sup_\psi \boldsymbol{P}\left\{\|\zeta(0) + M^{1-t}\psi\| < b\rho^{-(t-1)}\right\} &\overset{c)}{=} \sup_\alpha \int_{\{\zeta : \|\zeta + \alpha\| < b\rho^{-(t-1)}\}} p(\zeta) d\zeta.
\end{aligned}$$

Here a) is justified by $\psi := \Delta + \phi$, b) holds since $M^{t-1}\zeta(0)$ and $\Delta_{1,t}^{-c} = \sum_{\theta=1}^{t-1} M^{t-1-\theta}\zeta(\theta)$ are independent, and c) results from putting $\alpha := M^{1-t}\psi$ and invoking that $p(\zeta)$ is the probability density of $\zeta(\theta)$. The proof is completed by noting that the last integral goes to 0 as $t \to \infty$ uniformly over $\alpha \in \mathbb{R}^n$ because so does the volume of the set $\{\zeta : \|\zeta + \alpha\| < b\rho^{-(t-1)}\}$ (see, e.g., [191, Theorem 10.3.2]).

2. Now let $\rho = 1$. We focus on the case where M is the rotation matrix $M = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$; in the case $M = \pm 1$, the proof is similar.

Let η_t^{-c} denote the first component of Δ_t^{-c} and $\zeta(\theta) = \textbf{col}\left[\zeta^{(1)}(\theta), \zeta^{(2)}(\theta)\right]$. The distribution of

$$\eta_t^{-c} = \sum_{\theta=0}^{t-1} \left[\cos(\theta\varphi)\zeta^{(1)}(t-1-\theta) - \sin(\theta\varphi)\zeta^{(2)}(t-1-\theta)\right]$$

is equal to that of

$$\widetilde{\eta}_t^{-c} = \sum_{\theta=0}^{t-1} \varpi(\theta), \quad \text{where} \quad \varpi(\theta) := \cos(\theta\varphi)\zeta^{(1)}(\theta) - \sin(\theta\varphi)\zeta^{(2)}(\theta)$$

are mutually independent random quantities with zero mean and

$$\begin{aligned}
\mu_3(\theta) &:= \boldsymbol{E}\left|\varpi(\theta)\right|^3 \leq \boldsymbol{E}\left[\left|\zeta^{(1)}(\theta)\right| + \left|\zeta^{(2)}(\theta)\right|\right]^3 \\
&\overset{a)}{\leq} 2^{3/2}\boldsymbol{E}\left(\left[\zeta^{(1)}(\theta)\right)^2 + \left(\zeta^{(2)}(\theta)\right]^2\right)^{3/2} = 2^{3/2}\int_{\mathbb{R}^2} \|\zeta\|^3 p(\zeta)\, d\zeta =: \mu \overset{b)}{<} \infty.
\end{aligned}$$

Here a) follows from the elementary inequality $(a+b)^3 \leq 2^{3/2}(a^2+b^2)^{3/2}\ \forall a, b \geq 0$ and b) holds by (8.8.2). Since the function

$$\psi \mapsto f(\psi) := \int \left|\zeta_1 \cos\psi - \zeta_2 \sin\psi\right|^2 p(\zeta_1, \zeta_2)\, d\zeta_1 d\zeta_2 \in (0, +\infty)$$

is continuous due to the Lebesgue's dominated convergence theorem, we have

$$\sigma_-^2 := \min_{\psi} f(\psi) > 0, \qquad \sigma_\theta^2 := \boldsymbol{E}\varpi(\theta)^2 = f(\theta\varphi) \geq \sigma_-^2.$$

By the central limit theorem, the normalized sum $s_t^{-1}\widetilde{\eta}_t^{-c}, s_t^2 = \sigma_0^2 + \cdots + \sigma_{t-1}^2$ of independent unbiased variables $\varpi(0), \ldots, \varpi(t-1)$ is distributed approximately normally. Moreover by a variant of the Berry–Esséen theorem (see [53, Theorem 2, Sec. 5, Ch. XVI]), the distribution function of this sum $F_t(\eta) := \boldsymbol{P}\{s_t^{-1}\widetilde{\eta}_t^{-c} < \eta\} = \boldsymbol{P}\{s_t^{-1}\eta_t^{-c} < \eta\}$ approaches the distribution function $\mathfrak{N}(\eta)$ of the normal distribution with zero mean and variance 1 uniformly over $\eta \in \mathbb{R}$:

$$|F_t(\eta) - \mathfrak{N}(\eta)| \leq 6\frac{\mu_3(0) + \cdots + \mu_3(t-1)}{s_t^3} \leq 6\frac{t\mu}{t^{3/2}\sigma_-^3} = \frac{6\mu}{\sqrt{t}\sigma_-^3}.$$

This implies that

$$\boldsymbol{P}\{\eta_- < s_t^{-1}\eta_t^{-c} < \eta_+\} \leq \frac{12\mu}{\sqrt{t}\sigma_-^3} + \frac{1}{\sqrt{2\pi}}\int_{\eta_-}^{\eta_+} e^{-\frac{y^2}{2}}\, dy \tag{8.8.4}$$

whenever $\eta_- \leq \eta_+$. The proof is completed as follows:

$$\sup_{\phi\in\mathbb{R}^2} \boldsymbol{P}\left\{\|\Delta_t^{-c} + \phi\| < b\right\} \overset{a)}{\leq} \sup_{\phi_1\in\mathbb{R}} \boldsymbol{P}\left\{|\eta_t^{-c} + \phi_1| < b\right\}$$

$$= \sup_{\phi_1\in\mathbb{R}} \boldsymbol{P}\left\{|s_t^{-1}\eta_t^{-c} + s_t^{-1}\phi_1| < s_t^{-1}b\right\} \overset{\eta:=s_t^{-1}\phi_1}{=\!=\!=} \sup_{\eta\in\mathbb{R}} \boldsymbol{P}\left\{|s_t^{-1}\eta_t^{-c} + \eta| < s_t^{-1}b\right\}$$

$$\overset{b)}{\leq} \sup_{\eta\in\mathbb{R}} \boldsymbol{P}\left\{|s_t^{-1}\eta_t^{-c} + \eta| < \frac{b}{\sqrt{t}\sigma_-}\right\} \overset{(8.8.4)}{\leq} \frac{12\mu}{\sqrt{t}\sigma_-^3} + \frac{1}{\sqrt{2\pi}} \sup_{\eta\in\mathbb{R}} \int_{\eta-\frac{b}{\sqrt{t}\sigma_-}}^{\eta+\frac{b}{\sqrt{t}\sigma_-}} e^{-\frac{y^2}{2}}\, dy.$$

Here a) holds since $\|\Delta_t^{-c} + \phi\| < b \Rightarrow |\eta_t^{-c} + \phi_1| < b$, where ϕ_1 is the first component of ϕ, and b) follows because $s_t^{-1}b \leq \frac{b}{\sqrt{t}\sigma_-}$. The proof is completed by noting that the last integral converges to 0 as $t \to \infty$ since so does the Lebesgue measure $\frac{2b}{\sqrt{t}\sigma_-}$ of the interval $\left[x\eta - \frac{b}{\sqrt{t}\sigma_-}; \eta + \frac{b}{\sqrt{t}\sigma_-}\right]$ and the function $y \mapsto e^{-y^2/2}$ is summable with respect to this measure (see, e.g., [191, Theorem 10.3.2]). □

Proof of Lemma 8.8.1. Sufficiency. Let the eigenvalues of M lie in the open unit disk. Then $\|M^t\| \leq c\rho^t, t = 0, 1, \ldots$ for some $c > 0$ and $\rho \in (0, 1)$. It follows that

$$\|\Delta(t)\| = \left\|M^t\Delta_0 + \sum_{\theta=0}^{t-1} M^{t-1-\theta}\zeta(\theta)\right\|$$

$$\leq \left\|M^t\right\| \|\Delta_0\| + \sum_{\theta=0}^{t-1} \left\|M^{t-1-\theta}\right\| \|\zeta(\theta)\| \leq cD_0 + \frac{D_\zeta}{1-\rho} < \infty.$$

Thus the sequence $\{\Delta(t)\}$ is bounded almost surely and so in probability.

Necessity. Now we suppose that the sequence $\{\Delta(t)\}$ is bounded in probability. We need to show that the eigenvalues of M lie in the open unit disk. Suppose the contrary. Then by Lemma 8.8.2, relation (8.8.3) holds. Note that $\Delta(t) = M^t\Delta_0 + \Delta_t^{-c}$, where the addends are independent. We denote by $\boldsymbol{P}(d\Delta)$ the probability distribution of Δ_0 and for any $b > 0$, have

$$\begin{aligned}\boldsymbol{P}\big[\|\Delta(t)\| \le b\big] &= \boldsymbol{P}\big[\|M^t\Delta_0 + \Delta_t^{-c}\| \le b\big] \\ &= \int_{\mathbb{R}^n} \boldsymbol{P}\big[\|M^t\Delta_0 + \Delta_t^{-c}| \le b\big|\Delta_0 = \Delta\big]\boldsymbol{P}(d\Delta) \\ &\overset{\text{(a)}}{\le} \int_{\mathbb{R}^n} \sup_{\phi\in\mathbb{R}^n} \boldsymbol{P}\big[\|\phi + \Delta_t^{-c}\| \le b\big|\Delta_0 = \Delta\big]\boldsymbol{P}(d\Delta) \overset{(8.8.3)}{=\!=} \omega_t(b).\end{aligned}$$

Here (a) results from putting $\phi := M^t\Delta$. Thus $\boldsymbol{P}\big[\|\Delta(t)\| \le b\big] \to 0$ as $t \to \infty$ for any b, in violation of Definition 8.2.1. The contradiction obtained proves that the eigenvalues of M lie in the open unit disk. $\square$

9

Decentralized Stabilization of Linear Systems via Limited Capacity Communication Networks

9.1 Introduction

In the previous chapters, the issue of state estimation and control under communication constraints was addressed for networks with the simplest topologies. Basically, they contained only one "sensor-controller" and "controller-actuator" channel. In Chap. 3, the centralized stabilization problem was examined, where multiple sensors were linked via direct channels with a controller, which directly acts on the plant. However, many modern control systems are implemented in a decentralized fashion, which results in a less trivial topology with multiple spatially distributed sensors, controllers, and actuators communicating over a serial digital network.

Such a situation is examined in this chapter. We consider a general network with spatially distributed communicating elements. Any element is endowed with a computing capability, which is used to convert the incoming information streams into outgoing data flows. Some elements are also endowed with a sensing capability: They are able to partially observe an outer (with respect to the network) unstable process. These elements are called *sensors*. Some other elements are able to directly affect the process; they are called *actuators*. The remaining elements act as intermediate *controllers* taking part in transformation of the sensor data into controls in a decentralized and distributed fashion. The algorithms of data processing at the elements should be designed so that the closed-loop system is stable. This system is composed by the outer process and the elements communicating via the network.

We suppose that the network is given. This means that it is indicated between which elements information can be communicated, at which rates, and in which way. Considered are the cases where the transmitted messages may incur delays, be lost and corrupted, interfere, and collide with each other. So messages arriving at a given element may depend on the packets dispatched from many (up to all) elements, including one at hand. In other words, we examine general networks with arbitrary topology. Moreover, we assume that this topology may be dynamically altered by authorized elements. A simple example is a switch of a communication channel from service (connection) of one pair of elements to service of some other pair.

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*, doi: 10.1007/978-0-8176-4607-3_9,

We consider the case where arbitrary restrictions are imposed on data processing algorithms admissible by the controllers. These restrictions may express limitations on the memory size, variety of executable operations, the processor speed, etc. Not excluded is the case where the set of admissible algorithms contains only one algorithm. Then the corresponding controller is given. An example is a network with a data storage dynamically served in accordance with an a priori given rule. Another example is a switching channel with a given protocol of switching. At the same time, we suppose that the sensors and actuators accept arbitrary causal algorithms.

We still consider the case where the outer process is described by linear discrete-time equations. The process is subjected to a bounded additive exogenous disturbance, and there are bounded noises in the sensors. The objective is to find the conditions under which the elements can be equipped with admissible data processing algorithms so that the closed-loop system is stable. In the case where these conditions are satisfied, these algorithms should be explicitly described.

We provide an evidence that this problem is reduced to the long-standing standard problem of the information sciences: Finding the so-called rate (capacity) domain of the network as well as block codes (functions) transmitting data at the rates from this domain. We recall that the capacity domain describes how much information can be transmitted from an element (network node) or set of nodes to another node or set of nodes. The necessary and sufficient conditions for stabilizability established in this chapter are given in terms of this domain. These conditions are tight: For stabilizablity, it is necessary and sufficient that a certain vector characterizing the rate of instability of the outer process belongs to this domain and its interior, respectively. Design of the stabilizing algorithm is ultimately reduced to construction of the block function transmitting data at the rates matching the entries of this vector.

We show that the rate domain of the original network cannot be put in use here and another network should be employed. It is obtained by including new nodes and infinite capacity channels into the original network. These channels are of three kinds. The channels of the first kind go from every actuator to all sensors influenced by this actuator. These channels explicitly express the view of the control loop as a link transmitting information (see Subsects. 6.11.2 and 7.8.1 starting on pp. 185 and 233, respectively). The channels of the second kind broadcast messages from artificial data sources associated with the unstable modes of the process to all sensors detecting the corresponding mode. The channels of the third kind are additive interference channels delivering data to a set of new nodes each associated with an unstable mode as well. Via such a channel, every such node collects data from all actuators that affect the corresponding mode. The decoding algorithms at these nodes are limited to merely projecting the received real signal onto the integer grid. The above capacity domain answers the question: How much data can be transmitted from the sources to the respective new nodes?

The material of this chapter is technically extensive. In order to avoid further complications and highlight the ideas, the exposition does not concern certain points that were discussed in the previous chapters. Unlike Chaps. 6–8, we do not deal with stochastic phenomena in communication and focus attention on deterministic networks. Unlike Chaps. 7 and 8, we suppose that the bounds on the initial state and the plant and sensor noises are known. This permits us to simplify the design of the stabilizing controller by omitting the "zoom-out"–"zoom-in" procedure. As compared with Chap. 3, we impose more restrictive assumptions on the plant, which permits us to avoid the technically demanding sequential stabilization approach (see p. 64). The simplest way to achieve this goal would be to assume that all unstable eigenvalues of the plant are real and distinct. Unfortunately, this casts away at the outlet many very natural cases, like the discrete-time counterparts of the systems $\ddot{x} = u$ and $\ddot{x} + x = u$. So we had no courage to impose such a strong assumption and adopted a weaker one. However, the reader interested more in ideas than in generality is welcomed to project the exposition onto that simplest case, which makes many technical facts evident. Certainly, the above simplifications in the problem statement could be discarded by uniting the approaches from this and the previous chapters. This however would make this rather large chapter much larger. For the same reason, we do not focus attention on the possibility of achieving stability by means of a controller with a limited algebraic complexity, unlike the previous chapters. In the design of the stabilizing decentralized controller, attention is focused on a method to generate a synchronized quantized data based on observations from sensors with independent noises (see Subsects. 9.7.1 and 9.7.2 starting on pp. 330 and 333, respectively).

A survey of the literature concerned with stabilization of systems with many sensors each linked by its own communication channel to a common controller was offered in Sect. 3.1 (starting on p. 37). The case of multiple sensors and actuators, where each sensor is linked with every actuator by a limited capacity perfect channel with time-varying data rate, was considered in [139] for real-diagonalizable systems. Separate necessary and sufficient conditions for stabilizability were obtained. In general, they are not tight [139]. In the case where the system is stabilizable by every actuator and detectable by every sensor, a single necessary and sufficient criterion was established in [139]. Stabilization over switching channels with unlimited capacities was addressed in, e.g., [71, 75]. In [86], minimum necessary data rates required to permit reconstruction of the global track estimates to a given level of accuracy are presented in the case when the state can be represented by a Gauss–Markov process.

The main results of this chapter were originally published in [114]. Also, some relevant results can be found in [122].

The body of the chapter is organized as follows. We first illustrate the general model of a deterministic network to be considered by examples in Sect. 9.2. The model itself is presented in Sect. 9.3. Section 9.4 offers the statements of the stabilization problem and the main result. In Sect. 9.5, this result is illustrated by examples, and several simple technical facts supporting computation of capacity domains are given. Sections 9.6 and 9.7 are devoted to the proofs of the necessity and sufficiency parts of the main result, respectively. The concluding Sect. 9.8 contains the proofs of supporting technical facts.

9.2 Examples Illustrating the Problem Statement

We start with the particular samples of the problem to be studied. Their main purpose is to illustrate both the general model of the network, which will be formally described in Sect. 9.3 and employed throughout the chapter, and the variety of particular cases encompassed by this model.

Example 1. Let multiple spatially distributed sensors observe a common unstable process affected by a set of spatially distributed actuators. The locations of the actuators differ from those of the sensors. Data from a given sensor are communicated to a variety of actuators; this variety depends on the sensor (see Fig. 9.1). Communication is via limited capacity channels each connecting a sensor with a particular actuator. There is no information exchange within both the sensor and the actuator sets. So a given actuator has no direct knowledge about the actions of the other actuators. The channels are given. The objective is to stabilize the process.

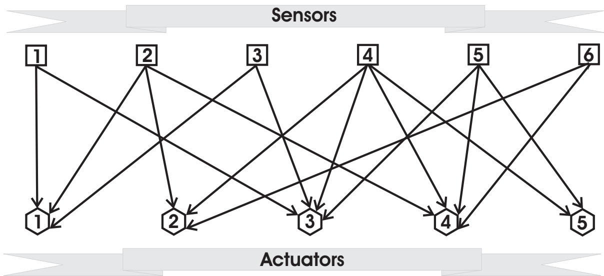


Fig. 9.1. Control over the simplest communication network.

In this example, we consider both perfect and delayed channels. For the unit time, the *perfect channel* correctly transmits messages e taken from the *channel alphabet* $\mathfrak{E}$; i.e., $s(t) = e(t-1)$, where e and s is the input and output of the channel, respectively. The d-*delayed channel* does the same with the delay d; i.e., $s(t) = e(t-d)$. Every channel is given by its alphabet and delay d [1] (see Fig. 9.2, where the channels from sensors 1 and 2 to actuators 3 and 4, respectively, are delayed.) The capacity of the channel equals $\mathfrak{c} = \log_2 |\mathfrak{E}|$ and is finite if $|\mathfrak{E}| < \infty$.

We assume that the sensor outputs are vectors from Euclidean spaces. They cannot be transmitted in full via finite capacity channels. To be sent, the measurements should be first converted into time sequences of symbols from the alphabets of the channels. This is done by special system components called the *coders*. Every sensor is equipped with its own coder. Based on the prior observations from this sensor, the coder generates messages dispatched over the outgoing channels. Each actuator produces its own control on the basis of the prior messages received over the incoming channels. The *control strategy* is the set of the encoding and control algorithms. It endows every sensor and actuator with an algorithm of producing messages to be

[1] The default delay value $d = 1$.

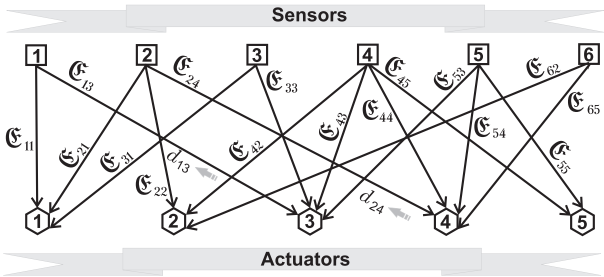


Fig. 9.2. Alphabets and delays of the channels.

emitted into the channels and controls, respectively. This strategy should be designed to achieve stability of the closed-loop system.

Remark 9.2.1. The important channel parameter (namely, the channel capacity) is determined by the number of elements in the channel alphabet, whereas the “physical nature” of these elements is inessential for the problems to be considered. So the channel will often be specified by indicating only its capacity or the above number.

Explanation 9.2.2. In Figs. 9.1 and 9.2 and throughout this chapter, the sensors and actuators are symbolized by squares and hexagons, respectively.

Remark 9.2.3. The d-delayed channel can be viewed as a recursive channel.

Indeed, define the current channel memory content (state) to be

$$n(t) := [e(t-1), \ldots, e(t-d+1)] \tag{9.2.1}$$

(see Fig. 9.3). Then the channel operation is in accordance with recursive equations

$$s(t+1) = \mathcal{S}[n(t), e(t)] \quad (:= e[t-d+1]),$$
$$n(t+1) = \mathcal{N}[n(t), e(t)] \quad (:= [e(t), \ldots, e(t-d+2)]). \tag{9.2.2}$$

The initial state is given by $n(0) = [\circledast, \ldots, \circledast]$, where $\circledast$ is the “void” symbol. So “nothing” arrives at the receiving end of the channel at times $t = 0, 1, \ldots, d-1$.

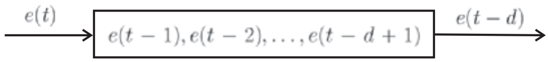


Fig. 9.3. Delayed channel.

Example 2: Network of perfect and delayed channels. This is an extension of the previous situation. Unlike Example 1,

- data are not directly sent from the sensors to actuators, but they pass through intermediate processors, which transform incoming data flows, distribute them over outgoing channels, and can be viewed as spatially distributed controllers;

- the information goes not necessarily only in one direction (from sensors to actuators through controllers): There may be backward data flows (from actuators to controllers and sensors, as well as from controllers to sensors) and an information exchange within the sets of sensors, controllers, and actuators.

Remark 9.2.4. In this case, the sensors generate outgoing messages on the basis of not only observations but also information received over the incoming channels.

Formally, let a directed finite graph be given. Its nodes represent spatially distributed *elements*, each endowed with a memory as well as with computing and communicating capabilities. Some elements (sensors) also have a sensing capability, whereas some other elements (actuators) are able to influence the unstable external process at hand. The elements different from sensors and actuators are interpreted as intermediate data processors or *controllers*. The edges of the graph represent communication channels transmitting data only in one direction. Each channel is either perfect or delayed; these channels are given. This situation is illustrated in Fig. 9.4, where specifications of the channels (capacity and delay) are omitted for simplicity.

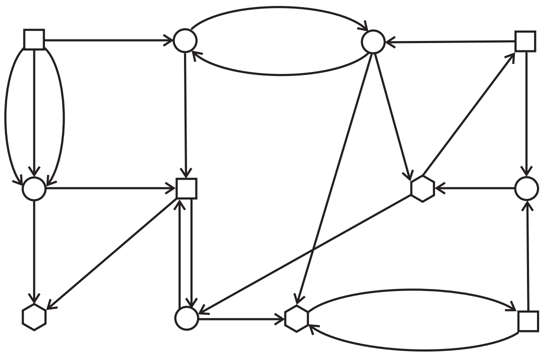


Fig. 9.4. Communication network.

Explanation 9.2.5. Intermediate controllers are symbolized by circles.

Remark 9.2.6. The situation where two elements are connected with several equally directed channels is not excluded.

The control strategy is given by endowing each element with an algorithm to generate its current output O from the prior inputs I. This strategy should be chosen so that the closed-loop system is stable.

Explanation 9.2.7. For all elements, the output O lists the messages dispatched over all channels outgoing from this element and also includes the control if the element is an actuator. The input I is constituted by the messages received over all incoming channels and, if the element is a sensor, the measurement.

Remark 9.2.8. For technical convenience, we suppose that O and I belong to sets whose cardinalities do not exceed that of the real numbers $2^{\aleph_0}$.

Example 3: Network with multiple access and broadcasting channels and interference. The *multiple access channel* [5] models the situation where multiple informants send messages to a common recipient by means of a common communication medium or facility. The recipient receives something that is different from the dispatched messages due to their collisions and interference: $s(t) = \mathfrak{R}[e_1(t-1), \ldots, e_p(t-1)]$, where e_i is the message from the ith informant (see Fig. 9.5a, where $p = 5$).

Explanation 9.2.9. Stars symbolize arbitrary elements; i.e., those not specified to be either sensors, controllers, or actuators.

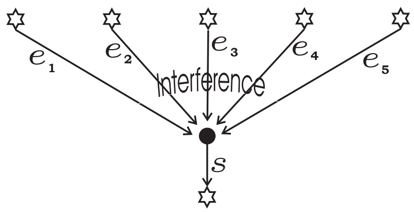


Fig. 9.5(a). Multiple access channel.

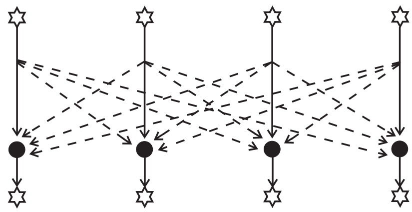


Fig. 9.5(b). Interfering messages.

The same model is relevant to the case where on the way from the informant to the recipient, the message is corrupted due to collisions with messages from other informants, which are addressed to other recipients. This situation is illustrated in Fig. 9.5b, where each informant has its own intended recipient.

An extreme extension of these examples is the deterministic network with interference [158, Sec. 1]. It can be viewed as the network from Example 2 (see Fig. 9.4), where, however, the output s_j of any channel is a deterministic function of not only the input e_j of this channel but also the inputs e_i of all other channels in the network:

$$s_j(t) = \mathfrak{R}_j[e_1(t-1), \ldots, e_K(t-1)]. \tag{9.2.3}$$

A particular instance of the multiple access channel is the *additive channel* [16, 31, 207]: e_i are elements of a commutative group (typically $\mathbb{R}^s$ or the group of integers $\bmod m$) and $s(t) = e_1(t-1) + \cdots + e_K(t-1)$. Another instance is the *binary switching* multiple access channel [212]: Two senders dispatch binary data $e_i = 0, 1$, whereas $s(t) = \frac{e_1(t-1)}{e_2(t-1)}$ (and $\frac{e}{0} := \circledast \, \forall e$).

The *broadcasting channel* delivers a common message to several recipients $s_j(t) = e(t-1), j \in J$ (see Fig. 9.6a). The model (9.2.3) captures the effect of broadcasting since several functions $\mathfrak{R}_j(\cdot)$ may be identical and depend on not all but a single input e_i. This model also accounts for broadcasting with interference.

The formal setup of the stabilization problem is the generalization of that from Example 2, where the network is not assumed to be composed of perfect and delayed

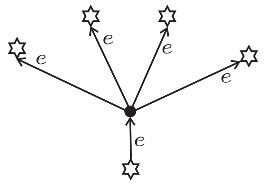


Fig. 9.6(a). Broadcasting channel.

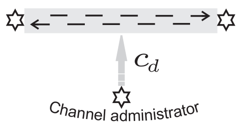


Fig. 9.6(b). Two-way channel.

channels any longer but, instead, is deterministic with interference (9.2.3). A further generalization concerns such a network with after-effects

$$s_j(t) = \mathcal{R}_j \left[e_1(\theta)|_{\theta=t-d}^{t-1}, \ldots, e_K(\theta)|_{\theta=t-d}^{t-1} \right].$$

Example 4: Network with switching channels. The *two-way switching channel* is able to transmit messages $e \in \mathfrak{E}$ between two elements in both directions. However, only one direction may be activated at a given time. This direction is associated with a control variable $c_d = \pm$ and dynamically chosen by a certain element called the channel *administrator* (see Fig. 9.6b). The value of c_d is generated from the knowledge currently available to the administrator and is a part of its output O.

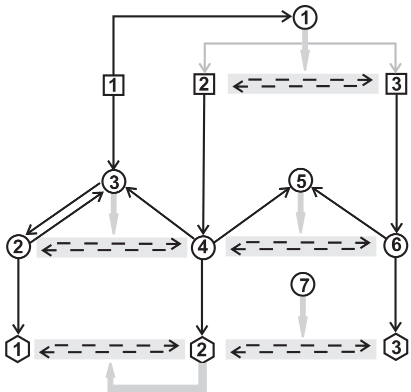


Fig. 9.7. Network with two-way channels.

A network with two-way channels is illustrated in Fig. 9.7. Controllers 1, 5, and 7 are pure administrators, whereas controller 3 and actuator 2 are also engaged in primal data processing. The two-way channel between actuators 1 and 2 is administered by one of the connected elements. All other such channels are administered by side elements. Controller 5 judges between requests for service from the end elements of the channel.[2] Controller 3 does the same using a side information received

[2]If the capacity of the channel is large enough, the end element may append a "justification" to its request.

from sensor 1. This controller may notify controller 2 about the state of the two-way channel. Controller 1 uses only side information. Sensors 2 and 3 are able to detect the current state (direction) of the channel. Controller 7 administers the channel in accordance with a given program in the open-loop fashion.

The two-way channel is a particular sample of the *switching channel*. Such a channel serves a finite set of possible informant–recipient pairs of elements. At any time, only one pair selected by the administrator is in fact connected. In other words, the output of the administrator includes the control variable whose value determines which possible pair is served by the channel. Fig. 9.8a illustrates a switching channel with disjoint sets of informants and recipients in which any informant can be connected with any recipient. In general, these two features may fail to hold. For example, for the two-way channel, the sets of informants and recipients are the same $\{a, b\}$, and the possible connections $a \mapsto b, b \mapsto a$ do not include $a \mapsto a, b \mapsto b$.

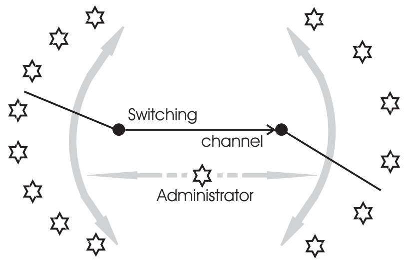


Fig. 9.8(a). Switching channel.

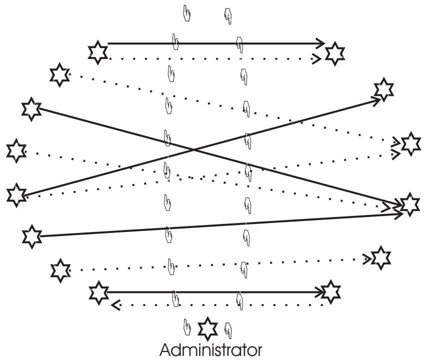


Fig. 9.8(b). Multipath switching channel.

A more general situation holds for the *multipath switching channel*. Suppose that several “possible” communication channels (e.g., channels from Fig. 9.4) share a common facility or medium so that only a special selection of them can be activated at a given time. (The channels outside the selected group are out of service.) Not every selection is possible; the current selection from the set of possible ones is taken by the administrator. A particular example concerns the case where there are p possible channels of equal capacity $\mathfrak{c}$, and at any time arbitrary $q < p$ channels can be activated. This may be due to the fact that these possible (or virtual) channels use a common “physical” channel with capacity $\mathfrak{c}\, q$, which is thus able to accommodate simultaneously only q data streams each of the rate $\mathfrak{c}$.

Fig. 9.8b illustrates a multipath switching channel with two possible selections of active channels (indicated with different arrows). Note that both top and bottom elements are connected under both selections, although the capacities of the corresponding channels may depend on the selection.

The multipath switching channel can be described by equations of the form (9.2.3),[3] where s_j and e_i range over the outputs and inputs, respectively, of all "primal" channels concerned by the multipath switching one. It is also supposed that the control signals generated by the administrators are among the e_i's, which formally may be arranged by interpreting these signals as input messages for special "network control" channels.

Example 5: Multimode channels and networks.[3] The *multimode* (multistate) *channel* is that operating in several modes, symbolized by μ. The mode determines channel characteristics (e.g., the capacity, level of data distortion, etc.) and is chosen dynamically by an administrator. For example, there may be a choice between larger channel alphabets with more message corruption during transmission and smaller alphabets with lesser corruption. A more general situation is where a common mode simultaneously determines the characteristics of several channels. An example is depicted in Fig. 9.9a, where the combination $C := (\mathfrak{c}_1, \mathfrak{c}_2, \mathfrak{c}_3, \mathfrak{c}_4, \mathfrak{c}_5)$ of the capacities of five channels depends on the mode $C = C(\mu)$ and there is a choice between several possible combinations.

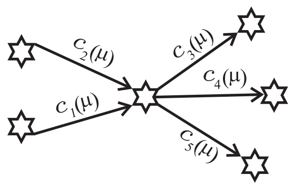


Fig. 9.9(a). Multimode channels.

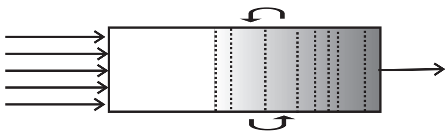


Fig. 9.9(b). Data storage with sorting.

A further generalization is a *multimode network*. In such a network, the mode determines parameters of not only separate channels but also the entire network, e.g., its topology (routes of data transfer) or types of channels (perfect, switched, with interference, etc.). Formally, the situation can be viewed as if each mode is associated with its own network. The networks associated with different modes have a common set of elements (nodes) but are not correlated in any other sense. There may be several current administrators (if, e.g., the mode is composed by several independent submodes each concerning its own part of the network). It is supposed that the mode activated at time $t-1$ is in service until t. (We recall that by assumption, data transmissions across the channels are completed for the unit time.) If such a network is deterministic, data transfer within it can still be described by equations of the form (9.2.3).

Example 6: Types of elements. The problem to be considered is to ensure stability of the closed-loop system by endowing each element in the network with a proper algorithm of producing its outputs O from the prior inputs I:

$$O(t) = \mathcal{O}[I(0), \dots, I(t), t]. \tag{9.2.4}$$

[3]To simplify matters, we do not discuss delay effects in this example.

The *type* is given by specifying the memory and computing capabilities of the element. Formally, this looks like putting forward certain requirements to the algorithm (9.2.4). The general problem statement to be adopted allows for confining each controller to a certain type. Now we give several illustrating examples of types.

★ *An element with unlimited memory and computational power.* No limitations on the right-hand side of (9.2.4) are imposed. This model is apposite if the restrictions on the memory and computational power of the element are neglectable.
★ *Memoryless element*: $O(t) = \mathcal{O}[I(t), t]$, where $\mathcal{O}[\cdot]$ is taken from a given set $\mathfrak{O}$.
★ *Memoryless stationary element*: $O(t) = \mathcal{O}[I(t)]$.
★ *Element with data storage of a limited size.* The state of the data storage device is represented by a symbol m from a given *memory alphabet* $\mathfrak{M}$. This symbol also determines the memory content. Operation of the element is in accordance with equations of the form

$$O(t) = \mathcal{O}[m(t), I(t), t], \qquad m(t+1) = \mathcal{M}[m(t), I(t), t]. \tag{9.2.5}$$

In this example, the computational power (the variety of executable operations, the number of basic operations per unit time, the size of the processor registers involved, the speed of the memory access, etc.) is not limited.
★ *Element with a limited memory and computational power.* The input is first converted into the form acceptable by the processor $i(t) := \mathcal{C}[I(t)]$. (For example, numerical data are quantified to match the numeral system adopted by the processor.) The conversion rule $\mathcal{C}[\cdot]$ is given. The processing algorithm is of the form

$$O(t) = \mathcal{O}[m(t), i(t), t], \qquad m(t+1) = \mathcal{M}[m(t), i(t), t]. \tag{9.2.6}$$

The choice of this algorithm is limited by the requirement

$$\big(\mathcal{O}[m, i, t], \mathcal{M}[m, i, t]\big) \in \mathfrak{R}(m, i) \qquad \forall m, i, t. \tag{9.2.7}$$

Here $\mathfrak{R}(m, i)$ is the set of all possible results that may be obtained by means of the processing unit at hand for the unit time, provided that its state is m and the input equals i. This set is determined by, e.g., the variety of basic commands (including memory access ones) and their sequences executable for the unit time, the sizes of the processor registers, etc.
★ *Multimode element.* The element may operate in several modes, which determine its computing capabilities and the size and organization of the memory. The mode, denoted by a symbol $p \in \mathfrak{P}$, cannot be changed dynamically and thus is a time-invariant design parameter. It should be chosen to achieve a certain objective. An example of such a situation is the case where the arithmetic used in the processor is not given a priori, and there is a choice between more and less memory consuming and precise arithmetics.
Within each mode, let the element operate like in the previous example. Then the processing algorithm is of the form

$$O(t) = \mathcal{O}[m(t), i(t), p, t], \; m(t+1) = \mathcal{M}[m(t), i(t), p, t], \; i(t) = \mathfrak{C}[I(t), p],$$
$$m(t) \in \mathfrak{M}(p), \quad p \in \mathfrak{P}, \quad \big(\mathcal{O}[m, i, p, t], \mathcal{M}[m, i, p, t]\big) \in \mathfrak{R}(m, i, p). \tag{9.2.8}$$

★ *Semirational element.* This is a time-invariant element of the form (9.2.5), where I, m, and O belong to products of certain Euclidean spaces and finite sets, the functions do not in fact depend on t, and their "finite-state" and "Euclidean" components are semialgebraic and semirational, respectively (see Definitions 3.4.11 and 3.4.12 on p. 47).

Now we revert to the general discussion. We assume that the *type* of the element is formally given by, first, the memory alphabet $\mathfrak{M}$ of the element and, second, the set $\mathfrak{A}$ of admissible data processing algorithms $\mathcal{A}$. In doing so, we suppose that the element acts in accordance with equations of the form (9.2.5), and so

$$\mathcal{A} \equiv [\mathcal{O}(\cdot), \mathcal{M}(\cdot), m^0], \tag{9.2.9}$$

where m^0 is the initial memory state. (Thus $\mathfrak{A}$ gives the set $\mathfrak{M}^0$ of admissible initial states.) It is clear that $\mathcal{A}$ uniquely determines the operation of the element provided that the time sequence of inputs is given.

Remark 9.2.10. For technical convenience, we suppose that the cardinality of the memory alphabet $\mathfrak{M}$ does not exceed that of the real numbers $2^{\aleph_0}$.

Now we discuss how the above examples fit this formalism.

In the case (9.2.4), the memory alphabet is constituted by all finite sequences of inputs $S = [I_0, I_1, \ldots, I_k]$, including the "empty" one $\circledast$. The set $\mathfrak{A}$ of admissible algorithms consists of all $\mathcal{A}$ for which $\circledast$ is the initial state and the function $\mathcal{M}(\cdot)$ of S, I, t acts by adding I to S from the right and by possible dropout of several entries of the resultant sequence. If the set $\mathfrak{A}$ is confined by the requirement that any entry outside the group of k concluding ones should be dropped necessarily, we get the description (type) of elements with the memory of length k.

Remark 9.2.11. It is easy to see that elements with unlimited memory and computational power (9.2.4) can also be characterized as elements (9.2.5) for which the memory alphabet has the maximal possible cardinality (that of the real numbers) and there are no restrictions on the right-hand sides of equations from (9.2.5).

For memoryless elements, the memory alphabet has size 1. For stationary such elements, admissible algorithms are given by time-invariant functions.

In the case (9.2.6), the admissible algorithms are associated with functions $\mathcal{O}(\cdot), \mathcal{M}(\cdot)$ that satisfy (9.2.7) and can be represented in the form

$$\mathcal{O}[m, I, t] = \mathcal{O}[m, \mathcal{C}(I), t], \quad \mathcal{M}[m, I, t] = \mathcal{M}[m, \mathcal{C}(I), t] \quad \forall m, I, t.$$

In the case of multimode element (9.2.8), the formal memory alphabet is defined to be $\mathfrak{M}_f := \{m_f = (m, p) : p \in \mathfrak{P}, m \in \mathfrak{M}(p)\}$. Thus m should be replaced by m_f in (9.2.5). The function $m_f' = \mathcal{M}(m, p, I, t)$ from (9.2.5) should not alter p; i.e., $\mathcal{M}(m, p, I, t) = [\mathcal{M}_*(m, p, I, t), p]$. In (9.2.9), the initial state should be taken from the set $\{m_f : p \in \mathfrak{P}, m \in \mathfrak{M}^0(p)\}$, where $\mathfrak{M}^0(p)$ is a (given) set of initial states admissible in the mode p. The other particulars of embedding this case into the formalism at hand are like in the previous example.

In the case of the semirational element, the memory alphabet is not prescribed and may equal any set of the form $\mathbb{R}^s \times \mathfrak{G}$, where $k := |\mathfrak{G}| < \infty$. Since the "physical nature" of the elements $g \in \mathfrak{G}$ is of no importance, $\mathfrak{G}$ can be replaced by $[1 : k]$ without any loss of generality. Then the semirational element can be regarded as multimode element with the modes $p = (s, k)$ and $\mathfrak{M}(p) = \mathbb{R}^s \times [1 : k]$. (The case $s = 0$ is not excluded, $\mathbb{R}^0 := \{0 \in \mathbb{R}\}$.) The functions in (9.2.5) are limited to those whose "finite-state" and "Euclidean" components are semialgebraic and semirational, respectively, within any mode.

Explanation 9.2.12. Certain requirements will be imposed on the types of the elements (see Assumptions 9.3.4, 9.4.20, and 9.4.22 on pp. 285, 298, and 299). Briefly, the network should be stationary and controllers should be able to reset their memories to the initial state (after a proper transient). At the same time, it will be assumed that the sensors and actuators have unlimited memories and computational powers.

Remark 9.2.13. The set of data processing algorithms $\mathfrak{A}$ (determining the type) may contain only one element $\mathcal{A}$.

Assigning such a type to a controller is identical to giving the controller. So in the discussed problem statement, some intermediate data processors (controllers) may be given a priori, provided they satisfy the requirements mentioned in Explanation 9.2.12. We conclude the section with two examples of such processors.

Autonomous stationary automaton, i.e., a time-invariant element with no input:

$$O(t) = \mathcal{O}[m(t)], \qquad m(t+1) = \mathcal{M}[m(t)].$$

The output sequence $\{O(t)\}$ becomes periodic since some time instant; any periodic sequence can be generated in this way. This element fits to model compulsory periodic changes in the network, e.g., a switching channel with prescribed periodic switching program (O determines the position of the switch) or a channel with a periodically varying alphabet (O determines the channel mode, which in turn determines the alphabet).

Explanation 9.2.14. To meet the requirements mentioned in Explanation 9.2.12, the initial memory state m^0 should be a recurrent element with respect to the map $\mathcal{M}(\cdot)$, i.e., such that $\mathcal{M}^p(m^0) = m^0$ for some $p = 1, 2, \ldots$. Then the reset to the initial state occurs automatically at any time that is a multiple of p.

Data storage with sorting. Consider a data storage (buffer) of a finite size B (see Fig. 9.9b on p. 278). It collects packages $e_1, \ldots, e_k$ constantly arriving over k channels. Any package e_i is characterized by the size $b(e_i)$ (typically the number of bits) and *priority* $\sharp(e_i) = 1, 2, \ldots$. Within the storage, packages are organized in numerical order (a sequence $P = \{p_1, \ldots, p_s\}$). The new data $e_1(t), \ldots, e_k(t)$ are added to P from the left if the buffer size is not exceeded $\sum_i b[e_i(t)] + \sum_j b[p_j] \leq B$; otherwise these data are dropped. Packages leave the storage in the order $p_s \mapsto p_{s-1} \mapsto \cdots \mapsto p_1$ at the constant data rate $\mathfrak{c}$. Thus $p_s/\mathfrak{c}$ units of time are required for the package p_s to be completely dispatched. Within the storage, packages are sorted to move those with higher priorities closer to the exit. The sorting algorithm compares a pair of adjacent packages and swaps them if the priority of the first package exceeds that of the second one. The pair is chosen via successive one-step shift along P from the left to the right until p_{s-1} is involved, then from the right to left until p_1 is concerned, and so on. The sort rate is r steps per unit time. We omit rather apparent technical consideration showing that this device can be formally described as a stationary automaton with inputs $E = (e_1, \dots, e_k)$.

9.3 General Model of a Deterministic Network

Now we introduce a general model of a deterministic network. It encompasses all examples considered in the previous section and will be employed further to pose the stabilization problem and state the main result.

9.3.1 How the General Model Arises from a Particular Example of a Network

We start with demonstrating this in order to promote comprehension of the model. Let us revert to Example 2 (starting on p. 273); i.e., consider a network of perfect and delayed channels. It is given by a finite directed graph whose nodes $h \in \mathfrak{H}$ and directed edges $J \in \mathfrak{J}$ are associated with elements and communication channels, respectively (see Fig. 9.4 on p. 274). Within the set of elements, the subsets $\mathfrak{H}_s$ and $\mathfrak{H}_a$ of sensors and actuators, respectively, are given.

Every element h receives messages from some other elements of the network. Taken as a whole, these messages constitute the *inner* (with respect to the network) *input* I_h of this element (see Fig. 9.10). Any sensor also has access to an outer input: the measurement $y_h \in \mathbb{R}^{n_{y,h}}$. Furthermore, any element emits messages to some other elements. These messages constitute the *inner output* O_h. Any actuator also produces an *outer output*: the control $u_h \in \mathbb{R}^{n_{u,h}}$.

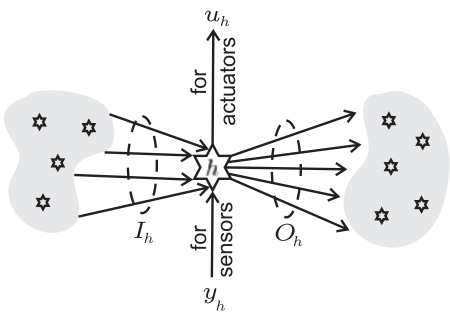


Fig. 9.10. Operation of an element.

Now we introduce four united ensembles of the inner and outer inputs and outputs, respectively:

$$\boldsymbol{O} := \{O_h\}_{h \in \mathfrak{H}}, \quad \boldsymbol{I} := \{I_h\}_{h \in \mathfrak{H}}, \quad \boldsymbol{Y} := \{y_h\}_{h \in \mathfrak{H}_s}, \quad \boldsymbol{U} := \{u_h\}_{h \in \mathfrak{H}_a}.$$

The roles of the network (communication facility) and control strategy (endowing of every element with a data processing algorithm) are as follows:

$$\boldsymbol{O} \xrightarrow{\text{network}} \boldsymbol{I}, \qquad [\boldsymbol{I}, \boldsymbol{Y}] \xrightarrow{\text{control strategy}} [\boldsymbol{O}, \boldsymbol{U}]. \tag{9.3.1}$$

The channels J are represented in the form (9.2.2) (on p. 273); i.e., as recursive channels with memories $n_J \in \mathfrak{N}_J$. We suppose that the type of every element h is specified; i.e., its memory alphabet $\mathfrak{M}_h$ with elements m_h and the set $\mathfrak{A}_h$ of admissible data processing algorithms $\mathcal{A}_h$ are given. Now we form the compound ensembles characterizing the memories of all channels and elements, and the control strategy, respectively:

$$\boldsymbol{N} := \{n_J\}_{J \in \mathfrak{J}}, \quad \boldsymbol{M} := \{m_h\}_{h \in \mathfrak{H}}, \quad \text{(control strategy)} \sim \boldsymbol{\mathcal{A}} := \{\mathcal{A}_h\}_{h \in \mathfrak{H}}.$$

Then (9.3.1) can be specified and complemented as follows:

$$\boldsymbol{O} \xrightarrow[\boldsymbol{N}]{\text{network}} \boldsymbol{I}, \quad [\boldsymbol{I}, \boldsymbol{Y}] \xrightarrow[\boldsymbol{M}]{\boldsymbol{\mathcal{A}}} [\boldsymbol{O}, \boldsymbol{U}], \quad \boldsymbol{O} \xrightarrow[\boldsymbol{N}]{\text{network}} \boldsymbol{N}^+, \quad [\boldsymbol{I}, \boldsymbol{Y}] \xrightarrow[\boldsymbol{M}]{\boldsymbol{\mathcal{A}}} \boldsymbol{M}^+,$$

where $^+$ symbolizes the value at the next time instant. Thanks to the deterministic nature of the network, these relations can be specified in the form

$$\begin{aligned} &\boldsymbol{I}(t{+}1) = \mathcal{I}[\boldsymbol{O}(t), \boldsymbol{N}(t)],\ \boldsymbol{N}(t{+}1) = \mathcal{N}[\boldsymbol{O}(t), \boldsymbol{N}(t)],\ \boldsymbol{I}(0) = \boldsymbol{I}_0,\ \boldsymbol{N}(0) = \boldsymbol{N}_0; \\ &\quad \boldsymbol{O}(t) = \mathcal{O}[\boldsymbol{I}(t), \boldsymbol{Y}(t), \boldsymbol{M}(t), \boldsymbol{\mathcal{A}}, t], \\ &\quad \boldsymbol{U}(t) = \mathcal{U}[\boldsymbol{I}(t), \boldsymbol{Y}(t), \boldsymbol{M}(t), \boldsymbol{\mathcal{A}}, t], \quad \boldsymbol{M}(0) = \mathcal{M}^0(\boldsymbol{\mathcal{A}}), \quad \boldsymbol{\mathcal{A}} \in \mathfrak{A}. \\ &\boldsymbol{M}(t+1) = \mathcal{M}[\boldsymbol{I}(t), \boldsymbol{Y}(t), \boldsymbol{M}(t), \boldsymbol{\mathcal{A}}, t], \end{aligned} \tag{9.3.2}$$

Here the penultimate relation results from (9.2.9) (on p. 280).

Roughly speaking, the controlled network will be further regarded as a facility transforming the time sequence of outer inputs $\boldsymbol{Y}(0), \boldsymbol{Y}(1), \boldsymbol{Y}(2), \ldots$ into the time sequence of outer outputs $\boldsymbol{U}(0), \boldsymbol{U}(1), \boldsymbol{U}(2), \ldots$ in accordance with equations of the form (9.3.2), where $\boldsymbol{\mathcal{A}}$ is a control parameter and all functions denoted by capital script letters are given. The parameter $\boldsymbol{\mathcal{A}}$ should be chosen from a given set $\mathfrak{A}$ to achieve a certain objective, e.g., to stabilize an outer process. Formally, such a facility is given by the above functions, the set $\mathfrak{A}$, the initial states $\boldsymbol{I}_0$ and $\boldsymbol{N}_0$, and the sets within which the variables from (9.3.2) range.

Remark 9.3.1. This point of view means that what is called the "network" describes not only the way to transmit information but also restrictions on the memory and computing capabilities of the elements.

Explanation 9.3.2. The relations from the first row in (9.3.2) describe how data are transmitted within the network, and they are called *communication equations*. The rest of the equations describe how data are processed by the elements and are called *data processing equations*.

9.3.2 Equations (9.3.2) for the Considered Example

The major purpose of this subsection is to prove formally that the network from the example at hand is described by equations of the form (9.3.2) indeed. The reader that does not need such a proof may skip this subsection since the particulars of the corresponding equations will be of minor importance for the subsequent discussion.

For any element (graph node) h, we denote by $\mathbf{IN}(h)$ and $\mathbf{OUT}(h)$ the sets of channels (graph edges) incoming to and outgoing from h, respectively. For any channel J, the symbols $d_J, \mathfrak{E}_J$, and $e_J(t)$ denote the channel delay, alphabet, and the message emitted into it at time t, respectively. With regard to (9.2.1) (on p. 273), the variables in (9.3.2) are of the form

$$\boldsymbol{O}(t) = \{O_h(t)\}_{h\in\mathfrak{H}}, \quad O_h(t) = \{e_J(t)\}_{J\in\mathbf{OUT}(h)};$$
$$\boldsymbol{I}(t) := \{I_h(t)\}_{h\in\mathfrak{H}}, \quad I_h(t) = \{e_J(t-d_J)\}_{J\in\mathbf{IN}(h)};$$
$$\boldsymbol{N}(t) = \{n_J(t)\}_{J\in\mathfrak{J}}, \quad n_J(t) = [e_J(t-1),\ldots,e_J(t-d_J+1)];$$
$$\boldsymbol{M}(t) := \{m_h(t)\}_{h\in\mathfrak{H}}, \quad \boldsymbol{Y}(t) := \{y_h(t)\}_{h\in\mathfrak{H}_s}, \quad \boldsymbol{U}(t) := \{u_h(t)\}_{h\in\mathfrak{H}_a}.$$

Explanation 9.3.3. If $d_J = 1$, the channel memory is void: $n_J = \circledast$. We also assume that $e_J(t) = \circledast$ for $t < 0$.

Let $\mathbf{LAST}(S)$ denote the last (right) entry of the sequence S, and let $\mathbf{DROPLAST}(S)$ denote the result of dropping this entry. Note that for every channel, $e_J = e_J(t)$ is a part (component) of $\boldsymbol{O} = \boldsymbol{O}(t) : e_J = \mathcal{P}_J[\boldsymbol{O}]$. For any h, the algorithm $\mathcal{A} = \mathcal{A}_h$ has the form (9.2.9) (on p. 280), where $\mathcal{O}(\cdot)$ encompasses both inner and outer output of the element. In other words,

$$\mathcal{A}_h \equiv \begin{cases} [\mathcal{O}_h(\cdot), \mathcal{U}_h(\cdot), \mathcal{M}_h(\cdot), m_h^0] & \text{if } h \text{ is an actuator} \\ [\mathcal{O}_h(\cdot), \mathcal{M}_h(\cdot), m_h^0] & \text{otherwise} \end{cases}. \tag{9.3.3}$$

Particular equations from (9.3.2) disintegrate into systems of equations.

$$\boldsymbol{I}(t+1) = \mathcal{I}[\boldsymbol{O}(t), \boldsymbol{N}(t)] \Longleftrightarrow I_h(t+1) = \{\mathcal{I}_J[\boldsymbol{O}(t), \boldsymbol{N}(t)]\}_{J\in\mathbf{IN}(h)} \quad \forall h \in \mathfrak{H},$$
$$\text{where} \quad \mathcal{I}_J[\boldsymbol{O}, \boldsymbol{N}] := \begin{cases} \mathbf{LAST}[n_J] & \text{if } d_J > 1 \\ \mathcal{P}_J[\boldsymbol{O}] & \text{otherwise} \end{cases};$$
$$\boldsymbol{N}(t+1) = \mathcal{N}[\boldsymbol{O}(t), \boldsymbol{N}(t)] \Longleftrightarrow n_J(t+1) = \mathcal{N}_J[\boldsymbol{O}(t), \boldsymbol{N}(t)] \quad \forall J,$$
$$\text{where} \quad \mathcal{N}_J[\boldsymbol{O}, \boldsymbol{N}] := \begin{cases} \mathbf{DROPLAST}[\mathcal{P}_J[\boldsymbol{O}], n_J] & \text{if } d_J > 1 \\ \circledast & \text{otherwise} \end{cases};$$
$$\boldsymbol{O}(t) = \mathcal{O}[\boldsymbol{I}(t), \boldsymbol{Y}(t), \boldsymbol{M}(t), \boldsymbol{\mathcal{A}}, t] \Longleftrightarrow O_h(t) = \mathcal{O}_{[h]}[I_h(t), \boldsymbol{Y}(t), m_h(t), \mathcal{A}_h, t]$$
$$\overset{(9.2.5),\,(9.3.3)}{=\!=\!=} \begin{cases} \mathcal{O}_h[m_h(t), I_h(t), y_h(t), t] & \text{if } h \text{ is a sensor} \\ \mathcal{O}_h[m_h(t), I_h(t), t] & \text{otherwise} \end{cases} \quad \forall h \in \mathfrak{H};$$

$$\boldsymbol{U}(t) = \mathcal{U}[\boldsymbol{I}(t), \boldsymbol{Y}(t), \boldsymbol{M}(t), \boldsymbol{\mathcal{A}}, t] \Longleftrightarrow u_h(t) = \mathcal{U}_{[h]}[I_h(t), \boldsymbol{Y}(t), m_h(t), \mathcal{A}_h, t]$$
$$\underset{\overline{\overline{}}}{\text{(9.2.5), (9.3.3)}} \begin{cases} \mathcal{U}_h[m_h(t), I_h(t), y_h(t), t] & \text{if } h \text{ is a sensor} \\ \mathcal{U}_h[m_h(t), I_h(t), t] & \text{otherwise} \end{cases} \quad \forall h \in \mathfrak{H}_a;$$
$$\boldsymbol{M}(t+1) = \mathcal{M}[\boldsymbol{I}(t), \boldsymbol{Y}(t), \boldsymbol{M}(t), \boldsymbol{\mathcal{A}}, t] \Longleftrightarrow m_h(t+1)$$
$$= \mathcal{M}_{[h]}[I_h(t), \boldsymbol{Y}(t), m_h(t), \mathcal{A}_h, t]$$
$$\underset{\overline{\overline{}}}{\text{(9.2.5), (9.3.3)}} \begin{cases} \mathcal{M}_h[m_h(t), I_h(t), y_h(t), t] & \text{if } h \text{ is a sensor,} \\ \mathcal{M}_h[m_h(t), I_h(t), t] & \text{otherwise} \end{cases} \quad \forall h \in \mathfrak{H}.$$

Furthermore in (9.3.2), $\boldsymbol{N}_0 = \{n_J^0\}_{J\in\mathfrak{J}}$, where n_J^0 is the void sequence $(\circledast, \dots, \circledast)$ of the proper length, $\boldsymbol{I}_0 = \{I_{h,0}\}_{h\in\mathfrak{H}}, I_{h,0} = \{\circledast\}_{J\in\mathbf{IN}(h)}$, $\boldsymbol{\mathcal{A}} \in \mathfrak{A} \Leftrightarrow \mathcal{A}_h \in \mathfrak{A}_h \; \forall h$, and $\{m_h\}_{h\in\mathfrak{H}} = \mathcal{M}_0(\boldsymbol{\mathcal{A}}) \Leftrightarrow m_h = m_h^0 \; \forall h$, where m_h^0 is the last component of $\mathcal{A}_h$ from (9.3.3).

9.3.3 General Model of the Communication Network

To obtain substantial results, we need to focus attention on a particular case of the extremely general model (9.3.2). First, we cannot ignore the difference among sensors, actuators, and controllers. Second, we shall basically deal with sensors and actuators with unlimited memories and computational powers. Formally, this results in viewing the network as a composition of sensors, actuators, and the "rest of the network" called its *interior* (see Fig. 9.11). We assume that these components process data separately, and the network outer input $\boldsymbol{Y}$ and output $\boldsymbol{U}$ from (9.3.2) are directly accessed and generated by the sensors and actuators, respectively. This imposes some structure on the processing equations from (9.3.2), whereas the communication ones may be arbitrary. All causal algorithms of data processing at the sensors and actuators are adopted, whereas restrictions may be imposed on the algorithm for the interior part of the network.

Formally, we shall consider networks described by equations of the form (9.3.2) and satisfying the following.

Assumption 9.3.4. *The following statements hold:*

A) *Two finite sets $\mathfrak{H}_s$ and $\mathfrak{H}_a$ are given;*
B) *To any $h \in \mathfrak{H}_+ := \mathfrak{H}_s \cup \mathfrak{H}_a$, assigned are the memory, input, and output alphabets $\mathfrak{M}_h, \mathfrak{I}_h$, and $\mathfrak{O}_h$ with elements denoted by m_h, I_h, and O_h, respectively;*
C) *Any $h \in \mathfrak{H}_a$ is also endowed with the outer output alphabet with elements u_h;*
D) *The network outer input $\boldsymbol{Y}$ consists of finitely many parts $\boldsymbol{Y} = \{\mathfrak{y}_\sigma\}_{\sigma\in\mathfrak{S}}$; each part $\mathfrak{y}_\sigma$ is generated by an independent* source *as an element of a given* source alphabet *and is distributed among the nodes $h \in \mathfrak{H}_s$ in accordance with a given* data injection scheme (prefix) $\{\mathfrak{H}_{\leftarrow\sigma}\}_{\sigma\in\mathfrak{S}}$*: The message $\mathfrak{y}_\sigma$ is immediately accessible at the nodes $h \in \mathfrak{H}_{\leftarrow\sigma} \subset \mathfrak{H}_s$ or, equivalently, the outer input of any node $h \in \mathfrak{H}_s$ is given by*

$$y_h(t) = \{\mathfrak{y}_\sigma(t)\}_{\sigma:h\in\mathfrak{H}_{\leftarrow\sigma}};$$

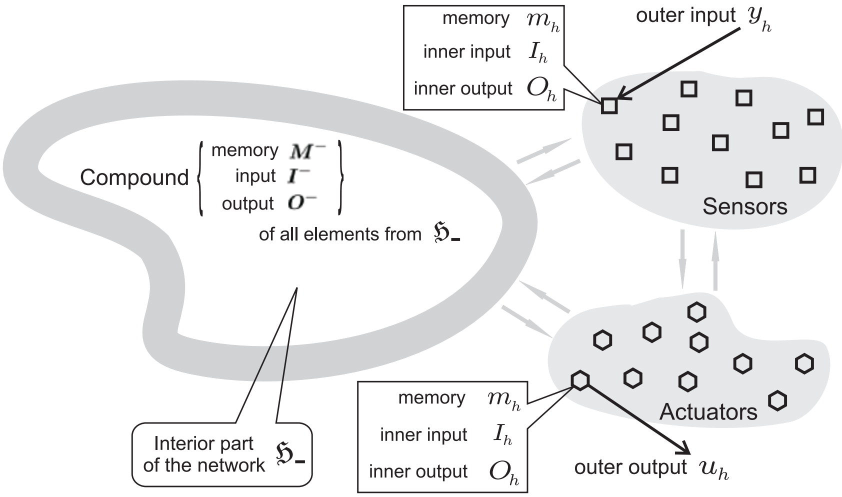


Fig. 9.11. Components of the network.

E) *The variables* $\boldsymbol{O}, \boldsymbol{I}, \boldsymbol{M}$*, and* $\boldsymbol{U}$ *from* (9.3.2) *have the special structure:*

$$\boldsymbol{O} = \Big[\{O_h\}_{h\in\mathfrak{H}_+}, \boldsymbol{O}^-\Big], \qquad \boldsymbol{I} = \Big[\{I_h\}_{h\in\mathfrak{H}_+}, \boldsymbol{I}^-\Big],$$
$$\boldsymbol{M} = \Big[\{m_h\}_{h\in\mathfrak{H}_+}, \boldsymbol{M}^-\Big], \qquad \boldsymbol{U} = \{u_h\}_{h\in\mathfrak{H}_a},$$

where $\boldsymbol{O}^-, \boldsymbol{I}^-$*, and* $\boldsymbol{M}^-$ *belong to given sets* $\mathfrak{O}^-, \mathfrak{I}^-$*, and* $\mathfrak{M}^-$*, respectively;*

F) *The data processing part of* (9.3.2) *disintegrates into separate equations concerning each sensor, each actuator, and the rest of the network, respectively:*

$$O_h(t) = \begin{cases} \mathcal{O}_h[I_h(t), y_h(t), m_h(t), \boldsymbol{\mathcal{A}}, t] & \text{if } h \in \mathfrak{H}_s \\ \mathcal{O}_h[I_h(t), m_h(t), \boldsymbol{\mathcal{A}}, t] & \text{if } h \in \mathfrak{H}_a \setminus \mathfrak{H}_s \end{cases};$$
$$m_h(t+1) = \begin{cases} \mathcal{M}_h[I_h(t), y_h(t), m_h(t), \boldsymbol{\mathcal{A}}, t] & \text{if } h \in \mathfrak{H}_s \\ \mathcal{M}_h[I_h(t), m_h(t), \boldsymbol{\mathcal{A}}, t] & \text{if } h \in \mathfrak{H}_a \setminus \mathfrak{H}_s \end{cases};$$
$$m_h(0) = \mathcal{M}_h^0(\boldsymbol{\mathcal{A}}) \qquad \forall h \in \mathfrak{H}_+;$$
$$u_h(t) = \begin{cases} \mathcal{U}_h[I_h(t), m_h(t), \boldsymbol{\mathcal{A}}, t] & \text{if } h \in \mathfrak{H}_a \setminus \mathfrak{H}_s \\ \mathcal{U}_h[I_h(t), y_h(t), m_h(t), \boldsymbol{\mathcal{A}}, t] & \text{if } \mathfrak{H}_a \cap \mathfrak{H}_s \end{cases}; \quad (9.3.4)$$

$$\boldsymbol{O}^-(t) = \mathcal{O}^-[\boldsymbol{I}^-(t), \boldsymbol{M}^-(t), \boldsymbol{\mathcal{A}}, t],$$
$$\boldsymbol{M}^-(t+1) = \mathcal{M}^-[\boldsymbol{I}^-(t), \boldsymbol{M}^-(t), \boldsymbol{\mathcal{A}}, t], \quad \boldsymbol{M}^-(0) = \mathcal{M}_0^-(\boldsymbol{\mathcal{A}}); \quad (9.3.5)$$

G) *The dependence of the right-hand sides of the equations from* (9.3.4) *on the control parameter* $\boldsymbol{\mathcal{A}} \in \mathfrak{A}$ *does not imply any restrictions on them: As* $\boldsymbol{\mathcal{A}}$ *runs over*

$\mathfrak{A}$, *these sides independently range over all functions (of the arguments indicated in* (9.3.4) *except for* $\mathcal{A}$*). Moreover, this holds even if the run of* $\mathcal{A}$ *is confined by the requirement to keep equations* (9.3.5) *of the inner part unchanged;*

H) *For any* $h \in \mathfrak{H}_+$*, the memory alphabets* $\mathfrak{M}_h$ *has the maximal possible cardinality (that of the real numbers due to Remark* 9.2.8 *on p.* 275*).*

Remark 9.3.5. **a)** The sets $\mathfrak{H}_s$ and $\mathfrak{H}_a$ may contain common elements.

b) The variables $\boldsymbol{O}^-, \boldsymbol{I}^-, \boldsymbol{M}^-$ represent the united ensembles of outputs, inputs, and memories of the elements from the interior part of the network, respectively.

c) The last claim from **G)** should hold with respect to equations (9.3.5) in their arbitrary possible form, i.e., form taken for at least one $\mathcal{A} \in \mathfrak{A}$.

d) The assumptions **G** and **H)** mean that $\mathcal{A}$ in fact influences only the algorithm of data processing within the interior part of the network, whereas the algorithms at sensors and actuators may be arbitrary.

e) Assumption 9.3.4 does not restrict the communication equations in any way.

f) Typically, any sensor node $h \in \mathfrak{H}_s$ is endowed with its own source and has access to the data from only this source: $\mathfrak{S} = \mathfrak{H}_s, \sigma = h, \mathfrak{H}_{\leftarrow h} = \{h\}$, and $y_h = \mathfrak{y}_h$ is the measurement of the outer process. In general, the data injection prefix can be viewed as an ensemble of channels, each instantaneously broadcasting the data from a particular source σ to all nodes $h \in \mathfrak{H}_{\leftarrow \sigma}$ (see Fig. 9.12).

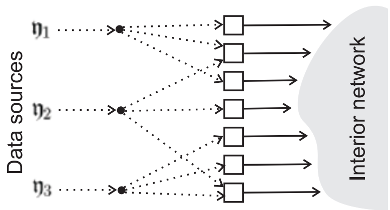


Fig. 9.12. Data injection prefix with three sources.

Definition 9.3.6. *The elements* $h \in \mathfrak{H}_+ = \mathfrak{H}_s \cup \mathfrak{H}_a$ *through which the network contacts its exterior are called the* contact nodes.

Since these nodes accept all causal algorithms of data processing due to **G** and **H)**, any change of the outer alphabets from **C** and **D)** keeps the network well defined. This shows that in fact the network can accept input data and produce outputs in any form. In other words, $\mathfrak{y}_\sigma$ and u_h may be elements of arbitrary given sets, not necessarily Euclidean spaces. So such a network can be used not only to generate controls from the sensor signals in order to stabilize the outer process, but can also be employed to transmit abstract information. The main result of the chapter states that the capacity of the network to stabilize the process is identical to its capacity for reliable information transmission.

Since the network will be regarded as not only a stabilizing but also a communicating facility, it is convenient to call the nodes from $\mathfrak{H}_s$ and $\mathfrak{H}_a$ not the "sensors" and "actuators," as before, but more neutrally *input* and *output* nodes, respectively. Similarly, $\mathcal{A}$ will be called *data processing strategy*.

Observation 9.3.7. *Let a time sequence* $\{\mathfrak{y}_\sigma(t)\}_{t\geq 0}$ *of messages be given for any source* σ*. Then given a data processing strategy* $\mathcal{A}$*, the output sequence* $\{\boldsymbol{U}(t)\}_{t\geq 0}$ *is determined uniquely.*

Definition 9.3.8. *A network described by equations of the form* (9.3.2) *and satisfying Assumption* 9.3.4 *is said to be* with unconstrained contact nodes.

9.3.4 Concluding Remarks

In the light of **e)** from Remark 9.3.5, it is clear that Examples 1–5 from Sect. 9.2 are particular cases of the network introduced in the previous subsection. It also encompasses the cases where the interior of the network includes elements of all types considered in Example 6.

In the remainder of the subsection, we discuss some formal operations with networks. In doing so, we consider only networks satisfying the requirements from Definition 9.3.8.

Deleting a contact is depriving the contact node of the outer input, output, or both. Then the node loses the status of an input, output, or both, respectively. As a result, the node may lose the status of the contact one. In this case, it should be considered as an element of the interior part of the network. So its processing equations should be removed to (9.3.5) within the system of equations from **F)** of Assumption 9.3.4.

If the outer input y_h (output u_h) is composed of several components, only one or some of them may be deleted. In this case, node h may keep its status unchanged.

Restricting the type of an inner node by imposing more restrictions on the memory and processing algorithm. Consider the case where the node from the previous consideration becomes inner. Then its memory and the set of admissible processing algorithms can be restricted arbitrarily without violation of Assumption 9.3.4.

Free union is interpreting several independent networks with disjoint sets of contact nodes as a single network. This situation is illustrated in Fig. 9.13a, where the resultant network has four input and three output nodes.

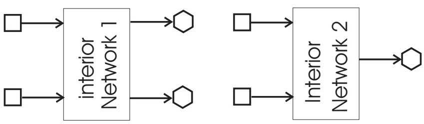


Fig. 9.13(a). Free union.

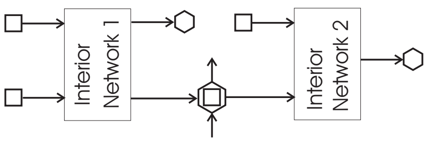


Fig. 9.13(b). Welding two nodes.

Welding several contact nodes is interpreting them as parts of a common node h^w (see Fig. 9.13b). The knowledge available to any of the original nodes is accessible by h^w. In other words, its memory and input encompasses the memories and inputs, respectively, of constituting nodes. (This concerns both inner and outer inputs.) Similarly, the output of h^w is the ensemble of outputs for all original nodes.

Explanation 9.3.9. Like for the original nodes, no restrictions are imposed on the processing algorithm for h^w. So any component of its output, say the output of an original node h, can be generated from the entire knowledge accessible by h^w, i.e., from the data available to all original nodes. This makes the difference with the situation before welding, where only the data available to h could be employed.

Remark 9.3.10. If the set of welded nodes contains an input (output) one, h^w is an input (output) node for the resultant network. So if this set contains both input and output nodes, h^w is simultaneously an input and output contact node, like in Fig. 9.13b.

Proceeding with the example in Fig. 9.13b, one may deprive this node of both contacts, thus making it inner (see Fig. 9.14a). The resultant node admits all data

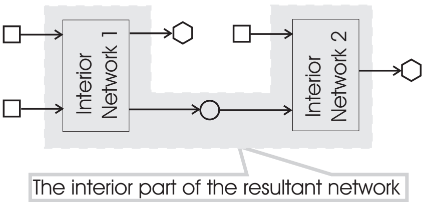


Fig. 9.14(a). Deleting contacts.

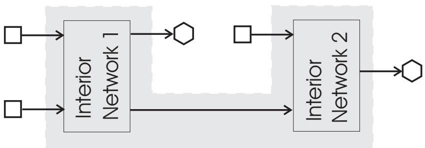


Fig. 9.14(b). Restricting the type.

processing algorithms and enjoys unlimited memory. Let us restrict its type to the memoryless node with the single admissible algorithm, i.e., merely forwarding the received message further. This in fact transforms the node into a part of a new deterministic compound communication channel (see Fig. 9.14b).

Now we consider three more operations that will be of special interest.

Connecting two contact nodes with an additional channel. This operation inserts one more channel into the network. The channel should go from a given contact node **A** to another given contact node **B** (see Fig. 9.15). For the definiteness, let this be the d-delayed channel with the alphabet $\mathfrak{E}$. It may be regarded as a simple network with one input and output node $\mathbf{A}'$ and $\mathbf{B}'$, respectively. The operation consists in 1) forming the free union of two networks, 2) welding **A** with $\mathbf{A}'$ and **B** with $\mathbf{B}'$, and 3) deleting the outer input of the former node $\mathbf{A}'$ and the outer output of the former node $\mathbf{B}'$. In the sequel, the case where the alphabet $\mathfrak{E}$ has the maximum possible cardinality $2^{\aleph_0}$ is of most interest. Then the additional channel makes the data available to node **A** at time t accessible by node **B** at time $t + d$.

Attaching a prefix is, in general, a change of the scheme of information exchange between the exterior of the network and its input nodes $h \in \mathfrak{H}_s$. This may

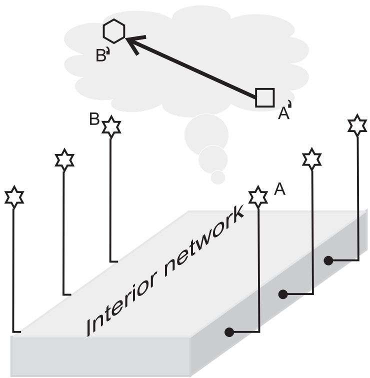


Fig. 9.15. Connecting two contact nodes.

involve transformation of some input nodes into output ones. Now we consider the simplest situation where this operation reduces to putting a new data injection prefix **PREF** with new data sources in use. In doing so, the original data sources and the corresponding outer inputs of the sensor nodes are deleted. The resultant network is denoted by **PREF** ⊞ **NW**.

Attaching a suffix (see Fig. 9.16). Several new output nodes are attached to **NW**, each collecting data from a given variety of original output nodes via given instantaneous interference channels. The ensemble **SUFF** of these channels is called the *suffix*. The outputs at the original output nodes are deleted, which makes these nodes inner. The resultant network is denoted by **NW** ⊞ **SUFF**.

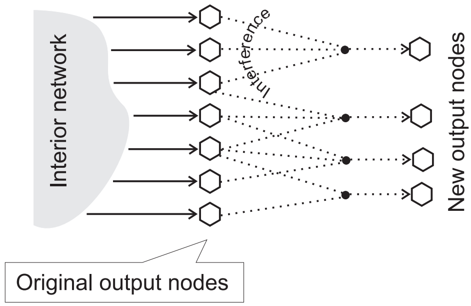


Fig. 9.16. Change of output nodes.

Explanation 9.3.11. To model the instantaneous interference channel

$$s(t) = \mathfrak{R}\big[\{o'_h(t)\}_{h \in \mathfrak{H}_*}\big] \tag{9.3.6}$$

with the output at the new node h' and inputs o'_h from the nodes $h \in \mathfrak{H}_* \subset \mathfrak{H}_a$ under the formalism of Assumption 9.3.4, the new node h' is granted access to the input data of all nodes $h \in \mathfrak{H}_*$. The memory of h' is structured $m_{h'} = \left[\{m_h\}_{h \in \mathfrak{H}_*}, \overline{m}_{h'}\right]$ to include the counterparts of the memory contents of all nodes $h \in \mathfrak{H}_*$. For h', the admissible data processing algorithms are confined to those that first process data related to a given $h \in \mathfrak{H}_*$, independently of all other accessible data, to generate a signal $o'_h(t)$ and update $m_h(t)$ and then produce the node output $u_{h'}(t)$ and update $\overline{m}_{h'}(t)$ in such a way that the use of the data associated with the nodes $h \in \mathfrak{H}_*$ is restricted to the use of the quantity (9.3.6).[4]

9.4 Decentralized Networked Stabilization with Communication Constraints: The Problem Statement and Main Result

This section offers the formulation of the main problem and results of this chapter.

9.4.1 Statement of the Stabilization Problem

The main problem addressed in this chapter is as follows. We consider linear discrete-time multiple sensor and actuator systems of the form:

$$x(t+1) = Ax(t) + \sum_{i=1}^{l} B_i u_i(t) + \xi(t), \qquad x(0) = x^0; \tag{9.4.1}$$

$$y_j(t) = C_j x(t) + \chi_j(t), \qquad j = 1, \dots, k. \tag{9.4.2}$$

Here l and k are the numbers of the actuators and sensors, respectively, and the variables are as follows:

- $x \in \mathbb{R}^{\dim(x)}$ – the state,
- $u_i \in \mathbb{R}^{\dim(u_i)}$ – the output of the ith actuator,
- $y_j \in \mathbb{R}^{\dim(y_j)}$ – the output of the jth sensor,
- $\xi(t) \in \mathbb{R}^{\dim(x)}$ – the exogenous disturbance,
- $\chi_j(t) \in \mathbb{R}^{\dim(y_j)}$ – the noise in the jth sensor.

The system is unstable: There is an eigenvalue λ of the matrix A with $|\lambda| \geq 1$. The objective is to stabilize the plant.

This should be accomplished by means of multiple decentralized controllers. Along with sensors and actuators, they are spatially distributed and linked via a given communication network. The algorithms of data processing at controllers, sensors, and actuators are to be designed so that the closed-loop system is stable.[5]

[4]At first sight, one more requirement is that the node h' should produce the memory content m_h identical to that generated at the node h. In fact, this requirement is superfluous. It can be always ensured by reorganizing the memory and data processing at h and h' so that both h and h' acquire two copies of m_h processed in accordance with the algorithms originally implemented at h and h', respectively.

[5]In this chapter, the term "actuator" denotes a device that is able not only to apply a given control to the plant but also to generate controls. So it may be more correct to call these devices "controllers directly acting upon the plant."

Formally, we suppose that a network **NW** with unconstrained contact nodes[6] is given. Its input and output nodes are associated with the sensors and actuators, respectively (see Fig. 9.17a). The jth sensor acts as the data source for the associated input node whose current outer input is thus equal to $y_j(t)$. Similarly, the control $u_i(t)$ is the current outer output generated by the corresponding output node. We recall that the algorithm of data processing within the network is denoted by the parameter $\boldsymbol{\mathcal{A}}$ in (9.3.2). This parameter should be chosen from the given set $\mathfrak{A}$ of admissible algorithms so that the closed-loop system is stable. The algorithm $\mathcal{A}$ will be also called the *(decentralized) control strategy*.

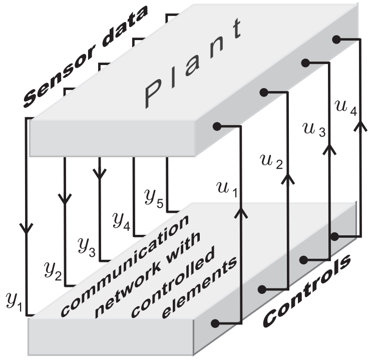


Fig. 9.17(a). Decentralized control.

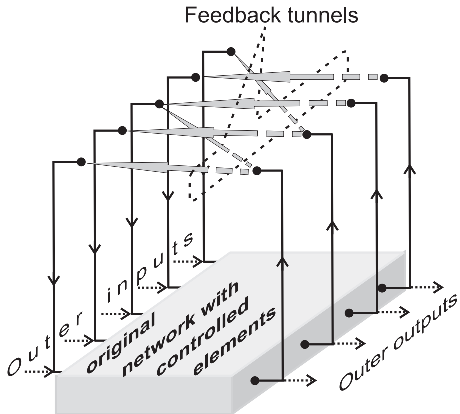


Fig. 9.17(b). Control-based extension.

Remark 9.4.1. It follows that in (9.3.2) (on p. 283), the compound vectors of the network inputs and outputs, respectively, are given by

$$\boldsymbol{Y}(t) = \mathbf{col}\,[y_1(t), \dots, y_k(t)]\,, \qquad \boldsymbol{U}(t) = \mathbf{col}\,[u_1(t), \dots, u_l(t)]\,.$$

The control strategy $\boldsymbol{\mathcal{A}}$, sequences of noises $\{\xi(t)\}$, $\{\chi_j(t)\}$, and initial state x^0 from (9.4.1) uniquely determine a process in the closed-loop system.

We consider bounded noises and initial states x^0:

$$\|\xi(t)\| \leq D < \infty, \qquad \|\chi_j(t)\| \leq D_j^y < \infty \;\forall t, j, \qquad \|x^0\| \leq D_x < \infty, \quad (9.4.3)$$

where $D, D_j^y, D_x > 0$, and accept that a control strategy succeeds if it keeps the stabilization error bounded:

$$\sup_t \sup_{\{\xi(\theta)\},\{\chi_j(\theta)\}\,,\,x^0} \|x(t)\| < \infty. \quad (9.4.4)$$

Here the second sup is over noises and initial states satisfying (9.4.3).

[6]See Definition 9.3.8 (on p. 288).

Definition 9.4.2. *A decentralized control strategy ensuring* (9.4.4) *is said to* stabilize *the system. If in addition* (9.4.4) *holds for any i with $u_i(t)$ substituted in place of $x(t)$, the strategy is said to* regularly stabilize *the system.*

Remark 9.4.3. We recall that arbitrary restrictions on the admissible algorithms for the controllers can be imposed within this problem statement. What is called the "network" captures not only the available ways to transmit information but also the restrictions on the possible ways of data processing by the controllers.

> *Which networks fit to stabilize a given unstable linear plant under a proper design of the control strategy?*

The main result of the chapter is that this question is reducible to the following standard question studied in the traditional information sciences.

> *How much data may be communicated from the input to output nodes across the network that results from a certain extension of the original one?*

All limitations on not only communication but also the data processing algorithms are inherited by the extended network. So the second question concerns communication under the restrictions imposed in the original network.

Remark 9.4.4. The extended network is an artificial and auxiliary object serving the answer to the primal question.

In the next four subsections, this network is introduced via three steps.

9.4.2 Control-Based Extension of the Network

As was shown in Subsect. 7.8.1 (on p. 233), there is a way to communicate as much information as desired via the plant. In this way, the information can be transmitted from any actuator to any sensor that is able to detect the actions of this actuator. To this end, the actuator should encode a message by producing a control that imparts to the system motion a certain specific feature. The sensor receives the message by observing the motion and detecting this feature.

The above phenomenon can be interpreted as if the plant hides several external[7] communication channels. The first step to extend the network is explicit incorporation of all these channels into it (see Fig. 9.17b). After this, the resultant network will be considered in the open loop (out of connection with the plant).

Now we come to details. To start with, we introduce the subspaces of states that are controllable and nonobservable by the ith actuator and jth sensor, respectively:

$$L_i^{+c} := \mathrm{Lin}\left[\bigcup_{\nu=0}^{n-1} \mathbf{Im}\,\left(A^\nu B_i\right)\right], \qquad L_j^{-o} := \bigcap_{\nu=0}^{n-1} \left\{x : C_j A^\nu x = 0\right\}, \tag{9.4.5}$$

where $n := \dim(x)$. The set of actuator–sensor pairs communicating via the plant is

[7] With respect to the network given in the problem statement.

$$\text{CVP} := \left\{(i,j) : L_i^{+c} \not\subset L_j^{-o}\right\}. \tag{9.4.6}$$

Since control u_i influences y_j with the delay

$$d_{i\to j} := \min\{d = 0, 1, \ldots : C_j A^d B_i \neq 0\} + 1, \tag{9.4.7}$$

data transmitted from the ith actuator to the jth sensor incur the same delay.

Now let us run over all pairs (i,j) from (9.4.6). For every pair, let us connect the network contact node associated with the ith actuator with that associated with the jth sensor. The connection is by means of $d_{i\to j}$-delayed channel $R_{i\to j}$ called the *feedback tunnel* (i.e., hidden channel) with the infinite alphabet of the maximal possible cardinality $2^{\aleph_0}$.

Insertion of all these tunnels gives rise to a new network called the *control-based extension* **CBE(NW)** of the original one.

Remark 9.4.5. The input and output nodes of **CBE(NW)** are identical to those of the original network.

9.4.3 Assumptions about the Plant

To proceed, we need some assumptions about the plant. They in fact come to a strengthened form of Assumption 3.4.24 (on p. 49) with respect to the sensors, and adopting its dual counterpart with respect to the actuators. The reason for imposing Assumption 3.4.24 was discussed in Sect. 3.9 (starting on p. 90). Its strengthening is motivated by the wish to simplify the design of the stabilizing strategy by avoiding the technically demanding sequential stabilization approach (see p. 64).

We first state the assumption about the plant in a preliminary form, which is easy to comprehend. Then we give its basic formulation, which is more complicated. The reader interested in rather ideas than generality may confine himself to the first form.

Assumption 9.4.6.* **(Restricted form.)** *The unstable* $|\lambda| \geq 1$ *eigenvalues* λ *of* A *are distinct.*

Assumption 9.4.6. *The unstable subspace* $M_{\text{unst}}(A)$ *of* A *can be decomposed into the direct sum* $M_{\text{unst}}(A) = \bigoplus_{\nu=1}^{g} M_\nu$ *of* A*-invariant (real) subspaces* M_ν *such that the following two claims hold:*

- **i)** *For any* ν*, the spectrum of* $A|_{M_\nu}$ *consists of only either one real eigenvalue* $\lambda_{[\nu]}$ *or a couple of conjugate complex ones* $\lambda_{[\nu]} = \overline{\lambda}_{[\nu]}$*;*
- **ii)** *The subspace* M_ν *is in simple relations with the sensors and actuators: Every subspace of the form* $L = L_i^{+c} \cap M_{\text{unst}}(A)$ *or* $L_j^{-o} \cap M_{\text{unst}}(A)$ *is a direct sum of several subspaces of the form* M_ν*.*

We recall that the direct sum of the empty group of subspaces is defined to be $\{0\}$.

Observation 9.4.7. *Assumption* 9.4.6* *clearly implies* 9.4.6*, with* M_ν *being the real invariant subspaces related to unstable eigenvalues.*

Due to Assumption 9.4.6, a proper linear change of the variables shapes x into

$$x = \mathbf{col}\Big(\underbrace{x_{-s},\ldots,x_0}_{\sim M_{\text{st}}(A)},\underbrace{x_1,\ldots,x_{g_1}}_{\sim M_1},\underbrace{\ldots}_{\sim M_2},\underbrace{\ldots}_{\sim M_3},\ldots,\underbrace{\ldots x_{n^+}}_{\sim M_g}\Big), \tag{9.4.8}$$

where $n^+ := \dim M_{\text{unst}}(A)$. In other words, $M_{\text{st}}(A)$ is the set of all x with $x_\alpha = 0\ \forall \alpha \geq 1$, and M_ν is the set of all x with zeros at all positions outside a certain group of coordinates. With a slight abuse of terminology, x_α with $\alpha \geq 1$ will be called *unstable modes*.

Remark 9.4.8. Any unstable mode x_α is related to a unique subspace M_ν, where $\nu = \nu(\alpha)$. So by **i)**, it is associated with an eigenvalue $\lambda_\alpha := \lambda_{[\nu(\alpha)]}$. This association is unique up to the conjugation in the case of the complex eigenvalue; the modulus $|\lambda_\alpha|$ is determined uniquely.

Remark 9.4.9. Given an unstable mode x_α and a sensor, this mode is either observed or ignored by this sensor. Similarly, a given actuator either controls the mode or does not affect it at all.

More specifically, the claim **ii)** implies that for any sensor j, the nondetectable subspace $L_j^{-o} \cap M_{\text{unst}}(A)$ is the set of all x with zero coordinates at all positions outside a certain group A_j with elements $\alpha \geq 1$. Any unstable mode x_α with $\alpha \in A_j$ does not affect the outputs of this sensor. Conversely, it can be restored from these outputs (the sensor observes the mode x_α) if $\alpha \notin A_j$. Similarly, the stabilizable subspace $L_i^{+c} \cap M_{\text{unst}}(A)$ of any actuator i is the set of all x with zero coordinates at all positions outside a certain group $A^{[i]}$ with elements $\alpha \geq 1$ thanks to **ii)**. Any unstable mode x_α with $\alpha \notin A^{[i]}$ is not affected by this actuator. Conversely, this mode is controlled by this actuator if $\alpha \in A^{[i]}$.

Remark 9.4.10. Assumption 9.4.6 is true whenever the system is detectable by each sensor and stabilizable by every actuator.

Indeed, then **ii)** holds for the decomposition of $M_{\text{unst}}(A)$ into the real invariant subspaces related to the unstable eigenvalues.

9.4.4 Mode-Wise Prefix

The next step to the new network is a change of the scheme of data injection into **CBE(NW)**. We introduce new data sources, each associated with a particular unstable mode x_α. The data related to mode x_α are instantaneously delivered to all sensors j that observe this mode, i.e., those from the set

$$\mathfrak{J}^{mw}_{\leftarrow\alpha} := \Big\{ j = 1,\ldots,k : M_{\nu(\alpha)} \cap L_j^{-o} = \{0\} \Big\}. \tag{9.4.9}$$

The *mode-wise prefix* $\mathbf{PREF_{mw}}$ is the described scheme of data injection (see Fig. 9.18a).

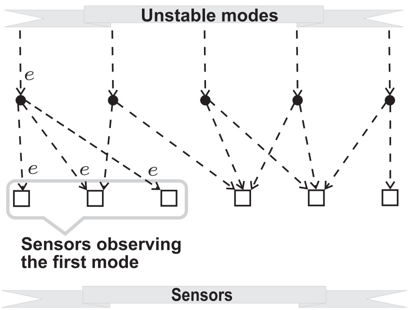


Fig. 9.18(a). Mode-wise prefix.

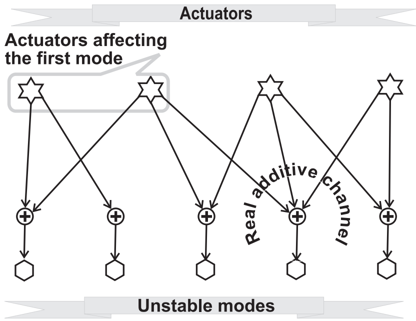


Fig. 9.18(b). Mode-wise suffix.

9.4.5 Mode-Wise Suffix and the Final Extended Network

This suffix displays similar dual relations between the actuators and unstable modes, and it is used to change the output nodes of the network. The new output nodes are also associated with unstable modes and attached to **CBE(NW)** via related interference channels (see Fig. 9.16 on p. 290). The channel related to x_α collects data from all actuators i that control this mode, i.e., those from the set

$$\mathfrak{I}^{mw}_{\to\alpha} := \{i : M_{\nu(\alpha)} \subset L_i^{+c}\}. \tag{9.4.10}$$

This channel is instantaneous and additive with the real alphabet $\mathbb{R}$. The *mode-wise suffix* is the ensemble $\mathbf{SUFF_{mw}}$ of all these additive real channels (see Fig. 9.18b). The *mode-wise suffix with quantization* $\mathbf{SUFF^q_{mw}}$ results from restriction the admissible algorithms at any terminal node (associated with an unstable mode) to those forming the current outer output as the projection of the currently transmitted real signal s into the nearest integer.[8]

Explanation 9.4.11. The locations of the source and the output node associated with a common unstable mode are regarded as different.

This is motivated by the double role of every unstable mode: It is an object of observation on the one hand and control on the other hand.

Explanation 9.4.12. The additive channel is chosen since the cumulative effect of the actuators on the plant (9.4.1) results from summation of the effects caused by each actuator.

The final extended network that will be of interest further results from attaching both the mode-wise prefix and mode-wise suffix with quantization

$$\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^q_{mw}} \tag{9.4.11}$$

to the control-based extension **CBE(NW)** of the original network **NW** (see Fig. 9.19).

[8]In the case of uncertainty ($v - 1/2$ is an integer), the void output $\circledast$ is produced.

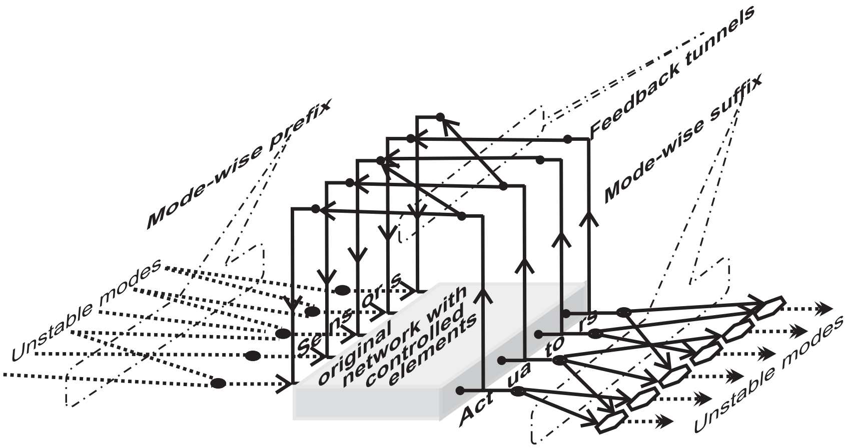


Fig. 9.19. Extended network.

Remark 9.4.13. The network (9.4.11) is disconnected with the plant.

Remark 9.4.14. Due to (9.4.6), (9.4.9), and (9.4.10), any pair (actuator node i)-(sensor node j) affecting and observing, respectively, at least one common unstable mode x_α is linked by a feedback tunnel:

$$\text{CVP} \supset \{(i, j) : \exists \alpha = 1, \ldots, n_+ \text{ such that } i \in \mathfrak{J}^{mw}_{\to\alpha} \text{ and } j \in \mathfrak{J}^{mw}_{\leftarrow\alpha}\}. \qquad (9.4.12)$$

9.4.6 Network Capacity Domain

As was remarked on p. 293, criterion for stabilizability involves an answer to the question: How much data can be communicated across the network (9.4.11)? Specifically, data from the informant associated with the unstable mode x_α should be sent to the output node related to the same mode.[9] The answers to such questions are traditionally given in terms of a fundamental concept of the classic information theory: The capacity (rate) domain.

Now we recall this notion for the particular situation at hand. Its critical feature is that the network data sources σ_α and output nodes $\mathbf{out}_\alpha$ are marked by the common index $\alpha \in [1 : n^+]$. The sources produce messages $\mathfrak{y}_\alpha$ independently. For any α, the message $\mathfrak{y}_\alpha$ should be transmitted to the equally enumerated node $\mathbf{out}_\alpha$, where it appears in the form of the outer output $\mathfrak{u}_\alpha$ of this node. The transmission is arranged by choosing an admissible algorithm $\mathcal{A} \in \mathfrak{A}$ of data processing within the network.

Definition 9.4.15. *A* networked block code *with block length N is an admissible data processing algorithm $\mathcal{A}$ that acts only during the time interval $[0 : N-1]$, serves sources producing constant message sequences $\mathfrak{y}_\alpha(t) \equiv \mathfrak{y}_\alpha \in [1 : F_\alpha]$ $\forall t$,*

[9] The data streams related to various α should go through the network simultaneously.

and generates the outputs in the matching form $\mathfrak{u}_\alpha(t) \in [1 : F_\alpha] \cup \{\circledast\}$.[10] *The* rate vector *of this code is defined to be*

$$\boldsymbol{r}_{\text{code}} := \mathbf{col}\left(\frac{\log_2 F_1}{N}, \ldots, \frac{\log_2 F_{n^+}}{N}\right). \tag{9.4.13}$$

Definition 9.4.16. *A networked block code is* errorless *if at the terminal time it correctly recognizes the messages from all sources* $\mathfrak{u}_\alpha(N-1) = \mathfrak{y}_\alpha \; \forall \alpha$ *irrespective of which messages* $\mathfrak{y}_\alpha \in [1 : F_\alpha]$ *were dispatched.*

Definition 9.4.17. *A vector* $\boldsymbol{r} \in \mathbb{R}^{n^+}$ *is called the* achievable rate vector *if for arbitrarily large* $\overline{N}$ *and small* $\epsilon > 0$, *an errorless networked block code with block length* $N \geq \overline{N}$ *exists whose rate vector* (9.4.13) *approaches* $\boldsymbol{r}$ *with accuracy* ϵ*; i.e.,* $\|\boldsymbol{r}_{\text{code}} - \boldsymbol{r}\| < \epsilon$. *The* capacity domain **CD** *is the set of all achievable rate vectors.*

Finding such a domain is a long-standing standard problem in the information theory. We refer the reader to [49, 214] for an excellent overview of achievements and troubles in its solution.

Remark 9.4.18. To underscore the capacity domain of which network is considered, the notation (in brackets) of the network may be added to **CD** from the right.

Remark 9.4.19. Since the suffix $\mathbf{SUFF}^{\mathbf{q}}_{\mathbf{mw}}$ with quantization admits a smaller range of data processing algorithms than $\mathbf{SUFF}_{\mathbf{mw}}$, the following inclusion holds:

$$\mathbf{CD}\left(\mathbf{PREF}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF}^{\mathbf{q}}_{\mathbf{mw}}\right) \subset \mathbf{CD}\left(\mathbf{PREF}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF}_{\mathbf{mw}}\right). \tag{9.4.14}$$

In some cases, these domains are equal. One of them will be discussed in Remark 9.5.20 (on p. 323).

9.4.7 Final Assumptions about the Network

In addition to Assumption 9.3.4 (on p. 285), we need more assumptions about the primal network **NW** involved in the stabilization problem setup. They basically mean that the network is stationary and can be reset to the initial state.

Assumption 9.4.20. *For any admissible data processing strategy* $\boldsymbol{\mathcal{A}} \in \mathfrak{A}$ *and time instant* $T = 0, 1, \ldots$, *another admissible strategy* $\boldsymbol{\mathcal{A}}_{\text{res}} \in \mathfrak{A}$ *exists such that the following properties hold:*

i) These strategies are identical until $t = T$*; i.e.,*

$$\mathcal{F}(\cdot, \boldsymbol{\mathcal{A}}, t) \equiv \mathcal{F}(\cdot, \boldsymbol{\mathcal{A}}_{\text{res}}, t) \qquad \text{for all} \quad t = 0, \ldots, T \quad \text{and} \quad \mathcal{F} := \mathcal{O}, \mathcal{U}, \mathcal{M}$$

and $\mathcal{M}^0(\boldsymbol{\mathcal{A}}) = \mathcal{M}^0(\boldsymbol{\mathcal{A}}_{\text{res}})$, *where* $\mathcal{O}(\cdot), \mathcal{U}(\cdot), \mathcal{M}(\cdot)$, *and* $\mathcal{M}^0(\cdot)$ *are the functions from* (9.3.2) *(on p.* 283*);*

[10] The output $\circledast$ means "no decision."

ii) The strategy $\boldsymbol{\mathcal{A}}_{\text{res}} \in \mathfrak{A}$ resets the network to the initial state at a time $T_* > T$:

$$\boldsymbol{N}(T_*) = \boldsymbol{N}_0, \qquad \boldsymbol{I}(T_*) = \boldsymbol{I}_0, \qquad \boldsymbol{M}(T_*) = \mathcal{M}^0(\boldsymbol{\mathcal{A}})$$

and can be chosen so that $T_* - T \leq \delta T_{\max}$, where the constant $\delta T_{\max}$ does not depend on $\boldsymbol{\mathcal{A}}$. Moreover, the memory content M can be driven to the state $\mathcal{M}^0(\boldsymbol{\mathcal{A}}_1)$ initial for any other a priori chosen strategy $\boldsymbol{\mathcal{A}}_1 \in \mathfrak{A}$ provided that it is equivalent to $\boldsymbol{\mathcal{A}}$ modulo a given finite partition (i.e., $\exists \nu : \boldsymbol{\mathcal{A}}_1, \boldsymbol{\mathcal{A}} \in \mathfrak{M}_\nu$) of the memory alphabet $\{M\} = \mathfrak{M} = \mathfrak{M}_1 \cup \ldots \cup \mathfrak{M}_Q$.

Explanation 9.4.21. It is tacitly assumed that ii) should be true irrespective of the outer inputs of the network. Typically, this means that the strategy $\boldsymbol{\mathcal{A}}_{\text{res}}$ ignores the inputs on the interval $[T + 1 : T_*]$.

Assumption 9.4.22. *For any two admissible data processing strategies $\boldsymbol{\mathcal{A}}_i \in \mathfrak{A}, i = 1, 2$ and time instant $T = 0, 1, \ldots$ such that the network driven by the strategy $\boldsymbol{\mathcal{A}}_1$ arrives at time T at the state initial for $\boldsymbol{\mathcal{A}}_2$*

$$\boldsymbol{N}(T) = \boldsymbol{N}_0, \qquad \boldsymbol{I}(T) = \boldsymbol{I}_0, \qquad \boldsymbol{M}(T) = \mathcal{M}^0(\boldsymbol{\mathcal{A}}_2),$$

another admissible strategy $\boldsymbol{\mathcal{A}} \in \mathfrak{A}$ exists that is identical to $\boldsymbol{\mathcal{A}}_1$ and $\boldsymbol{\mathcal{A}}_2$ on the time interval $[0 : T - 1]$ and afterward, respectively: $\mathcal{M}^0(\boldsymbol{\mathcal{A}}) = \mathcal{M}^0(\boldsymbol{\mathcal{A}}_1)$ and

$$\mathcal{F}(\cdot, \boldsymbol{\mathcal{A}}, t) \equiv \begin{cases} \mathcal{F}(\cdot, \boldsymbol{\mathcal{A}}_1, t) & \text{for all } t = 0, \ldots, T - 1 \\ \mathcal{F}(\cdot, \boldsymbol{\mathcal{A}}_2, t - T) & \text{for all } t \geq T \end{cases}, \quad \text{where} \quad \mathcal{F} := \mathcal{O}, \mathcal{U}, \mathcal{M}.$$

Explanation 9.4.23. With respect to the data processing algorithms hosted by the contact nodes, these properties are partly ensured by Assumption 9.3.4 (on p. 285).

To state the last assumption, we start with a technical definition.

Definition 9.4.24. *The data processing strategy $\boldsymbol{\mathcal{A}} \in \mathfrak{A}$ is said to be* τ-periodic *($\tau = 1, 2, \ldots$) if in* (9.3.2)*, the functions $\mathcal{O}(\cdot, \boldsymbol{\mathcal{A}}, t), \mathcal{U}(\cdot, \boldsymbol{\mathcal{A}}, t)$, and $\mathcal{M}(\cdot, \boldsymbol{\mathcal{A}}, t)$ are τ-periodic in time: $\mathcal{F}(\cdot, \boldsymbol{\mathcal{A}}, t + \tau) \equiv \mathcal{F}(\cdot, \boldsymbol{\mathcal{A}}, t + \tau)$ for all t and $\mathcal{F} := \mathcal{O}, \mathcal{U}, \mathcal{M}$.*

The last assumption means that any data processing strategy that drives the network into the initial state at some time instant τ can be extended from the interval $[0 : \tau - 1]$ on $[0 : \infty)$ as a τ-periodic strategy.

Explanation 9.4.25. The possibility of τ-periodic extension on any finite interval of the form $[0 : r\tau], r = 1, 2, \ldots$ follows from Assumption 9.4.22.

Assumption 9.4.26. *For any admissible data processing strategy $\boldsymbol{\mathcal{A}} \in \mathfrak{A}$ and time instant $\tau = 1, 2, \ldots$ such that the network is in the initial state $\boldsymbol{N}(\tau) = \boldsymbol{N}_0, \boldsymbol{I}(\tau) = \boldsymbol{I}_0, \boldsymbol{M}(\tau) = \mathcal{M}^0(\boldsymbol{\mathcal{A}})$ at time τ, a τ-periodic admissible data processing strategy $\boldsymbol{\mathcal{A}}_{\text{per}} \in \mathfrak{A}$ exists that is identical to $\boldsymbol{\mathcal{A}}$ on the time interval $[0 : \tau - 1]$.*

9.4.8 Criterion for Stabilizability

This criterion concerns the stabilization problem posed in Subsect. 9.4.1. We recall that $\mathrm{int}\, S$ is the interior of the set S.

Theorem 9.4.27. *Suppose that Assumptions* 9.3.4 (on p. 285), 9.4.6 (on p. 294), 9.4.20*,* 9.4.22*, and* 9.4.26 *hold. We consider the representation* (9.4.8) *of the state vector and associate any unstable mode* $x_\alpha, \alpha \geq 1$ *from* (9.4.8) *with an unstable eigenvalue* λ_α *in accordance with Remark* 9.4.8*. We also introduce the vector*

$$\Lambda := \mathbf{col}\Big(\log_2|\lambda_1|, \ldots, \log_2|\lambda_{n^+}|\Big). \tag{9.4.15}$$

Then the following implications are true:

a) $\Lambda \in \mathrm{int}\, \mathbf{CD}\big[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^{q}_{mw}}\big]$.

$\Downarrow$

b) The network **NW** hosts a time-periodic admissible control strategy that regularly stabilizes[11] the plant.

$\Downarrow$

c) The network **NW** hosts an admissible control strategy that stabilizes the plant.

$\Downarrow$

d) $\Lambda \in \mathbf{CD}\big[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^{q}_{mw}}\big]$.

The proof of this theorem will be given in Sects. 9.6 and 9.7. A stabilizing control strategy is described in Sect. 9.7.

The implication **b)** $\Rightarrow$ **c)** is evident. It is mentioned to stress that the claims **b)** and **c)** are included in a chain of implications with approximately identical extreme terms. So all claims from this chain are "almost equivalent."

The necessity part of Theorem 9.4.27 can also be slightly improved in the case where the vector (9.4.15) contains zero entries: $|\lambda_\alpha| = 1$ for some α.

Remark 9.4.28. If claim **c)** holds, the concerned capacity domain contains a vector with nonzero entries.

The proof of this minor remark will be given in Subsect. 9.8.4 (starting on p. 362).

Corollary 9.4.29. *Whenever* **c)** *holds, the plant is detectable and stabilizable by the entire set of sensors and actuators, respectively.*

Indeed otherwise, there would be an unstable mode x_α such that the corresponding either data source or output node of $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^{q}_{mw}}$ is not linked with the interior network **CBE(NW)**. Then it is impossible to transmit information between these nodes, and so $r_\alpha = 0$ for any achievable rate vector, in violation of Remark 9.4.28.

[11] See Definition 9.4.2 (on p. 293).

The claims **a)** and **d)** depend on the representation (9.4.8) satisfying Assumption 9.4.6. This representation is not unique in general. However its choice does not affect the contents of Theorem 9.4.27.

We demonstrate this under the simplified Assumption 9.4.6* (on p. 294).[12] For any two representations (9.4.8), a certain permutation $\alpha \in [1 : n_+] \mapsto \alpha' \in [1 : n_+]$ puts the eigenvalues in correspondence: $\lambda'_{\alpha'} \in \{\lambda_\alpha, \overline{\lambda}_\alpha\}$. It is easy to see that the matching permutation of the data sources and output nodes transforms the prefix $\mathbf{PREF}_{mw}$ and the suffix $\mathbf{SUFF}_{mw}$ related to one of these representations into, respectively, the prefix and suffix associated with the other representation. The entries of the corresponding vectors (9.4.15) are related by the same permutation. So it is apparent from Definitions 9.4.15 and 9.4.17 that the claims **a)** associated with these representations are equivalent, and so are the claims **d)**.

9.5 Examples and Some Properties of the Capacity Domain

Theorem 9.4.27 reduces the problem of stabilizability to the long-standing standard problem of the information sciences: Finding the capacity domain of the network. For an excellent survey of achievements and troubles in solution of the latter problem, we refer the reader to [49, 214]. In this section, we illustrate Theorem 9.4.27 by simple examples, where the capacity and so stabilizability domains are explicitly computed.

To serve this computation, we need to extend the definition of the capacity domain on networks with not necessarily a common number of sources and outputs, as well as some general facts about capacity domains. The extended definition and these facts are given in Subsects. 9.5.1 and 9.5.2, respectively. Subsection 9.5.2 is not intended to duplicate a special literature in providing a comprehensive theoretical environment for computing capacity domains. Conversely, it considers only a few simple and most probably well-known facts that are required to treat the subsequent examples. Proceeding from these facts, we also discuss some useful corollaries of Theorem 9.4.27 in Subsect. 9.5.2. Subsections 9.5.3 and 9.5.4 deal with examples.

9.5.1 Capacity Domain of Networks with Not Necessarily Equal Numbers of Data Sources and Outputs

In Subsect. 9.4.6, the capacity domain was defined in the case where the network has equal numbers of sources and output nodes marked by a common index, and data are transferred from every source to the equally enumerated output node. Now we consider a more general case, where these numbers are not necessarily equal, and more complicated schemes of data transmission are of interest.

Specifically, let a network **NW** with unconstrained contact nodes[13] be given. The *data communication scheme* for this network is given by a subset

[12] The discussed claim is not used anywhere in the book.

[13] See Definition 9.3.8 (on p. 288).

$$\mathfrak{T} \subset \mathfrak{S} \times \mathfrak{H}_a, \tag{9.5.16}$$

where $\mathfrak{S}$ and $\mathfrak{H}_a$ are the sets of data sources output nodes, respectively. The set $\mathfrak{T}$ lists the pairs that need communication (see Fig. 9.20a). In other words, data should be transmitted from every source σ to all output nodes

$$\mathbf{out} \in \mathfrak{T}^{\leftarrow}[\sigma] := \Big\{ \mathbf{out} : T := (\sigma, \mathbf{out}) \in \mathfrak{T} \Big\} = \Big\{ \mathbf{out}_\sigma^1, \ldots, \mathbf{out}_\sigma^{s_\sigma} \Big\}. \tag{9.5.17}$$

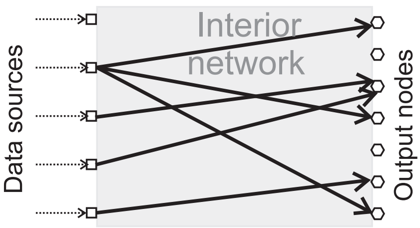


Fig. 9.20(a). Scheme of communication across the network.

Fig. 9.20(b). The simplest scheme.

Explanation 9.5.1. • If $\mathfrak{T}^{\leftarrow}[\sigma] = \emptyset$, there are no data to be sent from σ.

- Respective independent messages $\mu_\sigma^1, \ldots, \mu_\sigma^{s_\sigma}$ should be sent from σ to the output nodes listed in (9.5.17).
- Any output node $\mathbf{out}_\sigma^i$ from (9.5.17) should decode and receive only its own message μ_σ^i from σ: All other messages $\mu_\sigma^j, j \neq i$ are of no interest for $\mathbf{out}_\sigma^i$.
- In total, any output node **out** should receive messages from the sources

$$\sigma \in \mathfrak{T}^{\rightarrow}[\mathbf{out}] := \Big\{ \sigma : (\sigma, \mathbf{out}) \in \mathfrak{T} \Big\}. \tag{9.5.18}$$

- If $\mathfrak{T}^{\rightarrow}[\mathbf{out}] = \emptyset$, the node **out** should receive nothing.

Remark 9.5.2. The entire ensemble of transmitted messages can be enumerated $\big\{\mu_T\big\}_{T\in\mathfrak{T}}$ by the elements of the set (9.5.16).

How much data can be transmitted across the network in accordance with a given communication scheme, provided that the data streams related to various communicating pairs go through the network in parallel? The answer concerns transmitting ensembles $M = \big\{\mu_T\big\}_{T\in\mathfrak{T}}$ of messages of the form $\mu_T \in [1 : F_T]$. Associated with such an ensemble is a constant time sequence of messages from any source σ:

$$\mathfrak{y}_\sigma(t) \equiv \mathfrak{y}_\sigma := \begin{cases} \big\{\mu_{(\sigma,\mathbf{out})}\big\}_{\mathbf{out}\in\mathfrak{T}^{\leftarrow}[\sigma]} & \text{if} \quad \mathfrak{T}^{\leftarrow}[\sigma] \neq \emptyset \\ \circledast & \text{otherwise} \end{cases}.$$

The transmission is arranged by choosing an admissible algorithm $\boldsymbol{\mathcal{A}} \in \mathfrak{A}$ of data processing within the network.

Definition 9.5.3. *A* networked block code *with block length N serving the communication scheme* (9.5.16) *is an admissible data processing algorithm $\mathcal{A}$ that acts only during the time interval $[0 : N-1]$, deals with input data produced in the above way, and generates the outputs in the matching form*

$$\mathfrak{u}_{\mathbf{out}}(t) = \begin{cases} \left\{\mu'_{(\sigma,\mathbf{out})}(t)\right\}_{\sigma\in\mathfrak{T}^{\to}[\mathbf{out}]},\ \mu'_T(t) \in [1 : F_T] \cup \{\circledast\} & \text{if } \mathfrak{T}^{\to}[\mathbf{out}] \neq \emptyset \\ \circledast & \text{otherwise} \end{cases}.$$

The rate ensemble *of this code is defined to be*

$$\boldsymbol{r}_{\text{code}} := \left\{\frac{\log_2 F_T}{N}\right\}_{T\in\mathfrak{T}}. \tag{9.5.19}$$

Definition 9.5.4. *A networked block code is* errorless *if at the terminal time it correctly recognizes the messages for all communicating pairs: $\mu_T = \mu'_T(N-1)\ \forall T \in \mathfrak{T}$, irrespective of which messages $\mu_T \in [1 : F_T]$ were dispatched.*

Definition 9.5.5. *An ensemble $\boldsymbol{r} = \left\{r_T \geq 0\right\}_{T\in\mathfrak{T}}$ is called the* achievable rate ensemble *if for arbitrarily large $\overline{N}$ and small $\epsilon > 0$, an errorless networked block code with block length $N \geq \overline{N}$ exists whose rate ensemble* (9.5.19) *approaches $\boldsymbol{r}$ with accuracy ϵ; i.e., the matching entries differ at most by ϵ. The set of all achievable rate vectors is called the* capacity domain *for the communication scheme* (9.5.16) *and denoted by* $\mathbf{CD}[\mathbf{NW}\blacklozenge\mathfrak{T}]$.

In Subsect. 9.4.6, considered was the case where the numbers of sources and output nodes are equal, they are enumerated by a common index α, and the communication scheme dictates to transmit data from every source to the equally enumerated node (see Fig. 9.20b). Then the pairs from $\mathfrak{T}$ can be evidently enumerated by the same index α, and the definitions from this section come to those from Subsect. 9.4.6.

Conversely, the situation and definitions from this subsection can be reduced to those from Subsect. 9.4.6 up to minor formalities by enumerating the pairs from $\mathfrak{T} = \{T_\alpha\}$ and attaching a special prefix and suffix to the network **NW**. The channels from these prefix and suffix are enumerated by the same index and are instantaneous perfect channels with the alphabets of the maximal possible cardinalities coming to and departing from the first and second entries of the pair T_α from and to the source and output node associated with α, respectively.

We close the subsection with useful technical facts about the capacity domain.

Lemma 9.5.6. *Suppose that a network* **NW** *with unconstrained contact nodes*[14] *satisfies Assumptions* 9.4.20 *and* 9.4.22 *(on p.* 298*), and a data communication scheme* (9.5.16) *is given. Then the following claims are true:*

(i) The capacity domain $\mathbf{CD}[\mathbf{NW}\blacklozenge\mathfrak{T}]$ *is a closed set, which along with any element $\boldsymbol{r}$, contains all elements $0 \leq \boldsymbol{r}' \leq \boldsymbol{r}$, where the inequalities are meant component-wise;*

[14] See Definition 9.3.8 (on p. 288).

(ii) Whenever there exists an errorless block code whose rate ensemble has nonzero entries at the positions from some subset $\mathfrak{T}_* \subset \mathfrak{T}$, *the capacity domain contains an achievable rate ensemble with nonzero entries at the same positions;*

(iii) Let $\varepsilon > 0$ *be given. Any* $\boldsymbol{r} \in \mathbf{CD}[\mathbf{NW}\blacklozenge\mathfrak{T}]$ *can be approximated with accuracy* ε *by the rate ensembles* (9.5.19) *of errorless networked block codes* $\boldsymbol{\mathcal{A}}_m, m = 1, 2, \dots$ *of arbitrarily large block lengths* $N(m) \xrightarrow{m\to\infty} \infty$ *starting at a common initial state: In* (9.3.2) *(on p. 283),* $\mathcal{M}^0(\boldsymbol{\mathcal{A}}_{m'}) = \mathcal{M}^0(\boldsymbol{\mathcal{A}}_{m''})\ \forall m', m''$.

The proof of this lemma is given in Subsect. 9.8.1 (starting on p. 358). The arguments of the proof of (iii) also justify the following claim.

Remark 9.5.7. Statement (iii) holds for the capacity domain concerned in Theorem 9.4.27.

9.5.2 Estimates of the Capacity Domain from Theorem 9.4.27 and Relevant Facts

This domain is that of the network (9.4.11) (on p. 296), which results from attaching the mode-wise prefix and suffix to the control-based extension of the primal network **NW**. This suffix is composed by interference channels, which are not easy to deal with when computing the capacity domain. In this subsection, we give simple estimates of the above domain by domains related to networks with no new interference channels except for those that may be hidden in the original network **NW**.

The lower estimate to follow concerns the network

$$\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \tag{9.5.20}$$

(see Fig. 9.21 on p. 305) whose sources and output nodes are associated with the unstable modes and actuators, respectively. It is supplied with the transmission scheme

$$\mathfrak{T}_{\mathbf{mw}} := \left\{(\alpha, i) : i \in \mathfrak{I}^{mw}_{\to\alpha}\right\} \tag{9.5.21}$$

according to which data from every source $\sim x_\alpha$ should be transmitted to all actuators i controlling the corresponding mode x_α.

The upper estimate deals with the network $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^+_{mw}}$. Here the suffix $\mathbf{SUFF^+_{mw}}$ is obtained from the mode-wise one $\mathbf{SUFF_{mw}}$ (see Subsect. 9.4.5) by replacing every additive channel in $\mathbf{SUFF_{mw}}$ by several independent instantaneous channels with the alphabets of the maximal possible cardinalities. These channels connect the input nodes $i \in \mathfrak{I}^{mw}_{\to\alpha}$ of the former additive channel with its output node $\sim x_\alpha$. After this replacement, data going from these input nodes do not incur interference and so arrive at the output node uncorrupted and in full.

In the remainder of this subsection, the assumptions of Theorem 9.4.27 are assumed to hold.

Lemma 9.5.8. *The following inclusions are true:*

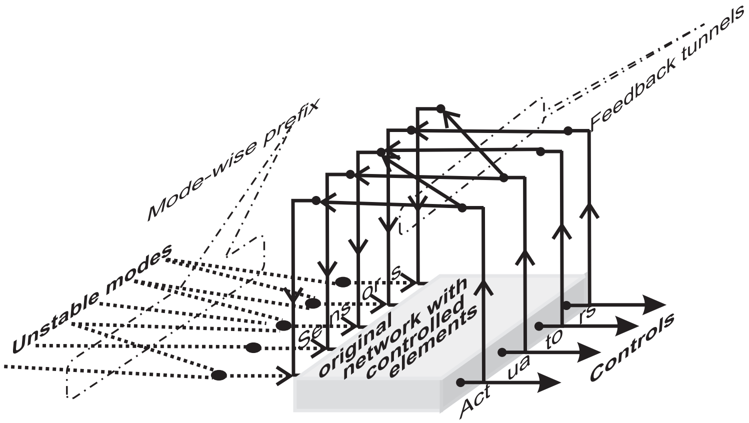


Fig. 9.21. $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)}$.

$$\Big\{ \{r_\alpha\}_{\alpha=1}^{n^+} : r_\alpha = \sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} r_{\alpha,i} \text{ for some}$$

$$\boldsymbol{r} = \big\{ r_{\alpha,i} \big\}_{(\alpha,i)\in\mathfrak{T}_{\mathbf{mw}}} \in \mathbf{CD}\big[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \mathfrak{T}_{\mathbf{mw}}\,\big]\Big\}$$

$$\subset \mathbf{CD}\big[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^{q}_{mw}}\big]$$

$$\subset \mathbf{CD}\big[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^{+}_{mw}}\big]. \quad (9.5.22)$$

The proof of this lemma is given in Subsect. 9.8.2 (starting on p. 360).

Corollary 9.5.9. *Suppose that different actuators do not influence common unstable modes* $\mathfrak{I}^{mw}_{\to\alpha'} \cap \mathfrak{I}^{mw}_{\to\alpha''} = \emptyset \; \forall \alpha' \neq \alpha''$. *Then* $\mathbf{SUFF^{+}_{mw}}$ *can be put in place of* $\mathbf{SUFF^{q}_{mw}}$ *in the statement of Theorem* 9.4.27.

This holds since in (9.5.22), the first and third domains are in fact the same under the circumstances.

Capacity Domain in the Case of Cloned Channels

Now we consider the case where $\mathbf{PREF_{mw}}$ and $\mathbf{SUFF^{+}_{mw}}$ contain cloned channels. Specifically, the unstable modes are partitioned into several groups

$$[1 : n^+] = G_1 \cup \ldots \cup G_s, \quad G_{\nu'} \cap G_{\nu''} = \emptyset \quad \forall \nu' \neq \nu'' \quad (9.5.23)$$

such that

$$\mathfrak{I}^{mw}_{\to\alpha'} = \mathfrak{I}^{mw}_{\to\alpha''} \quad \text{and} \quad \mathfrak{J}^{mw}_{\leftarrow\alpha'} = \mathfrak{J}^{mw}_{\leftarrow\alpha''} \qquad \forall \alpha', \alpha'' \in G_\nu. \quad (9.5.24)$$

Then the corresponding channels from $\mathbf{PREF_{mw}}$ broadcast from the sources $\sim x_{\alpha'}$ and $\sim x_{\alpha''}$, respectively, to a common set of sensor nodes. Similarly, the channels from $\mathbf{SUFF^{+}_{mw}}$ transfer data to the output nodes $\sim x_{\alpha'}$ and $\sim x_{\alpha''}$, respectively, from a common set of actuator nodes. In this case, the individual data rates r_α of communication $\alpha \mapsto \alpha$ between the clones $\alpha \in G_\nu$ are obtained by dividing $\sum_{\alpha \in G_\nu} r_\alpha = r_\nu$ the data rate r_ν achievable in the case where in each group of cloned channels, all channels except for one are discarded (see Fig. 9.22). Now we come to details.

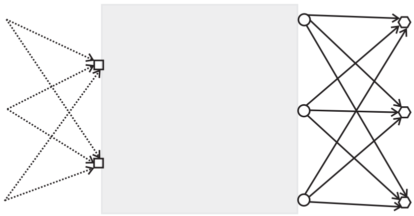


Fig. 9.22(a). Network with cloned channels.

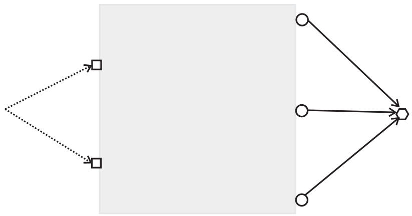


Fig. 9.22(b). Reduced network.

Lemma 9.5.10. *Let* (9.5.23) *and* (9.5.24) *hold.*

(i) Consider the prefix $\overline{\mathbf{PREF}}_{\mathbf{mw}}$ *and suffix* $\overline{\mathbf{SUFF}}^+_{\mathbf{mw}}$ *obtained from* $\mathbf{PREF}_{\mathbf{mw}}$ *and* $\mathbf{SUFF}^+_{\mathbf{mw}}$*, respectively, by discarding all channels except for one in every group* G_ν *and marking the remaining source and output node by the index* ν*. Then*

$$\mathbf{CD}\left[\mathbf{PREF}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF}^+_{\mathbf{mw}}\right] = \Bigg\{ \{r_\alpha\}_{\alpha=1}^{n^+} : r_\alpha \geq 0; \quad \left\{ \sum_{\alpha \in G_\nu} r_\alpha \right\}_{\nu=1}^{s} \in \mathbf{CD}\left[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}^+_{\mathbf{mw}}\right] \Bigg\}; \quad (9.5.25)$$

(ii) Consider the transmission scheme $\overline{\mathfrak{T}}_{\mathbf{mw}}$ *for the network* $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)}$ *obtained from* (9.5.21) *by discarding all pairs except for one in every group of the form* $(\alpha, i), \alpha \in G_\nu$ *and marking the remaining pair by the index* (ν, i)*. Then*

$$\mathbf{CD}\left[\mathbf{PREF}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \mathfrak{T}_{\mathbf{mw}}\right] = \Bigg\{ \{r_{\alpha,i}\}_{(\alpha,i) \in \mathfrak{T}_{\mathbf{mw}}} : r_{\alpha,i} \geq 0; \quad \left\{ \sum_{\alpha \in G_\nu} r_{\alpha,i} \right\}_{(\nu,i) \in \overline{\mathfrak{T}}_{\mathbf{mw}}} \in \mathbf{CD}\left[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \overline{\mathfrak{T}}_{\mathbf{mw}}\right] \Bigg\}.$$

The proof of this lemma is given in Subsect. 9.8.3 (starting on p. 360).

Aggregated Form of the Stabilizability Criterion

In some cases, the criterion from Theorem 9.4.27 can be rewritten in a more compact form on the basis of Lemma 9.5.10. Specifically, let we be given a decomposition

$$M_{\text{unst}}(A) = \bigoplus_{\nu=1}^{g} M_\nu \tag{9.5.26}$$

that satisfies the requirements of Assumption 9.4.6 (on p. 294) except for (i). Assumption 9.4.6 gives one such decomposition. However there may be other decompositions with a lesser number of subspaces.

Remark 9.5.11. Suppose that the plant is stabilizable and detectable by any actuator and sensor, respectively. Then the trivial decomposition $M_{\text{unst}}(A) = M_1$ with only one subspace can be employed.

The decomposition with the above properties gives rise to a partition (9.5.23) satisfying (9.5.24). Indeed, it is clear that any invariant subspace M_ν can be decomposed into invariant sub-subspaces satisfying (i) from Assumption 9.4.6.[15] Then for the corresponding state mode-wise representation (9.4.8) (on p. 295), every subspace M_ν is described as consisting of all vectors such that $x_\alpha = 0$ for all α outside a certain group G_ν of unstable modes. Then (9.5.23) and (9.5.24) evidently hold.

In this case,

- $\overline{\textbf{PREF}}_{\textbf{mw}}$ is composed by the channels associated with the subspaces from (9.5.26) and enumerated by $\nu = 1, \ldots, g$. The νth channel broadcasts from the source $\sim M_\nu$ to all sensors j that observe the subspace: $M_\nu \cap L_j^{-o} = \{0\}$.
- $\overline{\textbf{SUFF}}^{+}_{\textbf{mw}}$ is composed by g groups of channels associated with the subspaces from (9.5.26) and enumerated by $\nu = 1, \ldots, g$. The channels of the νth group arrive at a common output node $\sim M_\nu$ and depart from, respectively, all actuators $i \in \mathfrak{I}_{\to\nu} := \{i : M_\nu \subset L_i^{+c}\}$ that control the subspace at hand. These channels are instantaneous and have the alphabets of the maximal possible cardinalities.
- $\overline{\mathfrak{T}}_{\textbf{mw}}$ consists of all pairs (ν, i) such that the actuator i controls the subspace M_ν.

By observing that for the vector $\boldsymbol{r} = \{\log_2 |\lambda_\alpha|\}_{\alpha=1}^{n^+}$ from Theorem 9.4.27,

$$\sum_{\alpha \in G_\nu} r_\alpha = \log_2 |\det A|_{M_\nu}|,$$

we arrive at the following corollary of Lemmas 9.5.8 and 9.5.10.

Corollary 9.5.12. *Let a decomposition* (9.5.26) *satisfying Assumption* 9.4.6 *with* (i) *dropped be given. Then the sufficient* **a)** *and necessary* **d)** *conditions for stabilizability from Theorem* 9.4.27 *are ensured by and imply the following claims, respectively:*

a′) $\log_2 |\det A|_{M_\nu}| = \sum_{i \in \mathfrak{I}_{\to\nu}} r_{\nu,i} \; \forall \nu$ *for some ensemble* $\{r_{\nu,i}\} \in \operatorname{int} \mathbf{CD}\left[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \overline{\mathfrak{T}}_{\mathbf{mw}}\right]$;

d′) *The vector*

$$\Lambda_{\text{red}} := \mathbf{col}\left(\log_2 |\det A|_{M_1}|, \log_2 |\det A|_{M_2}|, \ldots, \log_2 |\det A|_{M_g}|\right) \tag{9.5.27}$$

belongs to $\mathbf{CD}\left[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}\right]$.

---

[15] Hence, the existence of such a decomposition is equivalent to this assumption.

Similarly, we get the following corollary of Lemma 9.5.10 and Corollary 9.5.9.

Corollary 9.5.13. *Suppose that different actuators do not influence common unstable modes* $\mathfrak{I}^{mw}_{\to\alpha'} \cap \mathfrak{I}^{mw}_{\to\alpha''} = \emptyset \; \forall \alpha' \neq \alpha''$*, and a decomposition* (9.5.26) *satisfying Assumption* 9.4.6 *with* (i) *dropped is given. Then in Theorem* 9.4.27*, the following inclusions* **a**$''$) *and* **d**$''$) *can be substituted in place of* **a)** *and* **d)***, respectively:*

a$''$) $\Lambda_{\rm red} \in \operatorname{int} \mathbf{CD}\left[\overline{\mathbf{PREF}_{\mathbf{mw}}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}^{+}_{\mathbf{mw}}}\right]$;
d$''$) $\Lambda_{\rm red} \in \mathbf{CD}\left[\overline{\mathbf{PREF}_{\mathbf{mw}}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}^{+}_{\mathbf{mw}}}\right]$.

9.5.3 Example 1: Platoon of Independent Agents

The main purpose of this rather special example is to highlight the typical phenomena that are encountered in networked stabilization and are caused by the possibility of information transmission by means of control.

We consider a platoon of s unstable controllable agents $\mathbf{AA}_i$ with independent dynamics

$$x_i(t+1) = A_i x_i(t) + B_i u_i(t) + \xi_i(t), \qquad i = 0, \ldots, s-1,$$

each equipped by a separate controller $\mathbf{C}_i$ (with an unlimited memory and computing capability) and subjected to a bounded exogenous disturbance $\xi_i(t)$. So the agents could stabilize their motions about the required trajectories, provided each of them observes its own state. However, any agent $\mathbf{AA}_i$ has access to noisy data only about the state of its "cyclic follower" $\mathbf{AA}_{i^-}$, $i^- := i-1(\mod s)$. These data are enough to detect the state of the follower with a bounded error and are sent to $\mathbf{AA}_i$ via a digital perfect channel with capacity c_{i^-} (see Fig. 9.23a, where $s = 4$). The agents do not communicate with each other. *Is it possible to stabilize the platoon motion?*

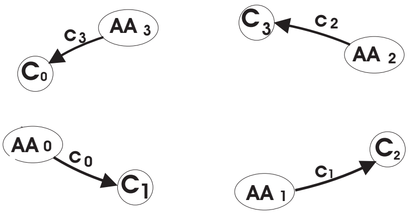


Fig. 9.23(a). Stabilization of autonomous agents.

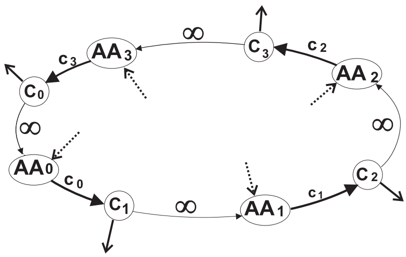


Fig. 9.23(b). Control-based extension.

Since none of the agents observes its own state, the answer seems to be in the negative at first sight. However, the following claim holds.

Proposition 9.5.14. *The platoon is stabiliazible whenever the unstable parts $A_i^+ := A_i|_{M_{\text{unst}}(A_i)}$ of the matrices A_i satisfy the inequality*

$$\log_2 |\det A_0^+| + \cdots + \log_2 |\det A_{s-1}^+| < \min_{i=0,\ldots,s-1} c_i.$$

This inequality is necessary for stabilizability, provided that $\leq$ is put in place of $<$.

This proposition is underlaid by the fact that data can be transmitted by means of controls. This possibility is explicitly displayed by the control-based extension of the network at hand (see Fig. 9.23b). It follows that data about the ith agent can be delivered to its controller $\mathbf{C}_i$ via the route depicted by dotted arrows in Fig. 9.24.

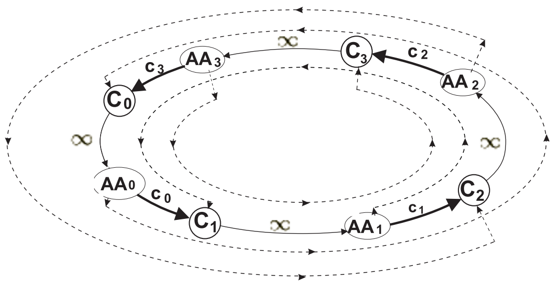


Fig. 9.24. The information exchange in the system.

Proof of Proposition 9.5.14. It is easy to see that Assumptions 9.3.4, 9.4.6, 9.4.20, 9.4.22, and 9.4.26 (on pp. 285, 294, 298, and 299) are satisfied. (The sensor and actuator nodes are associated with $\mathbf{AA}_i$ and $\mathbf{C}_i$, respectively. Initially, the memories of these nodes are empty.) The entire state of the system consists of the states of the agents $x = \mathbf{col}\,(x_0, \ldots, x_{s-1})$, and its dynamical matrix equals $A = \mathbf{diag}(A_0, \ldots, A_{s-1})$. By putting $M_\nu := \{x : x_i = 0 \;\forall\, i \neq \nu - 1,\; x_{\nu-1} \in M_{\text{unst}}(A_{\nu-1})\}$, we clearly get a decomposition (9.5.26) (with $g := s$) satisfying the requirements of Assumption 9.4.6 (on p. 294) except for (i). Since different actuators do not influence common unstable modes, we see that the situation is like in Corollary 9.5.13. By this corollary, it suffices to show that the capacity domain of the control-based extension of the network at hand (see Fig. 9.23b) is given by

$$\big\{\boldsymbol{r} = \{r_i\}_{i=0}^{s-1} : r_i \geq 0 \;\forall i, \quad r_0 + r_1 + \cdots + r_{s-1} \leq \min_{i=0,\ldots,s-1} c_i\big\}. \tag{9.5.28}$$

Here r_i is the rate at which data can be transmitted from $\mathbf{AA}_i$ to $\mathbf{C}_i$.

Sufficiency. If $\sum_i r_i < \min_i c_i$ and $r_i \geq 0$, the control-based extension is evidently able to accommodate simultaneously s continuous stationary fluid flows along the dotted directed arcs from Fig. 9.24 of respective intensities $r_0, \ldots, r_{s-1}$. Then the arguments from the first part of the proof of Theorem 3.1 [69] (based on a proper time division multiplexing) show that $\boldsymbol{r}$ is an achievable rate vector. Then Lemma 9.5.6 (on p. 303) implies that (9.5.28) is a subset of the capacity domain.

Necessity. The above "fluid" consideration is not relevant any longer.[16] Let $\boldsymbol{r}$ be an achievable rate vector. By Definition 9.4.17 (on p. 298), for any $\varepsilon > 0$, an errorless networked block code with block length N and input alphabets $[1 : F_i] = \{\mu_i\}$ exists such that $|r_i - N^{-1}\log_2 F_i| < \varepsilon \,\forall i$. Let A_i and C_i stand for the sequences of all input signals received by $\mathbf{AA}_i$ and $\mathbf{C}_i$, respectively, since $t = 0$ until $t = N - 1$. As follows from Fig. 9.23b,

$$C_{i^+} = \mathcal{C}_i(A_i), \quad \text{where} \quad i^+ := i + 1 \mod s, \qquad A_i = \mathcal{R}_i(C_i, \mu_i) \quad \forall i.$$

(We recall that capital script letters denote deterministic functions.) Since the code is errorless, $\mu_i = \mathcal{M}_i(C_i)$. Thus $C_{i^+} = \mathcal{C}_i\big\{\mathcal{R}_i[C_i, \mathcal{M}_i(C_i)]\big\}\,\forall i$; i.e., $\{C_j\}_{j=0}^{s-1} = \mathcal{F}_i(C_i)\,\forall i$ and so $(\mu_j)_{j=0}^{s-1} = \mathcal{P}_i(C_i)$ for all i and $\mu_i \in [1 : F_i]$. Since C_{i^+} is received over the channel of capacity c_i, it follows that $\sum_j \log_2 F_j \leq N c_i \,\forall i$. So $\sum_i r_i \leq s\varepsilon + \sum_j N^{-1}\log_2 F_j \leq s\varepsilon + \min_i c_i$. Letting $\varepsilon \to 0$ shows that the vector $\boldsymbol{r}$ belongs to (9.5.28). □

9.5.4 Example 2: Plant with Two Sensors and Actuators

In this subsection, we consider a particular case of the linear system (9.4.1), (9.4.2) (on p. 291) with two both sensors and actuators ($l = k = 2$), which are directly linked by perfect channels with given finite capacities (see Fig. 9.25a, where the sensors and actuators are marked by S1, S2 and A1, A2, respectively). The case where some of these channels do not occur is not excluded and identified with the situation where the corresponding capacity is zero $c_{ji} = 0$. The plant is assumed to be detectable and stabilizable by the entire sets of the sensors and actuators, respectively.

Actuators with Nonintersecting Zones of Influence

We start with the case where the actuators affect no common unstable mode. Assumption 9.4.6 (on p. 294) is still supposed to hold. Then the unstable subspace of the plant can be decomposed as follows:

$$M_{\text{unst}}(A) = M_{11} \oplus M_{\mathbf{b}1} \oplus M_{21} \oplus M_{12} \oplus M_{\mathbf{b}2} \oplus M_{22}, \tag{9.5.29}$$

where

- $M_{ji}, j \neq \mathbf{b}$ is the subspace of states controllable by the ith actuator, observable by the jth sensor, and nonobservable by the companion sensor;

[16] In [69], the existence of the continuous fluid flows with intensities matching the data rates of a given coding decoding scheme was justified only for the case of one recipient. The paper [4] offers extensions on the case where a single informant broadcasts to several recipients over a network and data processing strategies are restricted to the so-called alpha-codes. The authors are unaware of research, where relevant facts concerning the fluid-like treatment of information streams were justified for the case of several informants and recipients, as is required now.

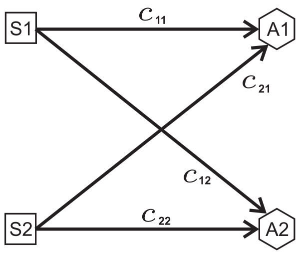


Fig. 9.25(a). System with two sensors and actuators.

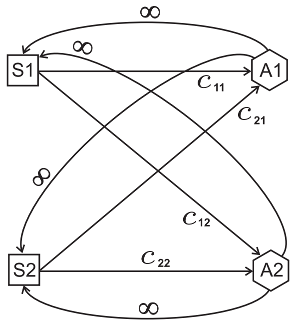


Fig. 9.25(b). Control-based extension.

- $M_{\mathbf{b}i}$ is the subspace of states controllable by the ith actuator and observable by both sensors.

The topological entropy of the related part of the open-loop system is denoted by

$$\boldsymbol{h}_{\nu i} := \log_2 |\det A|_{M_{\nu i}}| \qquad \nu = 1, 2, \mathbf{b}, i = 1, 2.$$

These notations are also explained in Table 9.1.

	Controllable by	
Observable by	Actuator 1	Actuator 2
Only sensor 1	$\boldsymbol{h}_{11}$	$\boldsymbol{h}_{12}$
Both sensors	$\boldsymbol{h}_{\mathbf{b}1}$	$\boldsymbol{h}_{\mathbf{b}2}$
Only sensor 2	$\boldsymbol{h}_{21}$	$\boldsymbol{h}_{22}$

Table 9.1. Entropies of various parts of the system.

Explanation 9.5.15. It is not excluded that some of the above parts may degenerate $M_{\nu i} = \{0\}$. Then $\boldsymbol{h}_{\nu i} := 0$.

To state conditions for stabilizability, we introduce the following operations over 2×2-matrices $M = \left(\begin{smallmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{smallmatrix}\right)$:

- $[M]^{\leftrightarrow} := m_{11} + m_{12}$ — the sum of the elements from the first row;
- $[M]_{\leftrightarrow} := m_{21} + m_{22}$ — the sum of the elements from the second row;
- ${}_{\updownarrow}[M] := m_{11} + m_{21}$ — the sum of the elements from the first column;
- $[M]_{\updownarrow} := m_{12} + m_{22}$ — the sum of the elements from the second column;
- $\Sigma_{-}^{ij}(M) := \sum_{(i',j') \neq (i,j)} m_{i',j'}$ — the sum of all entries except for (i, j)th one.

We also introduce the following matrices of entropies and capacities:

$$H := \begin{pmatrix} \boldsymbol{h}_{11} & \boldsymbol{h}_{12} \\ \boldsymbol{h}_{21} & \boldsymbol{h}_{22} \end{pmatrix}, \qquad C := \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}.$$

We start with the situation where the following property holds.

Assumption 9.5.16. *Any sensor is able to detect the actions of every actuator:* $L_i^{+c} \not\subset L_j^{-o} \; \forall i, j$.

Proposition 9.5.17. *Let Assumption* 9.5.16 *be true. Whenever the plant is stabilizable via the network at hand, the following relations hold:*

$$\begin{aligned} &[H]^{\leftrightarrow} \leq [C]^{\leftrightarrow}, && [H]_{\leftrightarrow} \leq [C]_{\leftrightarrow} \\ {}_{\updownarrow}[H + \mathbf{diag}(\boldsymbol{h}_{\mathbf{b}1}, \boldsymbol{h}_{\mathbf{b}2})] &\leq_{\updownarrow} [C], && [H + \mathbf{diag}(\boldsymbol{h}_{\mathbf{b}1}, \boldsymbol{h}_{\mathbf{b}2})]_{\updownarrow} \leq [C]_{\updownarrow} \\ & && \Sigma_{-}^{ij}(H) + \boldsymbol{h}_{\mathbf{b}j'} \leq \Sigma_{-}^{ij}(C) \qquad \forall i, j = 1, 2, \end{aligned} \tag{9.5.30}$$

where $j' := 1$ *if* $j = 2$ *and* $j' = 2$ *if* $j = 1$.

Conversely, if relations (9.5.30) *hold with the strict inequality signs, the plant is regularly stabilizable.*[17]

Proof of Proposition 9.5.17

By Corollary 9.5.13, the proof is reduced to computation of the capacity domain $\mathbf{CD}\left[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}_{\mathbf{mw}}^{+}\right]$ with respect to the decomposition (9.5.29). Thanks to Assumption 9.5.16, the control-based extension **CBE(NW)** looks as depicted in Fig. 9.25b. The reduced both prefix $\overline{\mathbf{PREF}}_{\mathbf{mw}}$ and suffix $\overline{\mathbf{SUFF}}_{\mathbf{mw}}^{+}$ involve six sources and output nodes, respectively. The source data and outputs are associated with the subspaces from (9.5.29) and marked by the corresponding indices $_{\nu i}$. The components of the rate vectors $\boldsymbol{r}$ are enumerated similarly $r_{\nu i}$, and the matrix $R = \left(\begin{smallmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{smallmatrix}\right)$ is associated with $\boldsymbol{r}$. The network $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}_{\mathbf{mw}}^{+}$ is shown in Fig. 9.26, where for conciseness, the cloned channels from the suffix and prefix are depicted as one channel with compound input or output. Due to Corollary 9.5.13, it suffices to show that the capacity domain of this network is described by the following set of inequalities, which the main part results from (9.5.30) by putting $r_{\nu i}$ in place of $\boldsymbol{h}_{\nu i}$:

$$\begin{aligned} &[R]^{\leftrightarrow} \leq [C]^{\leftrightarrow}, && [R]_{\leftrightarrow} \leq [C]_{\leftrightarrow} \\ {}_{\updownarrow}[R + \mathbf{diag}(r_{\mathbf{b}1}, r_{\mathbf{b}2})] &\leq_{\updownarrow} [C], && [R + \mathbf{diag}(r_{\mathbf{b}1}, r_{\mathbf{b}2})]_{\updownarrow} \leq [C]_{\updownarrow} \\ & && \Sigma_{-}^{ij}(R) + r_{\mathbf{b}j'} \leq \Sigma_{-}^{ij}(C) \qquad \forall i, j = 1, 2, \qquad r_{\nu i} \geq 0. \end{aligned}$$

In the expanded form, this system looks as follows:

$$\begin{aligned} r_{11} + r_{12} \leq c_{11} + c_{12}, &\qquad r_{21} + r_{22} \leq c_{21} + c_{22}; \\ r_{11} + r_{21} + r_{\mathbf{b}1} \leq c_{11} + c_{21}, &\qquad r_{12} + r_{22} + r_{\mathbf{b}2} \leq c_{12} + c_{22}; \\ r_{11} + r_{12} + r_{21} + r_{\mathbf{b}1} &\leq c_{11} + c_{12} + c_{21}; \\ r_{22} + r_{12} + r_{21} + r_{\mathbf{b}2} &\leq c_{22} + c_{12} + c_{21}; \\ r_{11} + r_{22} + r_{12} + r_{\mathbf{b}2} &\leq c_{11} + c_{22} + c_{12}; \\ r_{11} + r_{22} + r_{21} + r_{\mathbf{b}1} \leq c_{11} + c_{22} + c_{21}, &\qquad r_{\nu i} \geq 0. \end{aligned} \tag{9.5.31}$$

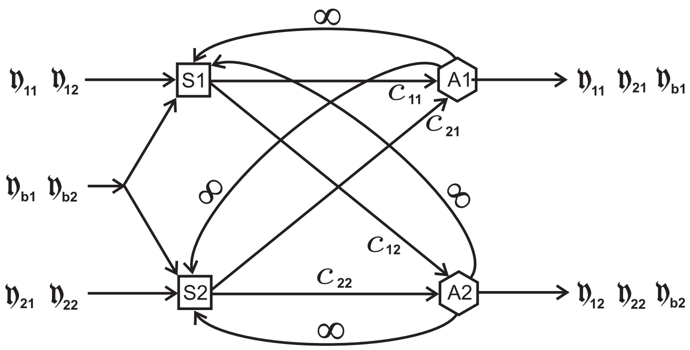


Fig. 9.26. $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus$ **CBE(NW)** $\boxplus \overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$.

Sufficiency. We are going to show that any vector $\boldsymbol{r}$ satisfying (9.5.31) is an achievable rate vector. Modulo the arguments from the first part of the proof of Theorem 3.1 [69] (based on a proper time division multiplexing), this goal will be achieved if we show that the network from Fig. 9.26 is able to transmit simultaneously six continuous stationary fluid flows of respective intensities $\mathfrak{y}_{\nu i} \sim r_{\nu i}, \nu = 1, 2, \mathbf{b}, i = 1, 2$. This will be accomplished via two steps. At the first step, we prove this claim assuming that the "central" input flows are discarded: $\mathfrak{y}_{\mathbf{b}i} := 0$. This shapes (9.5.31) into

$$
\begin{aligned}
&r_{11} + r_{12} \leq c_{11} + c_{12}, \qquad r_{21} + r_{22} \leq c_{21} + c_{22};\\
&r_{11} + r_{21} \leq c_{11} + c_{21}, \qquad r_{12} + r_{22} \leq c_{12} + c_{22};\\
&r_{11} + r_{12} + r_{21} \leq c_{11} + c_{12} + c_{21}, \quad r_{22} + r_{12} + r_{21} \leq c_{22} + c_{12} + c_{21};\\
&r_{11} + r_{22} + r_{12} \leq c_{11} + c_{22} + c_{12};\\
&r_{11} + r_{22} + r_{21} \leq c_{11} + c_{22} + c_{21}, \qquad r_{ij} \geq 0. \qquad (9.5.32)
\end{aligned}
$$

At the second step, we consider the general case of nonzero central flows.

Step 1. Consider a matrix R satisfying (9.5.32) and the distribution

$$
r_{ij} = r'_{ij} + r''_{ij}, \qquad r'_{ij} \geq 0, \quad r''_{ij} \geq 0 \qquad (9.5.33)
$$

of the input flows along the arcs of the network depicted in Fig. 9.27. For example, the flow r_{11} is divided into two subflows. One of them r'_{11} goes directly from S1 to A1. The other r''_{11} goes first to A2, then to S2 over the feedback tunnel, and ultimately from S2 to A1. The conditions necessary and sufficient for the network to be able to accommodate the subflows are as follows:

$$
\begin{aligned}
&r'_{11} + r''_{12} + r''_{21} \leq c_{11}, \quad r'_{22} + r''_{12} + r''_{21} \leq c_{22};\\
&r'_{12} + r''_{11} + r''_{22} \leq c_{12}, \quad r'_{21} + r''_{11} + r''_{22} \leq c_{21}. \qquad (9.5.34)
\end{aligned}
$$

[17] See Definition 9.4.2 (on p. 293).

Thus it should be demonstrated that whenever (9.5.32) holds, the quantities r_{ij} can be splitted (9.5.33) into addends satisfying (9.5.34).

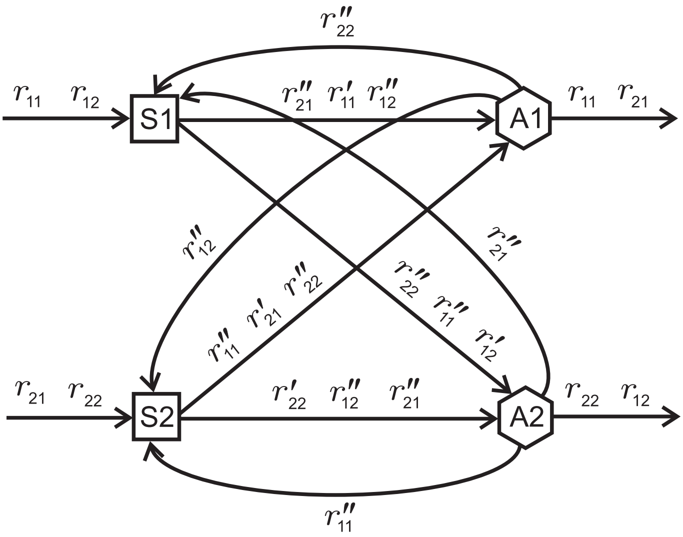


Fig. 9.27. Distribution of the flows.

We shall argue by contradiction. Suppose that this claim fails to be true. Then the convex cone

$$K := \{R : r_{ij} \leq c_{ij} \qquad \forall i, j = 1, 2\}$$

does not intersect the image P of the set described by (9.5.33) (where r_{ij} are given and satisfy (9.5.32)) under the linear operator

$$\left\{(r'_{ij}, r''_{ij})\right\}_{i,j=1,2} \mapsto R := \begin{pmatrix} r'_{11} + r''_{12} + r''_{21} & r'_{12} + r''_{11} + r''_{22} \\ r'_{21} + r''_{11} + r''_{22} & r'_{22} + r''_{12} + r''_{21} \end{pmatrix}. \qquad (9.5.35)$$

It follows that the convex cone K and compact convex set P can be strictly separated by a hyperplane: numbers $\theta_{ij}, i, j = 1, 2$ exist such that

$$\sup_{R \in K} \sum_{ij} \theta_{ij} r_{ij} < \inf_{R \in P} \sum_{ij} \theta_{ij} r_{ij} \qquad (9.5.36)$$

and $\sum_{ij} |\theta_{ij}| > 0$. Since $\sup$ is finite, we have $\theta_{ij} \geq 0$. Hence

$$\sup_{R \in K} \sum_{ij} \theta_{ij} r_{ij} = \sum_{ij} \theta_{ij} c_{ij}.$$

To compute $\inf$, we expand R in the form (9.5.35), invoke (9.5.33), and take into account that $\min_{r', r'' \geq 0 : r' + r'' = r} \theta' r' + \theta'' r'' = r \min\{\theta', \theta''\}$. It follows that (9.5.36) shapes into

$$\sum_{ij} \theta_{ij} c_{ij} < r_{11} \min\{\theta_{11}; \theta_{21} + \theta_{12}\} + r_{22} \min\{\theta_{22}; \theta_{21} + \theta_{12}\}$$
$$+ r_{12} \min\{\theta_{12}; \theta_{11} + \theta_{22}\} + r_{21} \min\{\theta_{21}; \theta_{11} + \theta_{22}\}. \quad (9.5.37)$$

From this, we see that the cone $\mathcal{K} := \{\{\theta_{ij}\}_{i,j=1,2} : \theta_{ij} \geq 0 \; \forall i, j\}$ contains a nonzero solution of the above inequality. This cone can be partitioned into a finite number of convex polyhedral subcones such that the right-hand side of (9.5.37) is linear on any subcone. So (9.5.37) must be satisfied on some extreme ray of some subcone. Any of them is bounded by a finite number of hyperplanes, each described by one of the following equations:

$$\theta_{11} = \theta_{21} + \theta_{12}, \quad \theta_{22} = \theta_{21} + \theta_{12}, \quad \theta_{12} = \theta_{11} + \theta_{22}, \quad \theta_{21} = \theta_{11} + \theta_{22};$$
$$\theta_{ij} = 0, \qquad i, j = 1, 2. \quad (9.5.38)$$

This implies [147, p. 104] that the extreme ray is described by a system of three such equations, which determines its solution uniquely up to multiplication by a scalar. Now we consider separately the possible crews of this system.

- *All three equations are of the form* $\theta_{ij} = 0$. Then the right-hand side of (9.5.37) is zero, whereas the left-hand side is non-negative. Thus we have arrived at a contradiction, as is required.
- *Only two equations are of the form* $\theta_{ij} = 0$. These two equations should be of the form $\theta_{ii} = 0, \theta_{\overline{i}\overline{j}} = 0, \overline{i} \neq \overline{j}$; otherwise, one more such equation is satisfied, and we return back to the first case. It is easy to see that any choice of the third equation either returns the situation back to the first case or implies that $\theta_{i'i'} = \theta_{\overline{i}'\overline{j}'} =: \theta > 0$, where $s' = 1, 2$ is the integer different from $s = 1, 2$. In the latter case, the left- and right-hand sides of (9.5.37) equal $\theta(c_{i'i'} + c_{\overline{i}'\overline{j}'})$ and $\theta(r_{i'i'} + r_{\overline{i}'\overline{j}'})$, respectively. Here we deal with the sums over a common row or column of the matrices C and R. So (9.5.37) contradicts one of the first four inequalities from (9.5.32).
- *Only one equation is the form* $\theta_{ij} = 0$. Since the systems (9.5.32) and (9.5.38) and inequality (9.5.37) are invariant under arbitrary change of the indices $i \mapsto i', j \mapsto j'$, it suffices to consider a particular equation $\theta_{ij} = 0$, say $\theta_{11} = 0$. The other two equations may be only $\theta_{12} = \theta_{11} + \theta_{22}, \theta_{21} = \theta_{11} + \theta_{22}$: Otherwise, one more variable θ_{ij} vanishes, and we return back the the second case. It follows that $\theta_{22} = \theta_{12} = \theta_{21} =: \theta > 0$. So (9.5.37) takes the form $\theta(c_{22} + c_{12} + c_{21}) < \theta(r_{22} + r_{12} + r_{21})$, in violation of the sixth inequality from (9.5.32).
- *There is no equation in the form* $\theta_{ij} = 0$. The equations describing the extreme ray encompass either the group $\theta_{11} = \theta_{21} + \theta_{12}, \theta_{22} = \theta_{21} + \theta_{12}$ or the group $\theta_{12} = \theta_{11} + \theta_{22}, \theta_{21} = \theta_{11} + \theta_{22}$. In the first case, $\theta_{11} = \theta_{22}$. Then depending on the third equation, which necessarily belongs to the second group, either $\theta_{12} = 2\theta_{11}$ or $\theta_{21} = 2\theta_{11}$. Since $\theta_{ij} \geq 0$ on the extreme ray, this contradicts the equations of the first group. The second case is considered likewise. Thus we see that no extreme ray is described in the way under consideration.

Thus (9.5.32) does describe admissible stationary continuous fluid flows through the network from Fig. 9.27.

Step 2. Now we revert to the general case, where the central flows are not necessarily zero (see Fig. 9.26). The idea is to split these flows

$$r_{\mathbf{b}i} = r'_{\mathbf{b}i} + r''_{\mathbf{b}i}, \qquad r'_{\mathbf{b}i}, r''_{\mathbf{b}i} \geq 0 \tag{9.5.39}$$

into subflows $r'_{\mathbf{b}i}, r''_{\mathbf{b}i}$ to the nodes S1 and S2, respectively. At these nodes, these subflows are added to the main flows from these nodes to the corresponding recipients:

$$r_{11} := r_{11} + r'_{\mathbf{b}1}, \quad r_{12} := r_{12} + r'_{\mathbf{b}2}, \quad r_{21} := r_{21} + r''_{\mathbf{b}1}, \quad r_{22} := r_{22} + r''_{\mathbf{b}2}.$$

By making these substitutions in (9.5.32), we get the conditions under which the static flow can be accommodated by the network from Fig. 9.26:

$$\begin{gathered}
r_{11} + r_{12} + r'_{\mathbf{b}1} + r'_{\mathbf{b}2} \leq c_{11} + c_{12}, \qquad r_{21} + r_{22} + r''_{\mathbf{b}1} + r''_{\mathbf{b}2} \leq c_{21} + c_{22}; \\
r_{11} + r_{21} + r_{\mathbf{b}1} \leq c_{11} + c_{21}, \qquad r_{12} + r_{22} + r_{\mathbf{b}2} \leq c_{12} + c_{22}; \\
r_{11} + r_{12} + r_{21} + r_{\mathbf{b}1} + r'_{\mathbf{b}2} \leq c_{11} + c_{12} + c_{21}; \\
r_{22} + r_{12} + r_{21} + r_{\mathbf{b}2} + r''_{\mathbf{b}1} \leq c_{22} + c_{12} + c_{21}; \\
r_{11} + r_{22} + r_{12} + r_{\mathbf{b}2} + r'_{\mathbf{b}1} \leq c_{11} + c_{22} + c_{12}; \\
r_{11} + r_{22} + r_{21} + r_{\mathbf{b}1} + r''_{\mathbf{b}2} \leq c_{11} + c_{22} + c_{21}.
\end{gathered} \tag{9.5.40}$$

Thus it should be shown that whenever (9.5.31) holds, the central flows can be decomposed (9.5.39) in such a way that (9.5.40) is true. The conditions imposed by (9.5.39) and the four last relations from (9.5.40) are equivalent to

$$\begin{aligned}
\max\Big\{0; r_{22} + r_{12} + r_{21} + r_{\mathbf{b}2} + r_{\mathbf{b}1} - c_{22} - c_{12} - c_{21}\Big\} &\leq r'_{\mathrm{b}1} \\
&\leq \min\Big\{r_{\mathbf{b}1}; c_{11} + c_{22} + c_{12} - r_{11} - r_{22} - r_{12} - r_{\mathbf{b}2}\Big\}; \\
\max\Big\{0; r_{11} + r_{22} + r_{21} + r_{\mathbf{b}1} + r_{\mathbf{b}2} - c_{11} - c_{22} - c_{21}\Big\} &\leq r'_{\mathrm{b}2} \\
&\leq \min\Big\{r_{\mathbf{b}2}; c_{11} + c_{12} + c_{21} - r_{11} - r_{12} - r_{21} - r_{\mathbf{b}1}\Big\}.
\end{aligned}$$

Here in both chains of inequalities, the starting term does not exceed the terminal one due to (9.5.31). So it should be checked that the last system of inequalities has a solution $r'_{\mathrm{b}1}, r'_{\mathrm{b}2}$ common with the first two inequalities from (9.5.40):

$$r_{21} + r_{22} + r_{\mathbf{b}1} + r_{\mathbf{b}2} - c_{21} - c_{22} \leq r'_{\mathbf{b}1} + r'_{\mathbf{b}2} \leq c_{11} + c_{12} - r_{11} - r_{12}.$$

This holds if and only if

$$\max\left\{0; r_{22} + r_{12} + r_{21} + r_{\mathbf{b}2} + r_{\mathbf{b}1} - c_{22} - c_{12} - c_{21}\right\}$$
$$+ \max\left\{0; r_{11} + r_{22} + r_{21} + r_{\mathbf{b}1} + r_{\mathbf{b}2} - c_{11} - c_{22} - c_{21}\right\}$$
$$\le c_{11} + c_{12} - r_{11} - r_{12} \qquad \text{and}$$
$$r_{21} + r_{22} + r_{\mathbf{b}1} + r_{\mathbf{b}2} - c_{21} - c_{22}$$
$$\le \min\left\{r_{\mathbf{b}1}; c_{11} + c_{22} + c_{12} - r_{11} - r_{22} - r_{12} - r_{\mathbf{b}2}\right\}$$
$$+ \min\left\{r_{\mathbf{b}2}; c_{11} + c_{12} + c_{21} - r_{11} - r_{12} - r_{21} - r_{\mathbf{b}1}\right\}.$$

At the same time, it is easy to check that these inequalities follow from (9.5.31). This completes the proof of the sufficiency part.

Necessity. Let $\boldsymbol{r}$ be an achievable rate vector. By Definition 9.4.17 (on p. 298), for any $\varepsilon > 0$, an errorless networked block code with block length N and input alphabets $[1 : F_{\nu i}]$ (with elements $\mathfrak{y}_{\nu i}$) exists such that

$$|r_{\nu i} - N^{-1} \log_2 F_{\nu i}| < \varepsilon \qquad \forall \nu = 1, 2, \mathbf{b}, i = 1, 2.$$

According to Fig. 9.26, data about $p = (\mathfrak{y}_{11}, \mathfrak{y}_{12})$ are injected into the network via the perfect channels S1 $\mapsto$ A1 and S1 $\mapsto$ A2 of the total capacity $c_{11} + c_{12}$. So if the total number of such pairs exceeds $2^{N(c_{11}+c_{12})}$, there are pairs that cannot be distinguished from the information taken from the network. Hence $F_{11} \times F_{12} \le 2^{N(c_{11}+c_{12})}$ and so

$$r_{11} + r_{12} \le N^{-1} [\log_2 F_{11} + \log_2 F_{12}] + 2\varepsilon \le c_{11} + c_{12} + 2\varepsilon. \tag{9.5.41}$$

Letting $\varepsilon \to 0$ results in the first inequality from (9.5.31). The second one is established likewise.

The third inequality is obtained via similar argument with focusing attention on the node A1, where $\mathfrak{y}_{11}, \mathfrak{y}_{21}, \mathfrak{y}_{\mathbf{b}1}$ are correctly decoded, and noting that all data accessible at this node ultimately arrive via two channels S1 $\mapsto$ A1 and S2 $\mapsto$ A1 of the total capacity $c_{11} + c_{21}$. The fourth inequality is established likewise.

To justify the fifth inequality, we consider data transmission under the assumption $\mathfrak{y}_{22} \equiv 1, \mathfrak{y}_{\mathbf{b}2} \equiv 1$. Let S_{ji} stand for the sequence of all messages received over the channel S$j \mapsto$ Ai since $t = 0$ until $t = N - 1$. As follows from Fig. 9.26,

$$\begin{array}{l} \mathfrak{y}_{11} = \mathcal{Y}_{11}[S_{11}, S_{21}], \;\; \mathfrak{y}_{12} = \mathcal{Y}_{12}[S_{12}, S_{22}] \\ \mathfrak{y}_{21} = \mathcal{Y}_{21}[S_{11}, S_{21}] \\ \mathfrak{y}_{\mathbf{b}1} = \mathcal{Y}_{\mathbf{b}1}[S_{11}, S_{21}] \end{array}, \quad S_{22} = \mathcal{S}_{22}[\mathfrak{y}_{21}, \mathfrak{y}_{\mathbf{b}1} S_{11}, S_{12}, S_{21}].$$

(We recall that capital script letters denote deterministic functions.) It follows that

$$S_{22} = \mathcal{S}_{22}\left\{\mathcal{Y}_{21}[S_{11}, S_{21}], \mathcal{Y}_{\mathbf{b}1}[S_{11}, S_{21}], S_{11}, S_{12}, S_{21}\right\},$$

and hence $(\mathfrak{y}_{11}, \mathfrak{y}_{12}, \mathfrak{y}_{21}, \mathfrak{y}_{\mathbf{b}1})$ is a function of (S_{11}, S_{12}, S_{21}). The first sequence ranges over the set of the size $F_{11} \times F_{12} \times F_{21} \times F_{\mathbf{b}1}$, whereas the second one belongs to a set of the size $\le 2^{N(c_{11}+c_{12}+c_{21})}$. Thus $F_{11} \times F_{12} \times F_{21} \times F_{\mathbf{b}1} \le 2^{N(c_{11}+c_{12}+c_{21})}$, which results in the fifth inequality from (9.5.31).

The remaining inequalities are established likewise. $\square$

Two Independent Agents

Up to this point, it was assumed that any sensor is able to detect the actions of every actuator. This implies four feedback tunnels in Fig. 9.25b (on p. 311). There also may be the cases where only three or two such tunnels occur. (One tunnel and no tunnels are impossible since the unstable plant is detectable and stabilizable by the entire sets of sensors and actuators, respectively.)[18] Now we consider the case of two feedback tunnels, still assuming that the zones of influence of the actuators are disjoint. By changing enumeration of the sensors, if necessary, this case can be reduced to the situation where sensor Si detects the actions of only the matching actuator Ai. Then the unstable subspace of the plant can be decomposed as follows:

$$M_{\text{unst}}(A) = M_1 \oplus M_2, \qquad (9.5.42)$$

where

- M_i is the subspace of states controllable by the ith actuator, observable by the ith sensor, and nonobservable by the companion sensor.

The observations are still transmitted via the network from Fig.9.25a.

The situation can be interpreted as if there are two independent agents with the state spaces M_1 and M_2, respectively, each equipped with its own actuator and sensor measuring the state of the owner. At first sight, it may seem that then the cross channels S1 $\mapsto$ A2 and S2 $\mapsto$ A1 are useless, and the conditions for stabilizability disintegrate into

$$\log_2 |\det A|_{M_i}| \leq c_{ii}, \qquad i = 1, 2.$$

However, the next proposition shows that this is not the case, and the system can be stabilized even if $c_{11} = c_{22} = 0$.

Proposition 9.5.18. *Whenever the plant is stabilizable via the network at hand, the following relations hold:*

$$\log_2 |\det A|_{M_1}| \leq c_{11}+\min\{c_{12}, c_{21}\}, \quad \log_2 |\det A|_{M_2}| \leq c_{22}+\min\{c_{12}, c_{21}\};$$
$$\log_2 |\det A|_{M_{\text{unst}}(A)}| \leq c_{11} + c_{22} + \min\{c_{12}, c_{21}\}. \qquad (9.5.43)$$

Conversely, if relations (9.5.43) *hold with the strict inequality signs, the plant is regularly stabilizable.*[19]

Proof of Proposition 9.5.18

By Corollary 9.5.13 (on p. 308), the proof is reduced to computation of the capacity domain $\mathbf{CD}\left[\overline{\mathbf{PREF}_{\mathbf{mw}}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}^{+}_{\mathbf{mw}}}\right]$ with respect to the decomposition (9.5.42). The control-based extension **CBE(NW)** is depicted in Fig. 9.28a.

[18] It is tacitly assumed that no actuator or sensor is useless for stabilization: $L_i^{+c} \not\subset M_{\text{st}}(A)$, $L_j^{-o} \not\supset M_{\text{unst}}(A)\ \forall i, j$.

[19] See Definition 9.4.2 (on p. 293).

The reduced prefix $\overline{\mathbf{PREF}}_{\mathbf{mw}}$ and suffix $\overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$ contain two sources and output nodes, respectively. The data from these sources and outputs at these nodes are associated with the subspaces from (9.5.42) and marked by the corresponding indices $\mathfrak{y}_i$. The components of the rate vectors $\boldsymbol{r}$ are enumerated similarly r_i. The network $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$ is shown in Fig. 9.28b.

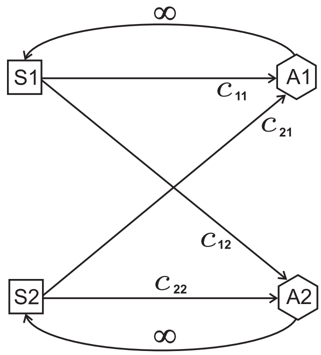


Fig. 9.28(a). Control-based extension.

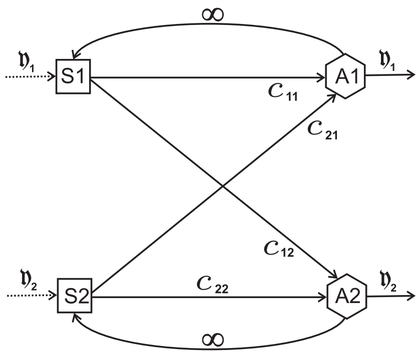


Fig. 9.28(b). The network with the prefix and suffix.

Due to Corollary 9.5.13, it suffices to show that the capacity domain of this network is described by the following set of inequalities:

$$r_1 \le c_{11} + \min\{c_{12}, c_{21}\}, \quad r_2 \le c_{22} + \min\{c_{12}, c_{21}\};$$
$$r_1 + r_2 \le c_{11} + c_{22} + \min\{c_{12}, c_{21}\}. \quad (9.5.44)$$

Sufficiency. We are going to show that any vector $\boldsymbol{r}$ satisfying (9.5.44) is an achievable rate vector. As before, this is reduced to showing that the network from Fig. 9.28b is able to transmit simultaneously two continuous stationary fluid flows of respective intensities $\mathfrak{y}_i \sim r_i, i = 1, 2$. Consider the distribution

$$r_i = r_i' + r_i'', \qquad r_i', r_i'' \ge 0 \quad (9.5.45)$$

of the input flows along the arcs of the network depicted in Fig. 9.29. The conditions necessary and sufficient for the network to be able to accommodate the subflows are

$$r_1' \le c_{11}, \quad r_2' \le c_{22}, \quad r_1'' + r_2'' \le \min\{c_{12}, c_{21}\}. \quad (9.5.46)$$

Let us pick

$$r_i' := \min\{c_{ii}, r_i\} \quad \text{and} \quad r_i'' := r_i - r_i'.$$

Then (9.5.45) and the first two relations from (9.5.46) are true. The third one takes the form

$$r_1 - \min\{c_{11}, r_1\} + r_2 - \min\{c_{22}, r_2\}$$
$$= \max\{r_1 - c_{11}, 0\} + \max\{r_2 - c_{22}, 0\} \le \min\{c_{12}, c_{21}\}$$

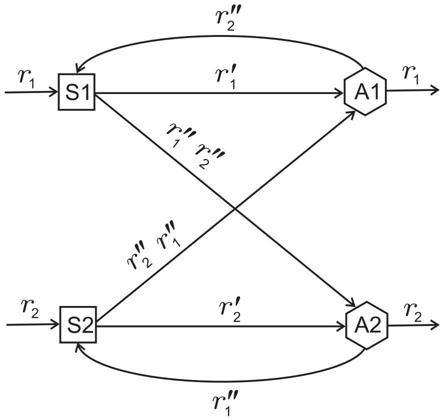


Fig. 9.29. Distribution of the flows.

and is immediate from (9.5.44).

Necessity. Let $\boldsymbol{r}$ be an achievable rate vector. By Definition 9.4.17 (on p. 298), for any $\varepsilon > 0$, an errorless networked block code with block length N and input alphabets $[1 : F_i]$ (with elements $\mathfrak{y}_i$) exists such that $|r_i - N^{-1} \log_2 F_i| < \varepsilon \; \forall i = 1, 2$. Let S_{ji} stand for the sequence of all messages received over the channel S$j \mapsto$ Ai since $t = 0$ until $t = N - 1$. According to Fig. 9.28b,

$$\begin{aligned} S_{11} &= \mathcal{S}_{11}[\mathfrak{y}_1, S_{21}], \quad & S_{12} &= \mathcal{S}_{12}[\mathfrak{y}_1, S_{21}], \quad & \mathfrak{y}_1 &= \mathcal{Y}_1[S_{11}, S_{21}]; \\ S_{21} &= \mathcal{S}_{21}[\mathfrak{y}_2, S_{12}], \quad & S_{22} &= \mathcal{S}_{22}[\mathfrak{y}_2, S_{12}], \quad & \mathfrak{y}_2 &= \mathcal{Y}_2[S_{12}, S_{22}]. \end{aligned}$$

Under the assumption $\mathfrak{y}_2 \equiv 1$,

$$S_{12} = \mathcal{S}_{12}\left[\mathcal{Y}_1[S_{11}, S_{21}], S_{21}\right] =: \overline{\mathcal{S}}_{12}[S_{11}, S_{21}], \quad S_{21} = \mathcal{S}_{21}[1, S_{12}] =: \overline{\mathcal{S}}_{21}[S_{12}].$$

Thus given S_{11}, the sequences S_{12} and S_{21} are computable from each other. They belong to sets of the sizes $2^{Nc_{12}}$ and $2^{Nc_{21}}$, respectively. Hence every of them in fact lies in a set of cardinality $2^{N \min\{c_{12}, c_{21}\}}$. It follows that the pair $[S_{11}, S_{21}]$ takes no more than $2^{Nc_{11}} \times 2^{N \min\{c_{12}, c_{21}\}}$ values. Since $\mathfrak{y}_1 \in [1 : F_1]$ is a function of this pair, we arrive at the first relation from (9.5.44) by letting $\varepsilon \to 0$, just like in (9.5.41). The second relation is established likewise by putting $\mathfrak{y}_1 \equiv 1$ and considering arbitrary $\mathfrak{y}_2 \in [1 : F_2]$. The third relation is justified by the same argument applied to arbitrary $\mathfrak{y}_i \in [1 : F_i], i = 1, 2$ by noting that then S_{12} and S_{21} are computable from each other given both S_{11} and S_{22}. □

Actuators with Identical Zones of Influence

In this case, the plant is stabilizable by any actuator (since the entire set of actuators stabilizes it). We still suppose that no sensor is useless for stabilization: $L_j^{-o} \not\supset M_{\text{unst}}(A)$. The unstable subspace is decomposed as follows:

$$M_{\text{unst}}(A) = M_1 \oplus M_{\mathbf{b}} \oplus M_2, \tag{9.5.47}$$

where

- $M_j, j \neq \mathbf{b}$ is the subspace of states observable by the jth sensor and nonobservable by the companion sensor;
- $M_{\mathbf{b}}$ is the subspace of states observable by both sensors.

Proposition 9.5.19. *Whenever the plant is stabilizable via the network at hand, the following relations hold:*

$$\log_2 |\det A|_{M_1}| \leq c_{11} + c_{12}, \quad \log_2 |\det A|_{M_2}| \leq c_{21} + c_{22};$$
$$\log_2 |\det A|_{M_{\text{unst}}(A)}| \leq c_{11} + c_{22} + c_{12} + c_{21}. \tag{9.5.48}$$

Conversely, if relations (9.5.48) *hold with the strict inequality signs, the plant is regularly stabilizable.*[20]

Proof of Proposition 9.5.19

Now there are four feedback tunnels, and the control-based extension of the network is depicted in Fig. 9.25b (on p. 311). We are going to employ Corollary 9.5.12 (on p. 307) related to the decomposition (9.5.47).

Sufficiency. Let (9.5.48) be true with the strict inequality signs. To prove stabilizability, it suffices to show that **a**′) of Corollary 9.5.12 (on p. 307) holds. In service of the data transmission scheme $\overline{\overline{\mathfrak{T}}}_{\mathbf{mw}}$ (see p. 307), the network $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)}$ from **a**′) is depicted in Fig. 9.26 (on p. 313). So the capacity domain $\mathbf{CD}\left[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \overline{\overline{\mathfrak{T}}}_{\mathbf{mw}}\right]$ is described by (9.5.31). So it suffices to show that the numbers $r_i := \log_2 |\det A|_{M_i}|$, $r_{\mathbf{b}} := \log_2 |\det A|_{M_{\mathbf{b}}}|$ can be decomposed

$$r_i := r_{i1} + r_{i2}, \quad r_{i1}, r_{i2} \geq 0, \qquad i = 1, 2, \mathbf{b} \tag{9.5.49}$$

into addends satisfying (9.5.31) with the strict inequality signs. Note that the initial assumption of this part of the proof means that

$$r_1 \leq c_{11} + c_{12}, \quad r_2 \leq c_{21} + c_{22}, \quad r_1 + r_2 + r_{\mathbf{b}} \leq c_{11} + c_{22} + c_{12} + c_{21}, \tag{9.5.50}$$

where all signs are strict. However, it suffices to show that (9.5.50) implies (9.5.31) in the case of nonstrict inequalities. (Then the claim for the strict inequalities follows by a small perturbation of c_{ij}'s.)

We decompose r_1 and r_2 as follows:

$$\begin{aligned} r_{11} &:= \min\{r_1; c_{11}\}, & r_{12} &:= \max\{0; r_1 - c_{11}\}; \\ r_{21} &:= \max\{0; r_2 - c_{22}\}, & r_{22} &:= \min\{r_2; c_{22}\}. \end{aligned} \tag{9.5.51}$$

This is a decomposition indeed since

[20] See Definition 9.4.2 (on p. 293).

$$r_{i1}+r_{i2} = \min\{r_i;c_{ii}\}+\max\{0;r_i-c_{ii}\} = \min\{r_i;c_{ii}\}+r_i+\max\{-r_i;-c_{ii}\}$$
$$= \min\{r_i;c_{ii}\}+r_i-\min\{r_i;c_{ii}\} = r_i.$$

Thanks to (9.5.49), the first two inequalities in (9.5.31) and (9.5.50) are identical. We rewrite the remaining inequalities from (9.5.31) as conditions on $r_{\mathbf{b}1}$ and $r_{\mathbf{b}2}$:

$$0 \le r_{\mathbf{b}1} \le \min\Big\{c_{11}+c_{21}-r_{11}-r_{21}; \quad c_{11}+c_{12}+c_{21}-r_{11}-r_{12}-r_{21}; \quad c_{11}+c_{22}+c_{21}-r_{11}-r_{22}-r_{21}\Big\};$$

$$0 \le r_{\mathbf{b}2} \le \min\Big\{c_{12}+c_{22}-r_{12}-r_{22}; \quad c_{22}+c_{12}+c_{21}-r_{22}-r_{12}-r_{21}; \quad c_{11}+c_{22}+c_{12}-r_{11}-r_{22}-r_{12}\Big\}.$$

Here the expressions on the right are non-negative, since due to (9.5.51),

$$r_{ii} \le c_{ii}, \quad r_{12} := \max\{0;r_1-c_{11}\} \overset{(9.5.50)}{\le} \max\{0;c_{11}+c_{12}-c_{11}\} = c_{12}; \quad \text{and} \quad r_{21} \le c_{21} \tag{9.5.52}$$

by the similar argument. It follows that r_{b} can be decomposed into $r_{\mathrm{b}1}$ and $r_{\mathrm{b}2}$ so that the remaining inequalities from (9.5.31) are satisfied if and only if

$$r_{\mathbf{b}} \le \min\Big\{c_{11}+c_{21}-r_{11}-r_{21}; \quad c_{11}+c_{12}+c_{21}-r_1-r_{21}; \quad c_{11}+c_{22}+c_{21}-r_{11}-r_2\Big\}$$
$$+\min\Big\{c_{12}+c_{22}-r_{12}-r_{22}; \quad c_{22}+c_{12}+c_{21}-r_2-r_{12}; \quad c_{11}+c_{22}+c_{12}-r_1-r_{22}\Big\}.$$

Thus it should be shown that $r_{\mathbf{b}}$ does not exceed any sum of the form $P_i+Q_j, i,j = 1,2,3$, where P_i and Q_i is the ith expression in the first and second min, respectively.

- For $i=j=1$, the inequality $r_{\mathbf{b}} \le P_i + Q_j$ is a reformulation of the third inequality in (9.5.50).
- For $i=j=2,3$, the inequality results from estimating r with the double index by (9.5.52) and the third inequality in (9.5.50).

- In any sum of the form $P_1 + Q_j$ or $Q_1 + P_j$ with $j \geq 2$, there are three r's with double indices. The sum of two of them equals either r_1 or r_2. The inequality results from estimating the remaining r with the double index by (9.5.52) and the third inequality in (9.5.50).
- $$P_2 + Q_3 = c_{11} + c_{12} + c_{21} - r_1 - r_{21} + c_{11} + c_{22} + c_{12} - r_1 - r_{22}$$
$$\left[c_{11} + c_{12} - r_1\right] + \left[c_{11} + c_{22} + c_{12} + c_{21} - r_1 - r_2\right] \overset{(9.5.50)}{\geq} r_{\mathbf{b}}.$$
- The sum $P_3 + Q_2$ is estimated likewise. □

Necessity. Suppose that the plant is stabilizable via the network at hand. By **d′)** of Corollary 9.5.12 (on p. 307), it suffices to show that the capacity domain $\mathbf{CD}\left[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}\right]$ related to the decomposition (9.5.47) lies in the set described by (9.5.50). The concerned network is depicted in Fig. 9.30. The first inequality in (9.5.50) holds since the data stream with the rate r_1 is injected into the network via the perfect channels S1 $\mapsto$ A1 and S1 $\mapsto$ A2 of the total capacity $c_{11} + c_{12}$. The second inequality is established likewise. The third one holds since all input data are correctly decoded from the messages received by A1 and A2 over four channels from Fig. 9.25a (on p. 311) of the total capacity $c_{11} + c_{12} + c_{21} + c_{22}$. □

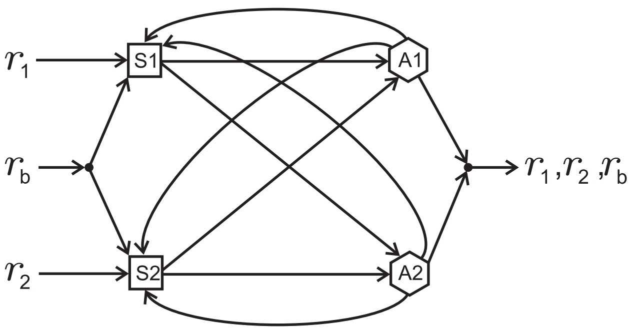


Fig. 9.30. $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$.

Remark 9.5.20. It follows that in the last example, the three domains from (9.5.22) (on p. 305) are identical, although the actuators have intersecting zones of influence. This also implies that (9.4.14) (on p. 298) holds with $=$ substituted in place of $\subset$.

Remark 9.5.21. The general situation in the case of four feedback tunnels, as well as the case of three tunnels can be considered likewise.

9.6 Proof of the Necessity Part of Theorem 9.4.27

In this section, we prove the **c)** $\Rightarrow$ **d)** part of this theorem stated on p. 300. So we assume that **c)** holds; i.e., a control strategy $\mathcal{A} \in \mathfrak{A}$ exists that makes the closed-loop system stable. To justify **d)**, this strategy will be transformed into a data processing strategy for the modified network, which ensures an errorless communication of data from multiple informants at the rates close to those from (9.4.15) (on p. 300). In doing so, we consider the closed-loop system with no noises and initial state x^0 with zero stable modes:

$$\xi(t) \equiv 0, \qquad \chi_j(t) \equiv 0, \qquad x^0_\alpha = 0 \quad \text{whenever} \quad \alpha \leq 0. \tag{9.6.1}$$

9.6.1 Plan on the Proof

The above transformation will be via three steps.

Step 1) We consider the network $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)}$ (see Fig. 9.31) assuming that

i) Its sources (associated with unstable modes x_α) produce constant messages $\mathfrak{y}_\alpha(t) \equiv x^0_\alpha$, which are the modes of the initial state x^0;

ii) The outputs of $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)}$ are controls $u_i(t)$, like in **NW**.

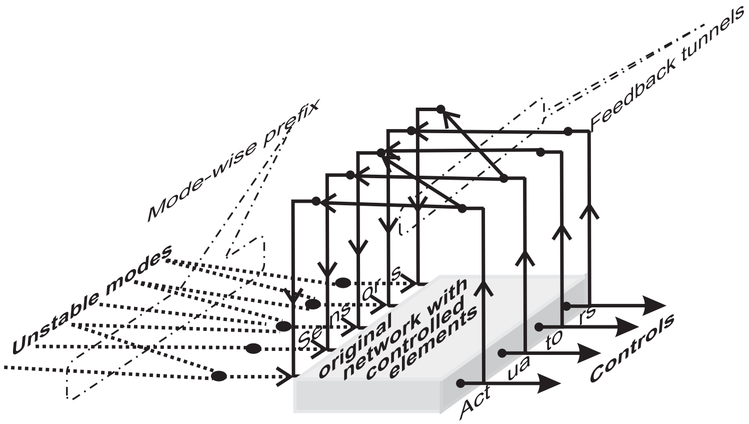


Fig. 9.31. Intermediate network.

We show that $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)}$ can be endowed with an admissible data processing strategy, which at any time t generates just the same controls as **NW** produces in the closed-loop interconnection with the plant.

Step 2) We consider the network $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF_{mw}}$ (see Fig. 9.19 on p. 297) assuming that i) still holds and the output nodes (associated with unstable modes x_α) produce estimates $\widehat{x}_\alpha(t)$ of the current diagonalized modes $x^d_\alpha(t)$ related to the open-loop plant ($u_i \equiv 0$). We show that this network can be endowed with an admissible data processing strategy for which the estimation error grows no faster than polynomially as time progresses.

Explanation 9.6.1. For the open-loop plant, the diagonalized unstable mode $x^d_\alpha(t)$ is defined to be

$$x_\alpha^d(t) := \lceil \lambda_\alpha \rceil^t \, x_\alpha^0, \quad \text{where } \lceil \lambda \rceil := \begin{cases} \lambda & \text{if } \lambda \text{ is real} \\ |\lambda| & \text{otherwise} \end{cases}. \tag{9.6.2}$$

Whenever the unstable eigenvalues are real and distinct, these modes are identical to the ordinary ones $x_\alpha^d(t) \equiv x_\alpha^0(t)$, where $x^0(t) := A^t x^0$.

Step 3) We convert this strategy into an errorless networked block code (for the network from **d)** on p. 300) with the rate vector arbitrarily close to that from **d)**.

After this, the proof of **d)** is completed by Definition 9.4.17 (on p. 298).

9.6.2 Step 1: Network Converting the Initial State into the Controls

The network $\mathbf{NW}_1 := \mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)}$ consists of

- The interior part of the original network **NW**;
- The sources of $\mathbf{NW}_1$ associated with the unstable modes x_α;
- The output actuator nodes of the original network;
- The input sensor nodes of the original network.

By Assumption 9.3.4 (on p. 285), operation algorithms for all these parts can be chosen separately. Specifically, we pick them as follows.

For the interior part of the original network NW, the algorithm of data processing is kept unchanged as compared with the original network **NW**.

Every source of $\mathbf{NW}_1$ constantly produces the mode value x_α^0. Thus all related sensor nodes of **NW** (given by (9.4.9) on p. 295) become aware of x_α^0.

Observation 9.6.2. *As follows from* **ii)** *of Assumption* 9.4.6 *and* (9.4.5)*,* (9.4.8)*,* (9.4.9)*, and* (9.6.1) *(on pp.* 293–295*, and* 324*), the jth sensor node does not receive only the modes that are ignored by this sensor.*

In other words, let $\boldsymbol{x}_j^0$ be the vector of the data received by the jth sensor node from all sources of $\mathbf{NW}_1$. This vector has the form (9.4.8) (on p. 295), where all missing data are replaced by zeros and all received modes are put in the proper places. Then

$$C_j A^t x^0 = C_j A^t \boldsymbol{x}_j^0 \qquad \forall t = 0, 1, \ldots. \tag{9.6.3}$$

For every output node of the original network, the algorithm is kept unchanged with respect to the memory update, the outer output (control $u_i(t)$), and the signal emitted into the original **NW**. Furthermore, the current control is emitted into all feedback tunnels (see Fig. 9.17b on p. 292) outgoing from this node.[21]

For every input node of the original network, the incoming data flows are like in the original network **NW** minus the former outer input $y_j(t)$ but plus $\boldsymbol{x}_j^0$ and data received over the feedback tunnels. Via these tunnels, the node receives u_i with the delay $d_{i \to j}$ for any $i : (i, j) \in$ CVP, where CVP and $d_{i \to j}$ are given by (9.4.6) and

[21] This can be accomplished since the tunnel alphabet has the maximal possible cardinality.

(9.4.7) (on p. 294), respectively. The idea is to design a data processing algorithm so that despite the changes in the incoming data, the output produced by this node and its memory do not change as compared with **NW**. To this end, it suffices to compute $y_j(t)$ from the available data.

This is possible since

$$
\begin{aligned}
y_j(t) &\overset{(9.4.1),\,(9.4.2),\,(9.6.1)}{=\!=\!=} C_j A^t x^0 + \sum_{\theta=0}^{t-1}\sum_{i=1}^{l} C_j A^{t-1-\theta} B_i u_i(\theta) \\
&\overset{(9.4.5),\,(9.4.6)}{=\!=\!=} C_j A^t x^0 + \sum_{i:(i,j)\in \mathrm{CVP}} \sum_{\theta=0}^{t-1} C_j A^{t-1-\theta} B_i u_i(\theta) \\
&\overset{(9.4.7),\,(9.6.3)}{=\!=\!=} C_j A^t \boldsymbol{x}_j^0 + C_j \left(\sum_{i:(i,j)\in \mathrm{CVP}} \sum_{\theta=0}^{t-d_{i\to j}} A^{t-1-\theta} B_i u_i(\theta) \right). \qquad (9.6.4)
\end{aligned}
$$

Remark 9.6.3. More formally, the memory of the node is reorganized to encompass the former one and accommodate two vectors $x_j^0(t), x_j^u(t) \in \mathbb{R}^{\dim(x)}$. Initially, $x_j^0(0) := \boldsymbol{x}_j^0, x_j^u(0) := 0$. These vectors are updated recursively:

$$
x_j^0(t{+}1) := A x_j^0(t), \quad x_j^u(t{+}1) := A x_j^u(t) + \sum_{i:(i,j)\in \mathrm{CVP}} A^{d_{i\to j}-1} B_i u_i(t{+}1{-}d_{i\to j}),
$$

where $u_i(\theta) := 0\, \forall \theta < 0$. The content m_j of the former memory and the node output are generated by the former rules (9.3.4) (on p. 286). To accomplish this, the former external input is computed in correspondence with (9.6.4)

$$
y_j(t) := C_j [x_j^0(t) + x_j^u(t)].
$$

Thus the data processing strategy for the network $\mathbf{PREF}_{\mathbf{mw}} \boxplus$ **CBE(NW)** is completely defined. Its properties are described by the following.

Lemma 9.6.4. *Whenever* (9.6.1) *holds, the process in the original network* **NW** *is coherent with that in* $\mathbf{PREF}_{\mathbf{mw}} \boxplus$ **CBE(NW)**. *Specifically, the controls* $u_i(t)$ *are the same, and both networks produce equal time sequences of variables related to the interior part* $\boldsymbol{O}^-, \boldsymbol{I}^-, \boldsymbol{M}^-$ *and any contact node* I_h, O_h, m_h *of the original network.*

Proof. The proof is by merely checking via the induction on t. □

9.6.3 Step 2: Networked Estimator of the Open-Loop Plant Modes

Preliminaries. We start with an explicit formula giving the relation between the initial state x^0 and the diagonalized unstable modes (9.6.2). In the case where the unstable eigenvalues are real and distinct, this relation is trivial $x_\alpha^d(t) = [A^t x^0]_\alpha$. Here $[x]_\alpha$ is the corresponding coordinate from the representation (9.4.8) (on p. 295) of the vector x. So the following lemma is of interest only if this case does not hold.

Lemma 9.6.5. *Put* $n := \dim(x)$*. An* $n \times n$*-matrix* W *exists such that* $[(WA)^t x^0]_\alpha = x^d_\alpha(t)\ \forall \alpha \in [1 : n^+]$*, the spectrum of* W *is a subset of the unit circle* $\{\lambda \in \mathbb{C} : |\lambda| = 1\}$ *of the complex plane,* $WA = AW$*, and any subspace* M_ν *from Assumption* 9.4.6 *(on p.* 294*) is* W*-invariant.*

Proof. It suffices to define the linear operator W by putting

$$Wx := \lceil \lambda_\nu \rceil \left(A|_{M_\nu}\right)^{-1} x \quad \text{for} \quad x \in M_\nu, \nu = 1, \dots, g,$$
$$Wx := x \quad \text{for} \quad x \in M_{\text{st}}(A). \qquad \square$$

Data processing strategy for the network $\mathbf{NW}_2 :=$ $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF_{mw}}$ should be designed so that its outputs (associated with unstable modes) $\mathfrak{u}_\alpha(t) =: \widehat{x}_\alpha(t)$ track the diagonalized open-loop modes (9.6.2).

This network includes that $\mathbf{NW}_1$ considered at Step 1: $\mathbf{NW}_2 = \mathbf{NW}_1 \boxplus \mathbf{SUFF_{mw}}$. However, the output nodes of $\mathbf{NW}_1$ (associated with actuators and producing controls u_i) are inner for $\mathbf{NW}_2$ and thus produce no outer outputs for $\mathbf{NW}_2$.

Within the $\mathbf{NW}_1$ part of $\mathbf{NW}_2$, the data processing strategy is inherited from Step 1 with adding the following extra operations at the actuator nodes.

For any actuator node, the memory is reorganized to include the former one m_i and accommodate a vector $x^{u,i} \in \mathbb{R}^{\dim(x)}$ updated recursively by the rule

$$x^{u,i}(t+1) = WAx^{u,i}(t) - W^{t+1} B_i u_i(t), \qquad x^{u,i}(0) = 0. \tag{9.6.5}$$

The vector $x^{u,i}(t)$ is decomposed into modes in accordance with (9.4.8) (on p. 295). Into every additive channel (from $\mathbf{SUFF_{mw}}$) that concerns this node, the corresponding mode $x^{u,i}_\alpha(t)$ is emitted. (By the definition of $\mathbf{SUFF_{mw}}$, these channels are enumerated by α such that $i \in \mathfrak{I}^{mw}_{\to\alpha}$, where $\mathfrak{I}^{mw}_{\to\alpha}$ is given by (9.4.10) on p. 296.)

The output node of $\mathbf{NW}_2$ associated with the mode x_α merely emits the signal received over the attached additive channel (from $\mathbf{SUFF_{mw}}$) as its outer output $\widehat{x}_\alpha(t)$.

Thus the data processing strategy for the network $\mathbf{NW}_2$ is completely defined. Its properties are described by the following.

Lemma 9.6.6. *Whenever* (9.6.1) *holds, the outputs* $\widehat{x}_\alpha(t)$ *of* $\mathbf{NW}_2$ *track the diagonalized open-loop unstable modes with errors of polynomial growths in time:*

$$\sup_{|x^0_1| \le D_x, \dots, |x^0_{n^+}| \le D_x} \left| \lceil \lambda_\alpha \rceil^t x^0_\alpha - \widehat{x}_\alpha(t) \right| \le c_\alpha t^{s_\alpha} \qquad \forall\, t \ge 0, \alpha \ge 1, \tag{9.6.6}$$

where D_x *is the constant from* (9.4.3) *(on p.* 292*)*

Explanation 9.6.7. Here and throughout the section, we employ the norm of a vector $\|x\| := \max_\alpha |x_\alpha|$, where x_α are taken from the representation (9.4.8) (on p. 295), and the related matrix norm $\|M\| := \max_{\|x\|=1} \|Mx\|$. The estimates (9.4.3) (on p. 292) on the plant noises and initial state remain true for this norm with, maybe, increased upper bounds.

Proof of Lemma 9.6.6. By Assumption 9.4.6 and (9.4.5) and (9.4.10) (on pp. 293 and 296), the mode $x_\alpha(t)$ is not influenced by controls u_i with $i \notin \mathfrak{I}^{mw}_{\to\alpha}$. Moreover, this property extends on $[W^t x(t)]_\alpha$ due to ii) of Assumption 9.4.6 and the last claim from Lemma 9.6.5. So

$$[W^t x(t)]_\alpha \overset{(9.4.1)}{=\!=} [W^t A^t x^0]_\alpha + \left[\sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} \sum_{\theta=0}^{t-1} W^t A^{t-1-\theta} B_i u_i(\theta) \right]_\alpha \overset{\text{Lemma 9.6.5}}{=\!=}$$

$$= \lceil \lambda_\alpha \rceil^t \, x^0_\alpha + \sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} \left[\sum_{\theta=0}^{t-1} (WA)^{t-1-\theta} W^{\theta+1} B_i u_i(\theta) \right]_\alpha$$

$$\overset{(9.6.5)}{=\!=} \lceil \lambda_\alpha \rceil^t \, x^0_\alpha - \sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} x^{u,i}_\alpha(t) = \lceil \lambda_\alpha \rceil^t \, x^0_\alpha - \widehat{x}_\alpha(t), \quad (9.6.7)$$

where the last equality follows from the definition of **SUFF$_{\mathbf{mw}}$**. So

$$\left| \lceil \lambda_\alpha \rceil^t \, x^0_\alpha - \widehat{x}_\alpha(t) \right| \leq \|W^t\| \|x(t)\|.$$

The proof is completed by noting that $\|x(t)\|$ is bounded by (9.4.4) (on p. 292), whereas $\|W^t\|$ is of a polynomial growth since the spectrum of W is a subset of the unit circle (see, e.g., [213]). □

Remark 9.6.8. As follows from the proof, the error is bounded, i.e., $s_\alpha = 0$ in (9.6.6) whenever the geometric and algebraic multiplicities of the eigenvalue λ_α are equal.

The next observation is required to prove Remark 9.4.28 (on p. 300) and is not used elsewhere.

Observation 9.6.9. *Suppose that the operations* (9.6.5) *at the actuator nodes are altered as follows:*

$$x^{u,i}(t+1) = Ax^{u,i}(t) - B_i u_i(t), \qquad x^{u,i}(0) = 0.$$

Then the arguments from the proof of Lemma 9.6.6 *show that*

$$\left[A^t x^0 - \sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} x^{u,i}(t) \right]_\alpha \in [-D_\infty, D_\infty] \qquad \forall \alpha, t,$$

where D_∞ denotes the supremum (9.4.4) *(on p.* 292*).*

Modified strategy for the network NW$_2$. The rate vector from **d)** of Theorem 9.4.27 assumes that the communication within any informant–recipient pair related to a mode x_α with $|\lambda_\alpha| = 1$ is at the zero rate. By Definition 9.4.15 (on p. 297), this is ensured by any strategy serving the constant output 1 at the source $\sim x_\alpha$.

Now we modify the strategy to make $\widehat{x}_\alpha(t) \equiv 1$ for $\alpha : |\lambda_\alpha| = 1$. To this end, the signal emitted from any actuator node $i \in \mathfrak{I}^{mw}_{\to\alpha}$ to the aditive channel coming to the output node $\sim x_\alpha$ is changed into $|\mathfrak{I}^{mw}_{\to\alpha}|^{-1}$. Note that this keeps (9.6.6) true.

9.6.4 Step 3: Completion of the Proof of c) ⇒ d) Part of Theorem 9.4.27

Now we convert the modified data processing algorithm introduced at Step 2 into a networked block code hosted by the network $\mathbf{NW}_3 = \mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^q_{mw}}$. It differs from that $\mathbf{NW}_2 = \mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF_{mw}}$ considered at Step 2 only by the suffix. The conversion will be via

- altering the data produced by the sources,
- changing the algorithms at the actuator nodes of **NW**,
- altering the algorithms at the output nodes of $\mathbf{SUFF_{mw}}$.

To start the conversion, we pick a time $\tau \geq 1$, which determines the length $\tau + 1$ of the networked block code to be constructed.

Change of data from the sources. The new signal $\mathfrak{y}_\alpha(t)$ from the source $\sim x_\alpha$ is chosen as an integer $\mathfrak{y}_\alpha(t) \equiv \mathfrak{y}_\alpha \in [1 : F_\alpha]$ to meet the requirements of Definition 9.4.15 (on p. 297). Here $F_\alpha := 1$ if $|\lambda_\alpha| = 1$, and otherwise, $F_\alpha := \lfloor (|\lambda_\alpha| - \varkappa_\alpha)^\tau \rfloor$, where $0 < \varkappa_\alpha < |\lambda_\alpha| - 1, \varkappa_\alpha \approx 0$. If $|\lambda_\alpha| = 1$, we set $\varkappa_\alpha := 0$. When getting the signals $\mathfrak{y}_\alpha \in [1 : F_\alpha]$, the sensor nodes of **NW** convert them into the inputs $x^0_\alpha(\mathfrak{y}_\alpha) := \mathfrak{y}_\alpha(|\lambda_\alpha| - \varkappa_\alpha)^{-\tau} D_x \leq D_x$ acceptable by $\mathbf{NW}_2$. Then these inputs are processed in accordance with the modified strategy of $\mathbf{NW}_2$ with the following alterations at the output nodes.

At the ith actuator node of NW, the signal emitted into the additive channel to any output node $\sim x_\alpha$ with $|\lambda_\alpha| > 1$ is amplified via multiplying by $\big(|\lambda_\alpha| - \varkappa_\alpha\big)^\tau \times \lceil \lambda_\alpha \rceil^{-\tau} D_x^{-1}$. This changes the output $\widehat{x}_\alpha(t)$ of this output node:

$$\textbf{original}\,[\widehat{x}_\alpha(t)] = \textbf{new}\,[\widehat{x}_\alpha(t)] \times \left(\frac{\lceil \lambda_\alpha \rceil}{|\lambda_\alpha| - \varkappa_\alpha} \right)^\tau D_x. \tag{9.6.8}$$

For any output node $\sim x_\alpha$ the algorithm is altered in accordance with the definition of $\mathbf{SUFF^q_{mw}}$. In other words, the output $\mathfrak{u}_\alpha(t)$ of $\mathbf{NW}_3$ is obtained by projecting the output $\widehat{x}_\alpha(t)$ of $\mathbf{NW}_2$ into the nearest integer (with producing the “no decision” output $\circledast$ if $\widehat{x}_\alpha(t) - 1/2$ is integer).

By the definition of the modified strategy for $\mathbf{NW}_2$, this ensures correct recognition $\mathfrak{u}_\alpha(\tau) = 1$ of the input message 1 if $|\lambda_\alpha| = 1$. The same is true if $|\lambda_\alpha| > 1$ and τ is large enough. Indeed, by invoking (9.6.6) and (9.6.8), we have

$$\left| \lceil \lambda_\alpha \rceil^\tau \, x^0_\alpha(\mathfrak{y}_\alpha) - \widehat{x}_\alpha(\tau) \times \left(\frac{\lceil \lambda_\alpha \rceil}{|\lambda_\alpha| - \varkappa_\alpha} \right)^\tau D_x \right| \leq c_\alpha \tau^{s_\alpha}.$$

With regard to the definition of $x^0_\alpha(\mathfrak{y}_\alpha)$, this implies that

$$|\mathfrak{y}_\alpha - \widehat{x}_\alpha(\tau)| \leq \frac{c_\alpha \tau^{s_\alpha}}{D_x} \times \left(\frac{|\lambda_\alpha| - \varkappa_\alpha}{|\lambda_\alpha|} \right)^\tau \to 0 \quad \text{as} \quad \tau \to \infty.$$

Thus for all sufficiently large τ the right-hand side of the last inequality is less than $1/2$. So the projection $\mathfrak{u}_\alpha(\tau)$ of $\widehat{x}_\alpha(\tau)$ into the nearest integer equals $\mathfrak{y}_\alpha$. This means that $\mathbf{NW}_3$ considered on $[0 : \tau]$ is an errorless networked block code. Thus we see that the rate vector of this code

$$\mathbf{col}\left(\frac{\log_2\lfloor(|\lambda_1|-\varkappa_1)^\tau\rfloor}{\tau+1},\ldots,\frac{\log_2\lfloor(|\lambda_{n^+}|-\varkappa_{n^+})^\tau\rfloor}{\tau+1}\right)$$
$$\xrightarrow{\tau\to\infty}\mathbf{col}\left(\log_2[|\lambda_1|-\varkappa_1],\ldots,\log_2[|\lambda_{n^+}|-\varkappa_{n^+}]\right).$$

Letting $\varkappa_\alpha \to 0$ whenever $|\lambda_\alpha| > 1$ and taking into account Definition 9.4.17 (on p. 298) completes the proof of **d)** from Theorem 9.4.27 (on p. 300). □

9.7 Proof of the Sufficiency Part of Theorem 9.4.27

Now we prove the **a)** ⇒ **b)** part of this theorem stated on p. 300. So we assume that (9.6.1) (on p. 324) does not hold any longer and **a)** is true. The latter means that there is a networked block code **BC** hosted by $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^q_{mw}}$ whose rate vector (9.4.13) (on p. 298)[22] dominates the vector (9.4.15) (on p. 300)

$$\frac{\log_2 F_\alpha}{N+1} > \log_2|\lambda_\alpha| + \delta \qquad \forall \alpha = 1,\ldots,n^+, \quad \text{where } \delta > 0. \tag{9.7.1}$$

To stabilize the system, this code will be used to communicate data from the sensors to actuators. The code accepts only discrete data, which every sensor should extract from its own measurements. Moreover the definition of $\mathbf{PREF_{mw}}$ (see Fig. 9.18a on p. 296) assumes that the transmission by means of the block code is successful only if all sensors observing a given unstable mode x_α extract identical discrete data about x_α.

Is this possible in the case of independent sensor noises? This question is addressed in Subsect. 9.7.1. We show that the answer is in the negative if there is no information exchange between the sensors. However it is reversed to the affirmative if at least one of the above sensors is able to communicate 2 bits of information to the others. In Subsect. 9.7.2, we show that the last possibility does occur and is ensured by the very existence of the block code **BC** with the mentioned properties. The main of them is that the rate vector has nonzero entries.

Subsection 9.7.3 offers a stabilizing control strategy hosted by the network control-based extension **CBE(NW)**. We recall that **CBE(NW)** is an artificially constructed network, which includes the real one **NW** but also contains imaginary channels (the feedback tunnels). These channels are not real beings but symbols to express the possibility to communicate data by means of control (see Subsect. 7.8.1 starting on p. 233). In concluding Subsect. 9.7.4, this possibility is explicitly employed, and the stabilizing strategy is converted into one hosted by the real network **NW**, which completes the proof of **b)**.

9.7.1 Synchronized Quantization of Signals from Independent Noisy Sensors

To highlight the ideas, we start with a simplified situation. It deals with several sensors observing $y_j = x + \chi_j$ a common scalar signal $x \in \mathbb{R}$ with independent noises

[22] In this section, N stands for the termination time of **BC**, which length is thus $N + 1$.

$|\chi_j| \leq D_j^y, D_j^y > 0$. (With a slight abuse, we employ the notations from the stabilization problem statement in the new but coherent senses.) We show first that when acting independently, the sensors cannot generate common quantized data $\mathcal{Q}_j(y_j) \in \mathfrak{Q}$, where $\mathfrak{Q}$ is a finite set.[23]

Lemma 9.7.1. *Whenever $j' \neq j''$ and $\mathcal{Q}_{j'}(x + \chi_{j'}) = \mathcal{Q}_{j''}(x + \chi_{j''})$ for all x and $\chi_{j'}, \chi_{j''}$ within the noise bounds $|\chi_j| \leq D_j^y,\ j = j', j''$, both quantizers assume only a single value: $\mathcal{Q}_j(x) \equiv q\ \forall x, j = j', j''$.*

Proof. By picking q such that $E_{j'}^q := \{x : \mathcal{Q}_{j'}(x) = q\} \neq \emptyset$, we have $E_{j'}^q + [-D_{j'}^y, D_{j'}^y] \subset E_{j'}^q \Rightarrow E_{j'}^q = \mathbb{R} \Rightarrow \mathcal{Q}_{j'}(x) \equiv q \equiv \mathcal{Q}_{j''}(x)\ \forall x$. □

Now we show that synchronized quantization is possible if one of the sensors is able to communicate 2 bits of information to the others. Moreover, the common quantized data may contain as much information as desired. To this end, we introduce (Γ, d, F)-**quantizer** $\mathcal{Q}_{\Gamma,d,F}(y)$. Here y is the measurement and

- Γ gives the quantizer effective domain $[-\Gamma, \Gamma]$ (see p. 71);
- F determines the number of the quantizer levels;
- $d \in \left(0, F^{-1}\Gamma\right)$ is the "thickness" of the boundary zone of any level set.

The interval $[-\Gamma, \Gamma]$ is partitioned into F subintervals $\Delta_s := \left[-\Gamma + 2F^{-1}\Gamma s, -\Gamma + 2F^{-1}\Gamma(s+1)\right)$ $(\cup\{\Gamma\}$ if $s = F - 1)$, where $s = 0, \ldots, F-1$. Within any interval Δ_s, the central Δ_s^0 and two boundary zones $\Delta_s^{\pm}$ are introduced as is shown in Fig. 9.32. If $|y| \leq \Gamma$, the interval $\Delta_s \ni y$ is found and then the output $q = \mathcal{Q}_{\Gamma,d,F}(y)$ is generated by the rule from Fig. 9.32. If $|y| > \Gamma$, the quantizer output is the alarm symbol ✠.

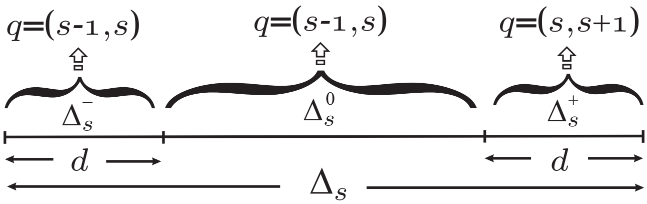


Fig. 9.32. The zones of the quantizer level domain.

Observation 9.7.2. *If $|y - x| < d$ and $|y| \leq \Gamma$, the quantized value $q := \mathcal{Q}_{\Gamma,d,F}(y)$ gives the interval $\Delta_q^x := \cup \Delta_s$ of length $\frac{4\Gamma}{F}$ containing x. Here $\cup$ is over s from q.*

By Lemma 9.7.1, two copies of this quantizer may transform the observations y_1 and y_2 of two independent noisy sensors into different outputs, even if $y_1, y_2 \approx x$. However if one of these sensors (say that fed by y_1) is aware of 2-bit information

[23] More precisely, the only common data are trivial and carry 0 bits of information.

about the companion measurement y_2, it is able to generate exactly the output of the companion quantizer. This information is given by $\beta := \beta(y_2)$, where

$$\beta(y) := \left\{ \begin{array}{ll} - & \text{if } y \text{ lies in the left boundary zone } \Delta_s^- \\ 0 & \text{if } y \text{ lies in the central zone } \Delta_s^0 \\ + & \text{if } y \text{ lies in the right boundary zone } \Delta_s^+ \\ \maltese & \text{if } y \text{ lies outside the effective domain } [-\varGamma, \varGamma] \end{array} \right\}. \tag{9.7.2}$$

The above generation is accomplished via employing the following.

$(\varGamma, d, F)$**-quantizer** $\mathfrak{Q}_{\varGamma,d,F}(y, \beta)$ **with side data**. For y from its effective domain $[-\varGamma, \varGamma]$, this quantizer first determines $\Delta_s \ni y$, as before, and proceeds as follows:

$$\mathfrak{Q}_{\varGamma,d,F}(y, \beta) := \begin{cases} (s-1, s) \text{ if } \left\{ \begin{array}{c} \beta = 0 \text{ or} \\ \beta = - \text{ and } y \in \Delta_s^- \cup \Delta_s^0 \text{ or} \\ \beta = + \text{ and } y \in \Delta_s^- \end{array} \right\} \\ (s, s+1) \text{ if } \left\{ \begin{array}{c} \beta = - \text{ and } y \in \Delta_s^+ \text{ or} \\ \beta = + \text{ and } y \in \Delta_s^0 \cup \Delta_s^+ \end{array} \right\} \\ \maltese \qquad\qquad \text{if } \beta = \maltese \text{ or } |y| > \varGamma \end{cases}.$$

Lemma 9.7.3. *If* $|y_i - x| < d/2, |y_i| \leq \varGamma, i = 1, 2,$ *and* $2F^{-1}\varGamma > 3d$*, the quantizer with the side data* $\beta = \beta(y_2)$ *transforms* y_1 *into the quantized value of* y_2*; i.e.,*

$$\mathfrak{Q}_{\varGamma,d,F}\big[y_1, \beta(y_2)\big] = \mathfrak{Q}_{\varGamma,d,F}(y_2) =: q, \quad \text{and} \quad x \in \Delta_q^x.$$

Proof. Note that $|y_1 - y_2| < d$. Let $\beta = 0$; i.e., $y_2 \in \Delta_{s'}^0$ for some s'. Since the interval $\Delta_s \ni y_1$ is separated from any $\Delta_{s'}^0$ with $s' \neq s$ by a distance $\geq d$, we have

$$s = s' \Rightarrow q_1 := \mathfrak{Q}_{\varGamma,d,F}\big[y_1, \beta(y_2)\big] = q_2 := \mathfrak{Q}_{\varGamma,d,F}(y_2).$$

Let $\beta = -$; i.e., $y_2 \in \Delta_{s'}^-$. Since $\Delta_s^- \cup \Delta_s^0$ is separated from $\Delta_{s'}^-$ by a distance $\geq d$ for any $s' \neq s$, we have

$$y_1 \in \Delta_s^- \cup \Delta_s^0 \Rightarrow s' = s \Rightarrow q_1 = q_2.$$

Since $\boldsymbol{V}\left(\Delta_s^0\right) = 2F^{-1}\varGamma - 2d > d$ by the assumptions of the lemma, Δ_s^+ is separated from $\Delta_{s'}^-$ by a distance $\geq d$ if $s' \neq s+1$. Hence $y_1 \in \Delta_s^+ \Rightarrow s' = s+1 \Rightarrow q_1 = q_2$.

The case $\beta = +$ is considered likewise. Observation 9.7.2 yields $x \in \Delta_q^x$. ☐

Remark 9.7.4. For $|y| \leq \varGamma$, the output q of both introduced quantizers is a pair $(s - 1, s)$, where $s = 0, \ldots, F$. From now on, we identify q with s; i.e., $q = 0, \ldots, F$. The center of the interval Δ_q^x is then given by $-\varGamma + 2\frac{\varGamma}{F}q$. This shapes the last inclusion in Lemma 9.7.3 into

$$\left| -\varGamma + 2\frac{\varGamma}{F}q - x \right| \leq 2F^{-1}\varGamma. \tag{9.7.3}$$

To meet the first requirement from Lemma 9.7.3, the parameter d should exceed the doubled noise level D_j^y of all sensors. The third one is ensured by picking $F := \left\lfloor \frac{2\varGamma}{3d} \right\rfloor - 1$. Letting $\varGamma \to \infty$ makes the number of levels $F + 1$ arbitrarily large. Note that the distortion $2F^{-1}\varGamma \approx 3d$ for $\varGamma \approx \infty$.

Remark 9.7.5. To communicate the quantized value to a remote location, $b = \log_2(F+2)$ bits are required. Let them be distributed over the sensors so that only a group of b_j bits is dispatched from the jth sensor site, where $\sum_j b_j = b$. If the quantized value is common for all sensors, such a way of transmission does not alter the distortion.

It should be remarked that conversely, the distortion may drastically degrade if some natural schemes of sensor cooperation in transmitting the bits of a common measured signal are employed. We close the subsection with a corresponding example.

Example 9.7.6. Let the observations $y_1 = x + \chi_1, y_2 = x$ from two sensors lie in the interval $[0, 1)$. The noise χ_1 is small $D_1^y \approx 0$ but nonzero $D_1^y > 0$. Let every sensor represent its measurement in the binary form and retain only the first $r = r_1 + r_2$ bits on the right of the binary point and let the first and second sensors dispatch the first r_1 and the next r_2 bits, respectively, to a remote recipient. It reconstructs the measurement by gathering the bits and putting them on proper positions (see Fig. 9.33).

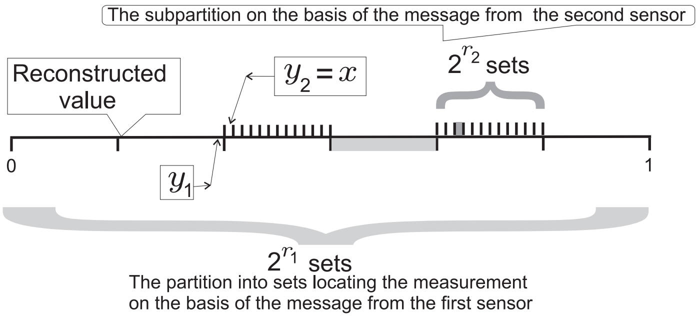


Fig. 9.33. A natural cooperation of two sensors.

Before transmission, the discrepancy between the true and quantized values does not exceed 2^{-r} for the second sensor and $2^{-r} + D_1^y (\approx 2^{-r}$ if $D_1^y \ll 2^{-r})$ for the first one. In the case from Fig. 9.33, the discrepancy after transmission and reconstruction $\approx 2^{-r_1}$, which is 2^{r_2} times worse than before the transmission.

Definition 9.7.7. *A member of a group of sensors that is able to and does transmit information to all other members of the group is called the* master *of this group.*

9.7.2 Master for Sensors Observing a Given Unstable Mode

Now we revert to the sensors j from the stabilization problem setup. The group $\mathfrak{J}_{\leftarrow\alpha}^{mw}$ of sensors able to observe (compute from their measurements with a bounded error) a given unstable mode x_α is described by (9.4.9) (on p. 295). Now we show that any such group contains a master sensor j_α^m, which can transmit as much information as desired to all other sensors from this group.

In this subsection, we deal with transmissions via **CBE(NW)**. It will be shown in Subsect. 9.7.4 (starting on p. 348) that they can also be arranged via the original network **NW**, provided that it interacts with the plant.

The transmissions of interest concern the network **CBE(NW)** modified as follows.

- All actuator nodes lose the status of output ones and are considered as inner. This may be viewed as attaching the "void" suffix ■ to the network (see Fig. 9.34);
- All sensor nodes outside the group $\mathfrak{J}^{mw}_{\leftarrow\alpha}$ lose the status of input ones and are considered as inner. The status of any node $j \in \mathfrak{J}^{mw}_{\leftarrow\alpha} \setminus \{j^m_\alpha\}$ is reversed from the input to output node. Thus there remains only one input node j^m_α getting data from a single source. These operations can be interpreted as attaching the *master–slave prefix* $\mathbf{PREF}^{\alpha}_{\mathbf{m\text{-}s}}$ (see Fig. 9.35).

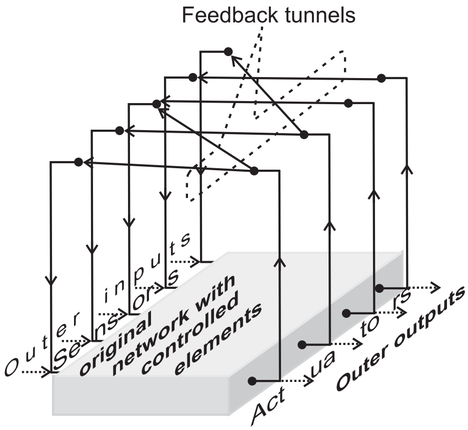


Fig. 9.34(a). Control-based extension.

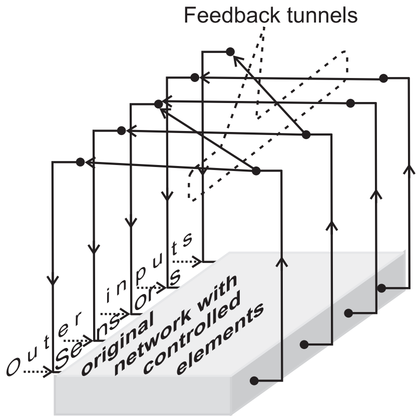


Fig. 9.34(b). Attaching the void suffix.

Lemma 9.7.8. *A choice of the master node $j^m_\alpha \in \mathfrak{J}^{mw}_{\leftarrow\alpha}$ exists under which the network* $\mathbf{PREF}^{\alpha}_{\mathbf{m\text{-}s}} \boxplus \mathbf{CBE(NW)} \boxplus$ ■ *hosts an errorless block code broadcasting one bit of information from the master j^m_α to all its slaves $j \in \mathfrak{J}^{mw}_{\leftarrow\alpha}, j \neq j^m_\alpha$. This block code can be picked so that the initial memory contents of the network elements*[24] *are equal to those for the code* **BC** *mentioned at the beginning of the section (see p.* 330*).*

Proof. The required block code for the network $\mathbf{PREF}^{\alpha}_{\mathbf{m\text{-}s}} \boxplus \mathbf{CBE(NW)} \boxplus$ ■ will be obtained via transformation of the block code **BC**. The latter is hosted by the network $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^q_{mw}}$ and fed by external inputs $\mathfrak{y}_{\alpha'} \in [1 : F_{\alpha'}]$.

Let N denote the length of **BC**, and $o_{i\to\alpha}(\mathfrak{y}_1, \ldots, \mathfrak{y}_{n^+})$ the signal emitted by **BC** at time $N-1$ from the ith actuator node into the additive channel (from $\mathbf{SUFF^q_{mw}}$) going to the output node $\sim x_\alpha$. This channel combines data from the actuator nodes

[24] Given by $\mathcal{M}^0(\boldsymbol{\mathcal{A}})$ in (9.3.2) (on p. 283).

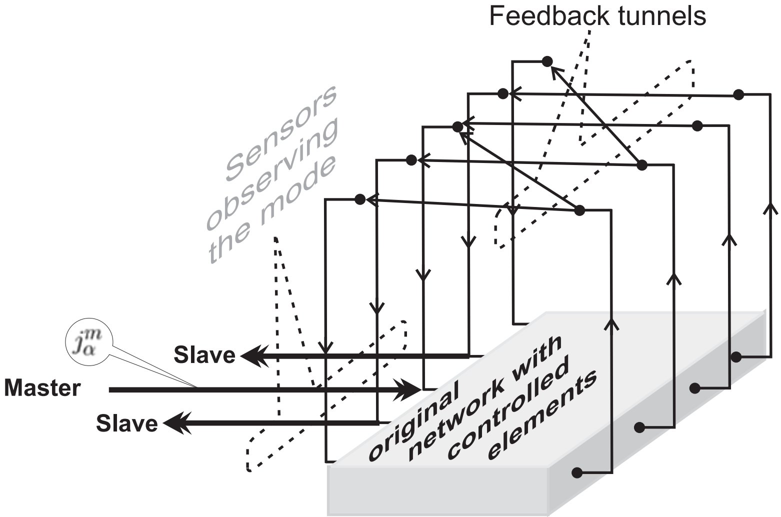


Fig. 9.35. Attaching the master-slave prefix.

$i \in \mathfrak{I}^{mw}_{\to\alpha}$ described by (9.4.10) (on p. 296). Since the code **BC** is errorless, $\mathfrak{y}_\alpha$ can be recognized from the knowledge of $\sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} o_{i\to\alpha}(1, \ldots 1, \mathfrak{y}_\alpha, 1, \ldots, 1)$. So there is an actuator $i^0 \in \mathfrak{I}^{mw}_{\to\alpha}$ for which the related addend is not constant in $\mathfrak{y}_\alpha \in [1 : F_\alpha]$.

Now we consider the operation of **BC** in the situation where $\mathfrak{y}_{\alpha'} := 1$ for all $\alpha' \neq \alpha$. By Assumption 9.3.4 (on p. 285), the sensor nodes process data separately. It follows that though the nodes $j \in \mathfrak{I}^{mw}_{\leftarrow\alpha}$ get a common input signal $\mathfrak{y}_\alpha$, they might operate in the situation where each of them deals with its own input $\mathfrak{y}^j_\alpha$. This would result in emitting a signal $o\big(\{\mathfrak{y}^j_\alpha\}_{j \in \mathfrak{I}^{mw}_{\leftarrow\alpha}}\big)$ from the actuator node i^0 into the above additive channel at time N. It is clear that $o_{i^0\to\alpha}(1, \ldots 1, \mathfrak{y}_\alpha, 1, \ldots, 1) = o\big(\{\mathfrak{y}_\alpha\}_{j \in \mathfrak{I}^{mw}_{\leftarrow\alpha}}\big)$. Since the function on the left assumes at least two different values, so does $o(\cdot)$. It follows that a sensor $j^m_\alpha \in \mathfrak{I}^{mw}_{\leftarrow\alpha}$ and inputs $\overline{\mathfrak{y}}^j_\alpha \in [1 : F_\alpha]$ for all other sensors $j \in \mathfrak{I}^{mw}_{\leftarrow\alpha}$ exist such that the function $o\big(\{\overline{\mathfrak{y}}^j_\alpha\}_{j \neq j^m_\alpha}, \{\mathfrak{y}_\alpha\}_{j = j^m_\alpha}\big)$ also assumes two different values for $\mathfrak{y}_\alpha = \mathfrak{y}^{[1]}_\alpha, \mathfrak{y}^{[2]}_\alpha \in [1 : F_\alpha]$, respectively.

The desired block code hosted by the network $\textbf{PREF}^{\alpha}_{\textbf{m-s}} \boxplus \textbf{CBE(NW)} \boxplus \blacksquare$ is as follows. After the sensor node $\sim j^m_\alpha$ accesses the single outer input $\mu = 1, 2$ of this network, the elements of **CBE(NW)** act until $t = N - 1$ just like they do in the initial block code **BC** when processing $\mathfrak{y}_{\alpha'} := 1\ \forall \alpha' \neq \alpha, \overline{\mathfrak{y}}^j_\alpha\ \forall j \in \mathfrak{I}^{mw}_{\leftarrow\alpha} \setminus \{j^m_\alpha\}$, and $\mathfrak{y}^{j^m_\alpha}_\alpha = \mathfrak{y}^{[\mu]}_\alpha$. It follows from the foregoing that the actuator node i^0 is able to recognize the input μ at $t = N$. Since this node influences x_α, it is linked with all sensor nodes $j \in \mathfrak{I}^{mw}_{\leftarrow\alpha}$ observing x_α by the feedback tunnels of **CBE(NW)** due to Remark 9.4.14 (on p. 297). The value of μ is emitted into each of these feedback tunnels except for that going to j^m_α. So μ will be received by all sensor nodes $j \in \mathfrak{I}^{mw}_{\leftarrow\alpha}$ after the delay $d := \max d_{i^0\to j}$, where $d_{i^0\to j}$ is the delay in the corresponding tunnel and $\max$ is over all tunnels concerned. It remains to increase the block length from N to $N + d$. □

Corollary 9.7.9. *For any natural* b*, the conclusion of Lemma* 9.7.8 *remains true if "*b *bits" is substituted in place of "one bit" in its statement.*

Indeed, the first bit may be transmitted by the block code from this lemma. Then the network should be reset to the initial state, which is possible by Assumption 9.4.20 (on p. 298). Then by Assumption 9.4.22 (on p. 299), the block code can be employed once more to transmit the second bit, and so on.

Being applied to the bits concerning different unstable modes x_α, the same argument gives rise to the following.

Corollary 9.7.10. *For any ensemble of naturals* $b_\alpha, \alpha = 1, \ldots, n^+$*, there exists a block code broadcasting without errors* b_α *bits of information from the master sensor node* j_α^m *to its* slaves $j \in \mathfrak{J}_{\leftarrow\alpha}^{mw}, j \neq j_\alpha^m$ *for all* $\alpha = 1, \ldots, n^+$*. This code can be picked so that both initial and terminal states of the network equal the initial state for the code* **BC** *mentioned at the beginning of the section (see p.* 330*).*

Explanation 9.7.11. Formally, this corollary deals with the network that results from **CBE(NW)** by attachment of the suffix ■ and the prefix $\mathbf{PREF}^{\alpha}_{\mathbf{m\text{-}s}}$ for all α. The operations related to different α's do not overwrite each other: Their results are combined.

9.7.3 Stabilization over the Control-Based Extension of the Original Network

Description of the Core Problem to be Solved

Modulo the discussions in Sects. 3.8 and 7.8 (starting on pp. 62 and 233), the core problem in design of a stabilizing control strategy concerns the networked block code **BC** mentioned at the beginning of this section. This code is what is given by **a)** in the implication **a)** $\Rightarrow$ **b)** to be proved. So this code is the basement on which the stabilizing control strategy should be constructed.

The core problem to be addressed further arises from the fact that this code is a data processing algorithm for the artificially constructed network $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^{q}_{mw}}$. Along with the real part **NW**, it includes imaginary channels, which do not exist in fact and serve only to explicitly express certain relations or possibilities. These are as follows:

1. The channels from $\mathbf{PREF_{mw}}$ broadcasting signals from imaginary informants;
2. The channels from $\mathbf{SUFF^{q}_{mw}}$ communicating signals to imaginary recipients;
3. The feedback tunnels introduced to construct the control-based extension and to explicitly express the possibility of data communication by means of control.

It should be demonstrated that everything required for stabilization and performed by **BC** via these imaginary channels can be done by real means. Specifically,

- Data received via each channel from $\mathbf{PREF_{mw}}$ and feedback tunnel can be accessed by the channel terminal element (sensor) in another and realistic way.
- The signals emitted from the actuator nodes into the imaginary channels from $\mathbf{SUFF^{q}_{mw}}$ should be used by actuators to generate controls.

This objective will be achieved successively with respect to channels mentioned in 1, 2, and 3. In this subsection, we deal with the first and second groups of channels. The result will be a control strategy stabilizing the plant via **CBE(NW)**. The channels from the third group will be treated in the next subsection. This will result in the final control strategy stabilizing the plant via the real network **NW**.

Outline of the Control Strategy Design

In design of the stabilizing strategy, we follow the lines of Sects. 3.8, 6.11, and 7.8. Within this framework, the role of the block code **BC** is to communicate quantized information about the unstable modes x_α from the sensors to actuators within a multirate stabilization scheme. The latter means that the time horizon is partitioned into operation epochs of equal duration r, and the major operations are performed only at the beginning of every epoch. These operations result in the control program and the quantized values of the modes x_α^+ of the state prognosis for the end of the epoch. The program and the value are carried out and communicated, respectively, within the forthcoming epoch. The role of **BC** is to serve this communication.

To make this possible, all sensors $j \in \mathfrak{J}^{mw}_{\leftarrow\alpha}$ observing the mode x_α should compute a common quantized value of the prognosis x_α^+. In its turn, the latter is possible if every sensor computes the prognosis with a bounded error depending on the sensor, as was shown in Subsects. 9.7.1 and 9.7.2. We start with a technical discussion aimed to show how this computation can be accomplished and with which error.

Prediction of the Values of Unstable Modes at the End of the Epoch

Now we revert to the plant equations (9.4.1) and increase the sample period up to r:

$$x[(\theta+1)r] = A^r x[\theta r] + v(\theta) + \zeta(\theta), \quad v(\theta) := \sum_{i=1}^{l} v_i(\theta);$$
$$v_i(\theta) := \sum_{t=0}^{r-1} A^{r-1-t} B_i u_i[\theta r + t], \quad \zeta(\theta) := \sum_{t=0}^{r-1} A^{r-1-t} \xi[\theta r + t]. \quad (9.7.4)$$

Here v_i and v express the action on the plant of, respectively, the ith and all control programs applied during the θth epoch, and ζ accounts for the plant noise. We recall that the operator $[\cdot]_\alpha$ extracts the αth coordinate from the vector (9.4.8) (on p. 295), and $\nu = \nu(\alpha)$ indicates the group of coordinates from (9.4.8) containing x_α.

To proceed, we adopt the following two temporary assumptions, which will be ultimately ensured by the design of the control strategy.

Assumption 9.7.12. *Put* $n := \dim(x)$. *During* $n-1$ *concluding time steps of any epoch, all controls are zero*

$$u_i(t) = 0 \qquad \forall t = r\theta - n + 1, \dots, r\theta - 1, \quad i = 1, \dots, l.$$

Assumption 9.7.13. *At the beginning θr of any epoch, all sensors $j \in \mathfrak{J}_{\leftarrow\alpha}^{mw}$ observing the mode x_α have access to the quantities $[v_i(\theta)]_\alpha$ from* (9.7.4) *expressing the influence on this mode of all actuators i that do affect it. (In other words, i runs over the set $\mathfrak{J}_{\rightarrow\alpha}^{mw}$ given by* (9.4.10) *on p.* 296*.)*

Then these sensors can predict $x_\alpha[(\theta+1)r]$ with bounded independent errors. Specifically, the following claim holds.

Proposition 9.7.14. *Suppose that Assumptions* 9.7.12 *and* 9.7.13 *hold. Then at the start $r\theta$ of any epoch $\theta \geq 1$ any sensor j observing x_α (i.e., $j \in \mathfrak{J}_{\leftarrow\alpha}^{mw}$) can compute an estimate $\widehat{x}_\alpha^j[(\theta+1)r]$ of $x_\alpha[(\theta+1)r]$ with the exactness*

$$\left|\widehat{x}_\alpha^j[(\theta+1)r] - x_\alpha[(\theta+1)r]\right| \leq \varphi_{\nu(\alpha)}(r)|\lambda_{[\nu(\alpha)]}|^r, \tag{9.7.5}$$

where $\varphi_\nu(\cdot)$ is a polynomial, and the notation $\lambda_{[\nu]}$ is taken from (i) of Assumption 9.4.6 *(on p.* 294*).*

To prove this claim, we note that by Assumption 9.4.6, x_α is among the nonzero components of the vectors from the A-invariant subspace $M_\nu, \nu = \nu(\alpha)$. The spectrum of $A|_{M_\nu}$ lies in the circle of the radius $|\lambda_{[\nu]}|$. Hence for some polynomial $\psi_\nu(\cdot)$, we have [213]

$$\left\|[A|_{M_\nu}]^t\right\| \leq \psi_\nu(t)|\lambda_{[\nu]}|^t \qquad \forall t = 0, 1, \dots. \tag{9.7.6}$$

Explanation 9.7.15. Here and throughout the section, we employ the norm of a vector $\|x\| := \max_\alpha |x_\alpha|$, where x_α are taken from the representation (9.4.8) (on p. 295), and the related matrix norm $\|M\| := \max_{\|x\|=1} \|Mx\|$.

Remark 9.7.16. The estimates (9.4.3) (on p. 292) on the plant noises and initial state remain true for this norm with, maybe, altered upper bounds.

Let $x_{[\nu]}$ be the projection of $x \in \mathbb{R}^n$ onto M_ν in correspondence with the decomposition $\mathbb{R}^n = M_{\text{st}}(A) \oplus \bigoplus_{\nu'=1}^g M_{\nu'}$ from Assumption 9.4.6 (on p. 294).

To prove Proposition 9.7.14, we first estimate the noise from (9.7.4).

Lemma 9.7.17. *For any $\nu = 1, \dots, g$, a polynomial $\phi_\nu(r)$ exists such that*

$$\|\zeta_{[\nu]}(\theta)\| \leq \phi_\nu(r)|\lambda_{[\nu]}|^r \qquad \forall \theta, r. \tag{9.7.7}$$

Proof. We have

$$\left\|\zeta_{[\nu]}(\theta)\right\| \overset{(9.7.4)}{=\!=} \left\|\sum_{t=0}^{r-1} (A|_{M_\nu})^{r-1-t}\, \xi_{[\nu]}[\theta r + t]\right\| \overset{(9.4.3)}{\leq} D \sum_{t=0}^{r-1} \left\|(A|_{M_\nu})^{r-1-t}\right\|$$
$$\overset{(9.7.6)}{\leq} D \sum_{t=0}^{r-1} \psi_\nu(t)|\lambda_{[\nu]}|^t \overset{|\lambda_{[\nu]}|\geq 1}{\leq} D|\lambda_{[\nu]}|^r \sum_{t=0}^{r-1} \psi_\nu(t) \leq |\lambda_{[\nu]}|^r \phi_\nu(r).$$

Here $\phi_\nu(r)$ is a polynomial in r, and the inequalities $\sum_{t=0}^{r-1} t^k \leq (k+1)^{-1} r^{k+1}$ are employed to estimate the sum in the penultimate expression. □

Proof of Proposition 9.7.14. By Assumption 9.4.6 (on p. 294), the subspaces $M_{\nu'}$ are A-invariant, and the unobservable subspace of the jth sensor is the direct sum of several $M_{\nu'}$'s. By applying the last remark from Explanation 7.7.4 (on p. 224) to the operator A acting in the sum of the remaining $M_{\nu'}$'s and taking into account Assumption 9.7.12, we see that proceeding from the observations $y_j(t), t = r\theta - n+1, \ldots, r\theta$, the jth sensor can compute an estimate $\widehat{x}^j_{[\nu]}(r\theta)$ of the current state $x_{[\nu]}(r\theta)$ ($\nu := \nu(\alpha)$) with an error bounded by a constant c_0 independent of r, j, α:

$$\left\|\widehat{x}^j_{[\nu]}(r\theta) - x_{[\nu]}(r\theta)\right\| \leq c_0 \qquad \forall\theta = 1, 2, \ldots.$$

The prognosis for the end of the epoch is generated as follows:

$$\widehat{x}^j_\alpha[(\theta+1)r] := \left[(A|_{M_\nu})^r\, \widehat{x}^j_{[\nu]}(r\theta)\right]_\alpha + \sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} [v_i(\theta)]_\alpha. \tag{9.7.8}$$

Now we observe that

$$\begin{aligned}
\left|\widehat{x}^j_\alpha[(\theta+1)r] - x_\alpha[(\theta+1)r]\right| &= \left|\left[(A|_{M_\nu})^r\, \widehat{x}^j_{[\nu]}(r\theta) + v_{[\nu]}(\theta) - x_{[\nu]}[(\theta+1)r]\right]_\alpha\right| \\
&\leq \left\|(A|_{M_\nu})^r\, \widehat{x}^j_{[\nu]}(r\theta) + v_{[\nu]}(\theta) - x_{[\nu]}[(\theta+1)r]\right\| \\
&\overset{(9.7.4)}{=\!=} \left\|(A|_{M_\nu})^r \left[\widehat{x}^j_{[\nu]}(r\theta) - x_{[\nu]}(r\theta)\right] - \zeta_{[\nu]}(\theta)\right\| \\
&\leq \|(A|_{M_\nu})^r\| \left\|\widehat{x}^j_{[\nu]}(r\theta) - x_{[\nu]}(r\theta)\right\| + \left\|\zeta_{[\nu]}(\theta)\right\| \\
&\leq c_0 \|(A|_{M_\nu})^r\| + \left\|\zeta_{[\nu]}(\theta)\right\| \overset{(9.7.6)}{\leq} c_0\psi_\nu(r)|\lambda_{[\nu]}|^r + \left\|\zeta_{[\nu]}(\theta)\right\|.
\end{aligned}$$

It remains to take into account (9.7.7). □

Description of the Control Strategy for CBE(NW)

We start with a partial description with the focus on the actions at the sensor nodes.

A) At the beginning of every epoch $\theta \geq 1$,
  - **1)** Any sensor j calculates the prognosis $\widehat{x}^j_\alpha[(\theta+1)r]$ for any observed unstable mode x_α (i.e., such that $j \in \mathfrak{I}^{mw}_{\leftarrow\alpha}$) from the measurements $y_j(t)|^{r\theta}_{t=r\theta-n+1}$;
  - **2)** For any unstable mode α, the corresponding master sensor node j^m_α applies the $(\Gamma, d, F_\alpha - 1)$-quantizer[25] (see p. 331) to its prognosis, thus producing the quantizer output $q^j_\alpha, j := j^m_\alpha$, and also calculates the companion 2-bits side data (9.7.2);

[25] Here F_α is the parameter of the initial block code **BC**, i.e., the number of its x_α-inputs. By Remark 9.7.4, the (Γ, d, F)-quantizer produces $F+1$ outputs to say nothing of ✠. Now we convert ✠ into 1, thus reducing the number of outputs to exactly $F+1$. (This will be commented in Explanation 9.7.18.) To meet the code rate capability, we pick $F := F_\alpha - 1$.

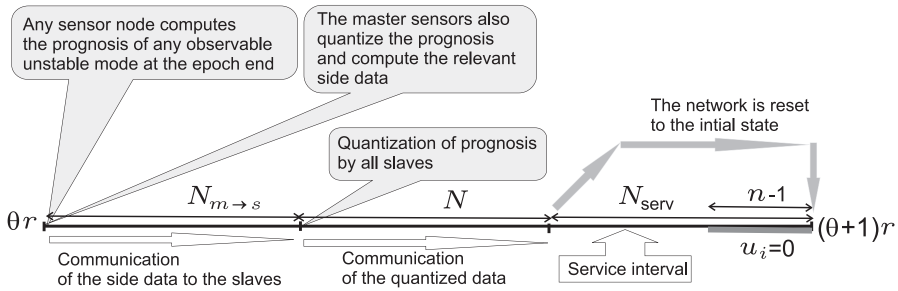


Fig. 9.36. Organization of the operation epoch with the focus on the sensors.

B) These side data are transmitted from every master to all its slaves by means of the networked block code from Corollary 9.7.10 (where $b_\alpha := 2\ \forall \alpha$) during the time interval of duration $N_{m \to s}$;

C) After receiving the side data at time $r\theta + N_{m \to s}$, any slave sensor j applies the $(\Gamma, d, F_\alpha - 1)$-quantizer with side data (see p. 332) to the related prognosis, thus producing the quantized value q_α^j;

D) The quantized data obtained by all sensor nodes are used to initiate the operation of the networked block code **BC** from the beginning of the section (see p. 330), which operates until $t = r\theta + N_{m \to s} + N$, where $N + 1$ is the length of the code;

E) The operation epoch is concluded with a *service interval* of duration N_{serv} (to be specified further), during which auxiliary technical operations are carried out.
  1) In particular, the network is reset to the initial state at the end $(\theta + 1)r$ of the epoch so that the above operations can be recursively iterated.

Explanation 9.7.18.

- In Fig. 9.36, the controls are taken to be zero during concluding $(n - 1)$ time steps of the epoch to ensure Assumption 9.7.12.
- In **A.1)**, the prognosis from Proposition 9.7.14 is computed. This is possible thanks to Assumptions 9.7.12 and 9.7.13.
- In **A.2)** and **C)**, the parameters Γ, d are taken common for all sensors and modes to simplify the notations. Their choice will be discussed further.
- A given sensor may act as a master for one mode and a slave for some other one.
- As will be shown, the quantizer output ✠ does not occur if the parameters Γ, d, N are chosen properly. This output was yet taken into account and converted into the form 1 acceptable by the block code for the sake of formal completeness of presentation.
- **D)** is possible thanks to Assumption 9.4.22 (on p. 299), since the block code from **B)** drives the network to the state initial for **BC** by Corollary 9.7.10.
- **E.1)** is possible due to Assumption 9.4.20 (on p. 298).

Remark 9.7.19. Let for all α, the assumptions of Lemma 9.7.3 (on p. 332) be satisfied when the sensors $j \in \mathfrak{J}_{\leftarrow \alpha}^{mw}$ are interpreted as devices producing noisy observations

$\widehat{x}^j_\alpha[(\theta+1)r]$ of $x_\alpha[(\theta+1)r]$. Then the slave sensors produce the quantized output q_α of the master: $q^j_\alpha = q_\alpha$. So the common value q_α appears at all nodes observing x_α. This is the situation under which the block code **BC** communicates data correctly.

To complete the description of the control strategy, it remains to specify the actions of the actuator nodes. To this end, we introduce the following.

Notation 9.7.20. *When* **BC** *completes the data transmission at time* $\theta r + N_{m\to s} + N$ *within the above scheme, every actuator node* i *generates a message to any affected unstable mode* x_α *(i.e., such that* $i \in \mathfrak{I}^{mw}_{\to\alpha}$*). This message is denoted by* $o_{i\to\alpha} \in \mathbb{R}$.

Formerly, the signals $o_{i\to\alpha}$ with $i \in \mathfrak{I}^{mw}_{\to\alpha}$ were emitted into the imaginary additive channel going to the imaginary output node $\sim x_\alpha$. The output of this channel

$$\sum_{i\in\mathfrak{I}^{mw}_{\to\alpha}} o_{i\to\alpha}$$

was projected at the nearest integer to form the output of this node by the definition of the mode-wise suffix with quantization (see p. 296). Under the assumption of Remark 9.7.19, the last output equals q_α. So the distance from the above sum to q_α does not exceed $1/2$. With regard to (9.7.3), this gives rise to the inequality

$$\left| -\Gamma + 2\frac{\Gamma}{F_\alpha - 1}\sum_{i\in\mathfrak{I}^{mw}_{\to\alpha}} o_{i\to\alpha} - x_\alpha[(\theta+1)r]\right| \leq \frac{3\Gamma}{F_\alpha - 1}.$$

Thus we arrive at the following.

Observation 9.7.21. *Whenever the assumption of Remark* 9.7.19 *holds,*

$$\left| \sum_{i\in\mathfrak{I}^{mw}_{\to\alpha}} \left[\frac{2\Gamma}{F_\alpha - 1} o_{i\to\alpha} - \frac{\Gamma}{|\mathfrak{I}^{mw}_{\to\alpha}|}\right] - x_\alpha[(\theta+1)r]\right| \leq \frac{3\Gamma}{F_\alpha - 1}. \tag{9.7.9}$$

We do not consider any longer the imaginary additive channel into which the message $o_{i\to\alpha}$ was emitted before. Now this message will be used by the ith actuator node to produce the control program for the next operation epoch. To this end, this message should be stored in the node memory until the start of the next epoch.

To specify the control, we denote by $\langle x\rangle_\alpha$ the vector x given by (9.4.8), where x is put at the position α and the remaining ones are filled by zeros. By (9.4.8) and (9.4.10) (on pp. 295 and 296),

$$\langle x\rangle_\alpha \in L^{+c}_i \qquad \forall x \quad \text{whenever} \quad i \in \mathfrak{I}^{mw}_{\to\alpha},$$

where L^{+c}_i is the subspace of states controllable by the ith actuator.

We also pick a deadbeat stabilizer, i.e., a linear transformation

$$x \in L^{+c}_i \overset{\mathfrak{L}_i}{\mapsto} \boldsymbol{U} = [u_i(0), u_i(1), \ldots, u_i(n-1), 0, 0, \ldots] \tag{9.7.10}$$

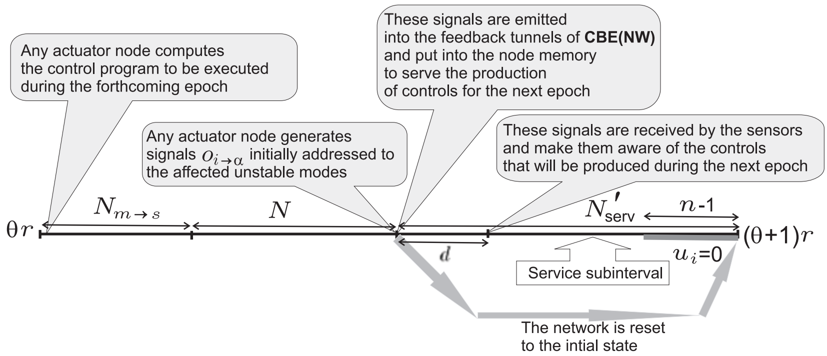


Fig. 9.37. Organization of the operation epoch with the focus on the actuators.

of the initial state x into a control program that drives the unperturbed ($\xi(t) \equiv 0$) plant (9.4.1) (on p. 291) with all actuators except for the ith one switched off from x at time $t = 0$ to 0 at time $t = n$.

Now we complete the description of the control strategy (see Fig. 9.37).

F) At the start θr of any epoch $\theta \geq 2$, any actuator node i generates the control program $\boldsymbol{U}_i(\theta) = \textbf{col}\,[u_i(\theta r), \ldots, u_i(\theta r + r - 1)]$ to be carried out during this epoch. The generation is via two steps.

1) For all affected unstable modes $\alpha : i \in \mathfrak{I}^{mw}_{\to\alpha}$, the node produces the control program $\boldsymbol{U}_i^\alpha$ that in the absence of the noise $\xi(\cdot) \equiv 0$ and actions of the companion actuators $u_{i'} \equiv 0 \; \forall i' \neq i$ simultaneously drives

- this mode to 0 at time $t = \theta r + n$ provided that it starts at $t = \theta r$ at the value given by the corresponding addend in (9.7.9);[26]
- All other modes (including the stable ones) from 0 to 0:

$$\boldsymbol{U}_i^\alpha(\theta) := \mathfrak{L}_i \left\{ \left\langle \frac{2\Gamma}{F_\alpha - 1} o_{i\to\alpha} - \frac{\Gamma}{|\mathfrak{I}^{mw}_{\to\alpha}|} \right\rangle_\alpha \right\}; \tag{9.7.11}$$

2) The generated control programs are summed up over the affected modes:

$$\boldsymbol{U}_i(\theta) := \sum_{\alpha : i \in \mathfrak{I}^{mw}_{\to\alpha}} \boldsymbol{U}_i^\alpha(\theta). \tag{9.7.12}$$

G) At time $\theta r + N_{m\to s} + N, \theta \geq 1$ when the block code **BC** completes its service for the current epoch, any actuator node i dispatches each of the newly generated signals $o_{i\to\alpha}$ to all sensor nodes j connected with i by feedback tunnels of **CBE(NW)**; i.e., such that $(i, j) \in \text{CVP}$, where CVP given by (9.4.6) (on p. 294). All these signals will reach the destinations no more than d time units later, where

$$d := \max_{(i,j)\in\text{CVP}} d_{i\to j}$$

and $d_{i\to j}$ is the delay in the tunnel, which is given by (9.4.7) (on p. 294).

[26]Since $o_{i\to\alpha}$ is the signal generated within the previous epoch, now $\theta := \theta - 1$ in (9.7.9).

H) At time $\theta r + N_{m\to s} + N + d$ after receiving these signals $o_{i\to\alpha}$, any sensor node j calculates the quantity $[v_i(\theta+1)]_\alpha$ from Assumption 9.7.13 for the next epoch for any observed unstable mode $x_\alpha : j \in \mathfrak{J}^{mw}_{\leftarrow\alpha}$ and every actuator i that does influence this mode $i \in \mathfrak{J}^{mw}_{\to\alpha}$.[27] This is accomplished via two steps:
1) The program $\boldsymbol{U}_i^\alpha(\theta+1) = \{u_i^\alpha(\tau)\}_{\tau\geq 0}$ is computed by duplicating (9.7.11);
2) In correspondence with (9.7.4),

$$[v_i(\theta+1)]_\alpha := \left[\sum_{t=0}^{r-1} A^{r-1-t} B_i u_i^\alpha(t)\right]_\alpha .$$

Explanation 9.7.22.

- **H.1)** is possible since the feedback tunnel connecting the concerned actuator and sensor nodes does exist due to (9.4.12) (on p. 297) and has the alphabet of the maximal possible cardinality $2^{\aleph_0}$.
- The initial control programs $\boldsymbol{U}_i^\alpha(0), \boldsymbol{U}_i^\alpha(1)$ are taken to be zero. The ensures Assumption 9.7.13 for $\theta = 0, 1$.

Note that **A)–H)** provide a complete description of a control strategy for the control-based extension **CBE(NW)** of the primal network **NW** during an epoch. Thanks to Assumption 9.4.26 (on p. 299), a periodic admissible strategy exists that operates over the infinite horizon and behaves as was described during any epoch.

Remark 9.7.23. Thanks to **F)–H)**, at the start of any epoch θr any sensor node j is able to compute all control programs $\boldsymbol{U}_i(\theta)$ that will affect its observations during this epoch, i.e., such that $(i, j) \in \text{CVP}$, where CVP is given by (9.4.6) (on p. 294).

Remark 9.7.24. Due to (9.7.10)–(9.7.12), $u_i(t) = 0$ for all $t \in [\theta r + n : (\theta+1)r)$. Hence Assumption 9.7.12 is satisfied if in Figs. 9.36 and 9.37

$$N_{\text{serv}} > d + 2n, \tag{9.7.13}$$

which is assumed to hold from now on.[28]

Observation 9.7.25. *Assumptions* 9.7.12 *and* 9.7.13 *are necessarily fulfilled for the control strategy* **A)–H)**.

This strategy depends of the quantizer parameters Γ, d, as well as the duration N_{serv} of the service interval and the length $N+1$ of of the primal block code **BC**.

Remark 9.7.26. The length N can be chosen arbitrarily large with keeping (9.7.1) (on p. 330) true, whereas $N_{m\to s}$ and N_{serv} can be kept independent of N and bounded $N_{\text{serv}} \leq N_*$, respectively.

[27] We recall that $[v_i(\theta+1)]_\alpha$ is the position to which the mode will be driven at the end of the next epoch from 0 at its start by the controls from the ith actuator if the plant noise is zero and the other actuators are switched off.

[28] This inequality will also be commented on in Explanation 9.7.35 and Remark 9.7.36 (on p. 352).

This follows from Remark 9.5.7 (on p. 304) and (ii) from Assumption 9.4.20 (on p. 298) provided that the block codes **BC** in Corollary 9.7.10 (on p. 336) and **D)** are taken from the common sequence mentioned in (iii) of Lemma 9.5.6 (on p. 303).

Remark 9.7.27. We recall that $\nu = \nu(\alpha)$ is the serial number of the group of coordinates from (9.4.8) (on p. 295) containing x_α, and $\lambda_\alpha = \lambda_{[\nu(\alpha)]}$. So in (9.7.1) (on p. 330), the right hand side does not change as α ranges so that $\nu = \nu(\alpha)$ is kept constant. So without any loss of generality, it can be assumed that the left-hand sides of (9.7.1) are also common for such α's:

$$F_\alpha = F_\nu \qquad \text{whenever} \quad \nu(\alpha) = \nu. \tag{9.7.14}$$

We close this subsection with the conditions under which the proposed strategy stabilizes the plant.

Proposition 9.7.28. *Suppose that for any $\nu \in [1:g]$, the following inequalities hold:*

$$2(F_\nu - 1)^{-1}\Gamma > 3d, \quad \varphi_\nu(r)|\lambda_{[\nu]}|^r < \frac{d}{2},$$
$$\left[\frac{3\Gamma}{F_\nu - 1}\psi_\nu(r) + D_\nu^*\psi_\nu(r) + 2\phi_\nu(r) + \varphi_\nu(r)\right] |\lambda_{[\nu]}|^r < \Gamma,$$
$$\text{where} \quad D_\nu^* := |\lambda_{[\nu]}|^r \left[D_x\psi_\nu(r) + \phi_\nu(r)\right]. \tag{9.7.15}$$

Here $\varphi_\nu(\cdot), \psi_\nu(\cdot)$, and $\phi_\nu(\cdot)$ are the polynomials from (9.7.5)*,* (9.7.6)*, and* (9.7.7)*, respectively, and D_x is the bound on the initial state from* (9.4.3) *(on p.* 292*). Then the strategy* **A)–H)** *regularly stabilizes*[29] *the plant.*

The proof of this proposition is given in the next part of the text. Then it will be shown (see p. 347) that whenever **a)** of Theorem 9.4.27 (on p. 300) holds, the conditions (9.7.15) can be satisfied by the proper choice of the parameters Γ, d, N, and so the strategy **A)–H)** does stabilize the plant via the artificial network **CBE(NW)**. In Subsect. 9.7.4 (starting on p. 348), this strategy will be converted into the stabilizing control policy for the real network **NW**.

Proof of Proposition 9.7.28

We recall that the choice of the norm $\|x\|$ was commented in Explanation 9.7.15 (on p. 338); $x_{[\nu]}$ is the projection of $x \in \mathbb{R}^n$ onto M_ν in correspondence with the decomposition $\mathbb{R}^n = M_{\text{st}}(A) \oplus \bigoplus_{\nu'=1}^g M_{\nu'}$ from Assumption 9.4.6 (on p. 294).

Lemma 9.7.29. *The conclusion of Proposition* 9.7.28 *is true if numbers*

$$D_\nu^+ > |\lambda_{[\nu]}|^r \left[\psi_\nu(r)D_\nu^* + \phi_\nu(r)\right] \qquad \nu = 1, \dots, g \tag{9.7.16}$$

exist such that the following claim holds:

[29] See Definition 9.4.2 on p. 293.

i) *Whenever at the end of some epoch* $\theta \geq 1$

$$|x_\alpha[(\theta+1)r]| < D_\nu^+ \quad \text{for} \quad \nu = \nu(\alpha) \quad \text{and all} \quad \alpha = 1, \ldots, n^+, \quad (9.7.17)$$

the same relations hold at the end of the next epoch ($\theta := \theta + 1$ *in* (9.7.17)).

Proof. We first prove that the states $x_{[\nu]}(\theta r)$ are bounded over $\theta \geq 0$ for all ν. In view of **i)** and (9.4.3) (on p. 292), it suffices to show that under the conditions (9.4.3), $\|x_{[\nu]}(r)\|$ is bounded and $\|x_{[\nu]}(2r)\| < D_\nu^+$. By Explanation 9.7.22 (on p. 343), the controls are zero for the first and second epochs. So with regard to (9.7.4) (on p. 337), we have for $\theta = 0, 1$

$$\left\|x_{[\nu]}[(\theta+1)r]\right\| = \left\|(A|_{M_\nu})^r x_{[\nu]}(\theta r) + \zeta_{[\nu]}(\theta)\right\|$$
$$\leq \|(A|_{M_\nu})^r\| \left\|x_{[\nu]}(\theta r)\right\| + \left\|\zeta_{[\nu]}(\theta)\right\| \overset{(9.7.6),\,(9.7.7)}{\leq} |\lambda_{[\nu]}|^r \left[\psi_\nu(r)\left\|x_{[\nu]}(\theta r)\right\| + \phi_\nu(r)\right];$$
$$\left\|x_{[\nu]}[r]\right\| \overset{(9.4.3)}{\leq} |\lambda_{[\nu]}|^r \left[D_x \psi_\nu(r) + \phi_\nu(r)\right] \overset{(9.7.15)}{=\!=} D_\nu^*,$$
$$\left\|x_{[\nu]}[2r]\right\| \leq |\lambda_{[\nu]}|^r \left[\psi_\nu(r) D_\nu^* + \phi_\nu(r)\right] \overset{(9.7.16)}{<} D_\nu^+.$$

The controls are also bounded. Indeed, they are given by (9.7.11) and (9.7.12), where the signals $o_{i\to\alpha}$ are produced by the deterministic block code **BC** from quantized data $q_\alpha^j \in [1 : F_\alpha]$ assuming only finitely many values. Hence the control programs $\boldsymbol{U}_i(\theta)$ range over finite sets and so $\|u_i(t)\| \leq D_i^u < \infty\ \forall t, i$.

Thus all controls $u_i(t)$ and the unstable parts of the state $x(\theta r)$ are bounded over $t \geq 0$ and $\theta \geq 0$, respectively. The standard arguments (see, e.g., (7.8.20) and (7.8.22) on pp. 244 and 245) show that then the states $x(t)$ are bounded over $t \geq 0$. Definition 9.4.2 (on p. 293) completes the proof. □

Thus it suffices to justify **i**). To this end, we pick the numbers D_ν^+ such that

$$\left[\frac{3\Gamma}{F_\nu - 1}\psi_\nu(r) + \phi_\nu(r)\right] |\lambda_{[\nu]}|^r < D_\nu^+, \qquad D_\nu^+ + \varphi_\nu(r)|\lambda_{[\nu]}|^r < \Gamma;$$
$$\left[D_\nu^* \psi_\nu(r) + \phi_\nu(r)\right] |\lambda_{[\nu]}|^r < D_\nu^+ \qquad \forall \nu. \quad (9.7.18)$$

This is possible. Indeed, let us take D_ν^+ greater but approximately equal to the maximum of the left-hand sides in the first and third inequalities. Then the second one follows from the third relation in (9.7.15).

Lemma 9.7.30. *The claim* **i)** *of Lemma* 9.7.29 *does hold.*

Proof. Let (9.7.17) hold for $\theta \geq 1$. Then the assumption of Remark 9.7.19 (on p. 340) is true due to (9.7.5), the first two inequalities in (9.7.15), and the estimates

$$\left|\widehat{x}_\alpha^j[(\theta+1)r]\right| \leq \left|\widehat{x}_\alpha^j[(\theta+1)r] - x_\alpha[(\theta+1)r]\right| + |x_\alpha[(\theta+1)r]|$$
$$\overset{(9.7.5),\,(9.7.17)}{\leq} \varphi_\nu(r)|\lambda_{[\nu]}|^r + D_\nu^+ < \Gamma,$$

where $\nu = \nu(\alpha)$ and the last inequality is taken from (9.7.18). So by Observation 9.7.21 (on p. 341), the signals $o_{i \to \alpha}$ generated within the epoch at hand satisfy (9.7.9).

For the control program $\boldsymbol{U}_i^\alpha = u_i^\alpha[(\theta+1)r], u_i^\alpha[(\theta+1)r+1], \ldots$ given by (9.7.11), we have due to the definition of the deadbeat stabilizer $\mathfrak{L}_i$ and (9.7.14),

$$A^r \left\langle \frac{2\Gamma}{F_\nu - 1} o_{i \to \alpha} - \frac{\Gamma}{|\mathfrak{I}_{\to\alpha}^{mw}|} \right\rangle_\alpha + \sum_{t=0}^{r-1} A^{r-1-t} B_i u_i^\alpha[(\theta+1)r+t] = 0.$$

Summing up over both modes $\alpha : i \in \mathfrak{I}_{\to\alpha}^{mw}$ affected by the actuator i and actuators i gives with regard to (9.7.12),

$$A^r \sum_{\alpha=1}^{n^+} \underbrace{\sum_{i \in \mathfrak{I}_{\to\alpha}^{mw}} \left\langle \frac{2\Gamma}{F_\nu - 1} o_{i \to \alpha} - \frac{\Gamma}{|\mathfrak{I}_{\to\alpha}^{mw}|} \right\rangle_\alpha}_{z(\alpha)} + \sum_{i=1}^{l} \sum_{t=0}^{r-1} A^{r-1-t} B_i u_i[(\theta+1)r+t] = 0. \quad (9.7.19)$$

By (9.7.9), $\|z(\alpha) - \langle x_\alpha[(\theta+1)r]\rangle_\alpha\| \leq \frac{3\Gamma}{F_\alpha - 1}$. Hence we have for the first sum

$$\begin{aligned}
&\left| \left[A^r \sum_{\alpha'=1}^{n^+} z(\alpha') - A^r x[(\theta+1)r] \right]_\alpha \right| \\
&\qquad = \left| \left[\sum_{\alpha'=1}^{n^+} A^r \left\{ z(\alpha') - \langle x_{\alpha'}[(\theta+1)r]\rangle_{\alpha'} \right\} \right]_\alpha \right| \\
&\qquad \leq \left\| \left[\sum_{\alpha'=1}^{n^+} A^r \left\{ z(\alpha') - \langle x_{\alpha'}[(\theta+1)r]\rangle_{\alpha'} \right\} \right]_{[\nu]} \right\| \\
&\qquad = \left\| \sum_{\alpha' : \nu(\alpha') = \nu} (A|_{M_\nu})^r \left\{ z(\alpha') - \langle x_{\alpha'}[(\theta+1)r]\rangle_{\alpha'} \right\} \right\| \\
&\qquad = \left\| (A|_{M_\nu})^r \sum_{\alpha' : \nu(\alpha') = \nu} \left\{ z(\alpha') - \langle x_{\alpha'}[(\theta+1)r]\rangle_{\alpha'} \right\} \right\| \\
&\qquad \leq \|(A|_{M_\nu})^r\| \left\| \sum_{\alpha' : \nu(\alpha') = \nu} \left\{ z(\alpha') - \langle x_{\alpha'}[(\theta+1)r]\rangle_{\alpha'} \right\} \right\|
\end{aligned}$$

$$= \|(A|_{M_\nu})^r\| \max_{\alpha':\nu(\alpha')=\nu} \|z(\alpha') - \langle x_{\alpha'}[(\theta+1)r]\rangle_{\alpha'}\|$$

$$\overset{(9.7.9)}{\leq} \|(A|_{M_\nu})^r\| \max_{\alpha':\nu(\alpha')=\nu} \frac{3\Gamma}{F_{\alpha'}-1}$$

$$\overset{(9.7.14)}{=\!=} \|(A|_{M_\nu})^r\| \frac{3\Gamma}{F_\nu - 1} \overset{(9.7.6)}{\leq} \frac{3\Gamma}{F_\nu - 1}\psi_\nu(r)|\lambda_{[\nu]}|^r.$$

The proof is completed as follows:

$$|x_\alpha[(\theta+2)r]|$$

$$\overset{(9.7.4)}{=\!=} \left| \left[A^r x[(\theta+1)r] + \sum_{i=1}^{l}\sum_{t=0}^{r-1} A^{r-1-t}B_i u_i[(\theta+1)r+t] + \zeta(\theta+1) \right]_\alpha \right|$$

$$\overset{(9.7.19)}{=\!=} \left| \left[A^r x[(\theta+1)r] - A^r \sum_{\alpha'=1}^{n^+} z(\alpha') + \zeta(\theta+1) \right]_\alpha \right|$$

$$\leq \frac{3\Gamma}{F_\nu - 1}\psi_\nu(r)|\lambda_{[\nu]}|^r + \left\|\zeta_{[\nu]}(\theta+1)\right\|$$

$$\overset{(9.7.7)}{\leq} \left[\frac{3\Gamma}{F_\nu - 1}\psi_\nu(r)| + \phi_\nu(r) \right] |\lambda_{[\nu]}|^r < D_\nu^+,$$

where the last inequality is taken from (9.7.18). □

Proposition 9.7.28 is immediate from Lemmas 9.7.29 and 9.7.30.

Analysis of the Sufficient Conditions for Stability (9.7.15)

Now we prove the following claim, which is the main result of the current subsection.

Proposition 9.7.31. *Whenever* **a)** *of Theorem* 9.4.27 *(on p.* 300*) holds, a choice of the parameters* Γ, d, N *exists under which the control strategy* **A)–H)** *(see pp.* 339–343*) regularly stabilizes the plant via the network* **CBE(NW)**.

Proof. By Proposition 9.7.28, it suffices to show that (9.7.15) can be ensured by a proper choce of the parameters. We first observe that (9.7.15) is true whenever

$$\frac{2}{3}\Gamma \min_{\nu=1,\dots,g} (F_\nu - 1)^{-1} > d > 2 \max_{\nu=1,\dots,g} \varphi_\nu(r)|\lambda_{[\nu]}|^r,$$

$$\Gamma\left[1 - 3\frac{\psi_\nu(r)|\lambda_{[\nu]}|^r}{F_\nu - 1}\right] > \left[D_\nu^*\psi_\nu(r) + 2\phi_\nu(r) + \varphi_\nu(r)\right]|\lambda_{[\nu]}|^r \qquad \forall \nu.$$

These inequalities can clearly be ensured by picking first Γ to be large enough and then d within the indicated bounds provided that

$$\frac{\psi_\nu(r)|\lambda_{[\nu]}|^r}{F_\nu - 1} < \frac{1}{3} \qquad \forall \nu. \tag{9.7.20}$$

Thus the objective focuses on ensuring (9.7.20).

Note that by Remark 9.7.26 (on p. 343), the epoch duration $r = N_{m\to s} + N + N_{\text{serv}}$ and the length $N+1$ of the block code (see Fig. 9.37) can be made arbitrarily large, whereas $N_{m\to s}$ does not depend on N and $N_{\text{serv}} \le N_*$ with N_* independent of N. By (9.7.1) (on p. 330) and (9.7.14) (on p. 344), $F_\nu > |\lambda_{[\nu]}|^N 2^{N\delta}$. By taking into account that the eigenvalue $\lambda_{[\nu]}$ is unstable $|\lambda_{[\nu]}| \ge 1$ and the polynomial $\psi_\nu(\cdot)$ can be upper bounded in the form $\psi_r(r) \le c_\nu(1 + r^{s_\nu})$, $s_\nu = \deg \psi_\nu$, we get

$$\frac{\psi_\nu(r)|\lambda_{[\nu]}|^r}{F_\nu - 1} \le \frac{c_\nu\left[1 + (N + N_{m\to s} + N_*)^{s_\nu}\right]|\lambda_{[\nu]}|^N|\lambda_{[\nu]}|^{N_{m\to s}+N_*}}{|\lambda_{[\nu]}|^N 2^{\delta N} - 1}$$

$$= \frac{c_\nu\left[1 + (N + N_{m\to s} + N_*)^{s_\nu}\right]|\lambda_{[\nu]}|^{N_{m\to s}+N_*}}{2^{\delta N} - |\lambda_{[\nu]}|^{-N}} \to 0 \quad \text{as} \quad N \to \infty.$$

Thus (9.7.20) can be ensured by picking N large enough. □

Remark 9.7.32. By our construction, the strategy at hand is periodic.[30]

9.7.4 Completing the Proof of b) from Theorem 9.4.27: Stabilization over the Original Network

We recall that the control-based extension **CBE(NW)** is composed of the original network **NW** and imaginary feedback tunnels. The latter symbolize the possibility to transmit data through the plant by means of control (see Fig. 9.38). Their explicit incorporation into the model is of sense when the network is considered in the open loop and its capacity domain is computed.

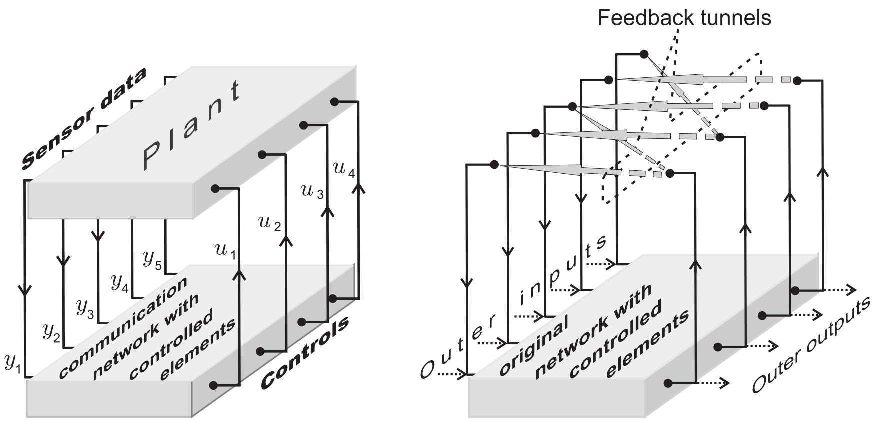


Fig. 9.38(a). Decentralized control. **Fig. 9.38(b).** Control-based extension.

It follows that the feedback tunnels become superfluous if **CBE(NW)** is put in the feedback interconnection with the plant. Exactly this situation holds whenever a

[30]See Definition 9.4.24 (on p. 299).

stabilizing control strategy for **CBE(NW)** is considered. Then all data streams that are interpreted as going through the imaginary feedback tunnels and so virtual can be arranged as real data transmissions by control actions upon the plant (see Fig. 9.39). The corresponding arrangements are proposed in this subsection and result in a stabilizing control strategy for the original network **NW**.

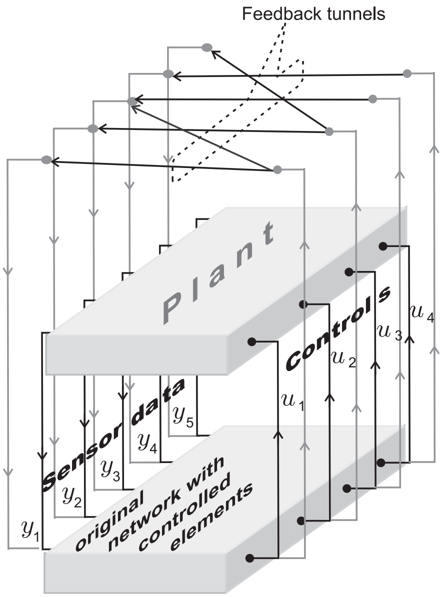


Fig. 9.39(a). Control over **CBE(NW)**.

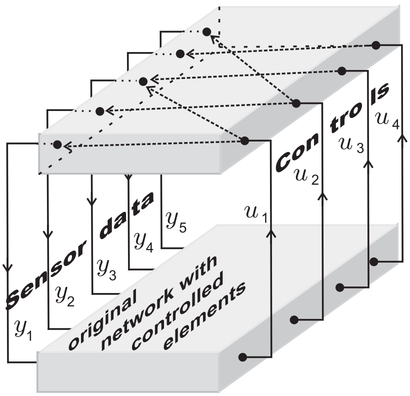


Fig. 9.39(b). Embedding the feedback tunnels back into the plant.

In this subsection, the stabilizing control strategy for **CBE(NW)** proposed in the previous one is said to be *initial*. When doing the above arrangements, the ideal objective might be to convert this strategy into a control strategy for the original network **NW** without any alteration in network outputs (controls) and inputs (observations). Unfortunately, this is impossible since the very idea of communication by means of control dictates to alter the controls by adding "information carrier" addends:

$$u_i(t) := u_i^{\mathrm{b}}(t) + u_i^{\mathrm{com}}(t), \tag{9.7.21}$$

where $u_i^{\mathrm{b}}(t)$ are initially generated controls organized into control programs $\boldsymbol{U}_i(\theta)$. This also alters the sensor signals.

To state the realistic objective, we recall that for the initial control strategy,

- The sensor data are collected during n time steps before the start $t = \theta r$ of any epoch $\theta \geq 1$ and utilized at $t = \theta r$ by the sensor nodes to compute the prognosis of the observed unstable modes $\widehat{x}_\alpha^j[(\theta + 1)r]$;

- The actuator nodes generate control programs

$$\boldsymbol{U}_i(\theta) = \mathbf{col}\left\{u_i^{\mathrm{b}}[\theta r], u_i^{\mathrm{b}}[\theta r+1], \ldots, u_i^{\mathrm{b}}[(\theta+1)r-1]\right\} \tag{9.7.22}$$

for the entire epoch at its beginning θr.

There is no other interaction between the network **CBE(NW)** and plant.

The objective that will be pursued is to alter the control strategy so that

S] The sensor nodes
1. Still compute the prognosis $\widehat{x}_\alpha^j[(\theta+1)r]$ at $t = \theta r$;
2. Get access to the data initially received over the imaginary feedback tunnels in another and realistic way; to serve this access, the processing algorithms at these nodes can be supplemented by additional operations;
3. Process the above data, estimates $\widehat{x}_\alpha^j[(\theta+1)r]$, and data received via the original network **NW** just like for the initial strategy;

I] The interior part of the network processes data just like the initial strategy;

A] The actuator nodes
1. Process the received data just like for the initial strategy in order to generate the control programs $\boldsymbol{U}_i(\theta)$ at the start of any epoch;
2. Do not send signals into imaginary feedback tunnels but instead, use these signals to generate the "communication" addends $u_i^{\mathrm{com}}(t)$ to the controls from **A.1]**;
3. Generate these addends in such a way that
   a) Their influence on the plant is annihilated at the end of the current epoch: the controls $u_i^{\mathrm{com}}[\theta r], \ldots, u_i^{\mathrm{com}}[(\theta+1)r-1]$ drive the unperturbed ($\xi(\cdot) \equiv 0$) plant from 0 at $t = \theta r$ to 0 at $t = (\theta+1)r$ provided that all actuators except for the ith one are inactive;
   b) They are bounded $\|u_i^{\mathrm{com}}(t)\| \leq c_u$ by a constant c_u independent of t, i;
   c) The overall controls given by (9.7.21) are zero n time steps before the start of any epoch.

Remark 9.7.33. Note that **S.1]** is possible whenever **S.2], A.3.a]**, and **A.3.c]** hold.

Indeed, **A.3.c)** implies Assumption 9.7.12 (on p. 337). At the same time, **S.2]** means that at $t = \theta r + N_{m\to s} + N + d$ (see Fig. 9.37), the sensors still have access to the data that make them aware about the required controls from **A.1]**. Due to **A.3.a]**, their effect on the plant during the entire epoch is identical to that of the overall controls (9.7.21). Thus the latter effect is computable by the sensor nodes, as is required by Assumption 9.7.13 (on p. 338). Hence the remark follows from Proposition 9.7.14 (on p. 338).

Lemma 9.7.34. *Any control strategy with the features* **S], I], A]** *regularly stabilizes*[31] *the plant via the original network* **NW**.

[31] See Definition 9.4.2 on p. 293.

Proof. By **S], I]**, and **A]**, this strategy generates just the same control programs $\boldsymbol{U}_i(\theta)$ as the initial strategy. At the same time, the "communication addends" from (9.7.21) dot not influence the states $x(\theta r)$ due to **A.3.a]**. Since the initial strategy is regularly stabilizing, it follows that the states $x(\theta r)$ and control programs $\boldsymbol{U}_i(\theta)$ are bounded over $\theta \geq 0$. Thanks to **A.3.b]**, this conclusion is extended on the overall controls $u_i(t)$ given by (9.7.21) and $t = 0, 1, \ldots$. Then the standard argument (see, e.g., (7.8.20) on p. 244) shows that the states $x(t)$ are bounded over $t \geq 0$, which completes the proof. □

Converting the initial control strategy into one with the features **S], I]**, and **A]** comes to arrangements for **S.2]** and **A.2]**. To this end, we recall that for the initial strategy, transmissions over the feedback tunnels were used to communicate (see Fig. 9.40)

1. Side data from the master sensor nodes to their slaves;
2. Quantized state prognosis from the sensor nodes to actuator nodes by means of the basic networked block code **BC** hosted by **CBE(NW)**;
3. Data $o_{i\to\alpha}$ driving control generation from the actuators to sensors.

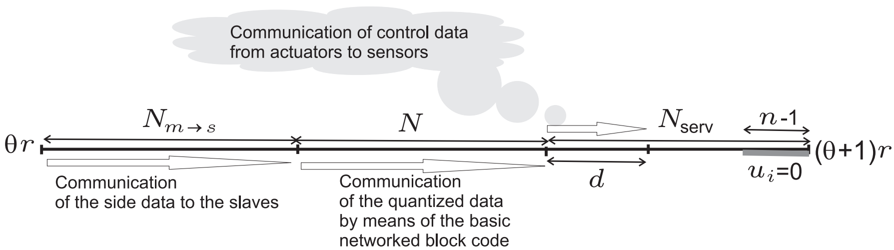


Fig. 9.40. Where the feedback tunnels are used.

Formally, these data are represented by, respectively,

1. A map associating every master sensor with the side data dispatched from this node and represented by a binary word of length 2;
2. A map associating every pair sensor-(observed unstable mode x_α) with a quantized data from the set $[1 : F_\alpha]$;
3. A map associating every pair (actuator i)-(affected unstable mode x_α) with a real-valued signal $o_{i\to\alpha}$.

In each case, these data are generated before the transmission is commenced and belong to a priori known finite sets. (In case 3, this holds since the signals $o_{i\to\alpha}$ are generated as outputs of the deterministic block code **BC** driven by the data from 2.)

Since the network **CBE(NW)** is deterministic, its operation during any of the above transmissions is uniquely determined by the transmitted data. It follows that at any time t the message emitted into any feedback tunnel belongs to an a priori known finite set $\mathfrak{R}_{i,j}(t)$. Here the pair (i,j) enumerates the tunnel and belongs to the set (9.4.6) (on p. 294). The corresponding tunnel transmits messages from the actuator node i to the sensor node j with the delay $d_{i \to j}$ given by (9.4.7) (on p. 294).

Thus the required arrangements focus on making the sensor j aware about the choice from a known finite set $\mathfrak{R}_{i,j}(t)$ that was made by the ith actuator $d_{i \to j}$ time steps before. This should be accomplished via generating a proper control u_i^{com}.

A method to achieve this goal was proposed in Subsect. 7.8.1 (starting on p. 233) for the case of one sensor and actuator. Generalization on the case of several sensors and actuators is straightforward; the required details are offered further (see p. 353).

According to this method, the initial control strategy hosted by **CBE(NW)** is converted into a control strategy hosted by the original **NW** as follows:

C1) The transmissions over the imaginary feedback tunnels are replaced by
- Generating controls at the actuator nodes in the form (9.7.21), where the first addend on the right is the corresponding entry of the control program from **A.1]**, and the second addend is taken from a finite control alphabet and encodes the message initially transmitted over the feedback tunnel;
- Decoding these messages at the sensor nodes on the basis of the prior observations and previously decoded messages.

C2) After the transmissions of all these messages are completed at time $t_\theta = \theta r + N_{m \to s} + N + d$ (see Fig. 9.40), the influence of the generated communication controls

$$\boldsymbol{U}_i^{\text{com}} = \textbf{col}\, \left(u_i^{\text{com}}[\theta r], \ldots, u_i^{\text{com}}[t_\theta - 1]\right)$$

on the plant is compensated by producing the "remainder" $u_i^{\text{com}}(t)\Big|_{t=t_\theta}^{(\theta+1)r-1}$ so that **A.3.a)** be satisfied.

Explanation 9.7.35. The latter is possible due to (9.7.13) (on p. 343). Moreover like in **ds.3)** on p. 240, the "remainder" can be produced as a function of $\boldsymbol{U}_i^{\text{com}}$ so that $u_i^{\text{com}}[t_\theta + n] = 0, \ldots, u_i^{\text{com}}[(\theta+1)r - 1] = 0$.

Remark 9.7.36. Thanks to (9.7.13), the last property implies that the overall controls, i.e., the right-hand sides of (9.7.21), are zero n time steps before the start of any epoch, which ensures **A.3.c)** and so by Remark 9.7.33 (on p. 350), **S.1]**.

Remark 9.7.37. The "remainder" from Explanation 9.7.35 assumes finitely many values since it is a function of the element $\boldsymbol{U}_i^{\text{com}}$ taken from a finite set. This ensures **A.3.b]**.

Observation 9.7.38. *For the strategy obtained via the above conversion, the claims* **S], I]**, *and* **A]** *are true. So this strategy regularly stabilizes the plant via the original network by Lemma* 9.7.34.

Observation 9.7.39. *By our construction, this strategy is r-periodic.*[32]

Remark 9.7.40. These two observations complete the proof of the **a)** $\Rightarrow$ **b)** part of Theorem 9.4.27 (on p. 300), provided that **C1)** is ensured.

We close the section with specifying the technical details related to **C.1)**.

[32]See Definition 9.4.24 (on p. 299).

Feedback Information Transmission by Means of Control in the Case of Multiple Sensors and Actuators

Now we deal with not necessarily the above control strategy but any strategy that satisfies the following requirements:

R.1) The controls are generated in the form (9.7.21);
R.2) At time $t = \theta r$, every sensor node j can compute the control programs (9.7.22) for all actuators $i : (i, j) \in \text{CVP}$ that affect the observations of this sensor;[33]
R.3) The claim **A.3.c)** on p. 350 holds.

The following technical fact [8] is immediate from **A.3.c]** and (9.4.3) (on p. 292).

Remark 9.7.41. At time $\theta r, \theta \geq 1$, any sensor node j can compute the state $x(\theta r)$ modulo the unobservable subspace L_j^{-o} with an accuracy D_j^{-o} that depends only on A, C_j and the bounds from (9.4.3).

The following lemma extends Lemma 7.8.4 (on p. 234) on the case of multiple sensors and actuators and displays the key fact that enables one to establish the required feedback communication by means of control.

Lemma 9.7.42. *Suppose that* **R.1)–R.3)** *hold and a natural number* $N_{i,j}$ *is assigned to any feedback tunnel (i.e., any* $(i, j) \in \text{CVP}$*, where* CVP *is given by* (9.4.6) *on p.* 294*). Then for any given sensor* j*, the related feedback tunnels can be equipped with finite control alphabets of the respective sizes* $N_{i,j}$

$$\mathfrak{U}_{i,j} = \Big\{u_{i,j}(1), \ldots, u_{i,j}(N_{i,j})\Big\} \subset \mathbb{R}^{\dim(u_i)} \tag{9.7.23}$$

so that the following property holds.

- *Whenever at a time* $t = \theta r + \tau \leq (\theta + 1)r - n, \theta \geq 1$ *the sensor node* j *is aware of the controls* $u_i(\theta r), \ldots, u_i(t - d_{i\to j} - 1)$ *for all actuators* $i : (i, j) \in \text{CVP}$ *affecting the observations of this sensor, this node is able to determine the second addend* $u_i^{\text{com}}(t - d_{i\to j})$ *in* (9.7.21) *for all concerned* i *provided that* $u_i^{\text{com}}(t - d_{i\to j}) \in \mathfrak{U}_{i,j}\ \forall i : (i, j) \in \text{CVP}$.

It follows that the sensor is able to determine the serial number s of the corresponding element $u_{i,j}(s) = u_i^{\text{com}}(t - d_{i\to j})$ of the alphabet (9.7.23). Thus the ith actuator can communicate any message $s \in [1 : N_{i,j}]$ to the jth sensor with delay $d_{i\to j}$.

Explanation 9.7.43. It is tacitly assumed that the communication controls are chosen from $\mathfrak{U}_{i,j}$ at not only time $t - d_{i,j}$ but also at all previous times to ensure prior communications. Then the knowledge of $u_i(\theta r), \ldots, u_i(t - d_{i,j} - 1)$ results from **R.1)** and **R.2)**, along with previous decoding.

Proof of Lemma 9.7.42. By Remark 9.7.41, the jth sensor is able to compute an estimate $\widehat{x}_j(\theta r)$ of $x(\theta r)$ such that

[33] See Remark 9.7.23 (on p. 343).

$$\|C_j A^s[\widehat{x}_j(\theta r) - x(\theta r)]\| \le D_j^- := \|C_j\|(\|A\|+1)^r D_j^{-o} \tag{9.7.24}$$

whenever $s \in [0:r]$. Now we observe that owing to (9.4.1) and (9.4.2) (on p. 291),

$$y_j(t) = C_j A^\tau x(\theta r) + \sum_{s=0}^{\tau-1} C_j A^{\tau-1-s}\left[\sum_{i=1}^{l} B_i u_i(\theta r + s) + \xi(\theta r + s)\right] + C_j \chi_j(t)$$

$$\overset{(9.4.6),\,(9.4.7)}{=\!=\!=} C_j A^\tau x(\theta r) + \sum_{i:(i,j)\in \mathrm{CVP}} \sum_{s=0}^{\tau-d_{i\to j}} C_j A^{\tau-1-s} B_i u_i(\theta r + s)$$

$$+ \sum_{s=0}^{\tau-1} C_j A^{\tau-1-s}\xi(\theta r + s) + C_j \chi_j(t)$$

$$\overset{(9.7.21)}{=\!=\!=} \left\{ \begin{array}{l} C_j A^\tau \widehat{x}_j(\theta r) + \displaystyle\sum_{i:(i,j)\in \mathrm{CVP}} \sum_{s=0}^{\tau-1-d_{i\to j}} C_j A^{\tau-1-s} B_i u_i(\theta r + s) \\ + \displaystyle\sum_{i:(i,j)\in \mathrm{CVP}} C_j A^{d_{i\to j}-1} B_i u_i^{\mathrm{b}}(t - d_{i\to j}) \end{array} \right\}$$

$$+ \left\{ \sum_{i:(i,j)\in \mathrm{CVP}} C_j A^{d_{i\to j}-1} B_i u_i^{\mathrm{com}}(t - d_{i\to j}) \right\}$$

$$+ \left\{ C_j A^\tau[x(\theta r) - \widehat{x}_j(\theta r)] + \sum_{s=0}^{\tau-1} C_j A^{\tau-1-s}\xi(\theta r + s) + C_j \chi_j(t) \right\}.$$

Here the discrepancy H_j between $y_j(t)$ and the first expression in the curly brackets is computable at the jth sensor node at time t. The norm of the third one is strictly bounded by a constant $D_*^{[0]}$ that does not depend on t, θ, j due to (9.4.3) (on p. 292) and (9.7.24). Hence

$$\left\| H_j - \sum_{i:(i,j)\in \mathrm{CVP}} C_j A^{d_{i\to j}-1} B_i u_i^{\mathrm{com}}(t - d_{i\to j}) \right\| < D_*^{[0]}. \tag{9.7.25}$$

Note that here $C_j A^{d_{i\to j}-1} B_i \neq 0$ due to (9.4.7) (on p. 294).

Now we put the set $\{i : (i,j) \in \mathrm{CVP}\}$ in order $\{i^1, \ldots, i^{p_j}\}$ and pick the alphabets (9.7.23) successively for $i = i^1, \ldots, i^{p_j}$. For $i = i^1$, a finite alphabet (9.7.23) (where $i := i^1$) is chosen so that the points

$$C_j A^{d_{i^1\to j}-1} B_{i^1} u_{i^1,j}(1), \ldots, C_j A^{d_{i^1\to j}-1} B_{i^1} u_{i^1,j}(N_{i^1,j})$$

are $2D_*^{[0]}$-separated.[34] Then the control $u_{i^1}^{\mathrm{com}}(t - d_{i^1\to j})$ from this alphabet can be uniquely determined from the knowledge of a vector $H_{j,1}$ such that

$$\left\| H_{j,1} - C_j A^{d_{i^1\to j}-1} B_{i^1} u_{i^1}^{\mathrm{com}}(t - d_{i^1\to j}) \right\| < D_*^{[0]}.$$

[34] See Definition 7.6.17 (on p. 217).

After the first alphabet is chosen, (9.7.25) implies

$$\left\| H_j - \sum_{i=i^2,\dots,i^{p_j}} C_j A^{d_{i\to j}-1} B_i u_i^{\text{com}}(t - d_{i\to j}) \right\| < D_*^{[1]}$$
$$:= D_*^{[0]} + \max_{s=1,\dots,N_{i^1,j}} \|C_j A^{d_{i^1\to j}-1} B_{i^1} u_{i^1,j}(s)\|. \quad (9.7.26)$$

For $i = i^2$, the alphabet (9.7.23) is chosen so that the points

$$C_j A^{d_{i^2\to j}-1} B_{i^2} u_{i^2,j}(1), \dots, C_j A^{d_{i^2\to j}-1} B_{i^2} u_{i^2,j}(N_{i^2,j})$$

are $2D_*^{[1]}$-separated. Then the control $u_{i^2}^{\text{com}}(t - d_{i^2\to j})$ from this alphabet can be uniquely determined from the knowledge of a vector $H_{j,2}$ such that

$$\left\| H_{j,2} - C_j A^{d_{i^2\to j}-1} B_{i^2} u_{i^2}^{\text{com}}(t - d_{i^2\to j}) \right\| < D_*^{[1]}.$$

Note that (9.7.26) implies

$$\left\| H_j - \sum_{i=i^3,\dots,i^{p_j}} C_j A^{d_{i\to j}-1} B_i u_i^{\text{com}}(t - d_{i\to j}) \right\| < D_*^{[2]}$$
$$:= D_*^{[1]} + \max_{s=1,\dots,N_{i^2,j}} \|C_j A^{d_{i^2\to j}-1} B_{i^2} u_{i^2,j}(s)\|.$$

By continuing likewise, we construct a sequence $D_*^{[0]} \leq D_*^{[1]} \leq \dots \leq D_*^{[p_j-1]}$ and the alphabets (9.7.23) for all $i = i^1, \dots, i^{p_j}$ such that for any $s = 1, \dots, p_j$,

i) the control $u_{i^s}^{\text{com}}(t - d_{i^s\to j})$ from the alphabet $\mathfrak{U}_{i^s\to j}$ can be uniquely determined from the knowledge of a vector $H_{j,s}$ such that

$$\left\| H_{j,s} - C_j A^{d_{i^s\to j}-1} B_{i^s} u_{i^s}^{\text{com}}(t - d_{i^s\to j}) \right\| < D_*^{[s-1]}. \quad (9.7.27)$$

ii) the following inequality holds:

$$\left\| H_j - \sum_{i=i^s,\dots,i^{p_j}} C_j A^{d_{i\to j}-1} B_i u_i^{\text{com}}(t - d_{i\to j}) \right\| < D_*^{[s-1]}. \quad (9.7.28)$$

For $s := p_j$, (9.7.28) has the form (9.7.27) with $H_{j,p_j} := H_j$ known. So $u_{i^s}^{\text{com}}(t - d_{i^s\to j})$ with $s = p_j$ can be determined. Then for $s = p_j - 1$, (9.7.28) takes the form (9.7.27) with known

$$H_{j,p_j-1} := H_j - C_j A^{d_{i^{p_j}\to j}-1} B_{i^{p_j}} u_{i^{p_j}}^{\text{com}}(t - d_{i^{p_j}\to j}).$$

This permits us to determine $u_{i^s}^{\text{com}}(t - d_{i^s\to j})$ with $s = p_j - 1$. Then we subtract from H_j the found addends

$$H_{j,p_j-2} := H_j - \sum_{i=i^{p_j-1},i^{p_j}} C_j A^{d_{i\to j}-1} B_i u_i^{\mathrm{com}}(t-d_{i\to j})$$

and find $u_{i^s}^{\mathrm{com}}(t-d_{i^s\to j})$ with $s = p_j - 2$. All required controls can be found by continuing likewise. □

Lemma 9.7.42 shows that the actuators can send as much information as desired by means of control and with the required delays to any given sensor. The next lemma extends this conclusion to the case where all sensors should receive data simultaneously.

Lemma 9.7.44. *Suppose that* **R1)–R3)** *hold and a finite set of messages* $\mathfrak{M}_{i,j}$ *is assigned to any feedback tunnel, which are enumerated by* $(i,j) \in \mathrm{CVP}$*, where* CVP *is given by* (9.4.6) *(on p. 294). Let at any time* $t \in [\theta r : (\theta+1)r-n-d]$*, where* $\theta \geq 1$ *and* $d := \max_{(i,j)\in\mathrm{CVP}} d_{i\to j}$*, every actuator* i *picks messages* $\mu_{i\to j}(t) \in \mathfrak{M}_{i,j}$ *to all concerned sensors* j*, i.e., such that* $(i,j) \in \mathrm{CVP}$.[35]

Then each actuator i *can be equipped with a rule converting the set of related messages*

$$\Big\{\mu_{i\to j}(t)\Big\}_{j:(i,j)\in\mathrm{CVP}} \mapsto u_i^{\mathrm{com}}(t) \tag{9.7.29}$$

into the second addend from (9.7.21) *(on p. 349) so that at any time* $t \in [\theta r : (\theta+1)r - n]$ *every sensor* j *can recognize all required messages*

$$\mu_{i\to j}\big(t-d_{i\to j}\big), \quad i:(i,j)\in\mathrm{CVP},\ t-d_{i\to j} \geq \theta r \tag{9.7.30}$$

from the knowledge available to this sensor. The conversion rules and the algorithms of recognition can be chosen independent of θ *and* t.

Remark 9.7.45. Since the required transmissions from **C1)** (on p. 352) concern messages taken from a priori-known finite sets according to the discussion preceding **C1)**, Lemma 9.7.44 justifies the possibility of **C1)** and completes the proof of the **a)** ⇒ **b)** part of Theorem 9.4.27 (on p. 300) by Remark 9.7.40.

Proof of Lemma 9.7.44. Let every actuator i first convert the set of all dispatched messages from (9.7.29) into a compound message $\mu_i(t)$, so that any original message $\mu_{i\to j}(t)$ from the set and its indices i, j are uniquely reconstructable from $\mu_i(t)$. This compound message will be delivered to every concerned sensor j, where the proper part will be extracted. The message $\mu_i(t)$ can be taken from a specific finite set $\mathfrak{M}_i$.

The remainder of the proof involves merely adjusting the arguments from the proof of Lemma 9.7.42 to the current context. In so doing, we pick the control alphabets (9.7.23) in a special way. We first observe that due to (9.4.7) (on p. 294),

$$C_j A^{d_{i\to j}-1} B_i \neq 0 \Leftrightarrow L_{i,j} := \ker C_j A^{d_{i\to j}-1} B_i \neq \mathbb{R}^{\dim(u_i)}, \qquad \forall (i,j) \in \mathrm{CVP}.$$

Hence the Lebesgue measure of the linear subspace $L_{i,j}$ is zero; and so is that of the union $\bigcup_{j:(i,j)\in\mathrm{CVP}} L_{i,j}$. Thus for any i, a control $u^i \in \mathbb{R}^{\dim(u_i)}$ exists such that

[35] It is assumed that every message $\mu_{i\to j}(t)$ should be delivered to sensor j with the delay $d_{i\to j}$ given by (9.4.7) (on p. 294). We set $\mu_{i\to j}(t) := \circledast$ for $t \in \big((\theta+1)r-n-d : (\theta+1)r\big)$.

$$C_j A^{d_{i\to j}-1} B_i u^i \neq 0, \qquad \forall j : (i,j) \in \text{CVP}. \tag{9.7.31}$$

To any actuator $i = 1, \dots, l$, we assign a finite control alphabet $\mathfrak{U}_i = \left\{u^i_\mu\right\}_{\mu \in \mathfrak{M}_i}$ whose elements are in a one-to-one correspondence with the messages from $\mathfrak{M}_i$ and have the following form:

$$u^i_\mu = a^i_\mu u^i, \qquad \text{where} \quad a^i_\mu \in \mathbb{R}. \tag{9.7.32}$$

As in the proof of Lemma 9.7.42, these alphabets will be chosen successively for $i = 1, \dots, l$. For $i = 1$, the numbers $a^1_\mu, \mu \in \mathfrak{M}_1$ are chosen so that the set

$$\left\{C_j A^{d_{1\to j}-1} B_1 u^1_\mu\right\}_{\mu \in \mathfrak{M}_1}$$

is $2D^{[0]}_*$-separated for any sensor j such that $(1,j) \in \text{CVP}$. Here $D^{[0]}_*$ is the constant from (9.7.25). This is possible thanks to (9.7.31) and (9.7.32). The numbers $a^2_\mu, \mu \in \mathfrak{M}_2$ are chosen so that the set

$$\left\{C_j A^{d_{2\to j}-1} B_2 u^2_\mu\right\}_{\mu \in \mathfrak{M}_2}$$

is $2D^{[1]}_*$-separated for any sensor $j : (2,j) \in \text{CVP}$. Here $D^{[1]}_*$ is defined as in (9.7.26) with $i^\nu := \nu$:

$$D^{[1]}_* := D^{[0]}_* + \max_{j=1,\dots,k} \max_{\mu \in \mathfrak{M}_1} \left\| C_j A^{d_{1\to j}-1} B_1 u^1_\mu \right\|.$$

Here and throughout, we assume that $d_{i\to j} := 1$ if $(i,j) \notin \text{CVP}$ and note that

$$C_j A^{d_{i\to j}-1} B_i = 0, \qquad \forall\, (i,j) \notin \text{CVP} \tag{9.7.33}$$

due to (9.4.6) (on p. 294). The numbers $a^3_\mu, \mu \in \mathfrak{M}_3$ are chosen so that the set

$$\left\{C_j A^{d_{3\to j}-1} B_3 u^3_\mu\right\}_{\mu \in \mathfrak{M}_3}$$

is $2D^{[2]}_*$-separated for any sensor $j : (3,j) \in \text{CVP}$, where

$$D^{[2]}_* := D^{[1]}_* + \max_{j=1,\dots,k} \max_{\mu \in \mathfrak{M}_2} \left\| C_j A^{d_{2\to j}-1} B_2 u^2_\mu \right\|.$$

By continuing likewise, we construct a sequence $D^{[0]}_* \leq \cdots \leq D^{[l-1]}_*$, control alphabets $\mathfrak{U}_i = \left\{u^i_\mu\right\}_{\mu \in \mathfrak{M}_i}$ for all actuators i, and ensure that the following claims hold:

i) For any actuator $i = 1, \dots, l$, any control $u_i \in \mathfrak{U}_i$ can be uniquely recognized by any sensor $j : (i,j) \in \text{CVP}$ from the knowledge of a vector $H_{i,j}$ such that

$$\left\| H_{i,j} - C_j A^{d_{i\to j}-1} B_i u_i \right\| < D^{[i-1]}_*;$$

ii) The numbers $D_*^{[i]}$ are related by the recursion

$$D_*^{[i]} := D_*^{[i-1]} + \max_{j=1,\ldots,k} \max_{\mu \in \mathfrak{M}_i} \left\| C_j A^{d_{i\to j}-1} B_i u_\mu^i \right\|. \tag{9.7.34}$$

Finally, we introduce the following conversion rule of the form (9.7.29)

$$\left\{ \mu_{i\to j}(t) \right\}_{j:(i,j)\in \text{CVP}} \mapsto \mu_i(t) \mapsto u_i^{\text{com}}(t) := u_{\mu_i(t)}^i(t) \in \mathfrak{U}_i. \tag{9.7.35}$$

To complete the proof, we show that for any $t \in [\theta r : (\theta+1)r - n]$, any sensor can recognize all required messages (9.7.30) and the controls $u_i(t - d_{i\to j})$ with $(i,j) \in$ CVP from the knowledge available to this sensor. The proof will be by induction on $t = \theta r, \theta r + 1, \ldots$.

For $t = \theta r$, the claim is evident: Since $n \geq d_{i\to j} \geq 1$ by (9.4.7) (on p. 294), no message from (9.7.30) should be recognized and $u_i(t - d_{i\to j}) = 0$ by **R3)** (on p. 353). Let the claim be true for all times $t' \geq \theta r$ not exceeding $t-1$, where $t \in (\theta r : (\theta+1)r - n]$. By retracing the arguments underlying (9.7.25) with regard to (9.7.33), we see that

$$\left\| H_j - \sum_{i=1}^{l} C_j A^{d_{i\to j}-1} B_i u_i^{\text{com}}(t - d_{i\to j}) \right\| < D_*^{[0]},$$

where H_j is a vector computable from the knowledge available to sensor j. Then (9.7.34) and (9.7.35) imply that

$$\left\| H_{i,j} - C_j A^{d_{i\to j}-1} B_i u_i^{\text{com}}(t - d_{i\to j}) \right\| < D_*^{[i-1]},$$

where

$$H_{i,j} := H_j - \sum_{i'>i} C_j A^{d_{i'\to j}-1} B_{i'} u_{i'}^{\text{com}}(t - d_{i'\to j}).$$

By putting $i := l$ here with regard to (9.7.33) and i), we see that sensor j can recognize $C_j A^{d_{l\to j}} B_l u_l^{\text{com}}(t - d_{l\to j})$ and if $(l,j) \in$ CVP, the control $u_l^{\text{com}}(t - d_{l\to j})$. After this, the sensor acquires the knowledge of $H_{l-1,j}$. Then by considering $i := l-1$ likewise, we see that sensor j can compute $C_j A^{d_{(l-1)\to j}} B_{l-1} u_{l-1}^{\text{com}}(t - d_{(l-1)\to j})$ and if $(l-1,j) \in$ CVP, the control $u_{l-1}^{\text{com}}(t - d_{(l-1)\to j})$. The proof is completed by continuing likewise. □

9.8 Proofs of the Lemmas from Subsect. 9.5.2 and Remark 9.4.28

9.8.1 Proof of Lemma 9.5.6 on p. 303

(i) is immediate from Definition 9.5.5 (on p. 303).

(ii) Let **BC** denote the corresponding networked block code, and N and $\boldsymbol{r}_{\text{code}} = \{r_T\}_{T\in\mathfrak{T}}$ stand for its length and rate ensemble, respectively. We recall that **BC** is an admissible data processing strategy $\boldsymbol{A} \in \mathfrak{A}$ considered on the time interval

$[0 : N-1]$ and processing outer inputs in the form $\mathfrak{y}_T \in [1 : F_T], T \in \mathfrak{T}$, and $r_T = N^{-1} \log_2 F_T$. By Definition 9.5.5 (on p. 303), it should be shown that an errorless block code of arbitrarily large length exists whose rate ensemble approaches an ensemble with positive entries at the positions from $\mathfrak{T}_*$.

To this end, we employ Assumption 9.4.20 (on p. 298) and prolong the strategy $\boldsymbol{\mathcal{A}}$ from $[0 : N-1]$ to $[0 : N_*] \supset [0 : N-1]$ to reset the network to the initial state at time $t = N_*$. Then by Assumption 9.4.22 (on p. 299) the resultant strategy can be repeated on the intervals $[0 : N_* - 1], [N_* : 2N_* - 1], \ldots, [(m-1)N_* : mN_* - 1]$, where m is arbitrary. For any $i = 0, \ldots, m-1$, the ith copy of **BC** located in $[iN_* : iN_* + N - 1]$ can be employed for errorless transmission of messages $\mathfrak{y}_T^{[i]} \in [1 : F_T], T \in \mathfrak{T}$. During the entire interval $[0 : mN_* - 1]$, whole ensembles of messages $\{\mathfrak{y}_T^{[i]}\}_{i=0,\ldots,m-1, T\in\mathfrak{T}}$ can be transmitted. They can be put into a one-to-one correspondence with the ensembles $\{\mathfrak{y}_T \in [1 : (F_T)^m]\}_{T\in\mathfrak{T}}$ by expansion

$$\mathfrak{y}_T =$$
$$= 1 + \left(\mathfrak{y}_T^{[0]} - 1\right) + \left(\mathfrak{y}_T^{[1]} - 1\right) F_T + \left(\mathfrak{y}_T^{[2]} - 1\right) (F_T)^2 + \cdots + \left(\mathfrak{y}_T^{[m-1]} - 1\right) (F_T)^{m-1}.$$

It follows that we have constructed an errorless block code with the length mN_* and the rate ensemble

$$\left\{ \frac{\log_2 (F_T)^m}{mN_*} \right\}_{T\in\mathfrak{T}} = \left\{ \frac{\log_2 F_T}{N_*} \right\}_{T\in\mathfrak{T}} = \left\{ \frac{N}{N_*} r_T \right\}_{T\in\mathfrak{T}}.$$

Since $r_T > 0 \forall\, T \in \mathfrak{T}_*$, letting $m \to \infty$ completes the proof.

(iii) By Definition 9.5.5 (on p. 303), $\boldsymbol{r} = \{r_T\}_{T\in\mathfrak{T}}$ can be approximated

$$|r_T - r_T'(m)| \leq \frac{\varepsilon}{2} \qquad \forall\, T \in \mathfrak{T} \tag{9.8.1}$$

by the rate ensembles $\left\{r_T'(m) = \frac{F_T(m)}{N'(m)}\right\}_{T\in\mathfrak{T}}$ of errorless networked block codes $\mathbf{BC}'(m) \sim \boldsymbol{\mathcal{A}}'(m), m = 0, 1, \ldots$ of arbitrarily large block lengths $N'(m) \xrightarrow{m\to\infty} \infty$. By passing to a subsequence, it can be ensured that the initial states $\mathcal{M}^0[\boldsymbol{\mathcal{A}}'(m)]$ of these codes are equivalent to each other modulo the partition from (ii) of Assumption 9.4.20 (on p. 298). For any $m = 1, 2, \ldots$, we construct the new code $\boldsymbol{\mathcal{A}}(m)$ as follows: First, the strategy $\boldsymbol{\mathcal{A}}'(0)$ is extended on a time interval of duration not exceeding $N'(0) + \delta T_{\max}$ to reset the network to the state initial for $\boldsymbol{\mathcal{A}}'(m)$, where $\delta T_{\max}$ is the constant from Assumption 9.4.20; and second, the resultant strategy is concatenated with $\boldsymbol{\mathcal{A}}'(m)$. These operations are possible thanks to Assumptions 9.4.20 and 9.4.22 (on p. 299). The required data transmission is in fact accomplished by the $\boldsymbol{\mathcal{A}}'(m)$ part of $\boldsymbol{\mathcal{A}}(m)$. It follows that this block code is errorless and $N'(m) \leq N(m) \leq N'(m) + \delta T_{\max} + N'(0)$, which implies that $N'(m)/N(m) \to 1$ as $m \to \infty$. For the rate ensemble $\left\{r_T(m) = \frac{F_T(m)}{N(m)}\right\}_{T\in\mathfrak{T}}$ of the constructed block code, we have

$$r_T(m) - r_T'(m) = r_T'(m) \left[\frac{N'(m)}{N(m)} - 1 \right] \xrightarrow{(9.8.1)} 0 \quad \text{as} \quad m \to \infty.$$

Since the block codes $\sim \mathcal{A}(m)$ have a common initial state by construction, the proof is completed by invoking (9.8.1). □

9.8.2 Proof of Lemma 9.5.8 on p. 304

The second inclusion from (9.5.22) (on p. 305) is evident. To prove the first one, we pick an element $\{r_\alpha\}$ from the set on the left. By (9.5.22), an ensemble

$$\boldsymbol{r} = \{r_{\alpha,i}\}_{(\alpha,i)\in\mathfrak{T}_{\mathbf{mw}}} \in \mathbf{CD}\Big[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \mathfrak{T}_{\mathbf{mw}}\Big]$$

exists such that $r_\alpha = \sum_{i\in\mathfrak{J}^{mw}_{\rightarrow\alpha}} r_{\alpha,i}\ \forall\alpha$. Let $0 < \delta < \min_{(\alpha,i):r_{\alpha,i}\neq 0} r_{\alpha,i}$ and $\overline{N} = 1,2,\ldots$ be given. By Definition 9.5.5 (on p. 303), there exists an errorless networked block code $\mathbf{BC}$ hosted by and serving the above network and transmission scheme, respectively, with block length $N \geq \overline{N}$ whose rate ensemble (9.5.19) (on p. 303) approaches $\boldsymbol{r}$ with accuracy δ.

Now we construct a block code $\mathbf{BC^{new}}$ serving $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^q_{mw}}$ and transmitting constant messages $\mu_\alpha \in [1 : \overline{F}_\alpha]$ from the source to the output node $\sim x_\alpha$ for all α. Here $\overline{F}_\alpha := \prod_{i\in\mathfrak{J}^{mw}_{\rightarrow\alpha}} F_{\alpha,i}$. Specifically when receiving the message μ_α from the source $\sim x_\alpha$, any sensor node $j \in \mathfrak{J}^{mw}_{\leftarrow\alpha}$ expands it in the form

$$\begin{aligned}\mu_\alpha = \mu_{\alpha,i^1_\alpha} + F_{\alpha,i^1_\alpha}(\mu_{\alpha,i^2_\alpha} - 1) + F_{\alpha,i^1_\alpha}F_{\alpha,i^2_\alpha}(\mu_{\alpha,i^3_\alpha} - 1) + \cdots \\ \cdots + F_{\alpha,i^1_\alpha}\cdots F_{\alpha,i^{p_\alpha-1}_\alpha}(\mu_{\alpha,i^{p_\alpha}_\alpha} - 1), \quad \mu_{\alpha,i^\sigma_\alpha} \in [1 : F_{\alpha,i^\sigma_\alpha}] \end{aligned} \tag{9.8.2}$$

and interprets μ_{α,i^σ} as the individual message from the source $\sim x_\alpha$ to the output node $i = i^\sigma$, thus creating the situation where $\mathbf{BC}$ can operate. Here $\{i^1_\alpha, \ldots, i^{p_\alpha}_\alpha\}$ is an enumeration of $\mathfrak{J}^{mw}_{\rightarrow\alpha}$. The signals $\mu_{\alpha,i}$ are processed by the rules of $\mathbf{BC}$. So every signal $\mu_{\alpha,i}$ is correctly decoded at time $N-1$ at the actuator node i. Into the additive channel from $\mathbf{SUFF^q_{mw}}$ going to the output node $\sim x_\alpha$, the actuator node emits the corresponding addend from (9.8.2). The addition is performed by the channel. So the correct message arrives at the output node $\sim x_\alpha$, and an errorless $\mathbf{BC^{new}}$ is constructed. The entries of its rate vector (9.4.13) (on p. 298)

$$\frac{\log_2 \overline{F}_\alpha}{N} = \sum_{i\in\mathfrak{J}^{mw}_{\rightarrow\alpha}} \frac{\log_2 F_{\alpha,i}}{N}$$

approach those $r_\alpha = \sum_{i\in\mathfrak{J}^{mw}_{\rightarrow\alpha}} r_{\alpha,i}$ of $\{r_\alpha\}$ with accuracy $\leq \delta|\mathfrak{J}^{mw}_{\rightarrow\alpha}|$, where $\delta > 0$ is arbitrary. The proof is completed by invoking Definition 9.4.17 (on p. 298). □

9.8.3 Proof of Lemma 9.5.10 on p. 306

(i) Let $\boldsymbol{r} = \{r_\alpha\}_{\alpha=1}^{n^+}$ belong to the set in the left-hand side of (9.5.25), and $\delta > 0, \overline{N} = 1,2,\ldots$ be given. By Definition 9.4.17 (on p. 298), an errorless $\mathbf{BC}$ hosted by $\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^+_{mw}}$ with block length $N \geq \overline{N}$ exists whose rate vector (9.4.13) (on p. 298) approaches $\boldsymbol{r}$ with accuracy δ.

We are going to transform it into a code $\mathbf{BC}^{\mathbf{new}}$ serving $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$. To this end, we introduce an enumeration $G_\nu = \{i_\nu^1, \dots, i_\nu^{p_\nu}\}$ for any ν. For $\mathbf{BC}^{\mathbf{new}}$, the νth source produces a constant message $\mu_\nu \in [1 : \overline{F}_\nu]$, where $\overline{F}_\nu := \prod_{\alpha \in G_\nu} F_\alpha$ and F_α is the size of the alphabet of the source $\sim x_\alpha$. On getting access to this message, each sensor node $j \in \mathfrak{J}^{mw}_{\leftarrow\alpha}$ (with $\alpha \in G_\nu$) expands it in the form (9.8.2) (where $\alpha := \nu$ and $F_{\nu,i} := F_i$). The quantities μ_{ν, i_ν^ϕ} are interpreted as messages from the node $\alpha = i_\nu^\phi, \phi = 1, \dots, p_\nu$ of $\mathbf{BC}$. These messages are then processed by the rules of the original $\mathbf{BC}$. So at $t = N - 1$, the message $\mu_{\nu,\alpha}$ is correctly decoded by the output node $\sim x_\alpha$. This node collects information from the actuator nodes $i \in \mathfrak{J}^{\mathrm{mw}}_{\rightarrow\alpha}$, which are common for all modes $\alpha \in G_\nu$ within a given group. Since the channels from $\mathbf{SUFF}^{+}_{\mathbf{mw}}$ have the maximal possible cardinalities, all data from the above actuator nodes can be removed to the single representative of the group of output nodes $\sim x_\alpha, \alpha \in G_\nu$ retained in $\overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$ as compared with $\mathbf{SUFF}^{+}_{\mathbf{mw}}$. It follows that all messages $\mu_{\nu,\alpha}$ with $\alpha \in G_\nu$ can be correctly decoded at this node. The original message μ_ν is restored by (9.8.2). Thus an errorless $\mathbf{BC}^{\mathbf{new}}$ is constructed.

The entries of its rate vector (9.4.13) (on p. 298)

$$\frac{\log_2 \overline{F}_\nu}{N} = \sum_{\alpha \in G_\nu} \frac{\log_2 F_\alpha}{N}$$

approach those of the vector $\left\{ \sum_{\alpha \in G_\nu} r_\alpha \right\}_{\nu=1}^{s}$ with accuracy $\leq \delta \max_\nu |G_\nu|$. Letting $\delta \to 0$ shows that the domain in the left-hand side of (9.5.25) (on p. 306) is a subset of that in the right-hand side by Definition 9.4.17 (on p. 298).

Conversely, let $\boldsymbol{r} = \{r_\alpha\}_{\alpha=1}^{n^+}$ belong to the set in the right-hand side of (9.5.25) (on p. 306), and let $0 < \delta < \min_{\alpha : r_\alpha > 0} r_\alpha, \overline{N} = 1, 2, \dots$ be given. By Definition 9.4.17 (on p. 298), an errorless block code $\mathbf{BC}$ hosted by $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$ with block length $N \geq \overline{N}$ exists whose rate vector $\left\{ \frac{\log_2 \overline{F}_\nu}{N} \right\}_{\nu=1}^{s}$ approaches the vector $\left\{ \sum_{\alpha \in G_\nu} r_\alpha \right\}_{\nu=1}^{s}$ with accuracy δ. We are going to transform it into a code $\mathbf{BC}^{\mathbf{new}}$ hosted by $\mathbf{PREF}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF}^{+}_{\mathbf{mw}}$.

Let the source $\sim \alpha$ of $\mathbf{PREF}_{\mathbf{mw}}$ produce constant messages $\mu_\alpha \in [1 : F_\alpha]$, where $F_\alpha := \max\left\{ \lfloor 2^{(r_\alpha - \delta)N} \rfloor ; 1 \right\}$. Any sensor node $j \in \mathfrak{J}^{\mathrm{mw}}_{\leftarrow\alpha}, \alpha \in G_\nu$ has access to all messages $\mu_{\alpha'}$ coming from the same group of input nodes $\alpha' \in G_\nu$. So all these sensor nodes may convert these messages into the common natural μ_ν by (9.8.2) (modified just as in the first part of the proof). Then μ_ν is interpreted as a common message broadcasted from the source $\sim \nu$, as is required by $\overline{\mathbf{PREF}}_{\mathbf{mw}}$. Moreover, this message is a legal input for $\mathbf{BC}$ since

$$\mu_\nu \leq \prod_{\alpha \in G_\nu} F_\alpha \leq \prod_{\alpha \in G_\nu} \left\lfloor 2^{(r_\alpha - \delta)N} \right\rfloor \leq \left\lfloor 2^{N\left(\sum_{\alpha \in G_\nu} r_\alpha - \delta |G_\nu|\right)} \right\rfloor \leq \overline{F}_\nu.$$

The messages μ_ν are then processed by the rules of $\mathbf{BC}$ and thus are correctly decoded at the output node $\sim \nu$ from the data received over the corresponding channel of $\overline{\mathbf{SUFF}}^{+}_{\mathbf{mw}}$. The clones of this channel are all channels from $\mathbf{SUFF}^{+}_{\mathbf{mw}}$ arriving at the nodes $\sim x_\alpha, \alpha \in G_\nu$. So at any of the latter nodes, μ_ν can also be correctly decoded and thus μ_α can be correctly restored via the expansion (9.8.2). Thus an errorless $\mathbf{BC}^{\mathbf{new}}$ is constructed.

For the entries of its rate vector (9.4.13) (on p. 298), we have

$$\frac{\log_2 F_\alpha}{N} = \max\left\{\frac{\log_2 \lfloor 2^{(r_\alpha-\delta)N}\rfloor}{N}; 0\right\}$$
$$\in \left[\frac{\log_2 \max\left\{2^{(r_\alpha-\delta)N}-1;1\right\}}{N}, \max\left\{r_\alpha-\delta;0\right\}\right].$$

Since $\delta < r_\alpha$ whenever $r_\alpha > 0$, the ends of this interval are close to r_α for $N \approx \infty$ and $\delta \approx 0$. It follows that the domain in the right-hand side of (9.5.25) is a subset of that in the left-hand side by Definition 9.4.17, which completes the proof.

(ii) Modulo apparent technical details, the proof is similar to that of (i). □

9.8.4 Proof of Remark 9.4.28 on p. 300

To prove this remark, we basically repeat the arguments from the proof of the necessity part of Theorem 9.4.27 (on p. 300) given in Sect. 9.6. This proof derived the required claim **d)** from stability under zero noises and stable modes. In other words, the stability assumption of **c)** was utilized only partly. Now we consider a more general situation where the plant noises may be nonzero within the first m time steps:

$$\xi(t) \equiv 0 \quad \forall t \geq m, \quad \chi_j(t) \equiv 0, \quad \xi_\alpha(t) = 0,\ x_\alpha^0 = 0 \quad \forall t \geq 0, \alpha \leq 0. \quad (9.8.3)$$

So the sequence of the plant noises is of the form

$$\Xi = \{\xi(0), \xi(1), \ldots, \xi(m-1), 0, 0, \ldots\}, \qquad \xi(t) \in M_{\text{unst}}(A). \qquad (9.8.4)$$

Retracing Steps 1 and 2 of the Initial Proof

They are retraced with the following modifications.

Step 1 (see Subsect. 9.6.2 starting on p. 325).

m.1) Unlike (i) on p. 324, the sources (associated with unstable modes x_α) produce not only the matching mode x_α^0 of the initial state but also the relevant part $\Xi_{[\nu]} := \{[\xi(t)]_{[\nu]}\}_{t\geq 0}$ of the noise sequence (9.8.4). Here $\nu := \nu(\alpha)$ is the serial number of the group from (9.4.8) (on p. 295) containing x_α.

m.2) The rule (9.6.4) (on p. 326) to restore the observation from the data currently available at the jth sensor node is modified as follows:

$$y_j(t) = C_j\left\{A^t \boldsymbol{x}_j^0 + \left(\sum_{i:(i,j)\in\text{CVP}} \sum_{\theta=0}^{t-d_{i\to j}} A^{t-1-\theta} B_i u_i(\theta)\right) + \sum_{\theta=0}^{t-1} A^{t-1-\theta} \boldsymbol{\xi}_j(\theta)\right\}.$$

Explanation 9.8.1. We recall that $\boldsymbol{x}_j^0$ is the vector of the data received by the jth sensor node from all initial sources of $\mathbf{NW}_1$. This vector has the form (9.4.8) (on p. 295), where all missing data are substituted by zeros and all modes received from the sourcers are put in the proper places. The vector $\boldsymbol{\xi}_j(\theta)$ is the similar partial reconstruction of the entry $\xi(t)$ from (9.8.4) based on the available source data.

Observation 9.8.2. *Since the above formula restores the observations correctly, Lemma* 9.6.4 *(on p.* 326*) remains true with* (9.8.3) *substituted in place of* (9.6.1)*.*

Step 2 (see Subsect. 9.6.3 starting on p. 326).

- The operation of the actuator nodes is changed as was discussed in Observation 9.6.9 (on p. 328).

With regard to the necessity to take into account the extra term

$$\sum_{\theta=0}^{t-1} A^{t-1-\theta}\xi(\theta)$$

in (9.6.7), the conclusion of this observation now transforms as follows:

$$\left[A^t x^0 + \sum_{\theta=0}^{t-1} A^{t-1-\theta}\xi(\theta) - \sum_{i \in \mathfrak{I}^{mw}_{\to\alpha}} x^{u,i}(t) \right]_\alpha \in [-D_\infty, D_\infty] \qquad \forall \alpha, t. \quad (9.8.5)$$

We recall that D_∞ denotes the supremum (9.4.4) (on p. 292).

As a result, the following claim is established.

Lemma 9.8.3. *Suppose that the network* $\mathbf{NW}_1 := \mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)}$ *is fed by the inputs described in* **m.1)** *on p.* 362*. Then this network can be equipped with an admissible data processing strategy such that the outputs $x^{u,i}(t)$ of the actuator nodes (marked by i) satisfy* (9.8.5) *whenever* (9.8.3) *holds and $\|x^0\| \le D_x, \|\xi(t)\| \le D$, where D_x, D are taken from* (9.4.3) *(on p.* 292*).*

Errorless Block Code for the Network $\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)}$ with the Communication Scheme $\overline{\mathfrak{T}}_{\mathbf{mw}}$

Now we consider the prefix $\overline{\mathbf{PREF}}_{\mathbf{mw}}$ and the data communication scheme $\overline{\mathfrak{T}}_{\mathbf{mw}}$ introduced on p. 307. We recall that the data sources of this prefix are associated with the subspaces from (9.5.26) (on p. 307) and enumerated by $\nu = 1, \ldots, g$. In accordance with the scheme $\overline{\mathfrak{T}}_{\mathbf{mw}}$, data should be transmitted from any such source $\sim M_\nu$ to all actuators i that control the subspace M_ν.

Lemma 9.8.4. *For any ν, there exists $i = i_\nu$ such that $(\nu, i_\nu) \in \overline{\mathfrak{T}}_{\mathbf{mw}}$ and the capacity domain* $\mathbf{CD}[\overline{\mathbf{PREF}}_{\mathbf{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \overline{\mathfrak{T}}_{\mathbf{mw}}]$ *contains a rate ensemble $\{r_T\}_{T \in \overline{\mathfrak{T}}_{\mathbf{mw}}}$ with $r_{\nu,i_\nu} > 0 \forall\, \nu$.*

Proof. Let ν be chosen. By applying Lemma 7.6.7 (on p. 212) to the plant $x_{[\nu]}(t+1) = A|_{M_\nu} x_{[\nu]}(t) + \xi_{[\nu]}(t)$, we see that m and two admissible disturbances

$$\Xi^s_{[\nu]} = \{\xi^s_{[\nu]}(0), \xi^s_{[\nu]}(1), \ldots, \xi^s_{[\nu]}(m-1), 0, 0, \ldots\}, \quad \|\xi^s_{[\nu]}(\theta)\| \le D, \qquad s = 1, 2$$

exist that drive this plant from 0 at $t = 0$ to the states $x^s_{[\nu]}(m)$ at $t = m$ such that $\|x^1_{[\nu]}(m) - x^2_{[\nu]}(m)\| > 2D_\infty$. Since we deal with the norm $\|\cdot\| := \|\cdot\|_\infty$in this chapter, an unstable mode x_α exists such that

$$\|x^1_\alpha(m) - x^2_\alpha(m)\| > 2D_\infty, \qquad \nu(\alpha) = \nu. \tag{9.8.6}$$

Note that $\mathfrak{I}^{mw}_{\to\alpha} = \mathfrak{I}^{mw}_{\to\nu(\alpha)}$, where $\mathfrak{I}^{mw}_{\to\nu}$ is the set of actuators controlling the subspace M_ν (see p. 307). So thanks to (9.8.5) and (9.8.6), the number $s = 1, 2$ is recognizable from the knowledge of

$$\sum_{i \in \mathfrak{I}^{mw}_{\to\nu}} x^{u,i}(t) \tag{9.8.7}$$

provided that the block code from Lemma 9.8.3 is fed by $x^0 := 0$ and $\Xi^s_{[\nu]}$ at any input node $\sim x_\alpha$ such that $\nu(\alpha) = \nu$, whereas all other nodes are fed by $\Xi_{[\nu(\alpha)]} := 0$. Then all input nodes from any group in (9.4.8) (on p. 295) are fed by identical data. These nodes are departing points for the cloned channels from $\mathbf{PREF_{mw}}$. So all these channels except for one may be discarded, which transforms $\mathbf{PREF_{mw}}$ into $\overline{\mathbf{PREF}}_{\mathbf{mw}}$. The source $\sim \nu$ produces $s = 0, 1$; the conversion $s \mapsto \mathit{\Xi}^s_{[\nu]}$ is performed at the related sensor nodes. All other sources produce the output 1, which is converted into the zero sequence at any concerned sensor node.

Since s is recognizable from the sum (9.8.7), it can be also recognized from the knowledge of some addend $x^{u,i}(t), i = i_\nu \in \mathfrak{I}_{\to\nu}$. By doing so, we obtain a block code transmitting without an error one bit from ν to i_ν, and zero bits from ν' to i' for all other pairs $(\nu', i') \in \overline{\mathfrak{T}}_{\mathbf{mw}}$. The network initial state for this code is identical to that for the strategy from Lemma 9.8.3. By invoking Assumption 9.4.20 (on p. 298), we prolong the block code to reset the network to this initial state. After this the block codes related to $\nu = 1, \ldots, g$ can be concatenated to form a block code transmitting without an error one bit from ν to i_ν for all ν, and zero bits from ν' to i' for the other pairs $(\nu', i') \in \overline{\mathfrak{T}}_{\mathbf{mw}}$. The proof is completed by (ii) of Lemma 9.5.6 (on p. 303). □

Completion of the Proof of Remark 9.4.28 on p. 300

By (ii) of Lemma 9.5.10 (on p. 306) and Lemma 9.8.4, the capacity domain $\mathbf{CD}\left[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \blacklozenge \mathfrak{T}_{\mathbf{mw}}\right]$ contains a rate ensemble $\{r_{\alpha,i}\}_{(\alpha,i)\in\mathfrak{T}_{\mathbf{mw}}}$ such that $r_{\alpha,i_\nu} > 0$ for all $\alpha \in G_\nu$ and $\nu = 1, \ldots, g$. Here G_ν is the νth group of modes from (9.4.8) (on p. 295). Then the first inclusion from (9.5.22) (on p. 305) implies that $\mathbf{CD}\left[\mathbf{PREF_{mw}} \boxplus \mathbf{CBE(NW)} \boxplus \mathbf{SUFF^q_{mw}}\right]$ contains a rate vector with positive entries, which completes the proof. □

10

H^∞ State Estimation via Communication Channels

10.1 Introduction

One approach to the problem of state estimation via limited capacity communication channels was proposed in [220] and developed in [45, 100, 132, 133, 184] (see also Chaps. 2–7 of this book). In this framework, the observation must be coded into a sequence of finite-valued symbols and transmitted via a digital communication channel. With the approach of [45, 100, 133, 184], classic Kalman and H^∞ state estimation theory cannot be applied since the estimator only observes the transmitted sequence of finite-valued symbols. In fact, we need to design a hybrid dynamical system that consists of two subsystems. The first subsystem, which is called *coder*, receives real-valued measurements and converts them into a finite-valued symbolic sequence that is sent over the limited capacity communication channel. The second subsystem (*decoder*) receives this symbolic sequence and converts it into a real-valued state estimate. In other words, such state estimators with bit-rate constraints form an important subclass of so-called hybrid dynamic systems. In general, hybrid systems are those that combine continuous and discrete event dynamics and involve both real and symbolic variables; see, e.g., [104, 174, 177, 185, 211].

Another approach was proposed in [175] where the bandwidth limitation constraint was modeled in a manner that the state estimator can communicate with only one of several sensors at any time instant (this approach will be presented in Chaps. 14 and 15 of this book). The main disadvantage of both these approaches is that the state estimation systems proposed in [45, 100, 132–135, 175, 184, 220, 221] are highly nonlinear and contain symbolic variables. In this chapter, we propose a new problem statement. In our new problem statement, the communication channel transmits a continuous time vector signal. The limited capacity of the channel means that the dimension of the signal transmitted by the channel is smaller than the dimension of the measured output of the system. Our goal is to design a linear time-invariant coder that transforms the measured output into the signal to be transmitted and a linear time-invariant decoder that transforms the transmitted signal into the state estimate. This approach simplifies the problem and allows for obtaining more constructive and understandable results. The main advantage is that estimation

doi: 10.1007/978-0-8176-4607-3_10,

system obtained in this chapter is linear and time-invariant. This allows for applying conventional linear control theory to the problem of state estimation with bandwidth limitation constraints. Unlike [45, 132, 133, 220] where the Kalman filtering problem was considered and [100, 175, 184] where the set-valued approach to state estimation was employed, in this chapter we consider the case of H^∞ state estimation. The main result of the chapter shows that the linear H^∞ control theory when suitably modified provides a good framework for the problem of limited communication state estimation. This result was originally published in [168].

It should be pointed out that the proposed state estimation method is computationally nonexpansive and easy to implement in real time. The obtained results can be extended to the case of uncertain linear systems.

The remainder of the chapter is organized as follows. In Sect. 10.2, we introduce the class of systems under consideration. The main results of the chapter and their proofs are given in Sect. 10.3.

10.2 Problem Statement

We consider the linear system defined on the infinite time interval $[0, \infty)$:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bw(t); \;\; x(0) = 0; \\ z(t) &= Kx(t); \\ y(t) &= Cx(t) + v(t), \end{aligned} \tag{10.2.1}$$

where $x(t) \in \mathbb{R}^n$ is the *state*; $w(t) \in \mathbb{R}^p$ and $v(t) \in \mathbb{R}^l$ are the *disturbance inputs*; $z(t) \in \mathbb{R}^q$ is the *estimated output*, and $y(t) \in \mathbb{R}^l$ is the *measured output*, and A, B, K, and C are given matrices.

The state estimation problem is to find an estimate $\hat{z}$ of z in some sense using the measurement of y. One of the most common frameworks for the state estimation problem is based on the H^∞ theory. More precisely, the problem of H^∞ state estimation can be stated as follows; e.g., see [14, 66, 130]:

Given a constant $\gamma > 0$, find a causal unbiased filter $\hat{z}(t) = \mathcal{F}[t, y(\cdot) \mid_0^t]$ if it exists such that

$$J := \sup_{[w(\cdot),v(\cdot)] \in \mathbf{L}_2[0,\infty)} \frac{\|z(\cdot) - \hat{z}(\cdot)\|_2^2}{\|w(\cdot)\|_2^2 + \|v(\cdot)\|_2^2} < \gamma^2. \tag{10.2.2}$$

Let $m < l$ be given. We will consider the problem of H^∞ state estimation via a communication channel of dimension m.

Suppose estimates of the output $z(t)$ are required at a distant location and are to be transmitted via a limited capacity communication channel such that only m real numbers may be sent at each time t. We consider a system that consists of the coder, the transmission channel, and the decoder. Using the measurement $y(\cdot) \mid_0^t$, the coder produces a vector $\hat{y}(t)$ of dimension m that is transmitted via the channel and then received by the decoder. In its turn, the decoder produces an estimate $\hat{z}(t)$

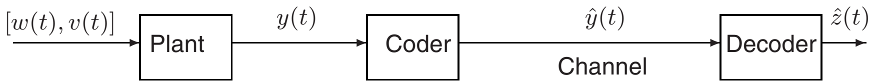


Fig. 10.1. H^∞ state estimation via communication channel.

that depends only on $\hat{y}(\cdot)\ |_0^t$. The block diagram of this state estimator is shown in Fig. 10.1.

We consider the class of linear time-invariant coders of the form

$$\begin{aligned} \dot{x}_c(t) &= A_c x_c(t) + B_c y(t); \;\; x_c(0) = 0; \\ \hat{y}(t) &= C_c x_c(t) + D_c y(t), \end{aligned} \tag{10.2.3}$$

where $\hat{y}(t) \in \mathbb{R}^m$ is the signal transmitted via the communication channel. Note that the dimension of the coder state vector $x_c(t)$ may be arbitrary.

Also, we consider linear time-invariant decoders of the form

$$\begin{aligned} \dot{x}_d(t) &= A x_d(t) + B_d \hat{y}(t); \;\; x_d(0) = 0; \\ \hat{z}(t) &= C x_d(t). \end{aligned} \tag{10.2.4}$$

The system (10.2.4) is an estimator for the system (10.2.1), and we wish to make the dynamics of (10.2.4) as close as possible to the dynamics of (10.2.1).

Definition 10.2.1. *Consider the system* (10.2.1)*. Let $\gamma > 0$ and $m < l$ be given. The H^∞ state estimation problem via a communication channel of dimension m with disturbance attenuation γ is said to have a solution if a coder of the form* (10.2.3) *and a decoder of the form* (10.2.4) *exist such that the condition* (10.2.2) *holds.*

The problem of H^∞ state estimation via limited capacity communication channel is to find matrix coefficients A_c, B_c, C_c, and B_d (if they exist) to satisfy the H^∞ requirement (10.2.2). This problem is difficult. In this chapter, we consider a simplified problem statement:

Suppose that B_d is given. Find the coefficients of the coder (10.2.3) *such that condition* (10.2.2) *holds.*

Definition 10.2.2. *Consider the system* (10.2.1)*. Let $\gamma > 0$, $m < l$, and B_d be given. The H^∞ state estimation problem via a communication channel of dimension m with disturbance attenuation γ is said to have a solution with the decoder* (10.2.4) *if a coder of the form* (10.2.3) *exists such that the requirement* (10.2.2) *holds.*

10.3 Linear State Estimator Design

The main result of this chapter requires the following assumptions.

Assumption 10.3.1. *The pair (A, B) is controllable.*

Assumption 10.3.2. *The pair* (A, K) *is observable.*

Assumption 10.3.3. *The pair* (A, C) *is observable.*

Our solution to the above problem involves the following Riccati algebraic equations:

$$AY + YA^{\mathrm{T}} + Y\left[\frac{1}{\gamma^2}K^{\mathrm{T}}K - C^{\mathrm{T}}C\right]Y + BB^{\mathrm{T}} = 0, \tag{10.3.5}$$

$$A^{\mathrm{T}}X_\epsilon + X_\epsilon A + X_\epsilon\left[\frac{1}{\epsilon}B_d B_d^{\mathrm{T}} - \frac{1}{\gamma^2}BB^{\mathrm{T}}\right]X_\epsilon + K^{\mathrm{T}}K = 0. \tag{10.3.6}$$

Now we are in a position to present the main result of this chapter.

Theorem 10.3.4. *Consider the system* (10.2.1)*, suppose that Assumptions* 10.3.1–10.3.3 *hold, and that the coefficient* B_d *of the decoder* (10.2.4) *is given. Let* $\gamma > 0$ *be a given constant and* m *be a given integer. Then, the following statements are equivalent:*

(i) The H^∞ *state estimation problem via a communication channel of dimension* m *with disturbance attenuation* γ *has a solution with the decoder* (10.2.4)*.*
(ii) A constant $\epsilon > 0$ *exists such that the algebraic Riccati equations* (10.3.5) *and* (10.3.6) *have stabilizing solutions* $Y \geq 0$ *and* $X_\epsilon \geq 0$ *such that* $\rho(YX_\epsilon) < \gamma^2$ *where* $\rho(\cdot)$ *denotes the spectral radius of a matrix.*

Furthermore, suppose that condition (ii) holds. Then, the coder (10.2.3) *with*

$$A_c = \begin{pmatrix} \tilde{A}_c & -\tilde{B}_c C \\ -B\tilde{C}_c & A \end{pmatrix}; \quad B_c = \begin{pmatrix} \tilde{B}_c \\ 0 \end{pmatrix}; \quad C_c = \begin{pmatrix} \tilde{C}_c & 0 \end{pmatrix};$$

$$\tilde{A}_c = A + B\tilde{C}_c - \tilde{B}_c C + \frac{1}{\gamma^2}BB^{\mathrm{T}}X_\epsilon;$$

$$\tilde{B}_c = \left[I - \frac{1}{\gamma^2}YX_\epsilon\right]^{-1}YC^{\mathrm{T}}, \quad \tilde{C}_c = \frac{1}{\epsilon}B_d^{\mathrm{T}}X_\epsilon \tag{10.3.7}$$

solves the H^∞ *state estimation problem via a communication channel of dimension* m *with disturbance attenuation* γ *with the decoder* (10.2.4)*.*

Proof of Theorem 10.3.4: Statement (i) $\Rightarrow$ *(ii).* Assume that condition (i) holds, and consider the decoder (10.2.4). Introduce a new vector variable $\tilde{x} := x(t) - x_d(t)$. Then $\tilde{x}(t)$ satisfies the equation

$$\dot{\tilde{x}}(t) = A\tilde{x}(t) + B\hat{y}(t) + Bw(t); \quad \tilde{x}(0) = 0. \tag{10.3.8}$$

Furthermore, introduce

$$\tilde{z}(t) := K\tilde{x}(t); \quad \tilde{y}(t) := C\tilde{x}(t) + v(t). \tag{10.3.9}$$

Now consider the linear system (10.3.8), (10.3.9) and suppose that $\hat{y}(t)$ is the control input, $w(t)$ and $v(t)$ are the disturbance inputs, $\tilde{y}(t)$ is the measured output, and $\tilde{z}(t)$ is the controlled output. The requirement (10.2.2) can now be rewritten as

$$J := \sup_{[w(\cdot),v(\cdot)]\in \mathbf{L}_2[0,\infty)} \frac{\|\tilde{z}(\cdot)\|_2^2}{\|w(\cdot)\|_2^2 + \|v(\cdot)\|_2^2} < \gamma^2. \tag{10.3.10}$$

Now the coder (10.2.3) can be considered as an output feedback controller with input $\tilde{y}(\cdot)$ and output $\hat{y}(\cdot)$. Furthermore, this controller solves the output feedback H^∞ control problem (10.3.10). Note that the H^∞ control problem (10.3.10) is a singular problem. The approach taken to solve this singular problem is a perturbation approach such as contained in [81]. As it was shown in [81], since the controller (10.2.3) solves the H^∞ control problem (10.3.10), a small $\epsilon > 0$ exists such that the same controller solves the nonsingular H^∞ control problem

$$J := \sup_{[w(\cdot),v(\cdot)]\in \mathbf{L}_2[0,\infty)} \frac{\|\tilde{z}(\cdot)\|_2^2 + \epsilon\|\tilde{y}(\cdot)\|_2^2}{\|w(\cdot)\|_2^2 + \|v(\cdot)\|_2^2} < \gamma^2. \tag{10.3.11}$$

Now condition (ii) follows from the standard H^∞ control theory; e.g., see [151], p.73. This completes the proof of this part of the theorem.

Statement (ii) ⇒ (i). Suppose that condition (ii) holds. Then, it follows from the standard H^∞ control theory (e.g., see [151], pp. 73 and 78) that the linear output feedback controller

$$\begin{aligned} \dot{\tilde{x}}_c(t) &= \tilde{A}_c\tilde{x}_c(t) + \tilde{B}_c\tilde{y}(t); \quad \tilde{x}_c(0) = 0; \\ \hat{y}(t) &= C_c\tilde{x}_c(t) \end{aligned} \tag{10.3.12}$$

with the coefficients defined by (10.3.7) solves the H^∞ control problem (10.3.11) for the system (10.3.8), (10.3.9). Since $\tilde{z} = z - \hat{z}$, the H^∞ requirement (10.2.2) follows from (10.2.2). Furthermore, since $\tilde{y}(t) = y(t) - Cx_d(t)$ and $x_d(t)$ is defined by (10.2.4), (10.3.12) can be rewritten in the form (10.2.3) with the coefficients defined by (10.3.7). This completes the proof of Theorem 10.3.4.

11

Kalman State Estimation and Optimal Control Based on Asynchronously and Irregularly Delayed Measurements

11.1 Introduction

In this chapter, we study discrete-time linear partially observed systems with Gaussian disturbances. Various sensor signals or even parts of a sensor output are communicated to the observer or controller over parallel channels with independent delays. Messages may arrive out of order; there may be periods when no information is received. Data transferred via a channel may be corrupted or even lost due to, e.g., noise in the communication medium and protocol malfunctions. The situation is complicated by the fact that many communication channels do not satisfy the time invariance condition [208, 232]. The reasons for it vary depending on the type of the channel. In some cases, this may be due to the relative motion of the transmitter and the receiver, as well as due to oscillator drifts and phase noise coupled with multipath effects [208]. For digital networked channels, the origins of the phenomenon may be traced back, in particular, to competition between channels for the network resources. As a result, the prognosis of the future states of the communication medium is often a hard task; and we suppose that the statistics of the transmission delays is not known. At the same time, it is characteristic of many channels that each message transferred is marked with a "time stamp" indicating the moment of the transfer beginning; and we assume this.

The minimum variance state estimation problem is solved and an analog of the Kalman filter is proposed. For a finite horizon linear-quadratic Gaussian (LQG) optimal control problem, we obtain the solution in the form of the above filter coupled with the controller optimal in the deterministic linear-quadratic optimization problem. We also derive the solution for the infinite-horizon problem of minimizing the average cost per unit time. To this end, we introduce a concept of observability via communication channels that may lose data. A number of criteria for such an observability are offered. The main results of this chapter were originally presented in [105, 106].

When dealing with the state estimation problem, we suppose that the control strategy is given. We also assume that there may be a difference between the control currently produced by the controller and that acting upon the plant. This difference

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_11,

may be caused by a variety of reasons, e.g., the noise, delays, and data dropout in the control loop. At the same time, this chapter addresses the optimal control problems in the case where the control loop is perfect; i.e., the control generated by the controller acts upon the plant immediately and with no distortion. Generalizations on the case of nonperfect control loop will be considered in the next chapter, and in doing so, the results on the state estimation from this chapter will be employed.

The issue of state estimation and control over communication channels with data dropouts and delays was addressed in, e.g., [13, 72, 92, 93, 97, 98, 131, 146, 163, 186, 187, 194–196, 209, 210, 225, 226, 231] (see also the literature therein). In [194, 195], the data losses are interpreted as sharp increases of the sensor noise level, which occur independently as time progresses and with a given probability. In [13], dropout in both observation and control loops was treated, and it was assumed that data losses happen independently and are equivalent to receiving the zero signal. An overview of practical advances in the area of estimation and control under data losses is offered in [196]. The paper [92] discusses a design of a dropout compensator. In [98], observability of linear systems is studied in the situation where the probability of loss of no more than m packets in every lot of $k > m$ ones is kept above a certain level. In [225, 226, 231], stabilization problems were considered under the assumption that the durations of the chains of successive packet losses are upper bounded. The paper [72] studies problems of optimal control over communication links with data dropout in the cases where the dropout acknowledgment signal is and is not available, respectively. The case of randomly delayed measurements was systematically studied by Ray et al. (see [93, 97, 209, 210] and the literature therein) under the assumption that the sensor delays τ do not exceed the sampling rate (i.e., they are binary $\tau \in \{0, 1\}$ in the discrete-time setting) and form a stationary sequence of mutually independent random quantities with a priori known statistics, and the communication channels do not lose signals.

The body of the chapter is organized as follows. In Sect. 11.2, the state estimation problem is posed. Its solution is given by the state estimator described in Sect. 11.3. Section 11.4 offers conditions for stability of this estimator. Sections 11.5 and 11.6 are devoted to finite and infinite horizon optimal control problems, respectively. Sections 11.7, 11.9, and 11.10 contain the proofs of the main results from Sects. 11.3, 11.5, 11.4, and 11.6. Section 11.8 offers the proofs of several propositions from Sect. 11.4, which in fact constitute criteria for observability via communication channels.

11.2 State Estimation Problem

We consider multiple sensor discrete-time linear systems of the form:

$$x(t+1) = A(t)x(t) + B(t)u(t) + \xi(t) \quad t = 0, \dots, T-1, \qquad x(0) = a; \tag{11.2.1}$$

$$y_\nu(t) = C_\nu(t)x(t) + \chi_\nu(t) \quad \nu = 1, \dots, l, \quad t = 0, \dots, T. \tag{11.2.2}$$

Here $x(t) \in \mathbb{R}^n$ is the state; $u(t) \in \mathbb{R}^m$ is the control; $\xi(t) \in \mathbb{R}^n$ is a process disturbance; $y_\nu(t) \in \mathbb{R}^{k_\nu}$ is a sensor output; and $\chi_\nu(t)$ is a noise. The measurement

$y_\nu(t)$ produced at a time t is sent to the estimator via a communication channel and arrives at the time $t + \tau_\nu(t) \geq t$. We put $\tau_\nu(t) := \infty$ if the signal $y_\nu(t)$ is lost. Otherwise the estimator receives the signal $y_\nu(t) + \lambda_\nu(t)$ at the time $t + \tau_\nu(t)$, where $\lambda_\nu(t)$ is the transfer error. We however shall not make an explicit use of $\lambda_\nu(t)$ and merely add $\lambda_\nu(t)$ to $\chi_\nu(t)$ in (11.2.2). So $\chi_\nu(t)$ is interpreted as the sum of the measurement and transfer errors. A particular vector y_ν may equal either the entire output of a sensor or only a part of it.

Denote by $Y(t)$ the vector incorporating all the observation signals that arrive at the estimator at the time t:

$$Y(t) := \Big(y_\nu[\theta]\Big)_{(\nu,\theta)\in S(t)}, \quad \text{where} \quad S(t) := \big\{(\nu,\theta) : \theta + \tau_\nu(\theta) = t\big\}. \tag{11.2.3}$$

(If no signal arrives, i.e., $S(t) = \emptyset$, the vector $Y(t)$ is defined to be $0 \in \mathbb{R}$ for consistency.) We assume that the producer (sensor) of any measurement received is recognizable. If the sensor output was partitioned into several portions to be transferred over parallel channels, there is a way to recognize which part of the output is represented by the message arrived. With regard to the messages time stamps, these assumptions mean that the set $S(t)$ becomes known at the time t. We also suppose that at this time the estimator is aware of the control $u(t-1)$. The problem is to find the minimum variance estimate of the state $x(t)$ based on the available data:

$$\mathfrak{Y}(t) := [Y(0), Y(1), \ldots, Y(t)], \quad \mathfrak{S}(t) := [S(0), S(1), \ldots, S(t)],$$
$$\mathfrak{U}(t-1) := \textbf{col}\,[u(0), \ldots, u(t-1)]. \tag{11.2.4}$$

This problem will be examined in the case where the control is generated by a given rule of the form

$$u(t) = \mathcal{U}\left[t, \mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{h}(t)\right]. \tag{11.2.5}$$

The variable $\mathfrak{h}(t)$ is introduced to express a possible difference between the current control generated by the controller and that acting upon the plant. For example, formula (11.2.5) may shape into

$$u(t) = \mathcal{U}_*\left[t, \mathfrak{Y}(t), \mathfrak{S}(t)\right] + \mathfrak{h}(t),$$

where the first and second addends on the right stand for the control generated by the controller and additive noise in the control loop, respectively. Another example is

$$u(t) = v[t - \mathfrak{h}(t)], \qquad v(t) = \mathcal{V}\left[t, \mathfrak{Y}(t), \mathfrak{S}(t), v(0), \ldots, v(t-1)\right],$$

where $v(t)$ is the control currently generated by the controller and $\mathfrak{h}(t)$ is the delay in the control loop. The case where $\mathfrak{h}(t)$ takes only two values $0, 1$ and $u(t) = \mathfrak{h}(t)v(t)$ models data dropout in the control loop.

The problem of state estimation will be studied under the following assumptions.

Assumption 11.2.1. *In* (11.2.1) *and* (11.2.2)*, the random vectors* $a, \xi(t)$*, and* $\chi_\nu(t)$*,* $\nu = 1, \ldots, l$ *are Gaussian and independent with* $\boldsymbol{E}\xi(t) = 0$ *and* $\boldsymbol{E}\chi_\nu(t) = 0$*. The mean* $\boldsymbol{E}a$ *and the correlation matrices*

$$R_{aa} := \boldsymbol{E}[a - \boldsymbol{E}a][a - \boldsymbol{E}a]^{\mathrm{T}}, \quad R_{\xi\xi}(t) := \boldsymbol{E}\xi(t)\xi(t)^{\mathrm{T}},$$
$$R^{\nu}_{\chi\chi}(t) := \boldsymbol{E}\chi_\nu(t)\chi_\nu(t)^{\mathrm{T}} \qquad (\text{where } \nu = 1, \dots, l) \quad (11.2.6)$$

are known. So are the matrices $A(t), B(t)$, and $C_\nu(t)$.

Assumption 11.2.2. *The measurement transmission delays are independent of $a, \xi(t)$, and $\chi_\nu(t)$ and bounded by a known constant: $\tau_\nu(t) \le \sigma$ whenever $\tau_\nu(t) \neq \infty$.*

Assumption 11.2.3. *At any time t, the estimator gets aware of the control $u(t-1)$.*

Assumption 11.2.4. *In* (11.2.5)*, the random variables $\mathfrak{h}(t)$ take values in a common finite set and, along with the transmission delays, are independent of $a, \xi(t)$, $\chi_\nu(t)$.*

In particular, this means that through the knowledge of $u(t-1)$, the estimator does not acquire an additional information about $x(t)$, as compared with that contained in $\mathfrak{Y}(t)$ and $\mathfrak{S}(t)$. Note in conclusion that (almost surely)

$$S(t_1) \cap S(t_2) = \emptyset \quad \text{whenever} \quad t_1 \neq t_2, \quad \text{and} \quad (\nu, \theta) \in S(t) \Rightarrow \theta \le t. \quad (11.2.7)$$

11.3 State Estimator

11.3.1 Pseudoinverse of a Square Matrix and Ensemble of Matrices

To describe the state estimator, we need some preliminaries. As is well known, any nonsingular $k \times k$ square matrix M has a unique inverse M^{-1}, i.e., an $k \times k$ matrix such that $MM^{-1} = M^{-1}M = I$. In some applications, an analog of the inverse matrix is required in the case where the initial matrix M is singular $\det M = 0$. One of such analogs is the *pseudoinverse* [25, 30, 157] of M, which is the matrix $\overset{+}{M}$ associated with the linear operator $M_0^{-1}\pi_M$. Here π_M is the orthogonal projection of $\mathbb{R}^k$ onto **Im** M, the space $\mathbb{R}^k$ is equipped with the standard inner product

$$\langle a, b \rangle = \sum_{i=1}^{k} a_i b_i \qquad a = \mathbf{col}\,(a_1, \dots, a_k), b = \mathbf{col}\,(b_1, \dots, b_k),$$

and the operator $M_0 : (\ker M)^{\perp} \to \mathbf{Im}\, M$ is obtained by restricting M on $(\ker M)^{\perp}$, where $^{\perp}$ stands for the orthogonal complement.

For any nonsingular matrix M, the pseudoinverse clearly equals the inverse. More generally, for any block matrix of the form

$$M = \begin{pmatrix} \mathcal{M} & 0 \\ 0 & 0 \end{pmatrix} \quad (11.3.1)$$

with the square nonsingular block $\mathcal{M}$, the following relation evidently holds:

$$\overset{+}{M} = \begin{pmatrix} \mathcal{M}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

This observation aids to compute the pseudoinverse of any normal $M^{\mathrm{T}}M = MM^{\mathrm{T}}$ (in particular, self-adjoint $M^{\mathrm{T}} = M$) matrix M. To this end, it suffices to note that first, such a matrix can be reduced to the form (11.3.1) by a similarity transformation $M \mapsto M_U := UMU^{-1}$ with an orthogonal matrix U and, second, the pseudoinverse is invariant with respect to this transformation:

$$M_U = UMU^{-1} \quad \Rightarrow \quad \overset{+}{M} = U^{-1}\overset{+}{M}_U U. \tag{11.3.2}$$

Now we proceed to the case introduced in the following.

Definition 11.3.1. *A* square ensemble Λ *of matrices over a finite set* $S = \{s\}$ *is composed of several matrices* $\Lambda_{s_1}^{s_2}$ *enumerated by the pairs* (s_1, s_2) *with* $s_1, s_2 \in S$, *each of the size* $k(s_1) \times k(s_2)$, *where* $k(s)$ *are some natural numbers.*

By picking an enumeration $\{s(1), \dots, s(q)\}$ of S, such ensembles can be put in the one-to-one correspondence $\Lambda \leftrightarrow M$ with square $k \times k$ block matrices

$$M = \begin{pmatrix} \Lambda_{s(1)}^{s(1)} & \Lambda_{s(1)}^{s(2)} & \Lambda_{s(1)}^{s(3)} & \dots & \Lambda_{s(1)}^{s(q)} \\ \Lambda_{s(2)}^{s(1)} & \Lambda_{s(2)}^{s(2)} & \Lambda_{s(2)}^{s(3)} & \dots & \Lambda_{s(2)}^{s(q)} \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ \Lambda_{s(q)}^{s(1)} & \Lambda_{s(q)}^{s(2)} & \Lambda_{s(q)}^{s(3)} & \dots & \Lambda_{s(q)}^{s(q)} \end{pmatrix}, \qquad k := \sum_{s \in S} k(s). \tag{11.3.3}$$

The ensemble $\overset{+}{\Lambda}$ corresponding to $\overset{+}{M}$ is called the *pseudoinverse* of Λ, and *inverse* if the matrix M is nonsingular. In the latter case, $\overset{+}{\Lambda}$ is also denoted by Λ^{-1}.

The matrices M produced by various enumerations can clearly be obtained from each other by a similarity transformation with a permutation matrix U. Since such a matrix U is orthogonal, the property (11.3.2) implies that the pseudoinverse $\overset{+}{\Lambda}$ does not depend on the enumeration of S.

Some further properties of the pseudoinverses are discussed in Appendix B.

11.3.2 Description of the State Estimator

We denote by $\widehat{x}(j|t)$ the minimum variance estimate of the state $x(j)$ based on

$$Y(0), \dots, Y(t), S(0), \dots, S(t), u(0), \dots, u(t-1),$$

where $Y(\theta)$ and $S(\theta)$ are given by (11.2.3). Being coupled with certain $n \times n$ matrices

$$P_{ij}(t), \overline{P}_{ij}(t), \qquad i, j = 0, \dots, \sigma$$

(where σ is taken from Assumption 11.2.2), the tuple of the estimates

$$\widehat{X}(t) = \Big[\widehat{x}(t|t), \widehat{x}(t-1|t), \dots, \widehat{x}(t-\sigma|t)\Big] \tag{11.3.4}$$

may be generated recursively by the following analog of the Kalman filter.

Recursive State Estimator

The next tuple $\widehat{X}(t+1)$ of the estimates is generated by equations

$$\widehat{x}(j|t+1) = \widehat{x}(j|t) + \sum_{(\nu,\theta)\in S(t+1)} K^{(\nu,\theta)}_{t+1-j}(t+1)\left[y_\nu(\theta) - \widehat{y}_\nu(\theta|t)\right], \quad (11.3.5)$$

where $j = t+1, t, t-1, \ldots, t+1-\sigma$,

$$\widehat{x}(t+1|t) := A(t)\widehat{x}(t|t) + B(t)u(t),$$

and

$$\widehat{y}_\nu(\theta|t) := C_\nu(\theta)\widehat{x}(\theta|t). \quad (11.3.6)$$

(We recall that any sum over the empty set is defined to be zero.) The gain matrices $K^s_j(t)$ are enumerated by the pairs $[j, s = (\nu,\theta)]$ with $j = 0,\ldots,\sigma$ and $s \in S(t)$ and have the size $n \times k_\nu$. These matrices are calculated as follows:

$$K^s_j(t) = \sum_{(\nu,\theta)\in S(t)} P_{j,t-\theta}(t)C_\nu(\theta)^{\mathrm{T}}\overset{+}{\Lambda}(t)^s_{(\nu,\theta)}. \quad (11.3.7)$$

Here $\overset{+}{\Lambda}(t)$ is the pseudoinverse of the following square ensemble $\Lambda = \Lambda(t)$ of matrices over the finite set $S(t)$:

$$\Lambda(t)^{s_2}_{s_1} = C_{\nu_1}(\theta_1)P_{t-\theta_1,t-\theta_2}(t)C_{\nu_2}(\theta_2)^{\mathrm{T}} + \varDelta^{s_2}_{s_1};$$
$$\varDelta^{s_2}_{s_1} := \begin{cases} R^{\nu_1}_{\chi\chi}(\theta_1) & \text{if} \quad s_1 = s_2 \\ 0 & \text{otherwise} \end{cases};$$
$$\forall s_1 = (\nu_1,\theta_1), s_2 = (\nu_2,\theta_2) \in S(t). \quad (11.3.8)$$

The square ensembles $\{P_{ij}(t)\}^\sigma_{i,j=0}$, $\{\overline{P}_{ij}(t)\}^\sigma_{i,j=0}$ of $n \times n$-matrices over the set $[0:\sigma]$ are generated recursively

$$\mapsto \Big\{P_{ij}(t)\Big\} \mapsto \Big\{\overline{P}_{ij}(t)\Big\} \mapsto \Big\{P_{ij}(t+1)\Big\} \mapsto; \quad (11.3.9)$$

$$\overline{P}_{ij}(t) := P_{ij}(t) - \sum_{(\nu,\theta)\in S(t)} K^{(\nu,\theta)}_i(t)C_\nu(\theta)P_{t-\theta,j}(t); \quad (11.3.10)$$

$$P_{ij}(t+1) := \begin{cases} A(t)\overline{P}_{ij}(t)A(t)^{\mathrm{T}} + R_{\xi\xi}(t) & \text{if } i = j = 0 \\ A(t)\overline{P}_{i,j-1}(t) & \text{if } i = 0, j \geq 1 \\ \overline{P}_{i-1,j}(t)A(t)^{\mathrm{T}} & \text{if } i \geq 1, j = 0 \\ \overline{P}_{i-1,j-1}(t) & \text{if } i,j \geq 1 \end{cases}. \quad (11.3.11)$$

The recursion (11.3.5), (11.3.9) *is initialized by the formulas:*

$$A(-1) := I, \quad B(-1) := 0, \quad u(-1) := 0, \quad \widehat{x}(-1|-1) := \boldsymbol{E}a;$$
$$\widehat{x}(-1-j|-1) := 0 \quad j = 1,\ldots,\sigma; \quad (11.3.12)$$

$$P_{ij}(0) = \begin{cases} 0 & \text{if} \quad i \geq 1 \text{ or } j \geq 1 \\ R_{aa} & \text{if} \quad i = j = 0 \end{cases}. \tag{11.3.13}$$

Note that in (11.3.5), (11.3.7), and (11.3.10), the sums are over the signals arrived at the moment under consideration.

If no information is received at time $t+1$, the recursion step (11.3.5), (11.3.10) becomes much simpler:

$$\overline{P}_{ij}(t+1) = P_{ij}(t+1), \qquad \widehat{x}(t+1|t+1) = A(t)\widehat{x}(t|t) + B(t)u(t);$$
$$\widehat{x}(j|t+1) = \widehat{x}(j|t) \quad \text{for} \quad j = t, t-1, \ldots, t+1-\sigma.$$

Remark 11.3.2. Since the above estimator generates estimates of not only the current but also past states, it has points of similarity with the fixed-lag smoother [125, 129].

11.3.3 The Major Properties of the State Estimator

Theorem 11.3.3. *Suppose that Assumptions* 11.2.1–11.2.4 *hold. Then the above estimator generates the sequence of minimum variance estimates; i.e.,*

$$\widehat{x}(j|t) = \boldsymbol{E}\left[x(j)|\mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)\right] \tag{11.3.14}$$

whenever $t - \sigma \leq j \leq t$ *and* $j \geq 0$. *Here* $\mathfrak{Y}(t)$, $\mathfrak{S}(t)$, *and* $\mathfrak{U}(t-1)$ *are defined by* (11.2.4)*, and* σ *is the constant from Assumption* 11.2.2.

The matrices $P_{ij}(t), \overline{P}_{ij}(t)$, $i, j = 0, \ldots, \sigma$ *generated by the estimator are the conditional covariance matrices of the estimation errors*

$$e(\theta|s) := \widehat{x}(\theta|s) - x(\theta). \tag{11.3.15}$$

More precisely, the following relations hold whenever $t - i \geq 0$ *and* $t - j \geq 0$*:*

$$\begin{aligned} \overline{P}_{ij}(t) &:= \boldsymbol{E}\big[e(t-i|t)e(t-j|t)^{\mathrm{T}}\big|\mathfrak{S}(t)\big]; \\ P_{ij}(t) &:= \boldsymbol{E}\big[e(t-i|t-1)e(t-j|t-1)^{\mathrm{T}}\big|\mathfrak{S}(t-1)\big]. \end{aligned} \tag{11.3.16}$$

Furthermore, the estimation errors (11.3.15) *do not depend on the control* (11.2.5) *(on p.* 373*).*

The proof of this theorem and the remark to follow will be given in Sect. 11.7.

The following remark presents useful technical facts.

Remark 11.3.4. (i) An access of the estimator to the quantity $\mathfrak{h}(t)$ from (11.2.5) and moreover, the knowledge of the entire sequences $S(\theta), 0 \leq \theta \leq T$ and $\mathfrak{h}(\theta), 0 \leq \theta \leq T-1$ at any time t do not alter the minimum variance state estimate:

$$\begin{aligned} \widehat{x}(j|t) &= \boldsymbol{E}\left[x(j)|\mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)\right] \\ &= \boldsymbol{E}\left[x(j)|\mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{h}(0), \ldots, \mathfrak{h}(t-1)\right] \\ &= \boldsymbol{E}\left[x(j)|\mathfrak{Y}(t), \mathfrak{S}(T), \mathfrak{h}(0), \ldots, \mathfrak{h}(T-1)\right] \end{aligned} \tag{11.3.17}$$

whenever $t - \sigma \leq j \leq t$ and $j \geq 0$.

(ii) Given $S(0), \ldots, S(T), \mathfrak{h}(0), \ldots, \mathfrak{h}(T-1)$, the estimation errors (11.3.15) (where $\max\{s-\sigma, 0\} \leq \theta \leq s$) are independent of the observations $Y(0), \ldots, Y(s)$.

11.4 Stability of the State Estimator

The stability of the estimator is essential when estimation is performed over a long period of time. As is well known, the observation delays are the source of potential instability [93,97,209,210]. In this section, we show that the proposed state estimator is stable provided natural assumptions hold. For the sake of conciseness, we confine ourselves to the case when the system (11.2.1), (11.2.2) (but not the communication channels) is stationary.

We start with definitions required to state the assumptions under which the stability issues will be addressed.

11.4.1 Almost Sure Observability via the Communication Channels

Definition 11.4.1. *Consider the plant* (11.2.1)*,* (11.2.2) *(on p. 372) with the noises and control removed*

$$x(t+1) = Ax(t), \quad y_\nu(t) = C_\nu x(t) \quad \nu = 1, \dots, l. \tag{11.4.1}$$

This system is said to be observable via the communication channels *on the interval* $[t_0, t_1]$ *(for a given realization of the random delays* $\{\tau_\nu(t)\}$*) if every state* $x(t_0)$ *can be determined from the knowledge of the observations that are both dispatched and received within this interval.*

Proposition 11.4.2. *The system* (11.4.1) *is observable via the communication channels on the interval* $[t_0, t_1]$ *(for a given realization of the random delays* $\{\tau_\nu(t)\}$*) if and only if the following symmetric nonnegative* $n \times n$*-matrix is positive-definite:*

$$M(t_0, t_1) := \sum_{\substack{(\nu,\theta): t_0 \le \theta, \\ \theta + \tau_\nu(\theta) \le t_1}} \left(A^{\theta - t_0}\right)^{\mathrm{T}} C_\nu^{\mathrm{T}} C_\nu A^{\theta - t_0}. \tag{11.4.2}$$

The proofs of the results formulated in this section will be given in Sect. 11.8.

If a realization of $\{\tau_\nu(t)\}$ is not fixed, the matrix (11.4.2) is random.

Definition 11.4.3. *Let an integer* $\sigma_* = 0, 1, \dots$ *exist such that* $M(t_0, t_1) > 0$ *almost surely for some, possibly random, time instant* $t_0 = t_0(t_1) \in [t_1 - \sigma_*, t_1]$ *whenever* $t_1 \ge \sigma_*$*. Then the system* (11.4.1) *is said to be* almost surely observable *via the communication channels.*

In fact, the matrix $M(t_0, t_1)$ can be replaced here by the easier computable matrix

$$\mathcal{M}(t_0, t_1) := \sum_{\theta = t_0}^{t_1} \sum_{\nu=1}^{l} \alpha_\nu(\theta) \left(A^{\theta - t_0}\right)^{\mathrm{T}} C_\nu^{\mathrm{T}} C_\nu A^{\theta - t_0}; \tag{11.4.3}$$

$$\alpha_\nu(t) := \begin{cases} 0 & \text{if the signal } y_\nu(t) \text{ is lost, i.e., } \tau_\nu(t) = \infty \\ 1 & \text{otherwise} \end{cases}$$

since $M(t_0, t_1) \leq \mathcal{M}(t_0, t_1) \leq M(t_0, t_1 + \sigma)$ due to Assumption 11.2.2. Thus verification of the above observability requires only the statistics of the data dropouts. It can be shown that whenever $\det A \neq 0$, the system (11.4.1) is observable via the communication channels if and only if $\mathcal{M}(T - \sigma_*, T) > 0 \ \forall T \geq \sigma_*$ almost surely for some $\sigma_* = 0, 1, \ldots$. Note that the random time instant t_0 is not employed here.

Observability via the communication channels apparently implies that

The system (11.4.1) *is observable in the standard sense.*

In the remainder of this subsection we consider only such systems. Then the system is clearly observable via the communication channels whenever they do not lose signals. To state other criteria, we say that a finite nonempty set

$$\mathcal{T} \subset \big\{(\nu, t) : \nu = 1, \ldots, l, t = 0, 1, \ldots\big\} \tag{11.4.4}$$

is *representative* for the system (11.4.1) if and only if

$$\bigcap_{(\nu,t)\in\mathcal{T}} \ker C_\nu A^{(t-t_-)} = \{0\}, \quad \text{where} \quad t_- := \min\{t : (\nu, t) \in \mathcal{T}\}. \tag{11.4.5}$$

Proposition 11.4.4. *The system* (11.4.1) *is observable via the communication channels if and only if an integer* $\sigma_* = 0, 1, \ldots$ *exists such that the set*

$$\mathfrak{T}(T) := \big\{(\nu, t) : T - \sigma_* \leq t \leq T \text{ and the signal } y_\nu(t) \text{ is not lost } \big\}$$

contains a representative subset almost surely for all $T \geq \sigma_*$.

Now we offer a number of facts concerning representative sets.

Proposition 11.4.5. *Denote by* $\mathcal{O}$ *the collection of all matrices* A *such that the system* (11.4.1) *is observable in the standard sense. For almost all* $A \in \mathcal{O}$ *(with respect to the Lebesgue measure), any finite set* (11.4.4) *such that the set* $\mathcal{T}_\nu := \{t : (\nu, t) \in \mathcal{T}\}$ *contains* n *elements (the* ν*th sensor output is not lost* n *times) for all* $\nu = 1, \ldots, l$ *is representative.*

By combining Propositions 11.4.4 and 11.4.5, we see that almost all observable (in the standard sense) systems are observable via the communication channels with "low rates" of data dropout. The last property means that for some integer $\sigma_* > n$, each channel does not lose (almost surely) at least n messages in every set of σ_* consecutively dispatched messages.

Proposition 11.4.6. *Any finite set of the form*

$$\mathcal{T} = \{(\nu, \theta) : \nu = 1, \ldots, l, \theta = t, t + 1, \ldots, t + n - 1\}$$

(the channels lose no sensor signal dispatched within some time interval of duration n*) is representative.*

Thus an observable (in the standard sense) system is observable via the communication channels if for some $\sigma_* > n$, any time interval of duration σ_* contains a subinterval of duration $n-1$ such that all messages sent via the channels within it arrive (sooner or later) at the destination point. Note that this subinterval is common for all channels. The next proposition shows that if $\det A \neq 0$, this subinterval may depend on the channel.

Proposition 11.4.7. *Suppose that* $\det A \neq 0$. *Any finite set* $\mathfrak{T}$ *such that the set* $\mathfrak{T}_\nu$ *is an interval* $t_\nu, t_\nu + 1, \ldots, t_\nu + n - 1$ *of the length* n *(consequent* n *outputs of the* ν*th sensor are not lost) for all* $\nu = 1, \ldots, l$ *is representative.*

Proposition 11.4.8. *Let* $a(\lambda) := \det(\lambda I - A)$ *be the characteristic polynomial of the matrix* A. *For* $t = 0, 1, \ldots$, *we denote by* $p_t(\lambda)$ *the remainder of* λ^t *when divided by* $a(\lambda)$. *Suppose that for any* $\nu = 1, \ldots, l$, *the set* $\mathfrak{T}_\nu$ *consists of* n *elements* $\mathfrak{T}_\nu = \{t_1 < t_2 < \ldots < t_n\}$ *and the polynomials*

$$p_{t_1 - t_-}(\lambda), p_{t_2 - t_-}(\lambda), \ldots, p_{t_n - t_-}(\lambda),$$

where t_- *is defined in* (11.4.5)*, are linearly independent. Then the set* $\mathfrak{T}$ *is representative.*

11.4.2 Conditions for Stability of the State Estimator

Now we suppose that the following three additional assumptions hold.

Assumption 11.4.9. *The coefficients of the system* (11.2.1), (11.2.2) *(on p.* 372*) do not vary in course of time:* $A(t) \equiv A$, $C_\nu(t) \equiv C_\nu$.

Assumption 11.4.10. *The process disturbance in* (11.2.1) *and the noises in* (11.2.2) *are statistically stationary and nonsingular:*

$$R_{\xi\xi}(t) \equiv R_{\xi\xi} > 0, \qquad R^\nu_{\chi\chi}(t) \equiv R^\nu_{\chi\chi} > 0.$$

Assumption 11.4.11. *The system* (11.4.1) *is either almost surely observable via the communication channels (see Definition* 11.4.1*) or stable.*

The following theorem is the main result of this section.

Theorem 11.4.12. *Suppose that Assumptions* 11.2.1– 11.2.4 *(on pp.* 373 *and* 374*), and* 11.4.9–11.4.11 *hold. Then the state estimator described in Subsect.* 11.3.2 *(starting on p.* 375*) is a.s. uniformly exponentially stable; i.e.,*

$$\sum_{j=0}^{\sigma} \|e(t-j|t)\| \leq c\rho^{t-t_0} \sum_{j=0}^{\sigma} \|e(t_0 - j|t_0)\| \tag{11.4.6}$$

whenever $t \geq t_0$. *Here* $e(\theta|t)$ *are the errors* (11.3.15) *that occur provided the noises are removed from* (11.2.1) *and* (11.2.2) *(on p.* 372*). The constants* $c > 0$ *and* $\rho \in (0, 1)$ *do not depend on* t, t_0 *and the initial states of both the system and the estimator.*

It is tacitly assumed that the estimator gain matrices from (11.3.5) are calculated in the presence of the noises for a specific initial data of the recursion (11.3.9) and fixed.

The proof of Theorem 11.4.12 will be given in Sect. 11.9.

Remark 11.4.13. Consider two initial random vectors a_1 and a_2 in (11.2.1). By (11.3.13), each of them a_i gives rise to its own sequence of the estimator gain matrices $\overset{i}{K}{}^s_j(t)$. It follows from (11.4.6), along with (11.3.15), (11.3.16) and (11.3.7), (11.3.8), that the discrepancy between these matrices is exponentially small

$$\left\| \overset{2}{K}{}^s_j(t) - \overset{1}{K}{}^s_j(t) \right\| \leq c_*(a_1, a_2)\rho_*^t \quad \text{a.s.},$$

where $\rho_* \in (0, 1)$. Thus the influence of the initial state a on the proposed state estimator vanishes in course of time.

11.5 Finite Horizon Linear-Quadratic Gaussian Optimal Control Problem

Now we revert to the controlled system (11.2.1), (11.2.2) (on p. 372) and consider the problem of minimizing the quadratic cost functional over a finite horizon

$$\mathfrak{I}_T := \boldsymbol{E} \sum_{t=0}^{T-1} \left[x(t+1)^{\mathrm{T}} Q(t+1) x(t+1) + u(t)^{\mathrm{T}} \Gamma(t) u(t) \right]. \tag{11.5.1}$$

Here $Q(t + 1) \geq 0$ and $\Gamma(t) > 0$ are symmetric $n \times n$ and $m \times m$ matrices, respectively.

Now the observations are sent to not the observer but to the controller. In this and the next sections, we consider the case where the control acts upon the plant immediately and with no distortion. This means that in (11.2.5) (on p. 373), the variable $\mathfrak{h}(t)$ takes only one value and so can be dropped. Hence the natural class of control strategies is given by

$$u(t) = \mathcal{U}\left[t, \mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)\right], \tag{11.5.2}$$

where $\mathcal{U}(\cdot)$ is a deterministic function and $\mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)$ are defined in (11.2.3) and (11.2.4) (on p. 373). It should be remarked that by recursive excluding $u(\theta)$ from the right-hand side, (11.5.2) can be shaped into the form $u(t) = \Psi\left[t, \mathfrak{Y}(t), \mathfrak{S}(t)\right]$ similar to (11.2.5).

In the other respects, the situation does not change. In particular, the measurement signals still incur random irregular delays and are sealed with time stamps.

To present the solution, we consider first the problem of minimizing the functional (11.5.1) subject to the constraints (11.2.1) in the case where the process disturbances are removed ($\xi(t) \equiv 0$) and the entire state $x(t)$ is immediately accessible by the controller. As is well known (see Appendix C), the solution of this simplified problem is given by the linear feedback $u(t) = -L(t)x(t)$, where

$$L(t) = F(t)^{-1} B(t)^{\mathrm{T}} \left[Q(t+1) + H(t+1)\right] A(t); \tag{11.5.3}$$

$$F(t) := \Gamma(t) + B(t)^{\mathrm{T}} \left[Q(t+1) + H(t+1)\right] B(t), \tag{11.5.4}$$

and the symmetric $n \times n$-matrices $H(T), H(T-1), \ldots, H(0)$ are calculated recursively in accordance with the following difference Ricatti equation:

$$H(t) = L(t)^{\mathrm{T}} \Gamma(t) L(t) + A_L(t)^{\mathrm{T}} \left[Q(t+1) + H(t+1)\right] A_L(t)$$
$$\text{where} \quad A_L(t) := A(t) - B(t)L(t). \tag{11.5.5}$$

Here $t = T-1, T-2, \ldots, 0$ and the recursion is initialized by putting $H(T) := 0$. It is easy to see that $H(t) \geq 0 \; \forall t$. So the matrix $F(t)^{-1}$ in (11.5.3) does exist since

$$F(t) = \Gamma(t) + B(t)^{\mathrm{T}} Q(t+1) B(t) + B(t)^{\mathrm{T}} H(t+1) B(t) \geq \Gamma(t) > 0. \tag{11.5.6}$$

Now we are in a position to state the main result of the section.

Theorem 11.5.1. *Suppose that Assumptions* 11.2.1 *and* 11.2.2 *(on p.* 374*) hold. Then the optimal strategy* (11.5.2) *exists and is given by the formula*

$$u(t) = -L(t)\widehat{x}(t|t) \quad \text{for} \quad t = 0, 1, \ldots, T-1. \tag{11.5.7}$$

Here the gain matrices $L(t)$ are determined by (11.5.3)–(11.5.5)*, and $\widehat{x}(t|t)$ is the estimation of the current state $x(t)$ generated by the estimator from Subsect.* 11.3.2.

Now this estimator should be implemented at the controller site.

The proof of Theorem 11.5.1 will be given in Sect. 11.7.

11.6 Infinite Horizon Linear-Quadratic Gaussian Optimal Control Problem

In this section, we consider the infinite-horizon time-invariant system (11.2.1), (11.2.2) and deal with the problem of minimizing the average cost per unit time

$$\overline{\lim_{T \to \infty}} \; T^{-1} \mathcal{I}_T. \tag{11.6.1}$$

Here $\mathcal{I}_T$ is defined by (11.5.1). Apart from Assumptions 11.2.1 and 11.2.2 (on p. 374), several additional assumptions are adopted. The first of them extends Assumption 11.4.9 (on p. 380).

Assumption 11.6.1. *The coefficients of both the system* (11.2.1), (11.2.2) *(on p.* 372*) and the functional* (11.5.1) *(on p.* 381*) do not vary as time progresses:*

$$A(t) \equiv A, \quad B(t) \equiv B, \quad C_\nu(t) \equiv C_\nu, \quad Q(t) \equiv Q, \quad \Gamma(t) \equiv \Gamma.$$

Assumption 11.6.2. *The pairs (A, B) and (A, Q) are stabilizable and detectable, respectively.*

The next assumption relaxes Assumption 11.4.10 (on p. 380) by dropping the requirement that the noises should be nonsingular.

Assumption 11.6.3. *The noises in* (11.2.1) *and* (11.2.2) *are statistically stationary:*

$$R_{\xi\xi}(t) \equiv R_{\xi\xi}, \qquad R^{\nu}_{\chi\chi}(t) \equiv R^{\nu}_{\chi\chi}.$$

Assumption 11.6.4. *A constant* $p \in (0, +\infty)$ *exists such that*

$$\boldsymbol{E}\,\mathbf{tr}\left[P_{00}(t) - \overline{P}_{00}(t)\right] \leq p \qquad \forall t = 0, 1, 2, \ldots. \tag{11.6.2}$$

Here $P_{ij}(t)$, $\overline{P}_{ij}(t)$ *are the symmetric* $n \times n$ *matrices generated by the estimator from Subsect.* 11.3.2 *(starting on p.* 375*).*

It should be remarked that the matrices $P_{ij}(t)$, $\overline{P}_{ij}(t)$ are generated by formulas (11.3.7)–(11.3.11), (11.3.13), which do not employ controls. So they are well defined in the current situation, where the controls have not been specified yet.

An informal interpretation of Assumption 11.6.4 proceeds from the fact that by (11.3.14), (11.3.15), and (11.3.16), $P_{00}(t)$ and $\overline{P}_{00}(t)$ are the covariance matrices of the current state estimation errors that occur, respectively, before and after the arrival of the current observation message. So the quantity $\boldsymbol{E}\,\mathbf{tr}\left[P_{00}(t) - \overline{P}_{00}(t)\right]$ from (11.6.2) evaluates to what extent this message improves the accuracy of the estimate. In view of this, we call this quantity the *efficiency* of the tth message. Thus Assumption 11.6.4 means that there is no observation message with arbitrarily high efficiency.

A criterion for Assumption 11.6.4 to hold will be offered at the end of this section.

Now we consider the algebraic Ricatti equation

$$H = L_H^{\mathrm{T}} \Gamma L_H + (A - BL_H)^{\mathrm{T}} (Q + H)(A - BL_H) \tag{11.6.3}$$

with respect to the unknown $n \times n$ symmetric matrix H. Here

$$L_H := F_H^{-1} B^{\mathrm{T}} (Q + H) A, \qquad F_H := \Gamma + B^{\mathrm{T}} (Q + H) B. \tag{11.6.4}$$

Under the circumstances, this equation is well known to have a solution H for which the matrix

$$A_H := A - BL_H \tag{11.6.5}$$

is asymptotically stable:

$$\|(A_H)^t\| \leq c\rho^t \qquad t = 0, 1, 2, \ldots \tag{11.6.6}$$

for some $c > 0, \rho \in (0, 1)$. This solution is unique and nonnegative. In particular, the last property implies that the inverse matrix F_H^{-1} in (11.6.4) does exist.

Now we are in a position to state the main result of the section.

Theorem 11.6.5. *Suppose that Assumptions* 11.2.1, 11.2.2 *and* 11.6.1–11.6.4 *(on pp.* 373 *and* 382*) hold. Then the feedback*

$$u(t) = -L_H \widehat{x}(t|t) \tag{11.6.7}$$

furnishes the minimum of the average cost for unit time (11.6.1) *over all control strategies* (11.5.2)*. In* (11.6.7)*, the estimate* $\widehat{x}(t|t)$ *is generated by the state estimator from Subsect.* 11.3.2 *(starting on p.* 375*).*

The proof of this theorem will be given in Sect. 11.10.

As will be shown in Sect. 11.10 (for details, see Lemma 11.10.1 on p. 397), the stationary control gain (11.6.7) can be viewed as the limit of the optimal nonstationary gain (11.5.7). At the same time, both gains are fed by the state estimate produced by the common and nonstationary estimation algorithm.

We close the section with sufficient conditions for Assumption 11.6.4 to hold and the control strategy (11.6.7) to stabilize the system.

Proposition 11.6.6. *Suppose that Assumptions* 11.2.1, 11.2.2 *and* 11.6.1–11.6.3 *(on pp.* 373 *and* 382*) are true. Assumption* 11.6.4 *(on p.* 383*) holds, and the control strategy* (11.6.7) *stabilizes the system*

$$\sup_{t=0,1,\dots} \boldsymbol{E}\|x(t)\|^2 < \infty$$

whenever at least one of the following claims is true:

(i) The uncontrolled plant (11.2.1)*,* (11.2.2) *(on p.* 372*) is asymptotically stable:* $\|A^t\| \le \gamma\mu^t$ *for* $t = 0, 1, 2, \dots$ *and some* $\gamma > 0, \mu \in (0, 1)$*;*
(ii) This plant is almost surely observable via the communication channels.[1]

The proof of this proposition will be given in Sect. 11.10.

11.7 Proofs of Theorems 11.3.3 and 11.5.1 and Remark 11.3.4

Proof of Theorem 11.3.3 and Remark 11.3.4 (on p. 377). Consider the linear space

$$\mathfrak{Z} := \left\{ Z = \{z_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} : z_{\nu,j} \in \mathbb{R}^{k_\nu} \ \forall \nu, j \right\}.$$

We recall that l is the number of sensors, k_ν is the dimension of the νth sensor output, and σ is taken from Assumption 11.2.2 (on p. 374). We also set

$$x(-\sigma) := x(-\sigma+1) := \cdots := x(-1) := 0$$

and introduce the vectors

[1]See Definition 11.4.3 on p. 378.

$$X(t) := \textbf{col}\,\big[x(t), x(t-1), \ldots, x(t-\sigma)\big];$$
$$Z(t) := \big\{z_{\nu,j}\big\} \in \mathfrak{Z}, \quad \text{where} \quad z_{\nu,j} := \begin{cases} y_\nu(t-j) & \text{if} \quad (\nu, t-j) \in S(t) \\ 0 & \text{otherwise} \end{cases}. \tag{11.7.1}$$

The vector $Z(t)$ can be interpreted as composed of the currently arrived observations $Y(t)$ from (11.2.3) (on p. 373) supplemented with several zeros. In terms of $X(t)$ and $Z(t)$, relations (11.2.1) and (11.2.2) (on p. 372) take the forms, respectively:

$$X(t+1) = \mathfrak{A}(t)X(t) + \mathfrak{B}(t)u(t) + \mathfrak{E}\,\xi(t) \qquad t = 0, \ldots, T-1; \tag{11.7.2}$$

$$X(0) = \mathfrak{a} := \textbf{col}\,(a, 0, \ldots, 0); \tag{11.7.3}$$

$$Z(t) = \mathfrak{C}\big[t, S(t)\big]X(t) + \Xi\big[t, S(t)\big] \qquad t = 0, \ldots, T. \tag{11.7.4}$$

Here

$$\mathfrak{A}(t) := \begin{pmatrix} A(t) & 0 & 0 & \cdots & 0 & 0 \\ I & 0 & 0 & \cdots & 0 & 0 \\ 0 & I & 0 & \cdots & 0 & 0 \\ 0 & 0 & I & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & I & 0 \end{pmatrix} \quad \text{and} \quad \begin{array}{l} \mathfrak{E} := \textbf{col}\,\big[I, 0, \ldots, 0\big] \\ \mathfrak{B}(t) := \textbf{col}\,\big[B(t), 0, \ldots, 0\big] \end{array}. \tag{11.7.5}$$

Furthermore for any $t \leq T$, $S \subset \{(\nu, j) : \nu = 1, \ldots, l, j = 0, \ldots, \sigma\}$, and $X = \textbf{col}\,(x_0, x_1, \ldots, x_\sigma) \in \mathbb{R}^{(\sigma+1)n}$, the following relations hold:

$$\mathfrak{C}[t, S]X := \{z_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} \in \mathfrak{Z},$$
$$\text{where} \quad z_{\nu,j} := \begin{cases} C_\nu(t-j)x_j & \text{if} \quad (\nu, t-j) \in S \\ 0 & \text{otherwise} \end{cases}; \tag{11.7.6}$$

$$\Xi[t, S] := \{\zeta_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} \in \mathfrak{Z},$$
$$\text{where} \quad \zeta_{\nu,j} := \begin{cases} \chi_\nu(t-j) & \text{if} \quad (\nu, t-j) \in S \\ 0 & \text{otherwise} \end{cases}. \tag{11.7.7}$$

Formula (11.2.5) (on p. 373) shapes into

$$u(t) = \mathfrak{U}\left[t, Z(0), \ldots, Z(t), S(0), \ldots, S(t), \mathfrak{h}(t)\right].$$

By Assumptions 11.2.2 and 11.2.4 (on p. 374), the quantities $\{\tau_\nu(t)\}$ and $\{\mathfrak{h}(t)\}$ are independent of a, $\{\xi(t)\}$, and $\{\chi_\nu(t)\}$. So evidently are the random sets $\{S(t)\}$ defined by (11.2.3) (on p. 373). Then (11.2.7) (on p. 374) implies that the vectors $a, \xi(t)$, and $\varXi[t, S(t)]$ conditioned over

$$S(0), \ldots, S(T), \mathfrak{h}(0), \ldots, \mathfrak{h}(T-1) \tag{11.7.8}$$

are (singular) Gaussian and independent. This yields that conditioning all random vectors over (11.7.8) reduces the system to that considered in the classic Kalman filtering theory. This theory gives the expression for the minimum variance estimate $\widehat{X}(t|t)$ of $X(t)$ based on $Z(0), \ldots, Z(t)$, and due to the preliminary conditioning, the quantities from (11.7.8). This estimate is generated recursively by the corresponding Kalman filter so that only the sets $S(0), \ldots, S(t)$ from (11.7.4) and $u(0), \ldots, u(t-1)$ are required to compute $\widehat{X}(t|t)$. This implies (i) of Remark 11.3.4 (on p. 377) and means that $\widehat{X}(t|t)$ is in fact the minimum variance estimate of $X(t)$ based on

$$Z(0), \ldots, Z(t), \quad S(0), \ldots, S(t), \quad \text{and} \quad u(0), \ldots, u(t-1).$$

Theorem 11.3.3 (on p. 377) results from putting (11.7.5)–(11.7.7) into the equations giving the solution of the classic minimum variance estimation problem; i.e., formulas (C.8)–(C.13) (on pp. 510 and 511), along with elementary transformations of the resultant formulas. This also gives the last claim of Theorem 11.3.3 and (ii) of Remark 11.3.4 (on p. 377) thanks to (ii)–(iv) of Theorem C.2 (on p. 511).

Proof of Theorem 11.5.1 (on p. 382). In terms of the vectors (11.7.1), the problem under consideration takes the form:

$$\text{minimize} \quad \boldsymbol{E} \sum_{t=0}^{T-1} \left[X(t+1)^{\mathrm{T}} \mathfrak{Q}(t+1) X(t+1) + u(t)^{\mathrm{T}} \Gamma(t) u(t) \right]; \tag{11.7.9}$$

$$\text{where} \qquad \mathfrak{Q}(t) := \begin{pmatrix} Q(t) & 0 \cdots & 0 \\ 0 & 0 \cdots & 0 \\ 0 & 0 \cdots & 0 \\ \cdot & \cdot \cdots & \cdot \\ 0 & 0 \cdots & 0 \end{pmatrix}, \tag{11.7.10}$$

subject to the constraints (11.7.2)–(11.7.4) and

$$u(t) = \mathfrak{U}\left[t, Z(0), \ldots, Z(t), S(0), S(1), \ldots, S(t), u(0), \ldots, u(t-1)\right]. \tag{11.7.11}$$

Now in (11.7.8), $\mathfrak{h}(t)$ take only one value and so may be ignored. As was shown, the vectors $a, \xi(t)$, and $\Xi[t, S(t)]$ conditioned over $S(0), \ldots, S(T-1)$ are Gaussian and independent. This yields that the problem resulting from (11.7.2)–(11.7.4), (11.7.9), and (11.7.11) by expanding the class of admissible controls as follows:

$$u(t) = \mathcal{U}\left[t, Z(0), \ldots, Z(t), S(0), \ldots, S(T-1), u(0), \ldots, u(t-1)\right]$$

reduces to a particular case of the standard LQG optimal control problem (see Appendix C starting on p. 509). By Theorem C.5 (on p. 513) from this appendix, the solution of the latter problem is given by the linear feedback $u(t) = -\mathfrak{F}(t)X(t)$ that is optimal for the problem of minimizing the functional (11.7.9) subject to the constraints (11.7.2) and (11.7.3) in the case where $\xi(t) \equiv 0$ and the entire state X is measured. This controller is fed by the minimum variance estimate $\widehat{X}(t|t)$ of $X(t)$ based on the observations $Z(0), \ldots, Z(t)$ and the known controls $u(0), \ldots, u(t-1)$.

This estimate is generated recursively by the Kalman filter so that only the sets $S(0), \dots, S(t)$ from (11.7.4) are in fact required to compute $\widehat{X}(t|t)$. Calculating the gain matrix $\mathfrak{F}(t)$ employs the coefficients of the functional (11.7.9) and the equation (11.7.2) for the time instants $t, t+1, \dots, T$. Since these coefficients do not depend on the sets $S(0), \dots, S(T-1)$ and are known in advance, the control appears to be of the form (11.7.11). So it furnishes the optimum over the strategies given by (11.7.11).

Theorem 11.5.1 (on p. 382) results from putting the expressions (11.7.5)–(11.7.7) and (11.7.10) into the equations giving the solutions of the minimum variance state estimate and the standard LQG control problems, respectively (i.e., formulas (C.8)–(C.13) and (C.18)–(C.21) on pp. 510–512 from Appendix C), along with elementary transformations of the resultant formulas.

11.8 Proofs of the Propositions from Subsect. 11.4.1

Proof of Proposition 11.4.2 (on p. 378). Given two processes

$$p_i = \left([x_i(t)]_{t=t_0}^{t_1}, [Y_i(t)]_{t=t_0}^{t_1} \right), \qquad i = 1, 2$$

in the system (11.4.1) (here $Y_i(t)$ is given by (11.2.3) (on p. 373), where $S(t)$ is replaced by $\widehat{S}(t) := \{(\nu, \theta) \in S(t) : \theta \geq t_0\}$), we have thanks to (11.2.3), (11.4.1), and (11.4.2),

$$[x_2(t_0) - x_1(t_0)]^{\mathrm{T}} M(t_0, t_1)[x_2(t_0) - x_1(t_0)] = \sum_{t=t_0}^{t_1} \|Y_2(t) - Y_1(t)\|^2.$$

So $\big(Y_1(t) = Y_2(t)\, \forall t = t_0, \dots, t_1\big) \Rightarrow p_1 = p_2$ if and only if $M(t_0, t_1) > 0$. □

Proof of Proposition 11.4.4 (on p. 379). Note first that whenever $t_0 \geq t_1 - \sigma_*$, (11.4.3) (on p. 378) yields

$$z^{\mathrm{T}} \mathcal{M}(t_0, t_1) z = \sum_{(\nu,\theta) \in \mathfrak{T}(t_1) : \theta \geq t_0} \left\| C_\nu A^{\theta - t_0} z \right\|^2. \tag{11.8.1}$$

If $\mathfrak{T}(t_1)$ contains a representative subset $\mathcal{T}$, then putting here $t_0 := t_-$ the constant t_- from (11.4.5) (on p. 379) makes the sufficiency part of the proposition apparent.

Conversely, let the system (11.4.1) be observable via the communication channels; i.e., there is $\sigma_* \geq 0$ such that $\mathcal{M}(t_0, t_1) > 0$ for some $t_0 \in [t_1 - \sigma_*, t_1]$ almost surely. Then (11.8.1) implies that the set $\mathcal{T} := \{(\nu, \theta) \in \mathfrak{T}(t_1) : \theta \geq t_0\}$ is representative since $t_0 < t_- \Rightarrow \det A \neq 0$. □

Proposition 11.4.6 (on p. 379) is evident.

Proof of Proposition 11.4.8 (on p. 380). By the definition of $p_t(\lambda)$,

$$\lambda^t = b_t(\lambda) a(\lambda) + p_t(\lambda), \tag{11.8.2}$$

where $b_t(\lambda)$ is a polynomial. Thanks to the assumptions of the proposition, $p_{s_1}(\lambda)$, $\ldots$, $p_{s_n}(\lambda)$ (where $s_i := t_i - t_-$) is the basis in the space of all polynomials of degree $\leq n-1$. So for $j = 0, \ldots, n-1$,

$$\lambda^j = \sum_{i=1}^{n} \gamma_i^{(j)} p_{s_i}(\lambda) = \sum_{i=1}^{n} \gamma_i^{(j)} \lambda^{s_i} - \left[\sum_{i=1}^{n} \gamma_i^{(j)} b_{s_i}(\lambda) \right] a(\lambda)$$

for some $\gamma_i^{(j)} \in \mathbb{R}$. Since $a(A) = 0$, we get

$$C_\nu A^j = \sum_{i=1}^{n} \gamma_i^{(j)} C_\nu A^{t_i - t_-}, \qquad \bigcap_{t \in \mathcal{T}_\nu} \ker C_\nu A^{t-t_-} = \bigcap_{i=1}^{n} \ker C_\nu A^{t_i - t_-}$$
$$\subset \bigcap_{j=0}^{n-1} \ker C_\nu A^j, \qquad \bigcap_{(\nu,t) \in \mathcal{T}} \ker C_\nu A^{t-t_-} \subset \bigcap_{j=0}^{n-1} \bigcap_{\nu=0}^{l} \ker C_\nu A^j = \{0\}. \qquad \square$$

Proof of Proposition 11.4.7 (on p. 380). Pick $\nu = 1, \ldots, l$. Suppose that

$$\sum_{j=0}^{n-1} \gamma_j p_{t_\nu + j - t_-}(\lambda) \equiv 0$$

for some $\gamma_j \in \mathbb{R}$, and denote $\gamma(\lambda) := \gamma_0 + \gamma_1 \lambda + \cdots \gamma_{n-1} \lambda^{n-1}$. Then by (11.8.2)

$$\gamma(\lambda) \lambda^{t_\nu - t_-} = \left(\sum_{j=0}^{n-1} \gamma_j b_{t_\nu + j - t_-}(\lambda) \right) a(\lambda).$$

Since $a(0) \neq 0$, the polynomials $a(\lambda)$ and $\lambda^{t_\nu - t_-}$ are coprime. So the polynomial $a(\lambda)$ of degree n must divide the polynomial $\gamma(\lambda)$ of degree $\leq n-1$. Hence $\gamma(\lambda) \equiv 0, \gamma_0 = \cdots = \gamma_{n-1} = 0$; i.e., the polynomials $p_{t_\nu - t_-}(\lambda), \ldots, p_{t_\nu + n - 1 - t_-}(\lambda)$ are linearly independent. Proposition 11.4.8 completes the proof. $\square$

Proof of Proposition 11.4.5 (on p. 379). Since

$$\mathcal{P}(A) := \det \sum_{\nu=1}^{l} \sum_{j=0}^{n-1} A^j C_\nu C_\nu^{\mathrm{T}} (A^{\mathrm{T}})^j$$

is evidently a nonzero polynomial in the entries of the matrix A, and $\mathcal{O} = \{A : \mathcal{P}(A) \neq 0\}$, the set $\mathcal{O}$ is of full Lebesgue measure in the space of all $n \times n$ matrices A. Now let a tuple $\mathfrak{t} = \{t_1, \ldots, t_n\}$ of integers $t_i \geq 0, t_i \neq t_j \; \forall \, i \neq j$ be given. It is easy to see that for any $t = 0, 1, \ldots$, the coefficients of the polynomial

$$p_t(\lambda) = p_t(0) + \cdots + p_t(n-1)\lambda^{n-1}$$

from Proposition 11.4.8 (on p. 380) are polynomials in the entries of A. So is

$$\mathfrak{p}_{\mathfrak{t}}(A) := \det \begin{pmatrix} p_{t_1}(0) & \ldots & p_{t_1}(n-1) \\ \ldots & \ldots & \ldots \\ p_{t_n}(0) & \ldots & p_{t_n}(n-1) \end{pmatrix}.$$

Note that $\mathfrak{p}_\mathfrak{t}(A) \not\equiv 0$. Indeed, otherwise, any polynomial $a(\lambda) = \lambda^n + \cdots$ of degree n would divide some nonzero polynomial of the form $\gamma_1\lambda^{t_1} + \cdots + \gamma_n\lambda^{t_n}$. In terms of the roots $\lambda_1, \ldots, \lambda_n$ of $a(\lambda)$, this implies

$$q(\lambda_1, \ldots, \lambda_n) := \det \begin{pmatrix} \lambda_1^{t_1} & \ldots & \lambda_n^{t_1} \\ \ldots & \ldots & \ldots \\ \lambda_1^{t_n} & \ldots & \lambda_n^{t_n} \end{pmatrix} = 0.$$

This relation holds at least for all real λ_i. At the same time,

$$q(\lambda_1, \ldots, \lambda_n) = \sum_\alpha (-1)^{\text{inv}(\alpha)} \lambda_1^{t_{\alpha(1)}} \cdots \lambda_n^{t_{\alpha(n)}}$$

(the sum is over all permutations $\alpha = [\alpha(1), \ldots, \alpha(n)]$ of $\{1, 2, \ldots, n\}$ and $\text{inv}(\alpha)$ is the number of inversions in the permutation) a nonzero polynomial and so $q(\lambda_1, \ldots, \lambda_n) \neq 0$ for some real λ_i. The contradiction obtained proves that $\mathfrak{p}_\mathfrak{t}(A) \not\equiv 0$. Hence the set $\mathcal{R}_\mathfrak{t} := \{A \in \mathcal{O} : \mathfrak{p}_\mathfrak{t}(A) \neq 0\}$ is of full measure. So is the set $\mathcal{R} := \bigcap_\mathfrak{t} \mathcal{R}_\mathfrak{t}$.

Now let $A \in \mathcal{R}$. It is easy to see that any set $\mathcal{T}$ with the properties described in Proposition 11.4.5 (on p. 379) satisfies the hypotheses of Proposition 11.4.8 (on p. 380), which completes the proof. □

11.9 Proof of Theorem 11.4.12 on p. 380

The state estimator introduced in Subsect. 11.3.2 (starting on p. 375) is in fact the conventional Kalman filter related to the representation of the system in the augmented form (11.7.2)–(11.7.4) (on p. 385). However in the vast literature on the Kalman filter, the authors failed to find a reference that proves Theorem 11.4.12. The reason is that the nonstationary (in the sensor part) system (11.7.2), (11.7.4) loses observability due to the state augmentation. In the case of unstable but observable in the sense of Assumption 11.4.11 (on p. 380) system (11.4.1), this causes violation of assumptions under which the stability of the Kalman filter was established in the studies known to the authors. In view of this, we adduce an independent proof. It employs a slightly refined standard technique.

From now on, Assumptions 11.2.1–11.2.4 (on pp. 373 and 374), and 11.4.9–11.4.11 (on p. 380) are supposed to hold. Introduce the following vector and matrices:

$$\mathcal{E}(t) := \begin{bmatrix} e(t|t) \\ \vdots \\ e(t-\sigma|t) \end{bmatrix}, \qquad \mathcal{P}(t) := \begin{pmatrix} P_{00}(t) \cdots P_{0\sigma}(t) \\ \vdots \\ P_{\sigma 0}(t) \cdots P_{\sigma\sigma}(t) \end{pmatrix},$$

$$\overline{\mathcal{P}}(t) := \begin{pmatrix} \overline{P}_{00}(t) \cdots \overline{P}_{0\sigma}(t) \\ \vdots \\ \overline{P}_{\sigma 0}(t) \cdots \overline{P}_{\sigma\sigma}(t) \end{pmatrix}. \quad (11.9.1)$$

Here $e(\theta|t)$ are the estimation errors (11.3.15) (on p. 377), $P_{ij}(t), \overline{P}_{ij}(t)$ are the matrices generated by the estimator, and σ is taken from Assumption 11.2.2 (on p. 374). By Theorem 11.3.3 (on p. 377) and (11.3.7)–(11.3.11), (11.3.13) (on p. 376), these errors and matrices do not depend on the control (11.2.5) (on p. 373). For simplicity, we assume that $u(t) \equiv 0$. In the remainder of the section, "constant" means "deterministic constant."

Lemma 11.9.1. *Constants $\beta > 0, \overline{\beta} > 0$ exist such that $\mathcal{P}(t) \leq \beta I$ and $\overline{\mathcal{P}}(t) \leq \overline{\beta} I$ almost surely for all t.*

Proof. Suppose that the first case from Assumption 11.4.11 (on p. 380) holds; i.e., the system (11.4.1) (on p. 378) is almost surely observable via the communication channels.[2] In view of (11.2.3) (on p. 373) and (11.4.2) (on p. 378), one can suppose that in Definition 11.4.3, $t_0(t_1)$ is a deterministic function of $t_1, S(0), \dots, S(t_1)$. Furthermore, the matrix $M[t_0(t_1), t_1]$ is determined by the sample sequence assumed by $\tau_\nu(t)$ over the time interval $[t_0(t_1), t_1]$ of a bounded duration $t_1 - t_0(t_1) \leq \sigma_*$. There are no more than finitely many such sample sequences (up to the change of the variable $t \mapsto t - t_0(t_1)$) even if t_1 is not fixed. This implies that for some $\mu > 0$,

$$\left\| M[t_0(t_1), t_1]^{-1} \right\| \leq \mu \quad \text{a.s. whenever} \quad t_1 \geq \sigma_*. \tag{11.9.2}$$

Choose $t_1 \geq \sigma_*$ and denote

$$t_0 := t_0(t_1), \qquad S(t_0, t_1) := \{(\nu, \theta) : t_0 \leq \theta, \theta + \tau_\nu(\theta) \leq t_1\}.$$

By (11.2.1) and (11.2.2) (on p. 372),

$$x(t) = A^{t-t_0} x(t_0) + \sum_{p=t_0}^{t-1} A^{t-1-p} \xi(p) \qquad \text{whenever} \quad t > t_0; \tag{11.9.3}$$

$$y_\nu(t) = C_\nu A^{t-t_0} x(t_0) + C_\nu \sum_{p=t_0}^{t-1} A^{t-1-p} \xi(p) + \chi_\nu(t) \qquad \text{whenever} \quad t \geq t_0;$$

$$\sum_{(\nu,\theta) \in S(t_0,t_1)} (A^{\mathrm{T}})^{\theta - t_0} C_\nu^{\mathrm{T}} y_\nu(\theta) = \underbrace{\sum_{(\nu,\theta) \in S(t_0,t_1)} (A^{\mathrm{T}})^{\theta - t_0} C_\nu^{\mathrm{T}} C_\nu A^{\theta - t_0}}_{M(t_0,t_1)} x(t_0)$$
$$+ \sum_{(\nu,\theta) \in S(t_0,t_1)} (A^{\mathrm{T}})^{\theta - t_0} C_\nu^{\mathrm{T}} \left[C_\nu \sum_{j=t_0}^{\theta-1} A^{\theta-1-j} \xi(j) + \chi_\nu(\theta) \right]. \tag{11.9.4}$$

Put $s := t_1 + \sigma + 1$ and for $j = 0, \dots, \sigma$, consider the following estimate of $x(s-j)$:

$$\widetilde{x}(s-j) := A^{s-j-t_0} M(t_0, t_1)^{-1} \sum_{(\nu,\theta) \in S(t_0,t_1)} (A^{\mathrm{T}})^{\theta - t_0} C_\nu^{\mathrm{T}} y_\nu(\theta).$$

[2]See Definition 11.4.3 on p. 378.

Now we invoke the notation $\mathfrak{S}(s-1)$ from (11.2.4) (on p. 373). Then (11.3.14)–(11.3.16) (on p. 377) imply

$$\mathbf{tr}\, P_{jj}(s) = \boldsymbol{E}\left[\|e(s-j|s-1)\|^2\middle|\mathfrak{S}(s-1)\right] \\ \leq \boldsymbol{E}\left[\|x(s-j) - \widetilde{x}(s-j)\|^2|\mathfrak{S}(s-1)\right]. \quad (11.9.5)$$

We recall that the symbol **tr** stands for the trace of a matrix. By (11.9.3) and (11.9.4),

$$x(s-j) - \widetilde{x}(s-j) = \sum_{i=t_0}^{s-1-j} A^{s-1-i-j}\xi(i)$$

$$-A^{s-t_0-j}M(t_0,t_1)^{-1} \sum_{(\nu,\theta)\in S(t_0,t_1)} (A^{\mathrm{T}})^{\theta-t_0} C_\nu^{\mathrm{T}} \left[C_\nu \sum_{j=t_0}^{\theta-1} A^{\theta-1-j}\xi(j) + \chi_\nu(\theta)\right].$$

We note that $t_1 - t_0 = t_1 - t_0(t_1) \leq \sigma_*$ by Definition 11.4.3 (on p. 378), and we denote

$$a := \max\{\|A^t\| : t = 0, \ldots, \sigma_* + \sigma + 1\}, \quad c := \max\{\|C_\nu\| : \nu = 1, \ldots, l\}.$$

Then in view of (11.9.2),

$$1/2\boldsymbol{E}\left[\|x(s-j) - \widetilde{x}(s-j)\|^2|\mathfrak{S}(s-1)\right] \leq [\sigma_* + \sigma]a^2\, \mathbf{tr}\, R_{\xi\xi} \\ + a^4\mu^2c^2\sigma_* l\left[c^2a^2\sigma_*^2 l\, \mathbf{tr}\, R_{\xi\xi} + \max_{\nu=1,\ldots,l} \mathbf{tr}\, R^\nu_{\chi\chi}\right].$$

Here $s = t_1 + \sigma + 1$ and $t_1 \geq \sigma_*$ is arbitrary. Thus $\mathbf{tr}\, P_{jj}(t)$ is bounded above. To complete the proof, we note that $\overline{P}_{jj}(t) \leq P_{jj}(t)$ thanks to (11.3.16) (on p. 377) and employ (11.9.1).

Now suppose that the second case from Assumption 11.4.11 (on p. 380) holds; i.e., the system (11.4.1) (on p. 378) is stable. By taking $\widetilde{x}(s') := 0\ \forall s'$ in (11.9.5) and invoking Assumption 11.2.2 (on p. 374), we see that $\mathbf{tr}\, P_{jj}(t)$ is still bounded above. After this, the proof is completed as in the previous case. □

From now on, we consider the vectors $\mathcal{E}(t)$ given by (11.9.1) in the case where the noises are removed from the plant equations (11.2.1) and (11.2.2) (on p. 372).

Lemma 11.9.2. *Constants $\alpha, \overline{\alpha}, \beta, \overline{\beta} > 0$, and t_* exist such that for $t \geq t_*$,*

$$\alpha I \leq \mathcal{P}(t) \leq \beta I, \quad \overline{\alpha} I \leq \overline{\mathcal{P}}(t) \leq \overline{\beta} I \qquad \text{a.s.}; \quad (11.9.6)$$

$$\mathcal{E}(t+1) = \mathfrak{A}\mathcal{E}(t) + V(t+1) \quad \text{a.s.}, \\ \text{where} \quad V(t+1) := \left[\overline{\mathcal{P}}(t+1)\mathcal{P}(t+1)^{-1} - I\right]\mathfrak{A}\mathcal{E}(t); \quad (11.9.7)$$

$$\mathcal{E}(t)^{\mathrm{T}}\overline{\mathcal{P}}(t)^{-1}\mathcal{E}(t) - \mathcal{E}(t-1)^{\mathrm{T}}\overline{\mathcal{P}}(t-1)^{-1}\mathcal{E}(t-1) \leq -V(t)^{\mathrm{T}}\mathcal{P}(t)^{-1}V(t) \\ - \sum_{(\nu,\theta)\in S(t)} e(\theta|t)^{\mathrm{T}} C_\nu^{\mathrm{T}}\left[R^\nu_{\chi\chi}\right]^{-1} C_\nu e(\theta|t) \quad \text{a.s.} \quad (11.9.8)$$

Proof. The upper bounds from (11.9.6) are established by Lemma 11.9.1. To obtain the lower ones, let us fix a realization of $\tau_\nu(t)$ by conditioning all random quantities over $\tau_\nu(t)$. It follows from (11.3.14)–(11.3.16) (on p. 377) and (11.9.1) that the error covariance matrices are no less $\mathcal{P}(t) \geq \mathcal{P}^f(t), \overline{\mathcal{P}}(t) \geq \overline{\mathcal{P}}^f(t)$ than those in the case where the observations are received without losses and delays; i.e., $S(t) = \{(\nu, t) : \nu = 1, \ldots, l\}$. Then with regard to Assumptions 11.2.1 and 11.4.10 (on pp. 373 and 380) and well-known facts about the asymptotic behaviour of the classic Kalman filter [123, p.100] applied to the augmented system, we conclude that the matrices $\mathcal{P}^f(t)$ and $\overline{\mathcal{P}}^f(t)$ are bounded from below by constant positive-definite matrices for large t. This completes the proof of (11.9.6).

Relations (11.9.7) and (11.9.8) are immediate from, e.g., (54)–(56) and (58) [43]. It was assumed in [43] that the noise Ξ in (11.7.4) (on p. 385) is nonsingular, and the matrix $\mathfrak{A}$ in (11.7.2) is invertable. The first requirement can be met by substituting $\widehat{\chi}_\nu(t)$ in place of zeros in (11.7.7) (on p. 385). Here $\widehat{\chi}_\nu(t) \in \mathbb{R}^{k_\nu}$ are Gaussian mutually independent and independent of $\tau_\nu(t), \xi(t), \chi_\nu(t)$ zero-mean random vectors with $\boldsymbol{E}\widehat{\chi}_\nu(t)\widehat{\chi}_\nu(t)^{\mathrm{T}} = I$. It is easy to check that this does not alter the Kalman filter formulas. The invertability of $\mathfrak{A}$ was employed in (57) [43] to ensure that

$$\boxed{D \geq 0 \,\&\, P > 0 \,\&\, \mathfrak{A}P\mathfrak{A}^{\mathrm{T}} + D > 0} \Rightarrow \mathfrak{A}^{\mathrm{T}}[\mathfrak{A}P\mathfrak{A}^{\mathrm{T}} + D]^{-1}\mathfrak{A} \leq P^{-1}.$$

However this implication holds for any matrix $\mathfrak{A}$, which can be proved by considering a nonsingular perturbation $\mathfrak{A}_\varepsilon, \det \mathfrak{A}_\varepsilon \neq 0, \mathfrak{A}_\varepsilon \to \mathfrak{A}$ as $\varepsilon \to 0$ of $\mathfrak{A}$ and letting $\varepsilon \to 0$ in the above inequality. □

By (11.9.6) and (11.9.8),

$$\overline{\beta}^{-1}\|\mathcal{E}\|^2 \leq W(t, \mathcal{E}) := \mathcal{E}^{\mathrm{T}}\overline{\mathcal{P}}(t)^{-1}\mathcal{E} \leq \overline{\alpha}^{-1}\|\mathcal{E}\|^2 \quad \text{a.s. whenever } t \geq t_*, \tag{11.9.9}$$

$$W[t+1, \mathcal{E}(t+1)] \leq W[t, \mathcal{E}(t)] \quad \text{a.s. whenever } t \geq t_*. \tag{11.9.10}$$

This implies that the state estimator is Lyapunov stable. The following lemma shows that the additional estimate

$$\lambda W[t+r, \mathcal{E}(t+r)] \leq W[t, \mathcal{E}(t)] \quad \text{a.s. whenever } t \geq \widehat{t} \tag{11.9.11}$$

implies the exponential stability. Here $\lambda > 1, \widehat{t} = t_*, t_* + 1, \ldots$, and $r = 1, 2, \ldots$ are constants.

Lemma 11.9.3. *Suppose that* (11.9.11) *holds. Then the conclusion of Theorem* 11.4.12 *(on p.* 380*) is true.*

Proof. By (11.9.6) and (11.9.7),

$$\|\mathcal{E}(t+1)\| \leq \overline{\beta}\alpha^{-1}\|\mathfrak{A}\|\|\mathcal{E}(t)\| \qquad \text{a.s.}$$

whenever $t \geq t_*$. For $t \leq t_*$, (11.2.1), (11.2.2), and (11.3.5)–(11.3.11) (on pp. 372, 376) yield $\mathcal{E}(t+1) = \mathfrak{R}(t)\mathcal{E}(t)$, where $\mathfrak{R}(t) = \mathcal{R}[t, S(0), \ldots, S(t+1)]$ and $\mathcal{R}$ is a deterministic function. The random sequence $S(0), \ldots, S(t_* + 1)$ assumes only finitely many samples. Hence $\|\mathfrak{R}(t)\| \leq f \; \forall t = 0, \ldots, t_*$ a.s. for some constant f. By the foregoing, there is $\mu > 1$ such that

$$\|\mathcal{E}(t_2)\| \leq \mu^{t_2 - t_1}\|\mathcal{E}(t_1)\| \quad \text{a.s. whenever} \quad t_2 \geq t_1. \tag{11.9.12}$$

Thanks to (11.9.10), (11.9.11), $\lambda W[t, \mathcal{E}(t)] \leq W[\theta, \mathcal{E}(\theta)]$ a.s. whenever $t - r \geq \theta := \max\{t_0, \widehat{t}\}$. Then invoking (11.9.9) gives

$$\|\mathcal{E}(t)\| \leq \sqrt{\overline{\beta}\overline{\alpha}^{-1}}\lambda^{-j/2}\|\mathcal{E}(\theta)\| \quad \text{a.s. whenever } t - jr \geq \theta = \max\{t_0, \widehat{t}\}.$$

Now let $t \geq t_0$. By employing the above inequality with $j \geq r^{-1}(t - \theta) - 1$ if $t \geq \widehat{t}$ and (11.9.12) otherwise, and putting $\varkappa := \sqrt{\overline{\beta}\overline{\alpha}^{-1}\lambda}$, $\rho := \lambda^{-r^{-1}/2} < 1$, we get

$$\|\mathcal{E}(t)\| \leq \begin{cases} \varkappa\rho^{t-\theta}\begin{cases} \|\mathcal{E}(t_0)\| & \text{if } t_0 \geq \widehat{t}, \\ \|\mathcal{E}(\widehat{t})\| \overset{(11.9.12)}{\leq} \mu^{\widehat{t}-t_0}\|\mathcal{E}(t_0)\| & \text{if } t_0 < \widehat{t}, \end{cases} & \text{and } t \geq \widehat{t}, \\ \mu^{t-t_0}\|\mathcal{E}(t_0)\| & \text{if } t < \widehat{t} \end{cases}$$

$$\leq \begin{Bmatrix} \varkappa & \text{if } t_0 \geq \widehat{t} \text{ and } t \geq \widehat{t}, \\ \varkappa\rho^{t_0-\widehat{t}}\mu^{\widehat{t}} & \text{if } t_0 < \widehat{t} \text{ and } t \geq \widehat{t}, \\ \mu^{\widehat{t}}\rho^{t_0-\widehat{t}} & \text{if } t < \widehat{t} \end{Bmatrix} \rho^{t-t_0}\|\mathcal{E}(t_0)\|.$$

This clearly implies (11.4.6) (on p. 380) with a properly chosen constant c. □

The remainder of the section is devoted to proving (11.9.11). In so doing, the following fact will play a key role.

Lemma 11.9.4. *Deterministic sequences* $\{\varkappa(i)\}_{i=0}^{\infty}, \{d(i)\}_{i=0}^{\infty} \subset (0, \infty)$ *exist such that*

$$W[t + k, \mathcal{E}(t + k)] \leq \varkappa(k)\|e(t|t)\|^2 + d(k)\sum_{j=t+1}^{t+k} V(j)^{\mathrm{T}}\mathcal{P}(j)^{-1}V(j) \tag{11.9.13}$$

a.s. whenever $k \geq \sigma$ *and* $t \geq t_*$. *If the system* (11.4.1) *(on p.* 378*) is stable, these sequences can be chosen so that*

$$\varkappa(k) \to 0 \quad \text{as} \quad k \to \infty \qquad \text{and} \qquad \overline{d} := \sup_k d(k) < \infty. \tag{11.9.14}$$

Proof. Owing to (11.9.7) and (11.9.9),

$$\mathcal{E}(t + k) = \mathfrak{A}^k\mathcal{E}(t) + \sum_{j=t+1}^{t+k}\mathfrak{A}^{t+k-j}V(j);$$

$$\overline{\alpha}W[t + k, \mathcal{E}(t + k)] \leq \|\mathcal{E}(t + k)\|^2 \leq 2\|\mathfrak{A}^k\mathcal{E}(t)\|^2 + 2\sum_{i=0}^{k-1}\left\|\mathfrak{A}^i\right\|^2\sum_{j=t+1}^{t+k}\|V(j)\|^2.$$

Here $\|V(j)\|^2 \leq \beta V(j)^{\mathrm{T}} \mathcal{P}(j)^{-1} V(j)$ by (11.9.6), and due to (11.7.5) (on p. 385),

$$\mathfrak{A}^k = \begin{pmatrix} A^k & 0\cdots 0 \\ \vdots & \vdots \\ A^{k-\sigma} & 0\cdots 0 \end{pmatrix}.$$

So $\|\mathfrak{A}^k \mathcal{E}(t)\| \leq \|\mathfrak{A}^k\| \|e(t|t)\|$ thanks to (11.9.1). Thus (11.9.13) does hold with

$$\varkappa(k) := 2\overline{\alpha}^{-1} \left\|\mathfrak{A}^k\right\|^2, \qquad d(k) := 2\overline{\alpha}^{-1}\beta \sum_{i=0}^{k-1} \left\|\mathfrak{A}^i\right\|^2. \tag{11.9.15}$$

If the system (11.4.1) (on p. 378) is stable, so evidently is (11.7.2) (on p. 385). Then (11.9.14) is immediate from (11.9.15). □

Proof of (11.9.11) in the case of stable system (11.4.1). Pick $\overline{t} \geq t_* + 1$ and $k \geq \sigma$. By neglecting the last summand on the right in (11.9.8) and summing the resultant inequalities over $t = \overline{t}, \ldots, \overline{t} + k$, we get

$$W[\overline{t} + k, \mathcal{E}(\overline{t} + k)] + \sum_{t=\overline{t}}^{\overline{t}+k} V(t)^{\mathrm{T}} \mathcal{P}(t)^{-1} V(t) \leq W[\overline{t} - 1, \mathcal{E}(\overline{t} - 1)]$$

a.s. Here thanks to (11.9.13) and (11.9.14),

$$\sum_{t=\overline{t}}^{\overline{t}+k} V(t)^{\mathrm{T}} \mathcal{P}(t)^{-1} V(t) \geq \overline{d}^{-1} \left\{ W[\overline{t} + k, \mathcal{E}(\overline{t} + k)] - \varkappa(k) \|e(\overline{t} - 1|\overline{t} - 1)\|^2 \right\}$$

a.s. Furthermore, due to (11.9.1) and (11.9.9),

$$\|e(\overline{t} - 1|\overline{t} - 1)\|^2 \leq \left\|\mathcal{E}(\overline{t} - 1)\right\|^2 \leq \overline{\beta} W[\overline{t} - 1, \mathcal{E}(\overline{t} - 1)] \quad \text{a.s.}$$

Thus

$$\lambda(k) W[\overline{t} + k, \mathcal{E}(\overline{t} + k)] \leq W[\overline{t} - 1, \mathcal{E}(\overline{t} - 1)] \quad \text{a.s.},$$

where

$$\lambda(k) := (1 + \overline{d}^{-1})[1 + \overline{d}^{-1} \overline{\beta} \varkappa(k)] \to 1 + \overline{d}^{-1} > 1 \quad \text{as} \quad k \to \infty$$

by (11.9.14). Picking k such that $\lambda(k) > 1$ yields (11.9.11) with $r := k + 1$. □

In the case where the system (11.4.1) is not stable, we need more preliminaries.

Lemma 11.9.5. *Suppose that $\mathcal{Q}_1(z)$ and $\mathcal{Q}_2(z)$ are nonnegative quadratic forms in $z \in \mathbb{R}^p$, and $\mathcal{Q}_2(z) = 0$ whenever $\mathcal{Q}_1(z) = 0$. A constant $\gamma > 0$ exists such that*

$$\mathcal{Q}_1(z) \geq \gamma \mathcal{Q}_2(z) \qquad \forall z.$$

Proof. For $i = 1, 2$, introduce the symmetric $p \times p$ matrix Ω_i related to $\mathcal{Q}_i(z) = z^{\mathrm{T}}\Omega_i z$, and denote by π the orthogonal projection onto $L := (\ker \Omega_1)^{\perp}$. Since $\mathcal{Q}_i(z) = 0 \Leftrightarrow \Omega_i z = 0$, then $z - \pi z \in \ker \Omega_1 \subset \ker \Omega_2 \, \forall z$. So

$$\mathcal{Q}_i(z) = [(z - \pi z) + \pi z]^{\mathrm{T}}\Omega_i[(z - \pi z) + \pi z] = (\pi z)^{\mathrm{T}}\Omega_i(\pi z) = \mathcal{Q}_i(\pi z).$$

The form $\mathcal{Q}_1(z)$ is positive-definite on L and hence $\mathcal{Q}_1(z) \geq \gamma \mathcal{Q}_2(z) \; \forall z \in L$ for some $\gamma > 0$. Thus $\mathcal{Q}_1(z) = \mathcal{Q}_1(\pi z) \geq \gamma \mathcal{Q}_2(\pi z) = \gamma \mathcal{Q}_2(z) \; \forall z$. □

Lemma 11.9.6. *A constant $\gamma > 0$ exists such that*

$$\mathcal{J}^{\bar{t}} := \sum_{t=t_0(\bar{t})}^{\bar{t}+\sigma} \left[\frac{1}{2} V(t)^{\mathrm{T}} \mathcal{P}(t)^{-1} V(t) + \sum_{(\nu,\theta) \in S(t)} e(\theta|t)^{\mathrm{T}} C_\nu^{\mathrm{T}} \left[R_{\chi\chi}^{\nu} \right]^{-1} C_\nu e(\theta|t) \right]$$
$$\geq \gamma \left\| e\left[t_0(\bar{t}) | t_0(\bar{t}) \right] \right\|^2 \quad \text{a.s.} \quad \forall \bar{t} \geq t_* + \sigma_* + 1. \qquad (11.9.16)$$

Here $\sigma_*, t_0(\bar{t})$, and t_* are the time instants from Definition 11.4.3 *(on p.* 378*) and Lemma* 11.9.2.

Proof. In view of (11.9.6) and the inequality $\bar{t} - t_0(\bar{t}) \leq \sigma_*$ from Definition 11.4.3, $V^{\mathrm{T}}\mathcal{P}(t)^{-1}V \geq \beta^{-1}\|V\|^2$ in (11.9.16). Furthermore $\chi^{\mathrm{T}}[R_{\chi\chi}^{\nu}]^{-1}\chi \geq \upsilon^{-1}\|\chi\|^2$, where $\upsilon := \max_\nu \|R_{\chi\chi}^{\nu}\|$. Hence $\mathcal{J}^{\bar{t}}$ is no less than the minimum value of the cost functional in the following optimization problem:

$$\text{minimize} \;\; \frac{1}{2\beta} \sum_{t=t_0(\bar{t})+1}^{\bar{t}} \|V(t)\|^2 + \upsilon^{-1} \sum_{t=t_0(\bar{t})}^{\bar{t}} \sum_{(\nu,\theta) \in S(t): \theta \geq t_0(\bar{t})} \varepsilon_{t-\theta}(t)^{\mathrm{T}} C_\nu^{\mathrm{T}} C_\nu \varepsilon_{t-\theta}(t)$$

subject to

$$\boldsymbol{\mathcal{E}}(t) = \textbf{col}\left[\varepsilon_0(t), \ldots, \varepsilon_\sigma(t)\right] = \mathfrak{A}\boldsymbol{\mathcal{E}}(t-1) + V(t), \quad t = t_0(\bar{t}) + 1, \ldots, \bar{t}$$

and $\boldsymbol{\mathcal{E}}[t_0(\bar{t})] = \mathcal{E}[t_0(\bar{t})]$. The change of the variable $t \mapsto t - t_0(\bar{t})$ shapes it into

$$\text{minimize} \;\; \widetilde{\mathcal{J}} := \frac{1}{2\beta} \sum_{t=1}^{k} \|V(t)\|^2 + \upsilon^{-1} \sum_{t=0}^{k} \sum_{\substack{(\nu,\theta):\theta \geq 0, \\ [\nu,\theta+t_0(\bar{t})] \in S[t+t_0(\bar{t})]}} \varepsilon_{t-\theta}(t)^{\mathrm{T}} C_\nu^{\mathrm{T}} C_\nu \varepsilon_{t-\theta}(t)$$

subject to

$$\boldsymbol{\mathcal{E}}(t) = \mathfrak{A}\boldsymbol{\mathcal{E}}(t-1) + V(t), \quad t = 1, \ldots, k, \qquad \boldsymbol{\mathcal{E}}(0) = Z.$$

Here $k := \bar{t} - t_0(\bar{t}) \leq \sigma_*$ and $Z = \text{col}\,(z_0, \ldots, z_\sigma) = \mathcal{E}[t_0(\bar{t})]$. As is well known, the above minimum value is a nonnegative quadratic form $\mathcal{Q}_1(Z)$ in Z. Let $\mathcal{Q}_1(Z) = 0$ for some Z. Then $V(t) = 0$ for $t = 1, \ldots, k$. So (11.7.5) (on p. 385) imply $\varepsilon_j(t) = A^{t-j} z_0$ whenever $t - j \geq 0$. Whence

$$0 = \upsilon\widetilde{\mathfrak{I}} = \sum_{t=0}^{k} \sum_{\substack{(\nu,\theta):\theta\geq 0,\\ [\nu,\theta+t_0(\overline{t})]\in S[t+t_0(\overline{t})]}} z_0^{\mathrm{T}}(A^{\mathrm{T}})^{\theta} C_\nu^{\mathrm{T}} C_\nu A^{\theta} z_0$$

$$\overset{(11.2.3)}{=\!=} \sum_{t=t_0(\overline{t})}^{\overline{t}} \sum_{\substack{(\nu,\theta):\theta\geq t_0(\overline{t}),\\ \tau_\nu(\theta)+\theta\leq t}} z_0^{\mathrm{T}}(A^{\mathrm{T}})^{\theta-t_0(\overline{t})} C_\nu^{\mathrm{T}} C_\nu A^{\theta-t_0(\overline{t})} z_0 \overset{(11.4.2)}{=\!=} z_0^{\mathrm{T}} M[t_0(\overline{t}),\overline{t}] z_0.$$

Due to Definition 11.4.3 (on p. 378), $z_0 = 0$ and so $\mathcal{Q}_2(Z) := \|z_0\|^2 = 0$. By Lemma 11.9.5, $\mathcal{Q}_1(Z) \geq \gamma\|z_0\|^2 \;\forall Z$. The number γ may depend on the sample assumed by the sequence $\widetilde{S}(0),\ldots,\widetilde{S}(k)$, where

$$\widetilde{S}(t) := \Big\{(\nu,\theta) : \theta\geq 0, [\nu,\theta+t_0(\overline{t})] \in S[t+t_0(\overline{t})]\Big\}.$$

Since $k\leq\sigma_*$ and $\widetilde{S}(t)\subset\{(\nu,\theta): 0\leq\theta\leq\sigma_*, 1\leq\nu\leq l\}$, there are only finitely many such samples. By minimizing γ over them, we arrive at the assertion of the lemma. □

Proof of (11.9.11) in the case where the system (11.4.1) is observable via the communication channels. Let $\overline{t}\geq\sigma_*+t_*+1$. By summing (11.9.8) over $t=t_0(\overline{t}),\ldots,\overline{t}+\sigma$, we get

$$W[\overline{t}+\sigma,\mathcal{E}(\overline{t}+\sigma)] + \frac{1}{2}\sum_{t=t_0(\overline{t})}^{\overline{t}+\sigma} V(t)^{\mathrm{T}}\mathcal{P}(t)^{-1}V(t) + \mathfrak{I}^{\overline{t}} \leq W\left(t_0(\overline{t})-1, \mathcal{E}\left[t_0(\overline{t})-1\right]\right).$$

By Lemma 11.9.6, the sum of the second and third summands on the left is no less than

$$\frac{1}{2}\sum_{t=t_0(\overline{t})}^{\overline{t}+\sigma} V(t)^{\mathrm{T}}\mathcal{P}(t)^{-1}V(t) + \gamma\left\|e\left[t_0(\overline{t})|t_0(\overline{t})\right]\right\|^2$$

$$\geq \overline{\gamma}\omega\left[\overline{t}+\sigma-t_0(\overline{t})\right] W[\overline{t}+\sigma,\mathcal{E}(\overline{t}+\sigma)].$$

Here $\omega(j) := (\varkappa[j]+d[j])^{-1}, \overline{\gamma} := \min\{\gamma,1/2\}$ and the inequality holds by Lemma 11.9.4. Since $\overline{t}-t_0(\overline{t})\leq\sigma_*$ by Definition 11.4.3,

$$\omega[\overline{t}+\sigma-t_0(\overline{t})] \geq \overline{\omega} := \min_{j=\sigma,\sigma+\sigma_*}\omega(j) > 0.$$

With regard to (11.9.10), we see that

$$(1+\overline{\gamma}\,\overline{\omega})W[\overline{t}+\sigma,\mathcal{E}(\overline{t}+\sigma)] \leq W[t_0(\overline{t})-1,\mathcal{E}(t_0(\overline{t})-1)] \leq W[\overline{t}-\sigma_*-1,\mathcal{E}(\overline{t}-\sigma_*-1)].$$

Thus (11.9.11) does hold with $r := \sigma+\sigma_*+1$. □

11.10 Proofs of Theorem 11.6.5 and Proposition 11.6.6 on p. 384

Throughout the section, Assumptions 11.2.1, 11.2.2 (on pp. 373 and 374), and 11.6.1–11.6.4 (on p. 383) are assumed to hold. We start with recalling well-known facts concerning Ricatti equations [8, 30, 85]. In so doing, we equip the solution of the Ricatti difference equation (11.5.5) (on p. 382) and the corresponding matrices (11.5.3) and (11.5.4) with the index T: $H_T(t), L_T(t), F_T(t)$.

Lemma 11.10.1. (i) *Infinite sequences of $m \times n$ and $m \times m$ matrices $L(0), L(1), \dots$ and $F(0), F(1), \dots$, respectively, exist such that*

$$L_T(t) = L(T-t), \quad F_T(t) = F(T-t)$$

whenever $t \leq T$. Furthermore

$$L(t) \to L_H, \quad F(t) \to F_H \qquad \text{as} \quad t \to \infty.$$

Here L_H and F_H are given by (11.6.4) *(on p.* 383*), where H is the solution of the algebraic Ricatti equation* (11.6.3) *(on p.* 383*) for which the matrix $A - BL_H$ is asymptotically stable.*

(ii) *Let $x \in \mathbb{R}^n, u \in \mathbb{R}^m$, $T = 1, 2, \dots$, and $t = 0, \dots, T-1$. Put $x_+ := Ax + Bu$. Then the following relation holds:*

$$\begin{aligned} x_+^{\mathrm{T}} H_T(t+1) x_+ - x^{\mathrm{T}} H_T(t) x + x_+^{\mathrm{T}} Q x_+ + u^{\mathrm{T}} \Gamma u \\ = \big[u + L_T(t)x\big]^{\mathrm{T}} F_T(t) \big[u + L_T(t)x\big]. \end{aligned} \quad (11.10.1)$$

We preface the proof of Theorem 11.6.5 (on p. 384) with three lemmas. The first of them gives a useful representation of the functional (11.5.1) (on p. 381).

Lemma 11.10.2. *Suppose that a control strategy* (11.5.2) *(on p.* 381*) is chosen. Consider the corresponding process and the output*

$$\widehat{X}(t), \quad \{P_{ij}(t)\}_{i,j=0}^{\sigma}, \quad \{\overline{P}_{ij}(t)\}_{i,j=0}^{\sigma} \qquad t = 0, 1, \dots \quad (11.10.2)$$

of the state estimator described in Subsect. 11.3.2 *(starting on p.* 375*). Then for any $T = 1, 2, \dots$, the value of the cost functional* (11.5.1) *is given by the formula*

$$\mathfrak{I}_T = \boldsymbol{E} \sum_{t=0}^{T-1} \Big[u(t) + L_T(t)\widehat{x}(t|t)\Big]^{\mathrm{T}} F_T(t) \Big[u(t) + L_T(t)\widehat{x}(t|t)\Big] + \Delta_T. \quad (11.10.3)$$

Here $\widehat{x}(t|t)$ is the first component of the output (11.3.4) *(on p.* 375*) of the estimator and*

$$\begin{aligned} \Delta_T := \sum_{t=0}^{T-1} \mathbf{tr}\ [\{H_T(t+1) + Q\} R_{\xi\xi}] + \mathbf{tr}\ [H_T(0) R_{aa}] \\ + \boldsymbol{E} \sum_{t=0}^{T-1} \mathbf{tr}\ \Big[L_T(t)^{\mathrm{T}} F_T(t) L_T(t) \overline{P}_{00}(t)\Big]. \end{aligned} \quad (11.10.4)$$

Remark 11.10.3. It follows from (11.3.7), (11.3.8), (11.3.10), (11.3.11), and (11.3.13) (on p. 376) that the quantity (11.10.4) does not depend on the control strategy (11.5.2).

Proof of Lemma 11.10.2. For $t = 0, \dots, T-1$, put $x := x(t)$ and $u := u(t)$ into (11.10.1). By (11.2.1) (on p. 372), $x_+ = x(t+1) - \xi(t)$ in (11.10.1). Summing the resultant formulas over t gives

$$\boldsymbol{E} \sum_{t=0}^{T-1} [x(t+1) - \xi(t)]^{\mathrm{T}} \underbrace{[H_T(t+1) + Q]}_{\mathcal{N}_T(t)} [x(t+1) - \xi(t)] \\ - \boldsymbol{E} \sum_{t=0}^{T-1} [x(t)^{\mathrm{T}} H_T(t) x(t) - u(t)^{\mathrm{T}} \Gamma u(t)] \\ = \sum_{t=0}^{T-1} \underbrace{\boldsymbol{E} [u(t) + L_T(t)x(t)]^{\mathrm{T}} F_T(t) [u(t) + L_T(t)x(t)]}_{\varkappa(t)}. \quad (11.10.5)$$

Here

$$[x(t+1) - \xi(t)]^{\mathrm{T}} \mathcal{N}_T(t) [x(t+1) - \xi(t)] = x(t+1)^{\mathrm{T}} \mathcal{N}_T(t) x(t+1) \\ - [x(t+1) - \xi(t)]^{\mathrm{T}} \mathcal{N}_T(t)\xi(t) - \xi(t)^{\mathrm{T}} \mathcal{N}_T(t) [x(t+1) - \xi(t)] - \xi(t)^{\mathrm{T}} \mathcal{N}_T(t)\xi(t)$$

and

$$x(t+1) - \xi(t) = Ax(t) + B\mathfrak{U}[t, \mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)]$$

is independent of $\xi(t)$. So

$$\boldsymbol{E} [x(t+1) - \xi(t)]^{\mathrm{T}} \mathcal{N}_T(t) [x(t+1) - \xi(t)] = \boldsymbol{E} x(t+1)^{\mathrm{T}} \mathcal{N}_T(t) x(t+1) \\ - \boldsymbol{E} \xi(t)^{\mathrm{T}} \mathcal{N}_T(t) \xi(t) = \boldsymbol{E} x(t+1)^{\mathrm{T}} \mathcal{N}_T(t) x(t+1) - \mathbf{tr}\, [\mathcal{N}_T(t) R_{\xi\xi}].$$

This and (11.2.1), (11.5.1) (on pp. 372 and 381) shape (11.10.5) into

$$\mathcal{I}_T - \sum_{t=0}^{T-1} \mathbf{tr}\, [\mathcal{N}_T(t) R_{\xi\xi}] - \underbrace{\boldsymbol{E} a^{\mathrm{T}} H_T(0) a}_{\mathbf{tr}\, [H_T(0) R_{aa}]} = \sum_{t=0}^{T-1} \varkappa(t).$$

By (11.3.15) (on p. 377),

$$u(t) + L_T(t)x(t) = \mathfrak{d}(t) - L_T(t) e(t|t),$$

where $\mathfrak{d}(t) := u(t) + L_T(t)\widehat{x}(t|t)$. So

$$\varkappa(t) = \boldsymbol{E} \mathfrak{d}(t)^{\mathrm{T}} F_T(t) \mathfrak{d}(t) - \boldsymbol{E} e(t|t)^{\mathrm{T}} L_T(t)^{\mathrm{T}} F_T(t) \mathfrak{d}(t) \\ - \boldsymbol{E} \mathfrak{d}(t)^{\mathrm{T}} F_T(t) L_T(t) e(t|t) + \boldsymbol{E} e(t|t)^{\mathrm{T}} \underbrace{L_T(t)^{\mathrm{T}} F_T(t) L_T(t)}_{\Omega_T(t)} e(t|t). \quad (11.10.6)$$

Due to (11.5.2) and (11.3.14) (on pp. 381 and 377),

$$\mathfrak{d}(t) = \mathfrak{D}[t, \mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)],$$

whereas

$$e(t|t) = \boldsymbol{E}\Big[x(t)|\mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)\Big] - x(t)$$

by (11.3.15) (on p. 377). This implies that in (11.10.6), the second and third summands on the right equal 0 [30][Lemma 1.9.9]. The last summand amounts to

$$\mathbf{tr}\left\{\Omega_T(t)\boldsymbol{E}\Big[e(t|t)e(t|t)^{\mathtt{T}}\Big]\right\} = \mathbf{tr}\left\{\Omega_T(t)\boldsymbol{E}\boldsymbol{E}\Big[e(t|t)e(t|t)^{\mathtt{T}}\Big|\mathfrak{S}(t)\Big]\right\}$$
$$\overset{(11.3.16)}{=\!=\!=} \mathbf{tr}\left[\Omega_T(t)\boldsymbol{E}\overline{P}_{00}(t)\right].$$

Summarizing, we arrive at (11.10.3). □

The next lemma establishes a technical fact about the matrices (11.3.7) and (11.3.8) (on p. 376).

Lemma 11.10.4. *For any* $t = 0, 1, \ldots,$

$$\sum_{s_1, s_2 \in S(t)} K_0^{s_1}(t)\Lambda(t)_{s_1}^{s_2} K_0^{s_2}(t)^{\mathtt{T}} = P_{00}(t) - \overline{P}_{00}(t). \qquad (11.10.7)$$

Proof. In the proof, we employ properties of the pseudoinverse $\overset{+}{\Lambda}(t)$ listed in Appendix B (starting on p. 507). Due to (11.3.7) (on p. 376),

$$\sum_{s_1 \in S(t)} K_0^{s_1}(t)\Lambda(t)_{s_1}^{s} = \sum_{s_1, (\nu_2, \theta_2) \in S(t)} P_{0,t-\theta_2}(t) C_{\nu_2}^{\mathtt{T}} \overset{+}{\Lambda}(t)_{(\nu_2,\theta_2)}^{s_1} \Lambda(t)_{s_1}^{s}$$
$$\overset{(a)}{=} \sum_{(\nu_2,\theta_2) \in S(t)} P_{0,t-\theta_2}(t) C_{\nu_2}^{\mathtt{T}} \left[\Lambda_\pi(t)\right]_{(\nu_2,\theta_2)}^{s}.$$

Here (a) holds by the first formula from (B.1) (on p. 508) and Λ_π is defined in Appendix B. Furthermore,

$$\sum_{s_2 \in S(t)} K_0^{s_2}(t)\left[\Lambda_\pi(t)\right]_{s_2}^{s} \overset{(11.3.7)}{=\!=\!=} \sum_{(\nu_1,\theta_1), s_2 \in S(t)} P_{0,t-\theta_1}(t) C_{\nu_1}^{\mathtt{T}} \overset{+}{\Lambda}(t)_{(\nu_1,\theta_1)}^{s_2} \left[\Lambda_\pi(t)\right]_{s_2}^{s}$$
$$\overset{(b)}{=} \sum_{(\nu_1,\theta_1) \in S(t)} P_{0,t-\theta_1}(t) C_{\nu_1}^{\mathtt{T}} \overset{+}{\Lambda}(t)_{(\nu_1,\theta_1)}^{s} \overset{(11.3.7)}{=\!=\!=} K_0^{s}(t). \qquad (11.10.8)$$

Here (b) holds by the third formula from (B.1) (on p. 508). By taking into account the second one and summarizing, we see that the left-hand side of (11.10.7) amounts to

$$\sum_{s_2 \in S(t)} \sum_{(\nu,\theta)\in S(t)} P_{0,t-\theta}(t) C_\nu^{\mathrm{T}} \left[\Lambda_\pi(t)\right]^{s_2}_{(\nu,\theta)} K_0^{s_2}(t)^{\mathrm{T}}$$
$$= \left\{ \sum_{(\nu,\theta)\in S(t)} \left[\sum_{s_2 \in S(t)} K_0^{s_2}(t) \left[\Lambda_\pi(t)\right]^{(\nu,\theta)}_{s_2} \right] C_\nu P_{0,t-\theta}(t)^{\mathrm{T}} \right\}^{\mathrm{T}}$$
$$\overset{(11.10.8),(11.3.16)}{=\!=\!=} \left\{ \sum_{(\nu,\theta)\in S(t)} K_0^{(\nu,\theta)}(t) C_\nu P_{t-\theta,0}(t) \right\}^{\mathrm{T}} \overset{(11.3.10),(11.3.16)}{=\!=\!=} P_{00}(t) - \overline{P}_{00}(t). \square$$

Lemma 11.10.5. *Consider the feedback* (11.6.7) *(on p.* 384*), the corresponding process, and the output* (11.3.4) *(on p.* 375*) of the state estimator described in Subsect.* 11.3.2 *(starting on p.* 375*) . Then*

$$\omega := \sup_{t=0,1,\dots} \boldsymbol{E} \|\widehat{x}(t|t)\|^2 < \infty. \tag{11.10.9}$$

Proof. Pick $t = 0, 1, \dots$ and denote

$$S := S(t+1), \qquad K^{(\nu,\theta)} := K_0^{(\nu,\theta)}(t+1).$$

By (11.3.5), (11.6.5), and (11.6.7) (on pp. 376, 383, and 384),

$$\widehat{x}(t+1|t+1) = A_H \widehat{x}(t|t) + \sum_{(\nu,\theta)\in S} K^{(\nu,\theta)} \Big[y_\nu(\theta) - \widehat{y}_\nu(\theta|t) \Big]$$
$$\overset{(11.2.1),(11.2.2),(11.3.6)}{=\!=\!=} A_H \widehat{x}(t|t) + \sum_{(\nu,\theta)\in S:\theta \le t} K^{(\nu,\theta)} \Big[C_\nu \Big(x(\theta) - \widehat{x}(\theta|t) \Big) + \chi_\nu(\theta) \Big]$$
$$+ \sum_{\nu:(\nu,t+1)\in S} K^{(\nu,t+1)} \Big[C_\nu \big\{ Ax(t) + Bu(t) + \xi(t) \big\} + \chi_\nu(t+1)$$
$$- C_\nu \big\{ A\widehat{x}(t|t) + Bu(t) \big\} \Big] \overset{(11.3.15)}{=\!=} A_H \widehat{x}(t|t) - \eta(t). \tag{11.10.10}$$

Here $\eta(t) := \eta_1(t) + \eta_2(t) + \eta_3(t) + \eta_4(t)$ and

$$\eta_1(t) := - \sum_{(\nu,\theta)\in S} K^{(\nu,\theta)} \chi_\nu(\theta), \qquad \eta_2(t) := - \sum_{\nu:(\nu,t+1)\in S} K^{(\nu,t+1)} C_\nu \xi(t),$$
$$\eta_3(t) := \sum_{(\nu,\theta)\in S:\theta\le t} K^{(\nu,\theta)} C_\nu e(\theta|t), \qquad \eta_4(t) := \sum_{\nu:(\nu,t+1)\in S} K^{(\nu,t+1)} C_\nu A e(t|t).$$

Let the symbol $\boldsymbol{E_c}$ stand for the conditional mathematical expectation with respect to $\mathfrak{S}(t+1) = [S(0), \dots, S(t+1)]$. From the point of view of their conditional distributions with respect to $\mathfrak{S}(t+1)$, the vectors $e(\theta'|t)$ with $\theta' \le t$ and both $\xi(t)$ and $\chi_\nu(\theta)$ with $\theta \in S(t+1)$ are independent due to (11.2.7), along with (11.2.1), (11.2.2), (11.3.14), and (11.3.15) (on pp. 372, 374, and 377). So for $i = 1, 2$ and $j = 3, 4$, we have $\boldsymbol{E_c} \eta_i(t) \eta_j(t)^{\mathrm{T}} = \boldsymbol{E}_c \eta_i(t) \boldsymbol{E_c} \eta_j(t)^{\mathrm{T}} = 0$. Likewise $\boldsymbol{E_c} \eta_1(t) \eta_2(t)^{\mathrm{T}} = 0$. Since the errors $e(\theta|t)$ with $\theta \leq t$ are functions of $\mathfrak{S}(t)$ and the vectors $a, \xi(i), \chi_\nu(i)$ that are independent of $\mathfrak{S}(t+1) = \left[\mathfrak{S}(t), S(t+1)\right]$, we have

$$\boldsymbol{E}_{\boldsymbol{c}} e(\theta'|t)e(\theta''|t)^{\mathrm{T}} = \boldsymbol{E}\left[e(\theta'|t)e(\theta''|t)^{\mathrm{T}}\middle|\mathfrak{S}(t)\right] \overset{(11.3.16)}{=\!=} \overline{P}_{t-\theta',t-\theta''}(t)$$

for $\theta', \theta'' = \max\{t-\sigma, 0\}, \ldots, t$. Hence putting

$$\overline{S} := \{\nu : (\nu, t+1) \in S(t+1)\},$$

we get

$$\begin{aligned}
\boldsymbol{E}_{\boldsymbol{c}}\eta(t)\eta(t)^{\mathrm{T}} &= \sum_{i=1}^{4} \boldsymbol{E}_{\boldsymbol{c}}\eta_i(t)\eta_i(t)^{\mathrm{T}} + \boldsymbol{E}_{\boldsymbol{c}}\eta_3(t)\eta_4(t)^{\mathrm{T}} + \boldsymbol{E}_{\boldsymbol{c}}\eta_4(t)\eta_3(t)^{\mathrm{T}} \\
&= \sum_{(\nu,\theta)\in S} K^{(\nu,\theta)} R^{\nu}_{\chi\chi}\left[K^{(\nu,\theta)}\right]^{\mathrm{T}} + \sum_{\nu_1,\nu_2\in\overline{S}} K^{(\nu_1,t+1)} C_{\nu_1} R_{\xi\xi} C_{\nu_2}^{\mathrm{T}}\left[K^{(\nu_2,t+1)}\right]^{\mathrm{T}} \\
&\quad + \sum_{(\nu_1,\theta_1),(\nu_2,\theta_2)\in S:\theta_i\leq t} K^{(\nu_1,\theta_1)} C_{\nu_1} \overline{P}_{t-\theta_1,t-\theta_2}(t) C_{\nu_2}^{\mathrm{T}}\left[K^{(\nu_2,\theta_2)}\right]^{\mathrm{T}} \\
&\quad + \sum_{\nu_1,\nu_2\in\overline{S}} K^{(\nu_1,t+1)} C_{\nu_1} A\overline{P}_{00}(t) A^{\mathrm{T}} C_{\nu_2}^{\mathrm{T}}\left[K^{(\nu_2,t+1)}\right]^{\mathrm{T}} \\
&\quad + \sum_{\nu_1\in\overline{S},(\nu,\theta)\in S:\theta\leq t} K^{(\nu_1,t+1)} C_{\nu_1} A\overline{P}_{0,t-\theta}(t) C_{\nu}^{\mathrm{T}}\left[K^{(\nu,\theta)}\right]^{\mathrm{T}} \\
&\quad + \sum_{\nu_1\in\overline{S},(\nu,\theta)\in S:\theta\leq t} K^{(\nu,\theta)} C_{\nu} \overline{P}_{t-\theta,0}(t) A^{\mathrm{T}} C_{\nu_1}^{\mathrm{T}}\left[K^{(\nu_1,t+1)}\right]^{\mathrm{T}}.
\end{aligned}$$

By invoking (11.3.11) (on p. 376), we get

$$\begin{aligned}
\boldsymbol{E}_{\boldsymbol{c}}\eta(t)\eta(t)^{\mathrm{T}} &= \sum_{(\nu_1,\theta_1),(\nu_2,\theta_2)\in S} K^{(\nu_1,\theta_1)} C_{\nu_1} P_{t+1-\theta_1,t+1-\theta_2}(t+1) C_{\nu_2}^{\mathrm{T}}\left[K^{(\nu_2,\theta_2)}\right]^{\mathrm{T}} \\
&\quad + \sum_{(\nu,\theta)\in S} K^{(\nu,\theta)} R^{\nu}_{\chi\chi}\left[K^{(\nu,\theta)}\right]^{\mathrm{T}} \\
&\overset{(11.3.8)}{=\!=} \sum_{s_1,s_2\in S} K^{s_1}\Lambda(t+1)^{s_2}_{s_1}\left[K^{s_2}\right]^{\mathrm{T}} \overset{(11.10.7)}{=\!=} P_{00}(t+1) - \overline{P}_{00}(t+1) =: \Delta P(t+1).
\end{aligned}$$

Then (11.6.2) (on p. 383) yields

$$\begin{aligned}
\boldsymbol{E}\|\eta(t)\|^2 = \mathbf{tr}\,\boldsymbol{E}\eta(t)\eta(t)^{\mathrm{T}} = \boldsymbol{E}\,\mathbf{tr}\,\boldsymbol{E}_{\boldsymbol{c}}\eta(t)\eta(t)^{\mathrm{T}} \\
= \boldsymbol{E}\,\mathbf{tr}\,\Delta P(t+1) \leq p. \quad (11.10.11)
\end{aligned}$$

To complete the proof, note that (11.10.10) and (11.6.6) (on p. 383) imply

$$\|\widehat{x}(t|t)\| \leq c\rho^t \|\widehat{x}(0|0)\| + c\sum_{j=0}^{t-1} \rho^{t-1-j}\|\eta(j)\|$$

$$\leq c\rho^t \|\widehat{x}(0|0)\| + c \underbrace{\left[\sum_{j=0}^{t-1} \rho^{t-1-j}\right]^{1/2}}_{\leq (1-\rho)^{-1/2}} \left[\sum_{j=0}^{t-1} \rho^{t-1-j}\|\eta(j)\|^2\right]^{1/2}.$$

We finish by employing (11.10.11) and the apparent inequality $(a+b)^2 \leq 2(a^2+b^2)$

$$\boldsymbol{E}\|\widehat{x}(t|t)\|^2 \leq 2c^2 \left[\rho^{2t}\boldsymbol{E}\|\widehat{x}(0|0)\|^2 + (1-\rho)^{-1} p \sum_{j=0}^{t-1} \rho^{t-1-j}\right]$$

$$\leq 2c^2 \left[\boldsymbol{E}\|\widehat{x}(0|0)\|^2 + \frac{p}{(1-\rho)^2}\right] < \infty. \qquad \square$$

Proof of Theorem 11.6.5. Due to (11.5.6) (on p. 382) and (11.10.3),

$$\overline{\lim_{T\to\infty}}\, T^{-1}\mathfrak{I}_T \geq \overline{\lim_{T\to\infty}}\, T^{-1}\Delta_T$$

for any control strategy (11.5.2) (on p. 381). So in view of Remark 11.10.3, it suffices to show that

$$d(T) := \frac{1}{T}\boldsymbol{E}\sum_{t=0}^{T-1} \Big[u(t) + L_T(t)\widehat{x}(t|t)\Big]^{\mathrm{T}} F_T(t) \Big[u(t) + L_T(t)\widehat{x}(t|t)\Big] \to 0$$

as $T \to \infty$ for the feedback (11.6.7) (on p. 384). Pick $\varepsilon > 0$. By **(i)** of Lemma 11.10.1,

$$\|L_T(t) - L_H\| \leq \gamma < \infty, \qquad \|F_T(t)\| \leq f < \infty$$

for all $T = 1, 2, \ldots,\ t = 0, 1, \ldots, T$, and an integer $N \geq 1$ exists such that

$$\|L_T(t) - L_H\| \leq \varepsilon$$

whenever $t = 0, 1, \ldots, T - N$ and $T = N, N+1, \ldots$. For $T \geq N + 1$, relation (11.6.7) implies

$$d(T) = \frac{1}{T}\sum_{t=0}^{T-1} \boldsymbol{E}\widehat{x}(t|t)^{\mathrm{T}} \big[L_T(t) - L_H\big]^{\mathrm{T}} F_T(t) \big[L_T(t) - L_H\big] \widehat{x}(t|t)$$

$$\leq \frac{1}{T}\sum_{t=0}^{T-1} \|F_T(t)\|\|L_T(t) - L_H\|^2 \boldsymbol{E}\|\widehat{x}(t|t)\|^2 \overset{a)}{\leq} \frac{\omega}{T}\sum_{t=0}^{T-N} \underbrace{\|F_T(t)\|}_{\leq f}\, \underbrace{\|L_T(t) - L_H\|^2}_{\leq \varepsilon^2}$$

$$+ \frac{\omega}{T}\sum_{t=T-N+1}^{T-1} \underbrace{\|F_T(t)\|}_{\leq f}\, \underbrace{\|L_T(t) - L_H\|^2}_{\leq \gamma^2} \leq \omega f \left[\varepsilon^2\left(1 - \frac{N}{T}\right) + \frac{N}{T}\gamma^2\right],$$

$$\overline{\lim_{T\to\infty}}\, d(T) \leq \omega f \varepsilon^2.$$

Here a) holds due to (11.10.9). The proof is completed by letting $\varepsilon \to 0$. $\square$

Proof of Proposition 11.6.6. Statement (i). By the last claim from Theorem 11.3.3 (on p. 377), the estimation errors (11.3.15) (on p. 377) are not affected by the control (11.5.2) (on p. 381). Put $\mathfrak{U}(\cdot) \equiv 0$ in (11.5.2). Then (11.2.1) (on p. 372) implies

$$\boldsymbol{E}[x(t) - \boldsymbol{E}x(t)][x(t) - \boldsymbol{E}x(t)]^{\mathrm{T}} = A^t R_{aa} \left(A^{\mathrm{T}}\right)^t + \sum_{j=0}^{t-1} A^{t-1-j} R_{\xi\xi} \left(A^{\mathrm{T}}\right)^{t-1-j}.$$

As is well known,

$$\boldsymbol{E}[x - \boldsymbol{E}(x|\mathfrak{V})][x - \boldsymbol{E}(x|\mathfrak{V})]^{\mathrm{T}} \leq \boldsymbol{E}[x - \boldsymbol{E}x][x - \boldsymbol{E}x]^{\mathrm{T}}$$

for any random vector x and quantity $\mathfrak{V}$. By putting here

$$x := x(s), \qquad \mathfrak{V} := [\mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{U}(t-1)]$$

and invoking (11.3.15) (on p. 377), we get

$$\boldsymbol{E}e(s|t)e(s|t)^{\mathrm{T}} \leq \boldsymbol{E}[x(s) - \boldsymbol{E}x(s)][x(s) - \boldsymbol{E}x(s)]^{\mathrm{T}} \\ \leq P := \sum_{j=0}^{\infty} \left[A^j \left(R_{aa} + R_{\xi\xi}\right) \left(A^{\mathrm{T}}\right)^j\right].$$

Then (11.3.16) (on p. 377) yields

$$0 \leq \boldsymbol{E}\overline{P}_{00}(t) \leq P, \quad 0 \leq \boldsymbol{E}P_{00}(t) \leq P.$$

In view of (11.3.15), (11.3.16), and (11.10.9) this completes the proof in the case (i).

Statement (ii). We still assume that $\mathfrak{U}(\cdot) \equiv 0$ in (11.5.2) (on p. 381). In view of (11.2.3) and (11.4.2) (on pp. 373 and 378), one can assume that in Definition 11.4.3 (on p. 378), $t_0(t_1)$ is a deterministic function of $t_1, S(0), \ldots, S(t_1)$. Furthermore, the matrix $M[t_0(t_1), t_1]$ is determined by the realization of $\tau_\nu(t)$ over the time interval $[t_0(t_1), t_1]$ of a bounded duration $t_1 - t_0(t_1) \leq \sigma_*$. There are no more than finitely many such realizations up to the change of the variable $t \mapsto t - t_0(t_1)$ even if t_1 is not fixed. This implies that for some constant $\mu > 0$,

$$\left\|M[t_0(t_1), t_1]^{-1}\right\| \leq \mu \quad \text{a.s. whenever} \quad t_1 \geq \sigma_*. \tag{11.10.12}$$

Choose $t_1 \geq \sigma_*$ and denote

$$t_0 := t_0(t_1), \quad S(t_0, t_1) := \{(\nu, \theta) : t_0 \leq \theta, \theta + \tau_\nu(\theta) \leq t_1\}.$$

Put $s := t_1 + \sigma + 1$ and for $j = 0, \ldots, \sigma$, consider the following estimation of $x(s-j)$:

$$\widetilde{x}(s-j) := A^{s-j-t_0} M(t_0, t_1)^{-1} \sum_{(\nu,\theta) \in S(t_0,t_1)} \left(A^{\mathrm{T}}\right)^{\theta - t_0} C_\nu^{\mathrm{T}} y_\nu(\theta).$$

Then (11.3.14)–(11.3.16) (on p. 377) imply

$$\mathbf{tr}\, P_{jj}(s) = \boldsymbol{E}\left[\|e(s-j|s-1)\|^2|\mathfrak{S}(s-1)\right] \\ \leq \boldsymbol{E}\left[\|x(s-j)-\widetilde{x}(s-j)\|^2|\mathfrak{S}(s-1)\right].$$

Here $\mathfrak{S}(s-1)$ is given by (11.2.4) (on p. 373). By (11.9.3) and (11.9.4) (on p. 390),

$$x(s-j)-\widetilde{x}(s-j) = \sum_{i=t_0}^{s-1-j} A^{s-1-i-j}\xi(i)$$

$$-A^{s-t_0-j}M(t_0,t_1)^{-1}\sum_{(\nu,\theta)\in S(t_0,t_1)}(A^{\mathrm{T}})^{\theta-t_0}C_\nu^{\mathrm{T}}\left[C_\nu\sum_{j=t_0}^{\theta-1}A^{\theta-1-j}\xi(j)+\chi_\nu(\theta)\right].$$

Note that $t_1 - t_0 = t_1 - t_0(t_1) \leq \sigma_*$ by Definition 11.4.3 (on p. 378). Denote

$$a := \max\{\|A^t\| : t = 0,\dots,\sigma_*+\sigma+1\}, \quad c := \max\{\|C_\nu\| : \nu = 1,\dots,l\}.$$

Then in view of (11.10.12),

$$1/2\boldsymbol{E}\left[\|x(s-j)-\widetilde{x}(s-j)\|^2|\mathfrak{S}(s-1)\right] \leq [\sigma_*+\sigma+1]a^2\,\mathbf{tr}\,R_{\xi\xi} \\ + a^4\mu^2c^2(\sigma_*+1)l\left[c^2a^2(\sigma_*+1)^2 l\,\mathbf{tr}\,R_{\xi\xi}+\max_{\nu=1,\dots,l}\mathbf{tr}\,R^\nu_{\chi\chi}\right]. \quad (11.10.13)$$

Here $s = t_1+\sigma+1$ and $t_1 \geq \sigma_*$ is arbitrary. Thus $\mathbf{tr}\,P_{jj}(t)$ is bounded from above. To complete the proof, we note that $\boldsymbol{E}\overline{P}_{jj}(t) \leq \boldsymbol{E}P_{jj}(t)$ thanks to (11.3.16) (on p. 377) and invoke (11.3.15) and (11.10.9) (on pp. 377 and 400). □

12

Optimal Computer Control via Asynchronous Communication Channels

12.1 Introduction

In this chapter, we proceed with studying problems of optimal control via delayed, lossy, and asynchronous communication channels. In the previous chapter we addressed such a problem under the assumption that unlike the observation channels, the control loop is perfect, and so the controller output acts upon the plant immediately. Now we focus on the case where the control loop is delayed and lossy. As for the observations transmission and plant, the situation is just like in Chap. 11.

Specifically, we consider finite-horizon linear-quadratic optimal control problems for discrete-time partially observed systems perturbed by white noises. Data are sent from the sensors and controller to the controller and actuators, respectively, over parallel randomly delayed channels. Various signals are transferred with independent and a priori unknown transmission times. The signals may arrive out of order; there may be periods where no signal is received. The transmitted data may be lost due to, e.g., noise in the communication medium and protocol malfunctions.

We still suppose that any transmitted message is equipped with a time stamp indicating the moment of the transfer beginning. Hence the observations transmission times become known to the controller at the moments when the messages arrive at it. Likewise, the control signal transmission times become known at the actuators sites. We also suppose that this information is sent back, maybe, with delays and not continuously to the controller via special feedback control channels. As a result, there is an awareness about the bygone states of the communication medium, whereas its future states are unknown.

In this chapter, the solution of an analog of the classic finite-horizon linear-quadratic Gaussian (LQG) optimal control problem (see Appendix C starting on p. 509) is obtained for two different problem settings.

In the first one, the actuators are basically capable only to execute the currently received control signal. (In other words, they are not equipped with computing or memory modules.) So the (central) controller bears the entire responsibility for achieving the optimal performance. In this case, we suppose that this controller is given an additional information: The statistics of the data delays and dropouts in

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_12,

the control channels is known in advance. However no such information is available for the channels carrying data from the sensors. Furthermore, we suppose that the feedback control channels do not drop data and provide delays not exceeding the sample period. Under certain technical assumptions, an optimal strategy to control the plant is obtained. It is shown that the optimal control results from feeding a linear feedback with a minimum variance state estimate, along with several past controls. To generate this estimate, the recursive state estimator from Subsect. 11.3.2 (starting on p. 375) is employed. Explicit formulas for the gain matrices of the optimal feedback are offered. The core of them is constituted by a finite set of coupled difference Riccati equations.

It should be remarked that many communication channels do not satisfy the time invariance condition, and a reliable prognosis of the future states of the communication medium is often a hard problem (see, e.g., [208], [232]). This is taken into account in the second setup of the LQG optimal control problem considered in this chapter. It is not assumed any longer that the statistics of data delays and dropouts is known in advance. Moreover, no restrictions on this statistics are imposed. Similarly, it is not assumed any longer that the delays in the feedback control channels do not exceed the sample period: These delays may be arbitrary. Another distinction with the previous problem setup is that now each actuator is endowed with a rather powerful computing module, which can be viewed as a local controller. The central controller remains in use. It collects data from the sensors and sends messages to the local controllers. We also suppose that the system disintegrates into multiple semi-independent subsystems. Each local controller serves its own subsystem, which do not interact. There also is an uncontrolled subsystem affecting all controlled ones.

We have in mind the situation where due to the limited bandwidths of the control channels, the local controllers cannot have access to the entire sensor data. This gives rise to the role of the central controller as a processor and compressor of these data. Ideally, this controller might generate the control for each subsystem. However its unawareness about the time that will be taken to transmit the control restricts its ability to achieve the best performance without an aid of the local controllers.

We show that this performance can be achieved via certain distribution of control functions between the central and local controllers. More precisely, we consider the performance best under the artificial assumption that all sensor data are resent from the central controller to the local ones. An outline of this distribution is as follows. Proceeding from the sensor data, the central controller forms a whole package of controls for any subsystem at each sampling time and sends these packages via the control channels. On arrival of such a package at a subsystem, its local controller chooses the proper member of the package, proceeding from its time stamp, and then corrects this member. The correction is generated recursively by the local controller. In doing so, it employs the information about the past actuator inputs for the corresponding subsystem but does not utilize the sensor data. As a result, the performance optimal in the idealized circumstances where the control channels bandwidths constraints are neglected is achieved without transmission the entire sensor data to the subsystems. The sizes of the control packages are determined explicitly. The crucial point is that under certain circumstances, these sizes may be far less than those in the case where all observations are actually transmitted to the subsystems. Anyhow these sizes must be compared with the actual bandwidths of the channels to decide whether the above scheme is acceptable. Furthermore, the influence of the package length on the transmission time should be also taken into account.

More specifically, we offer an explicit description of the optimal control strategy for the corresponding LQG optimal control. This strategy consumes minimum variance estimates of the current and several past system states, along with the states of an auxiliary random process. These estimates are produced by a filter similar in spirit to that from Subsect. 11.3.2 (starting on p. 375). However in view of the difference caused by the concern for the auxiliary process, we offer an independent description of the corresponding state estimator, as well as the proof of its optimality.

The main results of this chapter were originally presented in [105, 109].

The body of the chapter is organized as follows. Sections 12.2 and 12.4 present the first and second setups of the LQG optimal control problem, respectively. The corresponding optimal control strategies are described in Sects. 12.3 and 12.7, respectively. The strategy from Sect. 12.7 employs a minimum variance state estimator offered in Sect. 12.6, which is prefaced by an informal discussion of ways to distribute estimation functions over the central and local controllers in Sect. 12.5. The proofs of the results stated in Sects. 12.3, 12.6, and 12.7 are given in Sects. 12.8, 12.9, and 12.10, respectively.

12.2 The Problem of Linear-Quadratic Optimal Control via Asynchronous Communication Channels

12.2.1 Problem Statement

In this section, we proceed with studying discrete-time linear systems of the forms:

$$x(t+1) = A(t)x(t) + B(t)u(t) + \xi(t) \qquad t = 0, \dots, T-1; \qquad x(0) = a; \quad (12.2.1)$$

$$y_\nu(t) = C_\nu(t)x(t) + \chi_\nu(t) \qquad \nu = 1, \dots, l, \quad t = 0, \dots, T. \quad (12.2.2)$$

We recall that $x(t) \in \mathbb{R}^n$ is the state; $u(t) \in \mathbb{R}^m$ is the control; $\xi(t) \in \mathbb{R}^n$ is a process disturbance; $y_\nu(t) \in \mathbb{R}^{k_\nu}$ is a sensor output; and $\chi_\nu(t) \in \mathbb{R}^{k_\nu}$ is a noise. A particular vector y_ν may equal either the entire output of a sensor or only a part of it.

Measurements transmission. Like in Sect. 11.2, the observations y_ν are communicated from the sensors to the controller via noisy, lossy, and delayed *observation channels*. We do not specify a policy to deal with messages that simultaneously arrive at the controller via parallel observation channels. This policy may consist in either accepting all messages or choosing a selection of them in accordance with a certain algorithm. However, we assume that no signal is accepted twice. As a result, the controller consumes the observations $y_\nu(\theta)$ with $(\nu, \theta) \in S(t)$ at the current time t. Here set $S(t)$ of signals accepted at time t may be empty and satisfies (11.2.7):

$$S(t_1) \cap S(t_2) = \emptyset \quad \text{whenever} \quad t_1 \neq t_2,$$
$$\text{and} \quad (\nu, \theta) \in S(t) \Rightarrow \theta \in \{0, 1, \ldots, t\}. \quad (12.2.3)$$

Typically it consists of all pairs (ν, θ) for which the observation signal $y_\nu(\theta)$ produced at time θ arrives at the controller at time t.

Transmission of control signals. The system contains several actuators. So $u =$ **col** $\left(u_1, \ldots, u_q\right)$, where u_i is the input of the ith actuator. Unlike Sects. 11.5 and 11.6, this input is transmitted from the controller via a randomly delayed ith *control channel* (see Fig. 12.1). If several messages arrive out of order over this channel, accepted is the most updated (i.e., produced at the latest moment) of them. If no message arrives, the last accepted one is kept employed. Hence in (12.2.1),

$$u(t) = \textbf{col}\,[u_1(t), \ldots, u_q(t)], \qquad u_i(t) = v_i[t - \theta_i(t)] \in \mathbb{R}^{m_i}. \qquad (12.2.4)$$

Here $v_i(t)$ is the output currently emitted by the controller into the ith control channel. Furthermore, $\theta_i(t) \geq 0$ and

$$t_1 - \theta_i(t_1) \leq t_2 - \theta_i(t_2) \quad \text{whenever} \quad t_1 \leq t_2. \qquad (12.2.5)$$

(We assume that $u_i(t) := 0$ and so $\theta_i(t) := t + 1, v_i(-1) := 0$ in (12.2.4) if no message has arrived until t.)

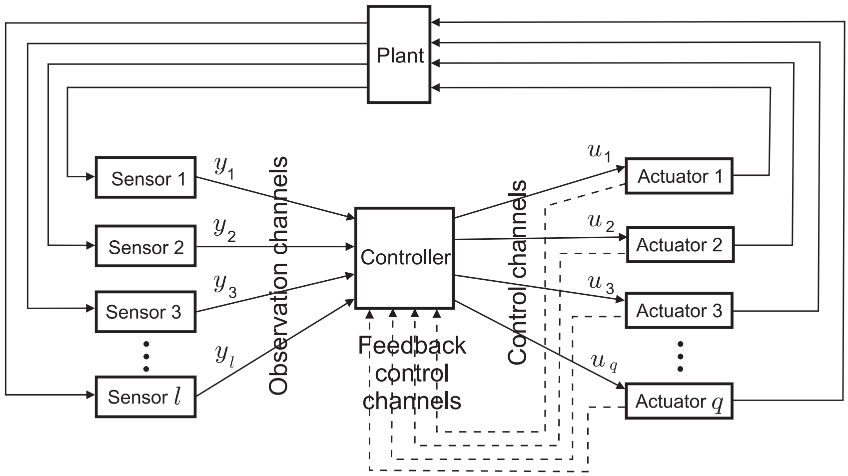


Fig. 12.1. Multiple sensor and actuator control system.

The problem is to minimize the quadratic cost functional

$$\mathcal{I}_T := \boldsymbol{E} \sum_{t=0}^{T-1} \mathcal{G}[t, x(t+1), v(t)],$$
$$\text{where} \quad \mathcal{G}[t, x, v] := x^{\mathrm{T}} Q(t+1) x + v^{\mathrm{T}} \Gamma(t) v. \quad (12.2.6)$$

Here $Q(t+1) \geq 0$ and $\Gamma(t) > 0$ are symmetric $n \times n$ and $m \times m$ matrices, respectively, and

$$v := \textbf{col}\left(v_1, \ldots, v_q\right). \tag{12.2.7}$$

The time taken by the communication channel to transfer a message becomes known a posteriori due to the time stamp. The producer of any measurement received is supposed to be recognizable. If the sensor output was partitioned into several portions, there is a way to recognize which part of the entire output is brought by the message arrived. In brief, this means that the set $S(t)$ and the delay $\theta_i(t)$ become known to the controller and at the ith actuator site, respectively, at time t.

Feedback control channels. The information about $\theta_i(t)$ is sent back to the controller via *feedback control channels*. This transmission may be with delays and not continuous. As a result, the controller becomes aware of $\theta_i(s)$ with $(i,s) \in \mathcal{T}(t)$ at time t, where

$$\mathcal{T}(t) \subset \{(i,s) : i = 1, \ldots, q, s = 0, \ldots, t\}.$$

The controller may also have access to a side information about the past states of the communication medium. This information received at the time t is represented by an element $\mathfrak{p}(t)$ of a finite set $\mathfrak{P}$. The natural class of control strategies is as follows:

$$v(t) = \textbf{col}\,[v_1(t), \ldots, v_q(t)] = \mathcal{U}\left[t, \mathfrak{Y}(t), \mathfrak{S}(t), \eta(0), \ldots, \eta(t)\right], \quad \text{where}$$
$$\mathfrak{Y}(t) := [Y(0), Y(1), \ldots, Y(t)], \qquad Y(\tau) := \Big\{y_\nu[\theta]\Big\}_{(\nu,\theta)\in S(\tau)};$$
$$\mathfrak{S}(t) := [S(0), S(1), \ldots, S(t)];$$
$$\eta(\tau) := \left[\mathcal{T}(\tau), \{\theta_i(s)\}_{(i,s)\in\mathcal{T}(\tau)}, \mathfrak{p}(\tau)\right], \tag{12.2.8}$$

and $\mathcal{U}(\cdot)$ is a deterministic function. (As before, $S(t) = \emptyset \Rightarrow Y(t) := 0 \in \mathbb{R}$.)

12.2.2 Assumptions

Now we state the assumptions under which the problem posed in the previous subsection will be studied. The first of them is a mere replica of Assumption 11.2.1 (on p. 373). We adduce it here once more for the convenience of the reader.

Assumption 12.2.1. *The random vectors* $a, \xi(t)$*, and* $\chi_\nu(t), \nu = 1, \ldots, l$ *from* (12.2.1) *and* (12.2.2) *are Gaussian and independent with* $\boldsymbol{E}\xi(t) = 0$ *and* $\boldsymbol{E}\chi_\nu(t) = 0$. *The mean* $\boldsymbol{E}a$ *and the correlation matrices*

$$R_{aa} := \boldsymbol{E}[a - \boldsymbol{E}a][a - \boldsymbol{E}a]^{\mathrm{T}}, \qquad R_{\xi\xi}(t) := \boldsymbol{E}\xi(t)\xi(t)^{\mathrm{T}},$$
$$R^{\nu}_{\chi\chi}(t) := \boldsymbol{E}\chi_\nu(t)\chi_\nu(t)^{\mathrm{T}}, \qquad \nu = 1, \ldots, l \tag{12.2.9}$$

are known. So are the matrices $A(t), B(t)$*, and* $C_\nu(t)$*.*

The following assumption generalizes the first part of Assumption 11.2.2 and, in fact, specifies and expends Assumption 11.2.4 (on p. 374).

Assumption 12.2.2. *The transmission delays in all communication channels as well as the side information accessible by the controller are independent of the plant. The measurement transmission delays are independent of those in feedforward and feedback control channels and the side information.*

In other words, the random sets $\{S(\theta)\}$ and quantities $\{\eta(t)\}$ from (12.2.8) *are independent of* a, $\{\xi(\tau)\}$, *and* $\{\chi_\nu(\tau)\}$. *The sets* $\{S(\theta)\}$ *are independent of* $\{\eta(t)\}$.

The last claim implies that the side information represented by $\mathfrak{p}(t)$ characterizes only the control channels and does not concern the observation ones.

The next assumption is similar to the second part of Assumption 11.2.2.

Assumption 12.2.3. *The delays in the communication channels are bounded by known constants:* $t-\theta \leq \sigma$ *if* $(\nu, \theta) \in S(t)$ *and* $\theta_i(t) \leq \sigma_*$ *for all* t *and* $i = 1, \dots, q$.

Assumption 12.2.4. *The delays in the feedback control channels do not exceed the sample period, and these channels do not lose signals:*

$$(i, t) \in \mathfrak{T}(t) \cup \mathfrak{T}(t+1) \quad \text{for all} \quad t \quad \text{and} \quad i = 1, \dots, q.$$

Thus the currently received information about the control channels can be represented by the tuple

$$\varkappa(t) := \left[I(t), \theta^+(t), \theta^-(t), \mathfrak{p}(t)\right]. \tag{12.2.10}$$

Here

$$I(t) := \{i = 1, \dots, q : (i, t) \in \mathfrak{T}(t)\}$$

and the vectors $\theta^\pm[t] := \mathbf{col}\left(\theta_1^\pm[t], \dots, \theta_q^\pm[t]\right)$ are given by

$$\theta_i^+(t) := \begin{cases} \theta_i(t) & \text{if} \quad i \in I(t) \\ 0 & \text{otherwise} \end{cases};$$

$$\theta_i^-(t) := \begin{cases} \theta_i(t-1) & \text{if} \quad i \notin I(t-1) \\ 0 & \text{otherwise} \end{cases}. \tag{12.2.11}$$

Notation 12.2.5. *The symbol* Υ *stands for the set of the values that can be taken by the tuple* (12.2.10).

Remark 12.2.6. We pick $\mathfrak{p}_* \in \mathfrak{P}$ and put $\mathfrak{p}(s) := \mathfrak{p}_*, I(s) := \emptyset, \theta_i(s) := 0$ for $s < 0$. Then the second formula from (12.2.11) is of sense for $t = 0$.

Remark 12.2.7. In terms of the tuples (12.2.10), the strategies (12.2.8) take the form

$$v(t) = \mathcal{V}\left[t, \mathfrak{Y}(t), \mathfrak{S}(t), \varkappa(0), \dots, \varkappa(t)\right]. \tag{12.2.12}$$

The last assumption to follow means that the control channels are systems with a finite aftereffect.

Assumption 12.2.8. *A known constant* $\overline{\sigma} = 0, 1, \ldots$ *exists such that the conditional distribution of the quantities* (12.2.10) *satisfies the following relation:*

$$\boldsymbol{P}\left[\varkappa(t+1) = \varkappa \middle| \varkappa(t) = \varkappa_t, \ldots, \varkappa(0) = \varkappa_0\right] \\ = \boldsymbol{P}\left[\varkappa(t+1) = \varkappa \middle| \varkappa(t) = \varkappa_t, \ldots, \varkappa(t-\widehat{\sigma}) = \varkappa_{t-\widehat{\sigma}}\right], \\ \text{where} \qquad \widehat{\sigma} := \min\{\overline{\sigma}, t\}, \quad (12.2.13)$$

for any $\varkappa, \varkappa_0, \ldots, \varkappa_t \in \Upsilon, t = 0, 1, \ldots$. *This distribution is known in advance.*

Remark 12.2.9. For $t \leq \overline{\sigma}$, the right-hand side of (12.2.13) is necessarily identical to its left-hand side.

Notation 12.2.10. *For any* $\varkappa = [I, \theta^+, \theta^-, \mathfrak{p}] \in \Upsilon$, *we set* $\theta^{\pm}(\varkappa) := \theta^{\pm}$.

Remark 12.2.11. It follows from (12.2.11) that the probability (12.2.13) equals 0 whenever the ith coordinate of both $\theta^-(\varkappa)$ and $\theta^+(\varkappa_t)$ is nonzero for some i.

Note that the side information $\mathfrak{p}(t)$ from (12.2.8) may, for example, concern the transmission times of the control signals overtaken by other ones in the course of transfer via control channels. Since any actuator employs the most updated control, this information is not incorporated into the sequences $\theta_i(0), \theta_i(1), \ldots$. However, it may be transmitted via the feedback control channels to the controller and be useful for statistical prognosis of future delays in the control channels.

12.3 Optimal Control Strategy

In this section, we present the solution of the problem posed in Subsect. 12.2.1. The main result of this section states that the strategy optimal in the class (12.2.12) exists and can be implemented by a feedback of the form:

$$v(t) = -L[t, \varkappa(t), \ldots, \varkappa(t-\overline{\sigma})]\widehat{x}(t|t) \\ - \sum_{j=1}^{\sigma_*} \mathfrak{L}_j[t, \varkappa(t), \ldots, \varkappa(t-\overline{\sigma})]v(t-j). \quad (12.3.1)$$

Here v is the controller output (12.2.7), σ_* and $\overline{\sigma}$ are the constants from Assumptions 12.2.3 and 12.2.8, respectively, $v(s) := 0 \; \forall s < 0$, and the tuple $\varkappa(t)$ is defined by (12.2.10). Furthermore, $\widehat{x}(t|t)$ is the estimate of the current state $x(t)$ generated by the estimator from Subsect. 11.3.2 (starting on p. 375). This estimator is implementable at the controller site since Assumption 12.2.4 implies that the controller is aware of the controls $u(\theta)$ with $\theta \leq t-1$ at the current time t. (In other words, Assumption 11.2.3 on p. 374 holds.)

Remark 12.3.1. Under the circumstances, the conclusions of Theorem 11.3.3 (on p. 377) hold. In particular, the estimator from Subsect. 11.3.2 produces minimum variance estimates $\widehat{x}(j|t)$ of $x(j)$ with $\max\{t-\sigma, 0\} \leq j \leq t$ based on $\mathfrak{Y}(t), \mathfrak{S}(t)$, and $u(0), \ldots, u(t-1)$.

The proofs of the claims stated in this section will be given in Sect. 12.8.

In (12.3.1), the $m \times n$ matrix $L(t, \overline{\varkappa})$ and $m \times m$ matrices $\mathfrak{L}_1(t, \overline{\varkappa}), \ldots, \mathfrak{L}_{\sigma_*}(t, \overline{\varkappa})$ are the functions of $t = 0, \ldots, T-1$ and the parameter

$$\overline{\varkappa} = (\varkappa_0, \ldots, \varkappa_{\overline{\sigma}}) \in \Upsilon^{\overline{\sigma}+1}. \quad (12.3.2)$$

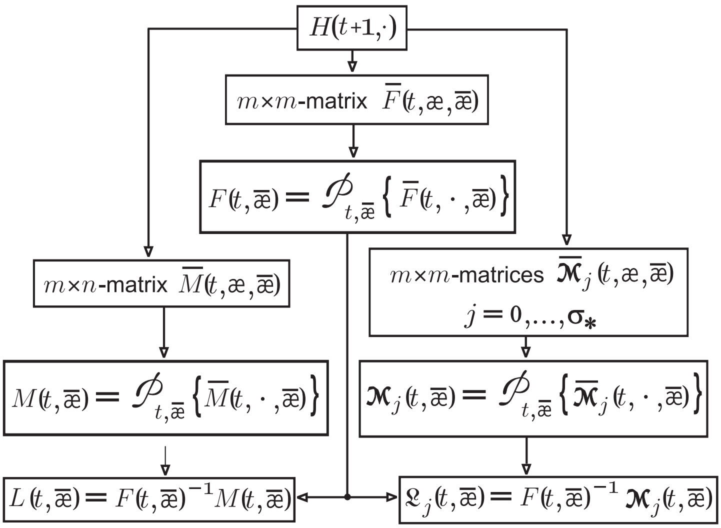
We recall that Υ is the set of the values that can be taken by the tuple (12.2.10). The matrix-functions corresponding to the optimal strategy can be calculated in advance. To this end, a certain $\overline{\varkappa}$-parametric system of difference Riccati equations should be solved. This system is with respect to the following matrices $H(t, \overline{\varkappa}), t = 0, \ldots, T$:

$$H(t, \overline{\varkappa}) = \begin{pmatrix} H_{xx}(t, \overline{\varkappa}) & H_{x1}(t, \overline{\varkappa}) & \ldots & H_{x\sigma_*}(t, \overline{\varkappa}) \\ H_{1x}(t, \overline{\varkappa}) & H_{11}(t, \overline{\varkappa}) & \ldots & H_{1\sigma_*}(t, \overline{\varkappa}) \\ H_{2x}(t, \overline{\varkappa}) & H_{21}(t, \overline{\varkappa}) & \ldots & H_{2\sigma_*}(t, \overline{\varkappa}) \\ \vdots & \vdots & \vdots & \vdots \\ H_{\sigma_* x}(t, \overline{\varkappa}) & H_{\sigma_* 1}(t, \overline{\varkappa}) & \ldots & H_{\sigma_*\sigma_*}(t, \overline{\varkappa}) \end{pmatrix}. \quad (12.3.3)$$

Here the sizes of the blocks are as follows:

$H_{xx}(t, \overline{\varkappa})$	$H_{xj}(t, \overline{\varkappa})$	$H_{jx}(t, \overline{\varkappa})$	$H_{ij}(t, \overline{\varkappa})$
$n \times n$	$n \times m$	$m \times n$	$m \times m$

Specifically, computation of the gain matrices from (12.3.1) follows the diagram



(12.3.4)

Here for any matrix function $D(\varkappa)$ of $\varkappa \in \Upsilon$ and the tuple (12.3.2),

$$\mathcal{P}_{t,\overline{\varkappa}}\{D(\cdot)\} := \sum_{\varkappa \in \Upsilon} P\left[\varkappa(t+1) = \varkappa \middle| \varkappa(t) = \varkappa_0, \ldots, \varkappa(t-\overline{\sigma}) = \varkappa_{\overline{\sigma}}\right] D(\varkappa), \quad (12.3.5)$$

where $P[\varkappa(t+1)=\varkappa|\ldots]$ is the conditional distribution from Assumption 12.2.8.

The other calculations depicted in (12.3.4) by arrows employ the matrix $\widehat{B}_i(t)(i = 1,\ldots,q)$ expressing the influence of the ith actuator on the dynamics:

$$\begin{aligned}&\widehat{B}_i(t) := B(t)J_i,\\ &\qquad J_i := \textbf{col}\,[0_{m_1\times m_i},\ldots,0_{m_{i-1}\times m_i}, I_{m_i}, 0_{m_{i+1}\times m_i},\ldots,0_{m_q\times m_i}],\end{aligned} \quad (12.3.6)$$

where $m_1,\ldots,m_q$ are the respective dimensions of the actuators inputs from (12.2.4) and I_{m_i} occupies the ith position. We also put for $\varkappa\in\Upsilon$ and the tuple (12.3.2),

$$\varkappa\oplus\overline{\varkappa} := (\varkappa,\varkappa_0,\ldots,\varkappa_{\overline{\sigma}-1}); \quad (12.3.7)$$

$$\mathcal{B}_\theta^j(t) := \sum_{i:\theta_i=j}\widehat{B}_i(t) \qquad \theta = \textbf{col}\,[\theta_1,\ldots,\theta_q]\in\mathbb{R}^q, j=0,\ldots,\sigma_*; \quad (12.3.8)$$

$$H_{xp}(t,\overline{\varkappa}) := 0_{n\times m},\quad H_{px}(t,\overline{\varkappa}) := 0_{m\times n},\quad H_{ip}(t,\overline{\varkappa}) := H_{pi}(t,\overline{\varkappa}) := 0_{m\times m}$$

for $p := \sigma_*+1$. We also invoke Notation 12.2.10 and set

$$\theta(\varkappa,\overline{\varkappa}) := \theta^-(\varkappa)+\theta^+(\varkappa_0), \quad (12.3.9)$$

where $\varkappa_0$ is the first component of the tuple (12.3.2). Then in (12.3.4),

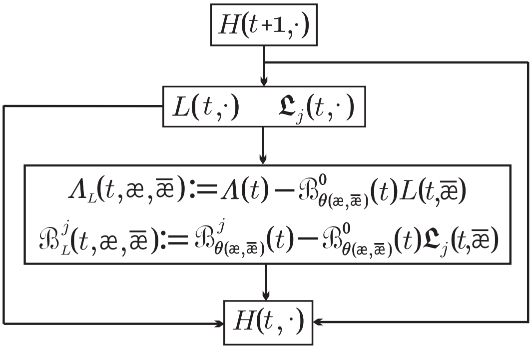
$$\begin{aligned}\overline{F}(t,\varkappa,\overline{\varkappa}) &:= \Gamma(t) + H_{11}(t+1,\varkappa\oplus\overline{\varkappa})\\ &\quad + \mathcal{B}^0_{\theta(\varkappa,\overline{\varkappa})}(t)^{\mathrm{T}}\,[H_{xx}(t+1,\varkappa\oplus\overline{\varkappa})+Q(t+1)]\,\mathcal{B}^0_{\theta(\varkappa,\overline{\varkappa})}(t)\\ &\quad + H_{1x}(t+1,\varkappa\oplus\overline{\varkappa})\mathcal{B}^0_{\theta(\varkappa,\overline{\varkappa})}(t) + \mathcal{B}^0_{\theta(\varkappa,\overline{\varkappa})}(t)^{\mathrm{T}}H_{x1}(t+1,\varkappa\oplus\overline{\varkappa});\end{aligned} \quad (12.3.10)$$

$$\begin{aligned}&\overline{M}(t,\varkappa,\overline{\varkappa}) := \mathcal{N}(t+1,\varkappa,\overline{\varkappa})A(t), \quad \text{where}\\ &\qquad \mathcal{N}(t+1,\varkappa,\overline{\varkappa}) := \mathcal{B}^0_{\theta(\varkappa,\overline{\varkappa})}(t)^{\mathrm{T}}\,[H_{xx}(t+1,\varkappa\oplus\overline{\varkappa})+Q(t+1)]\\ &\qquad\qquad + H_{1x}(t+1,\varkappa\oplus\overline{\varkappa});\end{aligned} \quad (12.3.11)$$

$$\begin{aligned}\overline{\mathcal{M}}_j(t,\varkappa,\overline{\varkappa}) &:= \mathcal{N}(t+1,\varkappa,\overline{\varkappa})\mathcal{B}^j_{\theta(\varkappa,\overline{\varkappa})}(t) + H_{1(j+1)}(t+1,\varkappa\oplus\overline{\varkappa})\\ &\quad + \mathcal{B}^0_{\theta(\varkappa,\overline{\varkappa})}(t)^{\mathrm{T}}H_{x(j+1)}(t+1,\varkappa\oplus\overline{\varkappa}).\end{aligned} \quad (12.3.12)$$

Here $Q(t),\Gamma(t)$ are the coefficients of the cost functional (12.2.6).

The matrices $H(T,\overline{\varkappa}), H(T-1,\overline{\varkappa}),\ldots,H(0,\overline{\varkappa})$ are generated recursively in accordance with the scheme



(12.3.13)

More precisely,

$$H_{ij}(t,\overline{\varkappa}) = \mathcal{P}_{t,\overline{\varkappa}}\Big\{\overline{H}_{ij}(t,\cdot,\overline{\varkappa}))\Big\} \qquad \forall i,j = x,1,\ldots,\sigma_*, \tag{12.3.14}$$

where the operator $\mathcal{P}_{t,\overline{\varkappa}}$ is defined by (12.3.5) and

$$\begin{aligned}\overline{H}_{xx}(t,\varkappa,\overline{\varkappa}) = A_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}\left[H_{xx}(t+1,\varkappa\oplus\overline{\varkappa}) + Q(t+1)\right]A_L(t,\varkappa,\overline{\varkappa})\\ - A_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}H_{x1}(t+1,\varkappa\oplus\overline{\varkappa})L(t,\overline{\varkappa})\\ - L(t,\overline{\varkappa})^{\mathrm{T}}H_{1x}(t+1,\varkappa\oplus\overline{\varkappa})A_L(t,\varkappa,\overline{\varkappa})\\ + L(t,\overline{\varkappa})^{\mathrm{T}}\left[H_{11}(t+1,\varkappa\oplus\overline{\varkappa}) + \Gamma(t)\right]L(t,\overline{\varkappa});\end{aligned} \tag{12.3.15}$$

$$\begin{aligned}\overline{H}_{xj}(t,\varkappa,\overline{\varkappa}) = A_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}\left[H_{xx}(t+1,\varkappa\oplus\overline{\varkappa}) + Q(t+1)\right]\mathcal{B}^j_L(t,\varkappa,\overline{\varkappa})\\ - A_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}H_{x1}(t+1,\varkappa\oplus\overline{\varkappa})\mathcal{L}_j(t,\overline{\varkappa}) - L(t,\overline{\varkappa})^{\mathrm{T}}H_{1x}(t+1,\varkappa\oplus\overline{\varkappa})\mathcal{B}^j_L(t,\varkappa,\overline{\varkappa})\\ + L(t,\overline{\varkappa})^{\mathrm{T}}\left[H_{11}(t+1,\varkappa\oplus\overline{\varkappa}) + \Gamma(t)\right]\mathcal{L}_j(t,\overline{\varkappa}) + A_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}H_{x(j+1)}(t+1,\varkappa,\overline{\varkappa})\\ - L(t,\overline{\varkappa})^{\mathrm{T}}H_{1(j+1)}(t+1,\varkappa\oplus\overline{\varkappa};\\ \overline{H}_{jx}(t,\varkappa,\overline{\varkappa}) = \overline{H}_{xj}(t,\varkappa,\overline{\varkappa})^{\mathrm{T}};\end{aligned} \tag{12.3.16}$$

$$\begin{aligned}\overline{H}_{ij}(t,\varkappa,\overline{\varkappa}) = \mathcal{B}^i_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}\left[H_{xx}(t+1,\varkappa\oplus\overline{\varkappa}) + Q(t+1)\right]\mathcal{B}^j_L(t,\varkappa,\overline{\varkappa})\\ - \mathcal{B}^i_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}H_{x1}(t+1,\varkappa\oplus\overline{\varkappa})\mathcal{L}_j(t,\overline{\varkappa}) - \mathcal{L}_i(t,\overline{\varkappa})^{\mathrm{T}}H_{1x}(t+1,\varkappa\oplus\overline{\varkappa})\mathcal{B}^j_L(t,\varkappa,\overline{\varkappa})\\ + \mathcal{L}_i(t,\overline{\varkappa})^{\mathrm{T}}\left[H_{11}(t+1,\varkappa\oplus\overline{\varkappa}) + \Gamma(t)\right]\mathcal{L}_j(t,\overline{\varkappa}) + \mathcal{B}^i_L(t,\varkappa,\overline{\varkappa})^{\mathrm{T}}H_{x(j+1)}(t+1,\varkappa\oplus\overline{\varkappa})\\ + H_{(i+1)x}(t+1,\varkappa\oplus\overline{\varkappa})\mathcal{B}^j_L(t,\varkappa,\overline{\varkappa}) - \mathcal{L}_i(t,\overline{\varkappa})^{\mathrm{T}}H_{1(j+1)}(t+1,\varkappa\oplus\overline{\varkappa})\\ - H_{(i+1)1}(t+1,\varkappa\oplus\overline{\varkappa})\mathcal{L}_j(t,\overline{\varkappa}) + H_{(i+1)(j+1)}(t+1,\varkappa\oplus\overline{\varkappa}).\end{aligned} \tag{12.3.17}$$

Here $i,j = 1,\ldots,\sigma_*$. The recursion is initialized by putting $H(T,\overline{\varkappa}) := 0$.

Remark 12.3.2. The matrix $F(t,\overline{\varkappa})$ given by (12.3.4) and (12.3.10) is positive-definite for all t and $\overline{\varkappa}$. Hence the inverse matrix $F(t,\overline{\varkappa})^{-1}$ employed in (12.3.4) does exist.

Now we are in a position to formulate the main result of the section.

Theorem 12.3.3. *Suppose that Assumptions* 12.2.1–12.2.4*, and* 12.2.8 *hold. Then a strategy optimal in the class* (12.2.8) *exists. This strategy is given by formula* (12.3.1)*.*

12.4 Problem of Optimal Control of Multiple Semi-Independent Subsystems

12.4.1 Problem Statement

In this section, we continue to treat the problem of minimizing the quadratic cost functional (12.2.6) for the linear plant (12.2.1), (12.2.2). However, now we deal with a different situation, as compared with the previous two sections of this chapter. The major distinctions are as follows:

(d1) We do not assume any longer that the statistics of the data delays in the control channels is known in advance. Moreover, now there are no assumptions about the statistical properties of these channels.
(d2) We do not assume any longer that the delays in the feedback control channels are less or equal to the sample period. Now arbitrary delays are admitted.
(d3) We suppose that the system consists of multiple subsystems, either controlled or uncontrolled. Thus in (12.2.1),

$$x = \textbf{col}\,(x_0, x_1, \dots, x_q), \quad x_i \in \mathbb{R}^{n_i},$$
$$u = \textbf{col}\,(u_1, \dots, u_q), \quad u_i \in \mathbb{R}^{m_i}, \quad (12.4.1)$$

where x_0 corresponds to the uncontrolled part, and x_i, u_i are the state and control for the ith subsystem. Every controlled subsystem does not influence any other subsystem.
(d4) As before, there is a *central controller* receiving data over the observation and feedback control channels and emitting messages into the (feedforward) control channels (see Fig. 12.2). However, now every controlled subsystem is equipped with its own *local controller*. Its output u_i acts upon the subsystem immediately.

The claims (d1) and (d2) mean that Assumptions 12.2.8 and 12.2.4 are dropped and relaxed, respectively. So with respect to the points concerned in these assumptions, now a more general situation is considered. At the same time, the situation is more special in another respect owing to (d3).

The properties (d1) and (d2) make a prognosis of the future delays in the control channels impossible and thus the side information $\mathfrak{p}(t)$ from (12.2.8) meaningless. So we do not assume any longer that it is available.

It should be also remarked that the last claim from (d3) concerns only the dynamics and is not extended on the observations. In particular, the sensors may produce characteristics of the relative motion of the subsystems.

Other features of the data transfer remain unchanged. In particular, the data received by the central controller at time t are represented by $S(t)$ and

$$Y(t) = \Big\{ y_\nu[\theta] \Big\}_{(\nu,\theta)\in S(t)}, \quad (12.4.2)$$

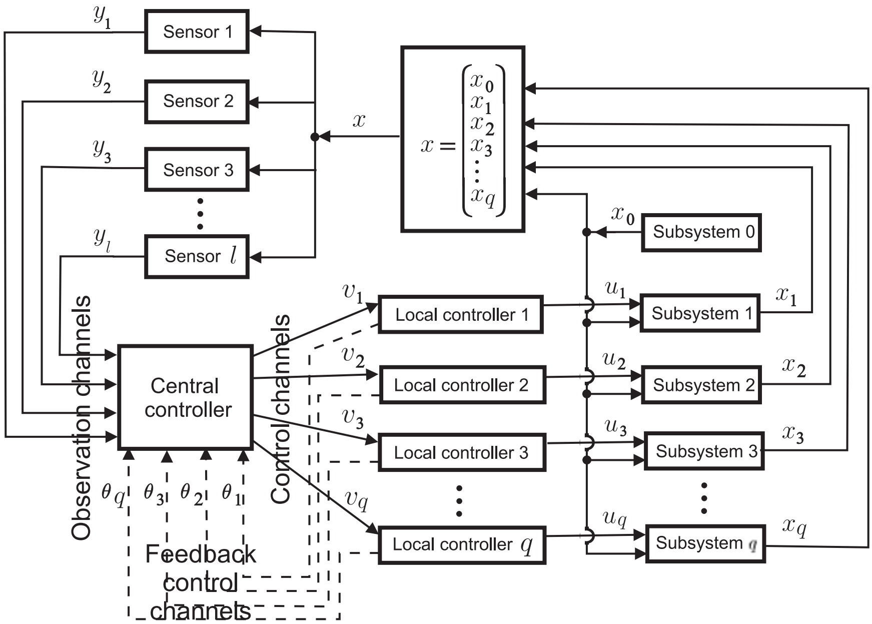


Fig. 12.2. Platoon of semi-independent subsystems.

where the set $S(t)$ of signals accepted at time t satisfies (12.2.3) (on p. 408). The messages v_i dispatched by the central controller via the ith control channel incur random delays. So at time t the ith local controller receives

$$u_i^r(t) = v_i[t - \theta_i(t)], \tag{12.4.3}$$

where $\theta_i(t) \geq 0$. If several messages arrive out of order, accepted is the most updated of them, as before. If no message arrives, the last accepted one is kept employed. So (12.2.5) still holds. (Now we put $u_i^r(t) := 0$ and so $\theta_i(t) := t+1, v_i(-1) := 0$ in (12.4.3) if no message has arrived until t. This choice of control will be altered since it is not optimal.) Thanks to the time stamps, the delay $\theta_i(t)$ becomes known to the ith local controller at time t. This information is sent back to the central controller via the delayed feedback control channels. Thus at the current time t, this controller gets aware of some delays $\theta_i(s)$ with $s \leq t$. The corresponding pairs (i, s) constitute a certain set, which is denoted by $\mathcal{T}(t)$.

To complete the problem statement, information constraints should be specified. If the local controllers do not take part in control generating, the following class of control strategies is natural:

$$v_i(t) = \mathcal{V}_i\left[t, \mathfrak{Y}(t), \mathfrak{S}(t), \mathcal{T}(0), \dots, \mathcal{T}(t), \Theta(0), \dots, \Theta(t)\right] \quad i = 1, \dots, q,$$
$$\text{where} \quad \Theta(t) := \left\{\theta_i(s)\right\}_{(i,s)\in\mathcal{T}(t)}, \tag{12.4.4}$$

$\mathfrak{Y}(t)$ and $\mathfrak{S}(t)$ are defined in (12.2.8) and $\mathcal{V}_i(\cdot)$ are deterministic functions. But we shall in fact consider the opposite extreme in the problem statement. It results from ignoring the control channels bandwidths constraints and assuming that all sensor data can be resent from the central controller to the local ones. Furthermore, it is supposed that there is an instantaneous exchange of information about the delays $\theta_i(t)$ in the control channels between the local controllers. In such artificial circumstances, these controllers acquire crucial advantage over the central one by having access to more information. (The central controller is not aware of the times that will be taken to transmit its outputs to the subsystems.) So it is beneficial to commission the local controllers to generate controls. We also suppose here that these controllers are able to manage it, having enough computational power to process the entire bulk of the sensor data. Then the delays in the control channels become in effect a part of the observation delays, and the following class of strategies is natural:

$$u_i(t) = \mathfrak{U}_i \left[t, \mathfrak{Y}(\tau), \mathfrak{S}(\tau), \left(\theta_j[s]\right)_{j=1}^{q} {}_{s=0}^{t}\right] \quad \tau := t - \theta_i(t),\ i \in [1 : q]. \tag{12.4.5}$$

As will be shown, the corresponding optimal strategy can be implemented in such a way that the bandwidth constraints are not violated (at least in some cases), and the above imaginary exchange of information about $\theta_i(t)$, along with other artificial assumptions, is not employed. Thus it can be implemented in the real circumstances.

In conclusion, we recall that if no message has arrived via the ith control channel until t, then $\tau = -1$ in (12.4.5) by the definition of $\theta_i(t)$. In view of this, we put $\mathfrak{S}(-1) := \emptyset$, $\mathfrak{Y}(-1) := 0 \in \mathbb{R}$ in order that (12.4.5) be true for all t.

12.4.2 Assumptions

Now we state the assumptions adopted to deal with the problem posed in the previous subsection. In particular, they more formally specify the properties (d1)–(d4) (on p. 415).

We suppose that Assumption 12.2.1 (on p. 409) is valid. So is Assumption 12.2.2 (on p. 410) with the references to the "side information" dropped (since now it is void).[1] The following assumption extends Assumptions 12.2.3 and 12.2.4.

Assumption 12.4.1. *The times taken by the communication channels to transmit messages are bounded from above by a known constant σ for any observation channel and σ_i for the ith control channel. A constant $\sigma_f \geq 1$ exists such that $(i, \theta) \in \mathfrak{T}(t)$ whenever $\theta \leq t - \sigma_f$ and $i = 1, \ldots, q$. This constant is known a priori.*

Due to the last claim, at the time t the central controller is aware of the local controls $u_i(s)$ with $s \leq t - \sigma_f$, e.g., by duplicating the calculations in accordance with (12.4.5). Furthermore

$$t - s \leq \sigma \quad \text{whenever} \quad (\nu, s) \in S(t). \tag{12.4.6}$$

Assumption 12.4.2. *The delays in the feedback control channels do not exceed those in the observation ones:*

[1]The variable $\mathfrak{p}(\tau)$ is also dropped in the last formula from (12.2.8) (on p. 409).

$$t - s \geq \sigma_f - 1 \quad \text{whenever} \quad (\nu, s) \in S(t).$$

At least one of any μ_i messages dispatched successively within a time interval of duration $\mu_i - 1$ via the ith control channel is not lost. Here $\mu_i \geq 1$ is a known constant for any i.

Assumption 12.4.3. *The nominal model of the plant*

$$x(t+1) = A(t)x + B(t)u$$

disintegrates into an uncontrolled and q independent controlled parts affected by the uncontrolled one:

$$x_0(t+1) = A_0(t)x_0(t),$$
$$x_i(t+1) = A_i(t)x_i(t) + \overline{A}_i(t)x_0(t) + B_i(t)u_i(t). \quad (12.4.7)$$

Neither the observations nor the noises disintegrate in a similar way. In particular, some characteristics of the relative motion of the subsystems may be observed.

Assumption 12.4.4. *The performance index* (12.2.6) *has a special structure:*

$$\mathcal{G}(t,x,u) = \sum_{i=1}^{q} \mathcal{G}_i(t,x,u), \quad \text{where}$$
$$\mathcal{G}_i(t,x,u) = \left[x_i^{\mathrm{T}} Q_i(t+1)x_i + 2x_0^{\mathrm{T}}\overline{Q}_i(t+1)x_i + u_i^{\mathrm{T}}\Gamma_i(t)u_i\right]. \quad (12.4.8)$$

Here $Q_i(t) \geq 0, \overline{Q}_i(t)$, $\Gamma_i(t) > 0$ are matrices of matching dimensions with $Q_i(t), \Gamma_i(t)$ symmetric.

12.5 Preliminary Discussion

As will be shown, the optimal control is determined, in harmony with Theorem C.5 (on p. 513), on the basis of the minimum variance estimate of the current state $x(t)$. In accordance with (12.4.5), the estimate is based on $\mathfrak{Y}(\tau), \mathfrak{S}(\tau), \tau = t - \theta_i(t)$, and $(\theta_j[s])_{j=0}^{q}{}_{s=0}^{t}$. This section offers an informal discussion of ways to distribute estimation functions over the central and local controllers.

The central controller receives the sensor data but is not aware of all controls (and delays θ_i) affecting the current state. On contrary, the local controllers are aware of these controls but have no direct access to the sensor data. So it seems reasonable to split the dynamics into "control-induced" part given by

$$x^c(t+1) = A(t)x^c(t) + B(t)u(t), \qquad x^c(0) = 0$$

and the remainder $x^u(t) := x(t) - x^c(t)$. Assumption 12.4.3 implies that for the ith subsystem, the first part of the dynamics obeys the equations

$$x_i^c(t+1) = A_i(t)x_i^c(t) + B_i(t)u_i(t), \qquad x_i^c(0) = 0 \tag{12.5.1}$$

and can be computed by the ith local controller. Hence the role of the central one can be confined to estimating $x^u(t)$ given, in particular, $\mathfrak{Y}(\tau), \mathfrak{S}(\tau)$, where $\tau = t - \theta_i(t)$.

To reach the ith subsystem at time t, the estimate must be dispatched at time τ. Thus at the moment τ, the central controller should produce a prognosis of the state $x^u(t)$. Unfortunately the time t is unknown at this moment. A way to cope with this trouble is to estimate the states for all $t = \tau, \ldots, \tau + \overline{\sigma}$, where $\overline{\sigma}$ is a known upper bound for θ_i, and to send all estimates to the subsystems, where the proper one can be chosen on the basis of the time stamp. The set of the estimates must be extended still further since the estimator (Kalman filter) compares currently arrived observations with their estimates. Now these observations $y_\nu(s), (\nu, s) \in S(\tau)$ concern time instants s in the past. This and (12.4.6) imply that the estimates of the states at $t = \tau - \sigma, \ldots, \tau - 1$ should be also produced.

The above scheme has at least two drawbacks. Firstly, the Kalman filter requires the estimates of $x(\theta)$ (not $x^u(\theta)$). Secondly, the local controllers are endowed with infinite dynamics, which may cause instability. Indeed, let, e.g., $\xi(t) \equiv 0$ in (12.2.1) and $\overline{A}_i(t) = 0$ in (12.4.7). Then the closed-loop ith subsystem is described by equations (12.4.5), (12.5.1), and

$$x_i(t+1) = A_i(t)x_i(t) + B_i(t)u_i(t).$$

This subsystem may be stable only if the uncontrolled ith subsystem is stable since $\delta x_i(t+1) = A_i(t)\delta x_i(t)$ for $\delta x_i(t) := x_i(t) - x_i^c(t)$.

A way to get rid of the above infinite dynamics is to commission the central controller to estimate the auxiliary process $\omega(t,\tau)$ that follows the uncontrolled dynamics only since some moment $\sigma_-(\tau) \le \tau$, i.e., $u(t) := 0$ for $t \ge \sigma_-(\tau)$ and is identical to $x(t)$ until it

$$\omega(t,\tau) := x(t) - \sum_{j=\sigma_-[\tau]}^{t-1} \prod_{\nu=j+1}^{t-1} A(\nu)B(j)u(j). \tag{12.5.2}$$

The above moment must be chosen so that

$$\sup_\tau [\tau - \sigma_-(\tau)] < \infty.$$

The corresponding estimates $\widehat{\omega}(t|\tau)$ can actually be produced if the process $\omega(t,\tau)$ is not affected by the controls unknown to the central controller. By the remark following Assumption 12.4.1, this is true if

$$\sigma_-(\tau) \le \max\{\tau - \sigma_f + 1, 0\}.$$

So the central controller sends the following package of the estimates to the ith subsystem:

$$\textbf{col}\left(\widehat{\omega}_i[0], \ldots, \widehat{\omega}_i[\overline{\sigma}]\right), \qquad \widehat{\omega}_i[j] := \widehat{\omega}_i[\tau + j|\tau]. \tag{12.5.3}$$

On its arrival at a time $t \geq \tau$, the estimate $\widehat{x}_i(t)$ of $x_i(t)$ can be computed by the ith local controller in accordance with (12.5.2), where $t = \tau + \theta_i(t)$, and Assumption 12.4.3

$$\widehat{x}_i(t) = \widehat{\omega}_i[\theta_i(t)] + \sum_{j=\sigma_-[t-\theta_i(t)]}^{t-1} \prod_{\nu=j+1}^{t-1} A_i(\nu) B_i(j) u_i(j). \tag{12.5.4}$$

The second summand in (12.5.4) can be generated especially simply if

$$\sigma_-(\tau) := \max\{\mathfrak{f}(\tau) - f, 0\}, \qquad \mathfrak{f}(t) := \max\{if : if \leq t, i = 0, 1, \ldots\}, \tag{12.5.5}$$

where $f \geq \sigma_f$ is an upper bound for the delays $\theta_i(t)$. Indeed then the lower limit in the sum equals $\mathfrak{f}(t) - f$ if $t - \theta_i(t) \geq \mathfrak{f}(t)$ and $\mathfrak{f}(t) - 2f$ otherwise. (The case $t \leq 2f$ is neglected here.) Thus

$$\widehat{x}_i(t) = \widehat{\omega}_i[\theta_i(t)] + \overline{x}_i^{(\eta)}(t),$$

where $\eta = 1$ if $\theta_i(t) \leq t - \mathfrak{f}(t)$ and $\eta = 2$ otherwise, and the vector

$$\overline{x}_i(t) := \mathbf{col}\,[\overline{x}_i^{(0)}(t), \overline{x}_i^{(1)}(t), \overline{x}_i^{(2)}(t)]$$

is given by

$$\overline{x}_i^{(\eta)}(t) := \sum_{j=\mathfrak{f}(t)-\eta f}^{t-1} \prod_{\nu=j+1}^{t-1} A_i(\nu) B_i(j) u_i(j). \tag{12.5.6}$$

Like $x_i^c(t)$, this vector can be generated by the ith local controller recursively:

$$\overline{x}_i^{(\eta)}(t+1) = A_i(t)\overline{x}_i^{(\eta)}(t) + B_i(t)u_i(t) \quad \text{whenever} \quad t+1 \neq \mathfrak{f}(t+1),$$
$$\text{otherwise}, \quad \overline{x}_i^{(0)}(t+1) = 0,$$
$$\overline{x}_i^{(j+1)}(t+1) = A_i(t)\overline{x}_i^{(j)}(t) + B_i(t)u_i(t), \quad j = 0, 1. \tag{12.5.7}$$

The disturbance of the state $\overline{x}_i(t)$ does not influence $\{\overline{x}_i(\theta)\}$ for $\theta \geq t+3f$ thanks to the second equation from (12.5.7). In this sense, the dynamics of (12.5.7) are finite.

To supply the Kalman filter with the estimates of the currently arrived observations, the estimates $\widehat{x}(s|\tau)$ of the states $x(s)$ at $s : \exists \nu, (\nu, s) \in S(\tau)$ are required. By Assumptions 12.4.1 and 12.4.2, $\tau - \sigma \leq s \leq \tau - \sigma_f + 1$. So these states are affected by the controls $u(0), \ldots, u(\tau - \sigma_f)$ known to the central controller at time τ by the remark following Assumption 12.4.1. Thus their estimates can actually be generated. Summarizing, we conclude that it is reasonable to commission the central controller to generate at time t the estimates

$$\widehat{X}(t) = \mathbf{col}\left[\widehat{x}(t-\sigma|t), \ldots, \widehat{x}(t-\sigma_f+1|t), \widehat{\omega}(t|t), \ldots, \widehat{\omega}(t+\overline{\sigma}|t)\right] \tag{12.5.8}$$

and focus the local controllers on producing the correction terms by (12.5.7).

As will be shown, the optimal control depends on the state estimates linearly with the coefficients known in advance. Since the dimension of the control is typically less than that of the state, it is beneficial to replace the estimates by their multiples by the above coefficients in the packages transferred via the control channels. After this, the entries of the packages can be considered as components of future controls. To form the ultimate control, each of them should be assembled with a complementary component generated by the local controller. With a slight abuse of exactness, it can be said that the control components generated by the central and local controllers are those determined by the observations and the history of the subsystem control, respectively.

The above distribution of estimation functions is not unique, and other schemes can be proposed. For example, it is easy to see that thanks to (12.4.7) and (12.5.2), the members of the package (12.5.3) are related with a simple recursion

$$\widehat{\omega}_0(s+1|\tau) = A_0(s)\widehat{\omega}_0(s|\tau),\ \widehat{\omega}_i(s+1|\tau) = A_i(s)\widehat{\omega}_i(s|\tau) + \overline{A}_i(s)\widehat{\omega}_0(s|\tau),\ s \geq \tau.$$

So only the launching members $\widehat{\omega}_i(\tau|\tau)$ and $\widehat{\omega}_0(\tau|\tau)$ may be sent over the ith control channel. On their arrival at time t, the above recursion should be executed for $s = \tau = t - \theta_i(t), \tau+1, \ldots, t-1$ by the ith local controller to produce the estimate $\widehat{\omega}_i(t|\tau)$ required. This approach may reduce the traffic over the control channel especially if the dimensions of the states x_i and x_0 are small enough. However this is for the expense of a larger amount of computations carried out by the local controller, which must execute the above recursion with renewed initial data at any time when new message arrives. This may be especially troublesome whenever the dimension n_0 of x_0 is large. In this chapter, we bear in mind the case where $n_0 >> 1$ and the local controllers have low computational powers. This forces avoidance of control algorithms for which the local controllers must process a large amount of data. In view of this, we focus on the packaged control scheme.

12.6 Minimum Variance State Estimator

In this section, we present a recursive algorithm by which the central controller can generate the estimates (12.5.8). It has much in common with the algorithm from Subsect. 11.3.2 (starting on p. 375).

Let a control strategy (12.4.5) and an integer $f \geq \sigma_f$ be chosen, where σ_f is taken from Assumption 12.4.1. We define $\sigma_-(t)$ by (12.5.5) and put

$$\delta_\eta(j|t) := \sum_{i=\sigma_-(t)+(\eta-1)f}^{\min\{t-\sigma_f, j-1\}} \prod_{\nu=i+1}^{j-1} A(\nu)B(i)u(i) \quad \eta = 1,2,$$
$$j = \sigma_-(t) + f + 1, \ldots, t + \overline{\sigma} + 1, \quad t \geq \max\{f - \overline{\sigma}, 0\}, \quad (12.6.1)$$

where $\overline{\sigma}$ is taken from (12.5.8). Coupled with these vectors and certain $n \times n$-matrices

$$P_{ij}(t), \quad \overline{P}_{ij}(t), \qquad i,j = -\overline{\sigma}, \ldots, \sigma,$$

the estimations (12.5.8) may be generated by the following analog of the Kalman filter.

Recursive State Estimator.

The next set $\widehat{X}(t+1)$ *of the estimates is given by equations*

$$\widehat{x}(i|t+1) = \widehat{x}(i|t)+\gamma_{t+1-i}(t+1)+\alpha(i|t), \quad i = t+1-\sigma,\dots,t+2-\sigma_f; \tag{12.6.2}$$

$$\widehat{\omega}(j|t+1) = \widehat{\omega}(j|t)+\gamma_{t+1-j}(t+1)+\begin{cases} 0 & \text{if} \quad t+1 \neq \nu f \ \forall \nu = 2,3,\dots \\ \delta_1(j|t)-\delta_2(j|t) & \text{otherwise} \end{cases} \tag{12.6.3}$$

for $j = t+1,\dots,t+1+\overline{\sigma}$.

Here

$$\widehat{x}(t+2-\sigma_f|t) := A(t+1-\sigma_f)\widehat{x}(t+1-\sigma_f|t),$$
$$\widehat{\omega}(t+1+\overline{\sigma}|t) := A(t+\overline{\sigma})\widehat{\omega}(t+\overline{\sigma}|t) \tag{12.6.4}$$

is the state prognosis without taking into account the newly arrived data,

$$\gamma_r(t+1) := \sum_{(\nu,\theta)\in S(t+1)} K_r^{(\nu,\theta)}(t+1)\left[y_\nu(\theta) - C_\nu(\theta)\widehat{x}(\theta|t)\right] \tag{12.6.5}$$

is the correction of the $(t+1-r)$*th state estimation on the basis of these data, and*

$$\alpha(j|t) := \begin{cases} \prod_{\nu=t-\sigma_f+2}^{j-1} A(\nu)B(\vartheta)u(\vartheta) & \text{if } j-1 \geq \vartheta := t+1-\sigma_f \\ 0 & \text{otherwise} \end{cases}. \tag{12.6.6}$$

The gain matrices $K_j^s(t)$ *are indexed by pairs* $[j,s]$ *with* $j = -\overline{\sigma},\dots,\sigma$ *and* $s = (\nu,\theta) \in S(t)$ *and have the dimension* $n \times k_\nu$. *Their computation is discussed further.*

The next set of the vectors (12.6.1) *is generated by the following equations, where*

$$j = \sigma_-(t+1)+f+1,\dots,t+\overline{\sigma}+2$$

and

$$\delta_\eta(t+\overline{\sigma}+2|t) := A(t+\overline{\sigma}+1)\delta_\eta(t+\overline{\sigma}+1|t);$$

$$\left.\begin{matrix} t+1 \neq \nu f \\ \forall \nu = 2,3,\dots \end{matrix}\right| \Rightarrow \left| \begin{matrix} \delta_1(j|t+1) = \begin{cases} \delta_1(j|t)+\alpha(j|t) & \text{if } 0 \leq t+1-\sigma_f \\ 0 & \text{otherwise} \end{cases} \\ \delta_2(j|t+1) = \begin{cases} \delta_2(j|t)+\alpha(j|t) & \text{if } \left\{ \begin{matrix} \sigma_-(t)+f \\ \leq t+1-\sigma_f \end{matrix} \right\} \\ 0 & \text{otherwise} \end{cases} \end{matrix}\right. .$$

$$\exists \nu = 2,3,\dots : t+1 = \nu f \ \Big| \Rightarrow \delta_1(j|t+1) = \delta_2(j|t)+\alpha(j|t), \delta_2(j|t+1) = 0. \tag{12.6.7}$$

The gain matrices from (12.6.5) *are calculated as follows:*

$$K_j^s(t) = \sum_{(\nu,\theta)\in S(t)} P_{j,t-\theta}(t) C_\nu(\theta)^{\mathrm{T}} \overset{+}{\Lambda}(t)^s_{(\nu,\theta)}. \tag{12.6.8}$$

Here $\overset{+}{\Lambda}(t)$ *is the pseudoinverse of the square ensemble of matrices*[2] $\Lambda(t) = \Lambda$ *over the finite set* $S(t)$ *that is given by*

$$\Lambda_{s_1}^{s_2} = C_{\nu_1}(\theta_1) P_{t-\theta_1, t-\theta_2}(t) C_{\nu_2}(\theta_2)^{\mathrm{T}} + \begin{cases} R_{\chi\chi}^{\nu_1}(\theta_1) & \text{if} \quad s_1 = s_2 \\ 0 & \text{otherwise} \end{cases}, \tag{12.6.9}$$

where $s_i = (\nu_i, \theta_i) \in S(t)$. *The matrices* $P_{ij}(t), \overline{P}_{ij}(t)$ *are generated recursively*

$$\cdots \longmapsto \{P_{ij}(t)\} \longmapsto \{\overline{P}_{ij}(t)\} \longmapsto \{P_{ij}(t+1)\} \longmapsto \cdots; \tag{12.6.10}$$

$$\overline{P}_{ij}(t) = P_{ij}(t) - \sum_{(\nu,\theta)\in S(t)} K_i^{(\nu,\theta)}(t) C_\nu(\theta) P_{t-\theta,j}(t); \tag{12.6.11}$$

$$P_{ij}(t+1) = \begin{cases} A(t+\overline{\sigma})\overline{P}_{ij}(t)A(t+\overline{\sigma})^{\mathrm{T}} + R_{\xi\xi}(t+\overline{\sigma}) & \text{if} \quad i = j = -\overline{\sigma} \\ A(t+\overline{\sigma})\overline{P}_{i,j-1}(t) & \text{if} \quad i = -\overline{\sigma}, j > -\overline{\sigma} \\ \overline{P}_{i-1,j}(t)A(t+\overline{\sigma})^{\mathrm{T}} & \text{if} \quad i > -\overline{\sigma}, j = -\overline{\sigma} \\ \overline{P}_{i-1,j-1}(t) & \text{if} \quad i, j > -\overline{\sigma} \end{cases}. \tag{12.6.12}$$

The recursion (12.6.7) *is initialized by* (12.6.1) *with* $t := \max\{f - \overline{\sigma}, 0\}$. *The recursion* (12.6.2), (12.6.3), (12.6.10) *is initialized by the formulas:*

$$\widehat{\omega}(j|-1) := \begin{cases} \widehat{z}(j) & \text{if } 0 \le j \le \overline{\sigma} - 1 \\ 0 & \text{if } j = -1 \end{cases} \quad \text{if } \overline{\sigma} \ge 1, \quad \widehat{\omega}(-1|-1) := \boldsymbol{E}a \quad \text{if } \overline{\sigma} = 0;$$

$$\widehat{x}(i|-1) := \begin{cases} \boldsymbol{E}a & \text{if } i = -\sigma_f \\ 0 & \text{otherwise} \end{cases}, \qquad P_{ij}(0) := P_{ij}^0(\overline{\sigma}).$$

Here $\widehat{z}(t), t = 0, \ldots, \overline{\sigma} - 1$ *and* $P_{ij}^0(t), t = 0, \ldots, \overline{\sigma}$ *are the solutions for the equations*

$$\widehat{z}(0) := \boldsymbol{E}a, \qquad \widehat{z}(t+1) = A(t)\widehat{z}(t), \tag{12.6.13}$$

$$P_{ij}^0(0) = \begin{cases} 0 & \text{if } i > -\overline{\sigma} \text{ or } j > -\overline{\sigma} \\ R_{aa} & \text{if } i = j = -\overline{\sigma}, \end{cases};$$

$$P_{ij}^0(t+1) = \begin{cases} A(t)P_{ij}^0(t)A(t)^{\mathrm{T}} + R_{\xi\xi}(t) & \text{if } i = j = -\overline{\sigma} \\ A(t)P_{i,j-1}^0(t) & \text{if } i = -\overline{\sigma}, j > -\overline{\sigma} \\ P_{i-1,j}^0(t)A(t)^{\mathrm{T}} & \text{if } i > -\overline{\sigma}, j = -\overline{\sigma} \\ P_{i-1,j-1}^0(t) & \text{if } i, j > -\overline{\sigma} \end{cases}.$$

(We put $u(\theta) := 0, A(\theta) := I, B(\theta) := 0$ *for* $\theta \le -1$.*)*

[2] See Subsect. 11.3.1 starting on p. 374.

Remark 12.6.1. Formula (12.6.9) is a replica of (11.3.8) (on p. 376). Formulas (12.6.8) and (12.6.12) differ from (11.3.7) and (11.3.11) (on p. 376), respectively, only by the fact that now the indices i and j range over wider sets.

Theorem 12.6.2. *Suppose that Assumptions* 12.2.1, 12.2.2, 12.4.1, and 12.4.2 *(on pp.* 409, 410, 417*, and* 417*) hold and that a control strategy* (12.4.5) *(on p.* 417*) is chosen. Then the above estimator generates the minimum variance estimates; i.e., whenever* $\max\{t-\sigma,0\} \le i \le t-\sigma_f+1$ *and* $\max\{t,0\} \le j \le t+\overline{\sigma}$*, we have*

$$\widehat{x}(i|t) = \boldsymbol{E}\left[x(i)\middle|\mathfrak{Y}(t),\mathfrak{S}(t),\Upsilon(t)\right], \qquad \widehat{\omega}(j|t) = \boldsymbol{E}\left[\omega(j,t)\middle|\mathfrak{Y}(t),\mathfrak{S}(t),\Upsilon(t)\right], \quad (12.6.14)$$

where $\mathfrak{Y}(t)$ *and* $\mathfrak{S}(t)$ *are defined in* (12.2.8)*, and*

$$\begin{aligned}\Upsilon(t) &:= \textbf{col}\left[\overline{\theta}(0),\ldots,\overline{\theta}(t-\sigma_f)\right] \\ \overline{\theta}(s) &:= \textbf{col}\left[\theta_1(s),\ldots,\theta_q(s)\right]\end{aligned} \qquad \text{if} \quad t \ge \sigma_f \quad (12.6.15)$$

and $\Upsilon(t) := 0 \in \mathbb{R}^q$ *otherwise. The matrices* $P_{ij}(t), \overline{P}_{ij}(t)$ *generated by the estimator are the conditional covariance matrices of the errors*

$$e(\theta|s) := \widehat{\omega}(\theta|s) - \omega(\theta,s) \quad \left(= \widehat{x}(\theta|s) - x(\theta) \text{ whenever } \theta \le s-\sigma_f+1\right). \quad (12.6.16)$$

More precisely, whenever $i,j = -\overline{\sigma},\ldots,\sigma$ *and* $t-i \ge 0, t-j \ge 0$*,*

$$\overline{P}_{ij}(t) := \boldsymbol{E}\left[e(t-i|t)e(t-j|t)^{\mathrm{T}}\middle|\mathfrak{S}(t)\right], \qquad P_{ij}(t) := \boldsymbol{E}\left[e(t-i|t-1)e(t-j|t-1)^{\mathrm{T}}\middle|\mathfrak{S}(t-1)\right]. \quad (12.6.17)$$

The proof of this theorem will be given in Sect. 12.9.

12.7 Solution of the Optimal Control Problem

Consider first the problem of minimizing the functional (12.2.6) (on p. 408) subject to the constraints (12.2.1) (on p. 407) in the case where the process disturbances are removed ($\xi(t) \equiv 0$) and the entire state x is accessible for "on-line" measurements. As was remarked in Sect. 11.5 (starting on p. 381), the solution of this problem is given by the feedback $u(t) = -L(t)x(t)$, where the gain matrix $L(t)$ is calculated in correspondence with (11.5.3)–(11.5.5) (on p. 382). Under Assumptions 12.4.3 and 12.4.4 (on p. 418), these relations take the form

$$u_i(t) = -L_i(t)x_i(t) - \overline{L}_i(t)x_0(t), \qquad \text{where} \quad i = 1,\ldots,q;$$

$$L_i(t) = F_i(t)^{-1}B_i(t)^{\mathrm{T}}\left[Q_i(t+1)+H_i(t+1)\right]A_i(t); \quad (12.7.1)$$

$$\overline{L}_i(t) = F_i(t)^{-1} B_i(t)^{\mathrm{T}} \Bigg\{ \left[\overline{Q}_i(t+1)^{\mathrm{T}} + \overline{H}_i(t+1)^{\mathrm{T}}\right] A_0(t) \\ + \left[Q_i(t+1) + H_i(t+1)\right] \overline{A}_i(t) \Bigg\};$$

$$F_i(t) := \Gamma_i(t) + B_i(t)^{\mathrm{T}} \left[Q_i(t+1) + H_i(t+1)\right] B_i(t). \qquad (12.7.2)$$

Here the matrices $H_i(t), \overline{H}_i(t)$ of dimensions $n_i \times n_i$ and $n_0 \times n_i$, respectively, are calculated recursively for $t = T, T-1, \dots, 0$ as follows:

$$H_i(T) := 0, \quad \overline{H}_i(T) := 0 \ \forall i, \qquad H_i(t) = L_i(t)^{\mathrm{T}} \Gamma_i(t) L_i(t) \\ + \left[A_i(t) - B_i(t)L_i(t)\right]^{\mathrm{T}} \left[Q_i(t+1) + H_i(t+1)\right] \left[A_i(t) - B_i(t)L_i(t)\right]; \quad (12.7.3)$$

$$\overline{H}_i(t) = \overline{L}_i(t)^{\mathrm{T}} \Gamma_i(t) L_i(t) \\ + A_0(t)^{\mathrm{T}} \left[\overline{Q}_i(t+1) + \overline{H}_i(t+1)\right] \left[A_i(t) - B_i(t)L_i(t)\right] \\ + \left[\overline{A}_i(t) - B_i(t)\overline{L}_i(t)\right]^{\mathrm{T}} \left[Q_i(t+1) + H_i(t+1)\right] \left[A_i(t) - B_i(t)L_i(t)\right]. \quad (12.7.4)$$

It is easy to see that $H_i(t) \geq 0 \ \forall i, t$. This and (12.7.2) imply that $F_i(t) \geq \Gamma_i(t) > 0$, and so the matrix $F_i(t)^{-1}$ in (12.7.1) does exist.

Corresponding to (12.4.1) (on p. 415) is the following partition:

$$\widehat{\omega}(j|t) = \textbf{col}\ \left[\widehat{\omega}_0(j|t), \dots, \widehat{\omega}_q(j|t)\right].$$

We also denote

$$\overline{\sigma}_i := \sigma_i + \mu_i - 1, \qquad \overline{\sigma} := \max\{\overline{\sigma}_1, \dots, \overline{\sigma}_q\}, \qquad f := \max\{\overline{\sigma}, \sigma_f\}, \quad (12.7.5)$$

where σ_i, σ_f, and μ_i are the constants from Assumptions 12.4.1 (on p. 417) and 12.4.2 (on p. 417), respectively.

Control Strategy

At the current time t, the **central controller** *produces the tuple* (12.5.8) *(on p.* 420*) of the estimates by the algorithm described in Sect.* 12.6 *(starting on p.* 421*). The constants $\overline{\sigma}$ in* (12.5.8) *and f in* (12.6.3) and (12.6.7) *are defined by* (12.7.5)*. For each $i = 1, \dots, q$, this controller sends via the ith control channel the control package*

$$p^{(i)} = \textbf{col}\ \left(v_0^{(i)}, \dots, v_{\overline{\sigma}_i}^{(i)}\right), \qquad (12.7.6)$$

where $v_j^{(i)} \in \mathbb{R}^{m_i}$ is given by

$$v_j^{(i)} := -L_i(t+j)\widehat{\omega}_i(t+j|t) - \overline{L}_i(t+j)\widehat{\omega}_0(t+j|t), \quad j \in [0 : \overline{\sigma}_i]. \quad (12.7.7)$$

In its memory, the ith **local controller** *stores, firstly, the last accepted package* (12.7.6)*; secondly, the time s_i of this package transmission; thirdly, the time τ_i elapsed since the arrival of this package; and fourthly, an auxiliary vector*

$$\overline{x}_i(t) = \textbf{col}\left[\overline{x}_i^{(0)}(t), \overline{x}_i^{(1)}(t), \overline{x}_i^{(2)}(t)\right]$$

whose dimension is thrice that of x_i. At the current time t, this controller, firstly, refreshes $p^{(i)}, s_i$ and sets $\tau_i := 0$ if an updated package[3] *arrives; secondly, forms the current output*

$$u_i(t) := v^{(i)}_{s_i+\tau_i} - L_i(t) \times \begin{cases} \overline{x}_i^{(0)}(t) & \text{if} \quad t < f \\ \overline{x}_i^{(2)}(t) & \text{if} \quad f \leq t - \theta_i - \tau_i < \mathfrak{f}(t) \\ \overline{x}_i^{(1)}(t) & \text{otherwise} \end{cases}, \tag{12.7.8}$$

where $\mathfrak{f}(t)$ is the quantity (12.5.5) *(on p.* 420*); and thirdly, sets $\tau_i := \tau_i + 1$ and calculates the next vector $\overline{x}_i(t+1)$ by formulas* (12.5.7) *(on p.* 420*). This controller is initialized at $t = 0$ by putting*

$$s_0 := 0, \quad \tau_0 := 0, \quad \overline{x}_i(0) := 0, \quad p^{(i)} := \textbf{col}\left(\widehat{v}_0^{(i)}, \ldots, \widehat{v}_{\overline{\sigma}_i}^{(i)}\right),$$

where

$$\widehat{v}_j^{(i)} := -L_i(j)\widehat{z}_i(j) - \overline{L}_i(j)\widehat{z}_0(j),$$

the vectors $\widehat{z}(j)$ are defined by (12.6.13) *(on p.* 423*), and* **col** $\left[\widehat{z}_0(j), \ldots, \widehat{z}_q(j))\right]$ *is the partition of $\widehat{z}(j)$ corresponding to* (12.4.1) *(on p.* 415*).*

Due to (12.7.1), (12.7.2), and (12.7.3), the gain matrices $L_i(t)$ employed by the ith local controller in (12.7.8) are calculated on the basis of the data related to only the ith subsystem and the uncontrolled part of the plant.

Now we are in a position to state the main result of this section.

Theorem 12.7.1. *Let Assumptions* 12.2.1, 12.2.2, and 12.4.1–12.4.4 *(on pp.* 409, 410, 417*, and* 418*) hold. Then the above control algorithm furnishes the optimum of the cost functional* (12.2.6) *(on p.* 408*) in the class of strategies given by* (12.4.5) *(on p.* 417*).*

The proofs of the results stated in this section will be given in Sect. 12.10.

This algorithm provides an optimal way to control not only the entire set of subsystems but also each of them. To specify this statement, suppose that all subsystems except for the ith one are controlled in accordance with some fixed strategies of the form (12.4.5). Consider the problem of optimal control of the ith subsystem:

$$\min \quad \boldsymbol{E} \sum_{t=0}^{T-1} \overline{\mathcal{G}}_i[t, x_i(t+1), u_i(t)] \quad \text{subject to} \quad x_i(0) = a_i;$$
$$x_i(t+1) = A_i(t)x_i(t) + \overline{A}_i(t)x_0(t) + B_i(t)u_i(t) + \xi_i(t) \tag{12.7.9}$$

[3] In other words, a package produced later than the currently employed one.

$t = 0, \ldots, T-1$ in the class (12.4.5) with i fixed. Here

$$\overline{\mathcal{G}}_i[t, x_i, u_i] := x_i^{\mathrm{T}} Q_i(t) x_i + 2x_0(t)^{\mathrm{T}} \overline{Q}_i(t) x_i + u_i^{\mathrm{T}} \Gamma_i(t) u_i$$

is in fact taken from (12.4.8) (on p. 418) and $\xi = \mathbf{col}\,(\xi_0, \ldots, \xi_q)$ is the partition of the disturbance from (12.2.1) (on p. 407) that corresponds to (12.4.1) (on p. 415). The solution for this problem is given by the following proposition.

Proposition 12.7.2. *The optimal way to control the ith subsystem is to employ the above algorithm restricted to the ith local controller. In other words, the parts of the algorithm that concern all local controllers except for the ith one are discarded, as well as the control packages are not sent from the central controller to them.*

Remark 12.7.3. So far we assumed that if several messages arrive at a subsystem out of order, accepted is the most updated of them. It can be shown that, in fact, this policy is optimal.

12.8 Proofs of Theorem 12.3.3 and Remark 12.3.1

Proof of Remark 12.3.1 (on p. 411). It suffices to justify the assumptions of Theorem 11.3.3 (on p. 377). We first show that the control given by (12.2.4) (on p. 408) and (12.2.12) (on p. 410) has the form (11.2.5) (on p. 373) with

$$\mathfrak{h}(t) := [\varkappa(0), \ldots, \varkappa(t+1)]\,. \tag{12.8.1}$$

Indeed, Assumption 12.2.4, Notation 12.2.10, and (12.2.10) and (12.2.11) (on pp. 410 and 411) imply that

$$\begin{aligned}\theta(t) := \mathbf{col}\ &[\theta_1(t), \ldots, \theta_q(t)] \\ &= \theta^+(t) + \theta^-(t+1) = \theta^+[\varkappa(t)] + \theta^-[\varkappa(t+1)].\end{aligned} \tag{12.8.2}$$

So $\theta_i(t) = \Theta_i[\mathfrak{h}(t)]$, and (12.2.4) and (12.2.12) yield that the control has the required form:

$$u(t) = \mathcal{U}\left[t, \mathfrak{Y}(t), \mathfrak{S}(t), \mathfrak{h}(t)\right].$$

Here the blocks $\mathcal{U}_i(\cdot)$ of the partition $\mathcal{U}(\cdot) = \mathbf{col}\left[\mathcal{U}_1(\cdot), \ldots, \mathcal{U}_q(\cdot)\right]$ matching that from (12.2.4) are defined as follows:

$$\begin{aligned}&\mathcal{U}_i\left[t, Y_0, \ldots, Y_t, S_0, \ldots, S_t, \mathfrak{h}\right] \\ &\qquad := \mathcal{V}_i\left[t-\theta_i, Y_0, \ldots, Y_{t-\theta_i}, S_0, \ldots, S_{t-\theta_i}, \varkappa_0, \ldots, \varkappa_{t-\theta_i}\right],\end{aligned}$$

where $\mathcal{V}_i(\cdot)$ and θ_i are the ith components of

$$\mathcal{V}(\cdot) = \mathbf{col}\left[\mathcal{V}_1(\cdot), \ldots, \mathcal{V}_q(\cdot)\right]$$

from (12.2.12) and $\Theta[\mathfrak{h}]$, respectively, and $\varkappa_j$ are the components of $\mathfrak{h} = \left\{\varkappa_s\right\}_{s=0}^{t+1}$.

To complete the proof, we note that Assumptions 12.2.1–12.2.4 (on pp. 409 and 410) imply Assumptions 11.2.1–11.2.4 (on p. 373) of Theorem 11.3.3 (on p. 377), and we apply this theorem. □

Since Assumptions 11.2.1–11.2.4 hold, not only the conclusion of Theorem 11.3.3 but also Remark 11.3.4 (on p. 377) is true. With regard to (12.8.1), this remark shapes into the following claim, which aids in proving Theorem 12.3.3.

Remark 12.8.1. (i) The minimum variance estimates $\widehat{x}(j|t)$ of $x(j)$ (where $\max\{t-\sigma,0\}\leq j\leq t$) based on

$$\mathfrak{Y}(t),\mathfrak{S}(t),u(0),\ldots,u(t-1)$$

are equal to those based on

$$\mathfrak{Y}(t),\mathfrak{S}(t),\varkappa(0),\ldots,\varkappa(t).$$

(ii) Given $S(0),\ldots,S(t),\varkappa(0),\ldots,\varkappa(t)$, the estimation errors

$$e(j|t):=\widehat{x}(j|t)-x(j), \tag{12.8.3}$$

where $\max\{t-\sigma,0\}\leq j\leq t$, are independent of the observations $Y(0),\ldots,Y(t)$.

We recall that the estimates $\widehat{x}(j|t)$ are produced by the estimator from Subsect. 11.3.2.

The remainder of the section is devoted to the proof of Theorem 12.3.3. This proof is broken into the string of four steps.

1) Like in Sect. 11.7 (starting on p. 384), we rewrite the problem in a more standard form by augmenting the system state.
2) We offer concise forms of relevant relations in terms of the coefficients from the augmented problem.
3) The cost functional is shaped into a form convenient for further analysis. (This is a key step.)
4) Theorem 12.3.3 is proved by accomplishing such an analysis.

1) *Augmenting the system state.* To this end, we introduce the linear space

$$\mathfrak{Z}:=\left\{Z=\{z_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma}:z_{\nu,j}\in\mathbb{R}^{k_\nu}\ \forall\nu,j\right\}$$

and put $x(\theta):=0,v(\theta):=0$ for $\theta\leq-1$. Here σ is taken from Assumption 12.2.3 (on p. 410). We also invoke the constant σ_* from this assumption and for $t=0,1,\ldots$, we introduce the vectors

$$X(t):=\textbf{col}\left[x(t),\ldots,x(t-\sigma),v(t-1),\ldots,v(t-\sigma_*)\right],$$
$$\theta(t):=\textbf{col}\left[\theta_1(t),\ldots,\theta_q(t)\right]; \tag{12.8.4}$$

$$Z(t):=\left\{z_{\nu,j}\right\}\in\mathfrak{Z},\quad\text{where}$$
$$z_{\nu,j}:=\begin{cases}y_\nu(t-j) & \text{if}\quad(\nu,t-j)\in S(t)\\ 0 & \text{otherwise}\end{cases}. \tag{12.8.5}$$

In terms of these vectors, the problem under consideration takes the form:

$$\text{minimize} \quad \mathfrak{I} := \boldsymbol{E} \sum_{t=0}^{T-1} \mathfrak{G}\left[t, X(t+1), v(t)\right], \quad \text{where}$$
$$\mathfrak{G}\left[t, X, v\right] := X^{\mathsf{T}} \mathfrak{Q}(t+1) X + v^{\mathsf{T}} \Gamma(t) v, \quad (12.8.6)$$

subject to the constraints

$$X(t+1) = \mathfrak{A}_{\theta(t)}(t) X(t) + \mathfrak{B}_{\theta(t)}(t) v(t) + \mathfrak{E}\, \xi(t) \qquad t = 0, \ldots, T-1; \quad (12.8.7)$$

$$X(0) = \mathfrak{a} := \textbf{col}\,(a, 0, \ldots, 0); \quad (12.8.8)$$

$$Z(t) = \mathfrak{C}\big[t, S(t)\big] X(t) + \Xi\big[t, S(t)\big] \qquad t = 0, \ldots, T; \quad (12.8.9)$$

$$v(t) = \mathcal{V}\left[t, Z(0), \ldots, Z(t), S(0), S(1), \ldots, S(t), \varkappa(0), \ldots, \varkappa(t)\right]. \quad (12.8.10)$$

Here for any $t = 0, \ldots, T$ and $\theta = \textbf{col}\,(\theta_1, \ldots, \theta_q), \theta_i = 0, \ldots, \sigma_*$,

$$\mathfrak{Q}(t) := \begin{pmatrix} Q(t) & 0 \cdots & 0 \\ 0 & 0 \cdots & 0 \\ 0 & 0 \cdots & 0 \\ \cdot & \cdot \cdots\cdot & \cdot \\ 0 & 0 \cdots & 0 \end{pmatrix}, \quad \begin{array}{c} \mathfrak{B}_\theta(t) := \textbf{col}\left[\mathcal{B}_\theta^0(t), 0, \ldots, 0 \big| I_m, 0, \ldots, 0\right], \\ \mathfrak{E} := \textbf{col}\left[I_n, 0, \ldots, 0\right]; \end{array}$$
$$(12.8.11)$$

$$\mathfrak{A}_\theta(t) := \left(\begin{array}{ccccc|cccc} A(t) & 0 & \cdots & 0 & 0 & \mathcal{B}_\theta^1(t) & \cdots & \mathcal{B}_\theta^{\sigma_*-1}(t) & \mathcal{B}_\theta^{\sigma_*}(t) \\ I_n & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & I_n & 0 & 0 & \cdots & 0 & 0 \\ \hline & & & & & 0 & \cdots & 0 & 0 \\ & & & & & I_m & \cdots & 0 & 0 \\ & & \mathbf{0} & & & 0 & \cdots & 0 & 0 \\ & & & & & \cdots & \cdots & \cdots & \cdots \\ & & & & & 0 & \cdots & I_m & 0 \end{array}\right). \quad (12.8.12)$$

The lines within matrices separate their "x"- and "v"- parts, and the matrices $\mathcal{B}_\theta^j(t)$ are defined by (12.3.6) and (12.3.8) (on p. 413). Furthermore for any

$$t \leq T, S \subset \{(\nu, j) : \nu \in [1:l], j \in [0:\sigma]\}, X = \textbf{col}\,(x_0, x_1, \ldots, x_\sigma, v_1, \ldots, v_{\sigma_*}),$$

the following relations hold:

$$\mathfrak{C}[t, S] X := \{z_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} \in \mathfrak{Z}, \quad \text{where}$$
$$z_{\nu,j} := \begin{cases} C_\nu(t-j) x_j & \text{if} \quad (\nu, t-j) \in S \\ 0 & \text{otherwise} \end{cases}; \quad (12.8.13)$$

$$\varXi[t,S] := \{\zeta_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} \in \mathfrak{Z}, \quad \text{where}$$

$$\zeta_{\nu,j} := \begin{cases} \chi_\nu(t-j) & \text{if} \quad (\nu, t-j) \in S \\ 0 & \text{otherwise} \end{cases}. \qquad (12.8.14)$$

When dealing with the augmented system, it is natural to employ the minimum variance estimate

$$\widehat{X}(t|t) = \boldsymbol{E}\left[X(t)\big|Z(0),\ldots,Z(t),S(0),\ldots,S(t),\varkappa(0),\ldots,\varkappa(t)\right]. \qquad (12.8.15)$$

Thanks to (i) of Remark 12.8.1, (12.8.4), (12.8.5), and (12.8.10),

$$\widehat{X}(t|t) = \textbf{col}\,[\widehat{x}(t|t), \widehat{x}(t-1|t), \ldots, \widehat{x}(t-\sigma|t), v(t-1), \ldots, v(t-\sigma_*)]. \qquad (12.8.16)$$

So the estimation error can be expressed in terms of (12.8.3):

$$\mathcal{E}(t) := \widehat{X}(t|t) - X(t) = \textbf{col}\,[e(t|t), e(t-1|t), \ldots, e(t-\sigma|t), 0, \ldots, 0]. \qquad (12.8.17)$$

2) *Concise form of relations* (12.3.10)–(12.3.17) *(on pp.* 413 *and* 414*)*. This form is in terms of the coefficients from (12.8.6) and (12.8.7), along with the following matrices:

$$\mathfrak{H}(t,\overline{\varkappa}) := \left(\begin{array}{cccc|ccc} H_{xx}(t,\overline{\varkappa}) & 0 & \cdots & 0 & H_{x1}(t,\overline{\varkappa}) & \cdots & H_{x\sigma_*}(t,\overline{\varkappa}) \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline H_{1x}(t,\overline{\varkappa}) & 0 & \ldots & 0 & H_{11}(t,\overline{\varkappa}) & \ldots & H_{1\sigma_*}(t,\overline{\varkappa}) \\ H_{2x}(t,\overline{\varkappa}) & 0 & \ldots & 0 & H_{21}(t,\overline{\varkappa}) & \ldots & H_{2\sigma_*}(t,\overline{\varkappa}) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{\sigma_* x}(t,\overline{\varkappa}) & 0 & \ldots & 0 & H_{\sigma_* 1}(t,\overline{\varkappa}) & \ldots & H_{\sigma_*\sigma_*}(t,\overline{\varkappa}) \end{array}\right); \qquad (12.8.18)$$

$$\mathfrak{L}(t,\overline{\varkappa}) := \left[L(t,\overline{\varkappa}), 0, \ldots, 0\big|\mathcal{L}_1(t,\overline{\varkappa}), \ldots, \mathcal{L}_{\sigma_*}(t,\overline{\varkappa})\right]. \qquad (12.8.19)$$

Here the partitions correspond to (12.8.12). We recall that $\overline{\varkappa}$ is the tuple (12.3.2) (on p. 412) and the operations $\mathcal{P}_{t,\overline{\varkappa}}$ and $\oplus$ are defined by (12.3.5) and (12.3.7) (on p. 413), respectively.

Lemma 12.8.2. *Suppose that* $t = 0, \ldots, T-1$ *and* $\overline{\varkappa} = [\varkappa_0, \ldots, \varkappa_{\overline{\sigma}}] \in \Upsilon^{\overline{\sigma}+1}$ *are given. Let the dot* $\bullet$ *stand for the variable* $\varkappa \in \Upsilon$. *We also invoke the notation* $\theta(\varkappa, \overline{\varkappa})$ *defined by* (12.3.9) *(on p.* 413*) and Notation* 12.2.10 *(on p.* 411*). Then the following statements hold:*

(i) The matrix $F(t,\overline{\varkappa})$ *introduced in* (12.3.4) *(on p.* 412*) is positive-definite and*

$$F(t,\overline{\varkappa}) = \mathcal{P}_{t,\overline{\varkappa}}\left\{\Gamma(t) + \mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)^{\mathrm{T}}\Big[\mathfrak{H}(t+1, \bullet \oplus \overline{\varkappa}) + \mathfrak{Q}(t+1)\Big]\mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)\right\}; \qquad (12.8.20)$$

$$F(t,\overline{\varkappa})\mathfrak{L}(t,\overline{\varkappa}) = \mathcal{P}_{t,\overline{\varkappa}}\Bigg\{\mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)^{\mathrm{T}}\Big[\mathfrak{H}(t+1,\bullet\oplus\overline{\varkappa}) \\ + \mathfrak{Q}(t+1)\Big]\mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t)\Bigg\}; \quad (12.8.21)$$

$$\mathfrak{H}(t,\overline{\varkappa}) = \mathfrak{L}(t,\overline{\varkappa})^{\mathrm{T}}\Gamma(t)\mathfrak{L}(t,\overline{\varkappa}) + \\ \mathcal{P}_{t,\overline{\varkappa}}\Bigg\{\Big[\mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t) - \mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)\mathfrak{L}(t,\overline{\varkappa})\Big]^{\mathrm{T}}\Big[\mathfrak{H}(t+1,\bullet\oplus\overline{\varkappa}) \\ + \mathfrak{Q}(t+1)\Big]\Big[\mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t) - \mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)\mathfrak{L}(t,\overline{\varkappa})\Big]\Bigg\}. \quad (12.8.22)$$

We recall that the matrices $\mathfrak{B}_\theta(t), \mathfrak{Q}(t+1)$*, and* $\mathfrak{A}_\theta(t)$ *are defined by* (12.8.11) *and* (12.8.12)*.*

(ii) For any $X = \mathbf{col}\,(x_0,\ldots,x_\sigma,v_1,\ldots,v_{\sigma_*})$ *and* $v \in \mathbb{R}^m$*, we put*

$$X_+(\bullet) := \mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t)X + \mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)v.$$

Then

$$\mathcal{P}_{t,\overline{\varkappa}}\Bigg\{X_+(\bullet)^{\mathrm{T}}\left[\mathfrak{H}(t+1,\bullet\oplus\overline{\varkappa}) + \mathfrak{Q}(t+1)\right]X_+(\bullet)\Bigg\} - X^{\mathrm{T}}\mathfrak{H}(t,\overline{\varkappa})X \\ + v^{\mathrm{T}}\Gamma(t)v = \left[v + \mathfrak{L}(t,\overline{\varkappa})X\right]^{\mathrm{T}}F(t,\overline{\varkappa})\left[v + \mathfrak{L}(t,\overline{\varkappa})X\right]. \quad (12.8.23)$$

Proof. Statement (i). Note first that (12.3.5) (on p. 412), (12.8.22), and the boundary condition $\mathfrak{H}(T,\overline{\varkappa}) = 0$ imply recursively that $\mathfrak{H}(t,\overline{\varkappa}) \geq 0$. Then (12.3.5) and (12.8.20) yield $F(t,\overline{\varkappa}) > 0$. Rewriting relations (12.8.20)–(12.8.22) in terms of the entries from (12.8.11), (12.8.12), (12.8.18), and (12.8.19) demonstrates that these relations are merely another form of (12.3.10)–(12.3.17) (on pp. 413 and 414).

Statement (ii) By (12.8.22),

$$\mathcal{P}_{t,\overline{\varkappa}}\Bigg\{\mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t)^{\mathrm{T}}\left[\mathfrak{H}(t+1,\bullet\oplus\overline{\varkappa}) + \mathfrak{Q}(t+1)\right]\mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t)\Bigg\} - \mathfrak{H}(t,\overline{\varkappa}) \\ = \mathfrak{L}(t,\overline{\varkappa})^{\mathrm{T}}\mathcal{P}_{t,\overline{\varkappa}}\Bigg\{\mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)^{\mathrm{T}}\left[\mathfrak{H}(t+1,\bullet\oplus\overline{\varkappa}) + \mathfrak{Q}(t+1)\right]\mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t)\Bigg\} \\ -\mathfrak{L}(t,\overline{\varkappa})^{\mathrm{T}}\mathcal{P}_{t,\overline{\varkappa}}\Bigg\{\Gamma(t)+\mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)^{\mathrm{T}}\left[\mathfrak{H}(t+1,\bullet\oplus\overline{\varkappa}) + \mathfrak{Q}(t+1)\right]\mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)\Bigg\}\mathfrak{L}(t,\overline{\varkappa}) \\ + \mathcal{P}_{t,\overline{\varkappa}}\Bigg\{\mathfrak{A}_{\theta(\cdot,\overline{\varkappa})}(t)^{\mathrm{T}}\left[\mathfrak{H}(t+1,\bullet\oplus\overline{\varkappa}) + \mathfrak{Q}(t+1)\right]\mathfrak{B}_{\theta(\cdot,\overline{\varkappa})}(t)\Bigg\}\mathfrak{L}(t,\overline{\varkappa}) \\ \overset{(12.8.20),\,(12.8.21)}{=\!=\!=} \mathfrak{L}(t,\overline{\varkappa})F(t,\overline{\varkappa})\mathfrak{L}(t,\overline{\varkappa})^{\mathrm{T}}.$$

In view of this, relationship (12.8.23) is immediate from (12.8.20) and (12.8.21). □

3) *Transformation of the cost functional.* This transformation plays a key role in proving Theorem 12.3.3.

Lemma 12.8.3. *Suppose that a control strategy* (12.8.10) *is chosen. Consider the estimates* (12.8.15) *and put*

$$\overline{\varkappa}(t) := [\varkappa(t), \ldots, \varkappa(t-\overline{\sigma})],$$

where $\overline{\sigma}$ is the constant from Assumption 12.2.8 *(on p.* 411*). We recall that $\varkappa(s)$ is defined for $s<0$ by Remark* 12.2.6 *and* (12.2.10) *(on p.* 410*). For any $t=0,\ldots,T$, the following representation of the cost functional from* (12.8.6) *holds:*

$$\begin{aligned}\mathfrak{I} = \boldsymbol{E}\sum_{s=0}^{t-1}\mathfrak{G}[s, X(s+1), v(s)] + \boldsymbol{E}X(t)^{\mathrm{T}}\mathfrak{H}[t,\overline{\varkappa}(t)]X(t)\\ +\boldsymbol{E}\sum_{s=t}^{T-1}\Big[v(s)+\mathfrak{L}[s,\overline{\varkappa}(s)]\widehat{X}(s|s)\Big]^{\mathrm{T}}F[s,\overline{\varkappa}(s)]\\ \times\Big[v(s)+\mathfrak{L}[s,\overline{\varkappa}(s)]\widehat{X}(s|s)\Big]+\Delta(t), \quad (12.8.24)\end{aligned}$$

where

$$\begin{aligned}\Delta(t) := \mathbf{tr}\sum_{s=t}^{T-1}\bigg\{\boldsymbol{E}\Big(L[s,\overline{\varkappa}(s)]^{\mathrm{T}}F[s,\overline{\varkappa}(s)]L[s,\overline{\varkappa}(s)]\overline{P}_{00}(s)\Big)\\ +\Big(Q(s+1)+\boldsymbol{E}H_{xx}[s+1,\overline{\varkappa}(s+1)]\Big)R_{\xi\xi}(s)\bigg\}. \quad (12.8.25)\end{aligned}$$

Here $\overline{P}_{00}(s)$ is the matrix generated by the estimator from Subsect. 11.3.2 *(starting on p.* 375 *), and $R_{\xi\xi}(t)$ is the disturbance correlation matrix from* (12.2.9) *(on p.* 409*).*

Proof. The proof will be by induction on $t=T,T-1,\ldots,0$. For $t=T$, (12.8.24) is immediate from (12.8.6), (12.8.18), and the boundary condition $H(T,\overline{\varkappa})=0 \Leftrightarrow \mathfrak{H}(T,\overline{\varkappa})=0$. Suppose that (12.8.24) holds for $t:=t+1\le T$; i.e.,

$$\begin{aligned}\mathfrak{I} = \boldsymbol{E}\sum_{s=0}^{t-1}\mathfrak{G}[s, X(s+1), v(s)] + \boldsymbol{E}v(t)^{\mathrm{T}}\Gamma(t)v(t)+\Delta(t+1)\\ +\boldsymbol{E}X(t+1)^{\mathrm{T}}\underbrace{\{\mathfrak{H}[t+1,\overline{\varkappa}(t+1)]+\mathfrak{Q}(t+1)\}}_{\mathfrak{N}(t+1)}X(t+1)\\ +\boldsymbol{E}\sum_{s=t+1}^{T-1}\Big[v(s)+\mathfrak{L}[s,\overline{\varkappa}(s)]\widehat{X}(s|s)\Big]^{\mathrm{T}}F[s,\overline{\varkappa}(s)]\Big[v(s)+\mathfrak{L}[s,\overline{\varkappa}(s)]\widehat{X}(s|s)\Big].\end{aligned}$$

By (12.8.7),

$$X(t+1) = X_+(t) + \mathfrak{E}\xi(t),$$

where the vector

$$X_+(t) := \mathfrak{A}_{\theta(t)}(t)X(t) + \mathfrak{B}_{\theta(t)}(t)v(t)$$

and the quantity $\overline{\varkappa}(t+1)$ are independent of $\xi(t)$ due to Assumptions 12.2.1, 12.2.2 (on pp. 409 and 410), (12.8.7), (12.8.9), and (12.8.10). Since $\boldsymbol{E}\xi(t) = 0$, this yields

$$\boldsymbol{E}X(t+1)^{\mathrm{T}}\mathfrak{N}(t+1)X(t+1) = \boldsymbol{E}X_+(t)^{\mathrm{T}}\mathfrak{N}(t+1)X_+(t) + \boldsymbol{E}\xi(t)^{\mathrm{T}}\mathfrak{E}^{\mathrm{T}}\mathfrak{N}(t+1)\mathfrak{E}\xi(t). \quad (12.8.26)$$

Owing to (12.2.9) (on p. 409), (12.8.11), and (12.8.18), the second summand on the right amounts to

$$\boldsymbol{E}\xi(t)^{\mathrm{T}}\Big\{H_{xx}[t+1,\overline{\varkappa}(t+1)] + Q(t+1)\Big\}\xi(t) = \mathbf{tr}\left[\Big\{Q(t+1) + \boldsymbol{E}H_{xx}[t+1,\overline{\varkappa}(t+1)]\Big\}R_{\xi\xi}(t)\right].$$

To compute the first one, note that by (12.3.9) (on p. 413) and (12.8.2), $\theta(t) = \theta[\varkappa(t+1),\overline{\varkappa}(t)]$. Due to (12.8.7), (12.8.9), and (12.8.10), $X(t)$ and $v(t)$ are deterministic functions of

$$\Pi := [a, \{\chi_\nu(0)\}_{\nu=1}^{l}, \ldots, \{\chi_\nu(t)\}_{\nu=1}^{l}, \xi(0), \ldots, \xi(t-1), S(0), \ldots, S(t)]$$

and $\varkappa(0), \ldots, \varkappa(t)$, where Π and $\{\varkappa(s)\}$ are independent thanks to Assumptions 12.2.1 and 12.2.2 (on pp. 409 and 410). Hence putting $\mathcal{K} := [\varkappa(\alpha), \ldots, \varkappa(t)]$, where $\alpha := \max\{t - \overline{\sigma}, 0\}$, we get

$$\boldsymbol{E}\left[X_+(t)^{\mathrm{T}}\mathfrak{N}(t+1)X_+(t)\middle|X(t) = \overline{X}, v(t) = \overline{v}, \mathcal{K}\right] =$$

$$\boldsymbol{E}\left\{\left[\mathfrak{A}_{\theta[\varkappa(t+1),\overline{\varkappa}(t)]}\overline{X} + \mathfrak{B}_{\theta[\varkappa(t+1),\overline{\varkappa}(t)]}\overline{v}\right]^{\mathrm{T}}\left[\mathfrak{H}[t+1,\overline{\varkappa}(t+1)] + \mathfrak{Q}(t+1)\right] \times \left[\mathfrak{A}_{\theta[\varkappa(t+1),\overline{\varkappa}(t)]}\overline{X} + \mathfrak{B}_{\theta[\varkappa(t+1),\overline{\varkappa}(t)]}\overline{v}\right]\middle|\mathcal{K}\right\} \overset{(12.3.7)}{=\!=}$$

$$\sum_{\varkappa\in\Upsilon}\Bigg\{P\left[\varkappa(t+1) = \varkappa\middle|\mathcal{K}\right]\left[\mathfrak{A}_{\theta[\varkappa,\overline{\varkappa}(t)]}\overline{X} + \mathfrak{B}_{\theta[\varkappa,\overline{\varkappa}(t)]}\overline{v}\right]^{\mathrm{T}}\left[\mathfrak{H}[t+1,\varkappa\oplus\overline{\varkappa}(t)] + \mathfrak{Q}(t+1)\right] \times \left[\mathfrak{A}_{\theta[\varkappa,\overline{\varkappa}(t)]}\overline{X} + \mathfrak{B}_{\theta[\varkappa,\overline{\varkappa}(t)]}\overline{v}\right]\Bigg\}.$$

Here $P\left[\varkappa(t+1) = \varkappa\middle|\mathcal{K}\right] = P\left[\varkappa(t+1) = \varkappa\middle|\overline{\varkappa}(t)\right]$ by Assumption 12.2.8 (on p. 411). So invoking (12.3.5) (on p. 412) shows that the conditional expectation under consideration equals

$$\mathcal{P}_{t,\overline{\varkappa}(t)}\Bigg\{\Big[\mathfrak{A}_{\theta[\cdot,\overline{\varkappa}(t)]}\overline{X}+\mathfrak{B}_{\theta[\cdot,\overline{\varkappa}(t)]}\overline{v}\Big]^{\mathrm{T}}\Big[\mathfrak{H}[t+1,\bullet\oplus\overline{\varkappa}(t)]+\mathfrak{Q}(t+1)\Big]$$
$$\times\Big[\mathfrak{A}_{\theta[\cdot,\overline{\varkappa}(t)]}\overline{X}+\mathfrak{B}_{\theta[\cdot,\overline{\varkappa}(t)]}\overline{v}\Big]\Bigg\}$$
$$\overset{(12.8.23)}{=\!=}\overline{X}^{\mathrm{T}}\mathfrak{H}\left[t,\overline{\varkappa}(t)\right]\overline{X}-\overline{v}^{\mathrm{T}}\Gamma(t)\overline{v}$$
$$+\left\{\overline{v}+\mathfrak{L}\left[t,\overline{\varkappa}(t)\right]\overline{X}\right\}^{\mathrm{T}}F\left[t,\overline{\varkappa}(t)\right]\left\{\overline{v}+\mathfrak{L}\left[t,\overline{\varkappa}(t)\right]\overline{X}\right\}.$$

Thus in (12.8.26), the first summand on the right equals

$$\boldsymbol{E}X(t)^{\mathrm{T}}\mathfrak{H}[t,\overline{\varkappa}(t)]X(t)-\boldsymbol{E}v(t)\Gamma(t)v(t)+\mu(t),\quad\text{where}$$
$$\mu(t):=\boldsymbol{E}\Big\{v(t)+\mathfrak{L}[t,\overline{\varkappa}(t)]X(t)\Big\}^{\mathrm{T}}F[t,\overline{\varkappa}(t)]\Big\{v(t)+\mathfrak{L}[t,\overline{\varkappa}(t)]X(t)\Big\}.$$

Now we employ the notation $\mathcal{E}(t)$ from (12.8.17) and denote

$$\Lambda:=[S(0),\dots,S(t),\varkappa(0),\dots,\varkappa(t)].\tag{12.8.27}$$

Then

$$\mu(t)=\boldsymbol{E}\Big\{\underbrace{v(t)+\mathfrak{L}[t,\overline{\varkappa}(t)]\widehat{X}(t|t)}_{w(t)}-\mathfrak{L}[t,\overline{\varkappa}(t)]\mathcal{E}(t)\Big\}^{\mathrm{T}}F[t,\overline{\varkappa}(t)]$$
$$\times\Big\{w(t)-\mathfrak{L}[t,\overline{\varkappa}(t)]\mathcal{E}(t)\Big\}$$
$$=\boldsymbol{E}w(t)^{\mathrm{T}}F[t,\overline{\varkappa}(t)]w(t)+\boldsymbol{E}\mathcal{E}(t)^{\mathrm{T}}\mathfrak{L}[t,\overline{\varkappa}(t)]^{\mathrm{T}}F[t,\overline{\varkappa}(t)]\mathfrak{L}[t,\overline{\varkappa}(t)]\mathcal{E}(t)$$
$$-2\boldsymbol{E}\boldsymbol{E}\Big[w(t)^{\mathrm{T}}F[t,\overline{\varkappa}(t)]\mathfrak{L}[t,\overline{\varkappa}(t)]\mathcal{E}(t)\Big|\Lambda\Big].\tag{12.8.28}$$

Due to (12.8.10) and (12.8.15), $w(t)$ is a deterministic function of Λ and $Z(0),\dots,$ $Z(t)$. At the same time, $\mathcal{E}(t)$ is independent of $Z(0),\dots,Z(t)$ given Λ, and $\boldsymbol{E}\mathcal{E}(t)=$ 0 due to (ii) of Remark 12.8.1, (12.8.5), and (12.8.17). Hence the last summand on the right in (12.8.28) is zero. The second one equals

$$\delta:=\ \mathbf{tr}\,\boldsymbol{E}\left\{\mathfrak{L}[t,\overline{\varkappa}(t)]^{\mathrm{T}}F[t,\overline{\varkappa}(t)]\mathfrak{L}[t,\overline{\varkappa}(t)]\boldsymbol{E}\left[\mathcal{E}(t)\mathcal{E}(t)^{\mathrm{T}}\middle|\Lambda\right]\right\}.$$

By Assumption 12.2.2 (on p. 410), the S and $\varkappa$-parts of (12.8.27) are independent. At the same time, Remark 12.3.1 (on p. 411) and the last claim of Theorem 11.3.3 (on p. 377) guarantee that the estimation error (12.8.17) is not affected by the controls. Hence it is not influenced by $\varkappa(0),\dots,\varkappa(t)$. It follows that

$$\boldsymbol{E}\left[\mathcal{E}(t)\mathcal{E}(t)^{\mathrm{T}}\middle|\Lambda\right]=\boldsymbol{E}\left[\mathcal{E}(t)\mathcal{E}(t)^{\mathrm{T}}\middle|S(0),\dots,S(t)\right].$$

By Remark 12.3.1, Theorem 11.3.3, and (12.8.17), the conditional expectation on the right amounts to

$$\left(\begin{array}{ccc|c} \overline{P}_{00}(t) & \cdots & \overline{P}_{0\sigma}(t) & \\ \cdots & \cdots & \cdots & \mathbf{0} \\ \overline{P}_{\sigma 0}(t) & \cdots & \overline{P}_{\sigma\sigma}(t) & \\ \hline & 0 & & 0 \end{array}\right).$$

In view of (12.8.19), this yields

$$\delta = \mathbf{tr}\,\boldsymbol{E}\left\{L[t,\overline{\varkappa}(t)]^{\mathrm{T}}F[t,\overline{\varkappa}(t)]L[t,\overline{\varkappa}(t)]\overline{P}_{00}(t)\right\}.$$

Summarizing, we arrive at (12.8.24). □

4) *Completion of the proof of Theorem 12.3.3* (on p. 414). By *(i)* of Lemma 12.8.2 and (12.8.8) and (12.8.18), relation (12.8.24) with $t := 0$ implies that for any control strategy (12.8.10),

$$\mathfrak{I} \geq \Delta(0) + \boldsymbol{E}a^{\mathrm{T}}H_{xx}[0,\overline{\varkappa}(0)]a.$$

By (12.8.25), the obtained lower bound does not depend on the control. This bound is evidently attained at the strategy $v(t) = -\mathfrak{L}[t,\overline{\varkappa}(t)]\widehat{X}(t|t)$, which equals (12.3.1) (on p. 411) thanks to (12.8.16) and (12.8.19). □

Concluding remarks. The representation (12.8.6)–(12.8.10) moves the problem under consideration into the area of LQG control of systems whose coefficients are determined by a finite-state Markov chain.[4] This area has received considerable attention (see, e.g., [1, 35, 36, 56, 79], we refer the reader to [36, 67, 78] for a survey). It however was mainly focused on continuous-time case, noise-free discrete-time systems, or systems with incomplete observation of the Markov parameters. In the corresponding literature, the authors failed to find a reference that directly proves Theorem 12.3.3. In view of this, we offered an independent proof, which is based on the standard technique.

12.9 Proof of Theorem 12.6.2 on p. 424

The proof is based on the state augmenting technique considered in Sect. 11.7 (starting on p. 384). Like in that section, we introduce the linear space

$$\mathfrak{Z} := \left\{Z = \{z_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} : z_{\nu,j} \in \mathbb{R}^{k_\nu}\ \forall \nu, j\right\},$$

where l and σ are taken from (12.2.2) (on p. 407) and Assumption 12.4.1 (on p. 417), respectively, and put

$$x(\theta) := 0, \quad u(\theta) := 0, \quad A(\theta) := I, \quad B(\theta) := 0 \qquad \forall \theta \leq -1.$$

Then formula (12.6.1) (on p. 421) makes sense for $j = t+1-\sigma, \ldots, t+1+\overline{\sigma}$ and $t = 0, 1, \ldots$. For any $\tau \geq 0$, we consider the process $\mathfrak{x}(t,\tau)$ that is identical to

[4] Indeed, the tuples $r(t) := [\varkappa(t+1), \varkappa(t), \ldots, \varkappa(t-\overline{\sigma}+1)]$ form such a chain due to Assumption 12.2.8 (on p. 411). At the same time, $\theta(t) = \Theta[r(t)]$ by (12.8.2), where $\Theta[\varkappa_0, \varkappa_1, \ldots, \varkappa_{\overline{\sigma}}] := \theta^-(\varkappa_0) + \theta^+(\varkappa_1)$ and $\theta^\pm(\cdot)$ is defined by (12.2.11) (on p. 410). So in (12.8.7), the coefficients $\mathfrak{A}_{\theta(t)}(t)$ and $\mathfrak{B}_{\theta(t)}(t)$ are deterministic functions of t and $r(t)$.

$\{x(t)\}$ until $t = \tau - \sigma_f + 1$ and then proceeds in correspondence with the system's uncontrolled dynamics. We also denote

$$X(t) := \textbf{col}\left[\mathfrak{x}(t+\overline{\sigma}, t), \ldots, \mathfrak{x}(t-\sigma, t)\right], \quad \Omega(t) := \textbf{col}\left[\omega(t+\overline{\sigma}, t), \ldots, \omega(t-\sigma, t)\right];$$
$$\Delta^\eta(t) := \textbf{col}\left[\delta_\eta(t+\overline{\sigma}+1|t), \ldots, \delta_\eta(t-\sigma+1|t)\right] \qquad \eta = 1,2;$$
$$W(t) := \left\{w_{\nu,j}\right\} \in \mathfrak{Z}, \qquad w_{\nu,j} := \begin{cases} y_\nu(t-j) \text{ if } & (\nu, t-j) \in S(t) \\ 0 & \text{otherwise} \end{cases}. \quad (12.9.1)$$

It is easy to check that the evolution of this vectors is governed by the equations

$$X(t+1) = \mathfrak{A}(t)X(t) + \mathfrak{B}(t)u(t+1-\sigma_f) + \mathfrak{E}\xi(t+\overline{\sigma}), \quad X(0) = \Omega_0; \quad (12.9.2)$$
$$\Omega(t+1) = \mathfrak{A}(t)\Omega(t) + \mathfrak{D}(t)\left[\Delta^1(t) - \Delta^2(t)\right] + \mathfrak{E}\xi(t+\overline{\sigma}), \quad \Omega(0) = \Omega_0; \quad (12.9.3)$$

$$\Delta^1(t+1) = \mathfrak{A}^{11}_\Delta(t)\Delta^1(t) + \mathfrak{A}^{12}_\Delta(t)\Delta^2(t) + \mathfrak{B}^1_\Delta(t)u(t+1-\sigma_f),$$
$$\Delta^1(0) = 0; \quad (12.9.4)$$

$$\Delta^2(t+1) = \mathfrak{A}^{22}_\Delta(t)\Delta^2(t) + \mathfrak{B}^2_\Delta(t)u(t+1-\sigma_f), \quad \Delta^2(0) = 0; \qquad (12.9.5)$$
$$W(t) = \mathfrak{C}\left[t, S(t)\right]X(t) + \Xi\left[t, S(t)\right] \qquad t = 0, \ldots, T. \qquad (12.9.6)$$

Here

$$\mathfrak{A}(t) := \begin{pmatrix} A(t+\overline{\sigma}) & 0 \cdots & 0 & 0 \\ I & 0 \cdots & 0 & 0 \\ 0 & I \cdots & 0 & 0 \\ \cdot & \cdot \cdots & \cdot & \cdot \\ 0 & 0 \cdots & I & 0 \end{pmatrix}, \quad \mathfrak{D}(t) := \begin{cases} \text{the unit matrix if} \\ t+1 = \mathfrak{f}(t+1) \\ t+1 \neq f \\ \text{and the zero matrix} \\ \text{otherwise} \end{cases}, \qquad (12.9.7)$$

where $\mathfrak{f}$ is defined by (12.5.5) (on p. 420) and the size of $\mathfrak{D}(t)$ equals that of $\mathfrak{A}(t)$. Furthermore,

$$\mathfrak{A}^{\eta\eta}_\Delta(t) := \begin{cases} \mathfrak{A}(t+1) & \text{if } \begin{Bmatrix} t \geq \sigma_f + (\eta-1)f - 1 \text{ and} \\ \left[t+1 \neq \mathfrak{f}(t+1) \text{ or } t+1 = f\right] \end{Bmatrix} \\ 0 & \text{if } \begin{Bmatrix} t+1 < \sigma_f + (\eta-1)f \text{ or} \\ t+1 = \nu f \text{ for some } \nu = 2,3,\ldots \end{Bmatrix} \end{cases};$$

$$\mathfrak{E} := \textbf{col}\,(I, 0, \ldots, 0), \quad \mathfrak{A}^{12}_\Delta(t) := \begin{cases} \mathfrak{A}(t+1) & \text{if} \quad t+1 = \mathfrak{f}(t+1),\ t+1 \neq f \\ 0 & \text{otherwise} \end{cases};$$

$$\mathfrak{B}(t) := \mathfrak{B}^{\langle\overline{\sigma}\rangle}(t), \qquad \mathfrak{B}^1_\Delta(t) := \mathfrak{B}^{\langle\overline{\sigma}+1\rangle}(t), \quad \text{where}$$
$$\mathfrak{B}^{\langle s\rangle}(t) := \textbf{col}\left[\mathcal{A}_s, \mathcal{A}_{s-1}, \ldots, \mathcal{A}_{2-\sigma_f}, I, 0, \ldots, 0\right]B(t+1-\sigma_f);$$
$$\mathcal{A}_\eta := \prod_{j=2-\sigma_f}^{\eta} A(t+j); \quad (12.9.8)$$

$$\mathfrak{B}^2_\Delta(t) := \begin{cases} \mathfrak{B}^1_\Delta(t) & \text{if } t+1 \geq \sigma_-(t+1) + f + \sigma_f \\ 0 & \text{otherwise} \end{cases};$$

$$\Omega_0 := \textbf{col}\,[x_0(\overline{\sigma}), x_0(\overline{\sigma}-1), \ldots, x_0(0), 0, \ldots, 0]\,. \tag{12.9.9}$$

Here $x_0(0), x_0(1), \ldots$ are determined recursively

$$x_0(0) := a, \qquad x_0(t+1) = A(t)x_0(t) + \xi(t)$$

and for any

$$t \leq T, S \subset \{(\nu, j) : \nu = 1, \ldots, l, j = 0, \ldots, \sigma\}, X = \textbf{col}\,(x_{-\overline{\sigma}}, x_{-\overline{\sigma}+1}, \ldots, x_\sigma),$$

the following relations hold:

$$\mathfrak{C}[t, S]X := \{w_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} \in \mathfrak{Z}, \text{ where}$$
$$w_{\nu,j} := \begin{cases} C_\nu(t-j)x_j & \text{if } (\nu, t-j) \in S \\ 0 & \text{otherwise} \end{cases}; \tag{12.9.10}$$

$$\Xi[t, S] := \{\zeta_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\sigma} \in \mathfrak{Z}, \quad \text{where}$$
$$\zeta_{\nu,j} := \begin{cases} \chi_\nu(t-j) & \text{if } (\nu, t-j) \in S \\ 0 & \text{otherwise} \end{cases}. \tag{12.9.11}$$

By (12.4.5) (on p. 417),

$$u_i(t_c) = \mathfrak{U}_i[t_c, \mathfrak{Y}(\tau), \mathfrak{S}(\tau), \overline{\theta}(0), \ldots, \overline{\theta}(t_c)] \quad \text{for} \quad t_c := t + 1 - \sigma_f,$$

where $\mathfrak{Y}(\tau), \mathfrak{S}(\tau)$ are defined in (12.2.8) (on p. 409), $\tau := t_c - \theta_i(t_c) \leq t$, and

$$\overline{\theta}(s) := \textbf{col}\,[\theta_1(s), \ldots, \theta_q(s)]\,.$$

Hence in (12.9.2)–(12.9.5),

$$u(t+1-\sigma_f) = \mathfrak{U}\left[t, W(0), \ldots, W(t), \mathfrak{S}(t), \overline{\theta}(0), \ldots, \overline{\theta}(t+1-\sigma_f)\right], \tag{12.9.12}$$

where $\mathfrak{U}(\cdot)$ is a deterministic function. By Assumption 12.2.2 (on p. 410), the random sets $\{S(t)\}$ from (12.4.2) (on p. 415) and the tuples of the delays $\{\overline{\theta}(t)\}$ are independent of a, $\{\xi(t)\}$, and $\{\chi_\nu(t)\}$. Then (12.2.3) (on p. 408) and (12.9.9) imply that the vectors Ω_0, $\{\xi(t+\overline{\sigma})\}$, and $\{\Xi[t, S(t)]\}$ conditioned over

$$S(0), \ldots, S(T-1) \quad \text{and} \quad \overline{\theta}(0), \ldots, \overline{\theta}(T-1)$$

are (singular) Gaussian and independent. This and (12.9.2), (12.9.6), and (12.9.12) imply that the estimate

$$\widehat{X}(t) := \boldsymbol{E}\left[X(t)\Big|W(0), \ldots, W(t), S(0), \ldots, S(T-1), \overline{\theta}(0), \ldots, \overline{\theta}(T-1)\right]$$

is generated by the ordinary Kalman filter (see Subsect. C.2 in Appendix C). By invoking the corresponding formulas (C.8)–(C.13) (on pp. 510 and 511) , it is easy to see that only the sets $S(0), \ldots, S(t)$ (and the delays $\overline{\theta}(0), \ldots, \overline{\theta}(t-\sigma_f)$ if $t \geq \sigma_f$) are used to compute $\widehat{X}(t)$. So in view of (12.4.2) (on p. 415), (12.9.1), and the definition of $\Upsilon(t)$ from (12.6.15) (on p. 424),

$$\widehat{Z}(t) = \boldsymbol{E}\left[Z(t)\middle|\mathfrak{Y}(t), \mathfrak{S}(t), \Upsilon(t)\right],$$

where $\mathfrak{Y}(t)$ and $\mathfrak{S}(t)$ are defined in (12.2.8) (on p. 409). Furthermore (12.9.2) and (12.9.3) yield

$$\Omega(t+1) - X(t+1) = \mathfrak{A}(t)\left[\Omega(t) - X(t)\right] \\ + \underbrace{\mathfrak{D}(t)\left[\Delta^1(t) - \Delta^2(t)\right] - \mathfrak{B}(t)u(t+1-\sigma_f)}_{\mathfrak{Z}^{t+1}}, \qquad \Omega(0) - X(0) = 0,$$

where $\mathfrak{Z}^t$ is a deterministic function of $\mathfrak{Y}(t), \mathfrak{S}(t)$ and $\Upsilon(t)$. Hence

$$\widehat{\Omega}(t+1) - \widehat{X}(t+1) = \mathfrak{A}(t)\left[\widehat{\Omega}(t) - \widehat{X}(t)\right] + \mathfrak{Z}^{t+1}, \quad \widehat{\Omega}(0) - \widehat{X}(0) = 0, \quad (12.9.13)$$

and $X(t) - \widehat{X}(t) = \Omega(t) - \widehat{\Omega}(t)$, where $\widehat{\Omega}(t) := \boldsymbol{E}\left[\Omega(t)\middle|\mathfrak{Y}(t), \mathfrak{S}(t), \Upsilon(t)\right]$. By the Kalman filter formulas (C.8)–(C.13), $\widehat{X}(t)$ obeys the first equation from (12.9.2) where the noise $\mathfrak{E}\xi(t+\overline{\sigma})$ is replaced by a correction term δ generated by the Kalman filter. Thanks to (12.9.13), $\widehat{\Omega}(t)$ obeys the first equation from (12.9.3) where the noise is replaced by the same term δ.

The statement of Theorem 12.6.2 (on p. 424) results from putting the expressions (12.9.1)–(12.9.11) into the equations (C.8)–(C.13) (on pp. 510 and 511) of the conventional Kalman filter, along with elementary transformations of the resultant formulas and taking into account (iv) of Theorem C.2 (on p. 511) in Appendix C.

12.10 Proofs of Theorem 12.7.1 and Proposition 12.7.2

From now on, Assumptions 12.2.1, 12.2.2, and 12.4.1–12.4.4 (on pp. 409, 410, 417, and 418) are supposed to hold. This section is organized as follows. Its first and second subsections contain auxiliary lemmas and the proof of Theorem 12.7.1 (on p. 426) in the case of a single subsystem, respectively. The proof of this theorem in the general case is offered in the third subsection, where Proposition 12.7.2 (on p. 427) is also proved.

12.10.1 Preliminaries

We recall that $\theta_i(t)$ is the time elapsed since the departure of the control message currently employed by the ith subsystem. If no message has arrived until t, then $\theta_i(t) := t + 1$. Furthermore, $\overline{\theta}(t) := \textbf{col}\left[\theta_1(t), \ldots, \theta_q(t)\right]$.

Lemma 12.10.1. *For any* $i = 1,\dots,q$ *and* $t = 0,\dots,T$*, the inequality* $\theta_i(t) \le \sigma_i + \mu_i - 1$ *holds. Here* σ_i *and* μ_i *are the constants from Assumptions* 12.4.1 *and* 12.4.2 *(on pp.* 417 and 417*), respectively.*

Proof. Put $s := t - \theta_i(t)$. Assumption 12.4.2 ensures that among the messages $\mathfrak{m}(\widetilde{s})$ dispatched via the ith control channel at times $\widetilde{s} = s+1,\dots,s+\mu_i$, there is at least one $\mathfrak{m}' = \mathfrak{m}(s')$ that reaches the ith subsystem. Let t' denote its arrival time. If $s = -1$, the subsystem has received no message until t and thus $t' > t$. Suppose that $s \ge 0$ and denote by $\mathfrak{m}$ the message employed by the subsystem at the time t. It was departed at $s < s'$, and so it is less updated than $\mathfrak{m}'$. Employed is the most updated message arrived until t. Thus $t' > t$; i.e., this inequality holds in any case. By Assumption 12.4.1 (on p. 417), $t' - s' \le \sigma_i$. Thus

$$s + \theta_i(t) = t < t' \le s' + \sigma_i \le s + \mu_i + \sigma_i \Rightarrow \theta_i(t) < \mu_i + \sigma_i,$$

which completes the proof. □

Lemma 12.10.2. *Suppose that a control strategy* (12.4.5) *(on p.* 417*) is chosen. Consider the corresponding process in the system* (12.2.1), (12.2.2) *(on p.* 407*), the output* (12.5.8) *(on p.* 420*) of the state estimator described in Sect.* 12.6*, and the minimum variance estimates*

$$\widehat{x}_i(t) := \boldsymbol{E}\left[x_i(t)\middle|\mathfrak{Y}(\tau), \mathfrak{S}(\tau), \overline{\theta}(0),\dots,\overline{\theta}(t)\right]; \tag{12.10.1}$$

$$\widehat{x}_{0|i}(t) := \boldsymbol{E}\left[x_0(t)\middle|\mathfrak{Y}(\tau), \mathfrak{S}(\tau), \overline{\theta}(0),\dots,\overline{\theta}(t)\right], \tag{12.10.2}$$

where $\tau := t - \theta_i(t)$*. Then for any* $i = 1,\dots,q, t = 0,\dots,T$*, we have*

$$\widehat{x}_{0|i}(t) = \widehat{\omega}_0[t|\tau], \qquad \widehat{x}_i(t) = \widehat{\omega}_i[t|\tau] + \begin{cases} \overline{x}_i^{(0)}(t) & \text{if} \quad t < f \\ \overline{x}_i^{(2)}(t) & \text{if} \quad f \le \tau < \mathfrak{f}(t) \\ \overline{x}_i^{(1)}(t) & \text{otherwise} \end{cases}. \tag{12.10.3}$$

Here **col** $[\widehat{\omega}_0(j|t),\dots,\widehat{\omega}_q(j|t)]$ *is the partition of* $\widehat{\omega}(j|t)$ *corresponding to* (12.4.1) *(on p.* 415*) and* $\mathfrak{f}(t), \overline{x}_i^{(\eta)}(t)$ *are defined by* (12.5.5) *and* (12.5.6) *(on p.* 420*), respectively.*

Proof. Pick $i = 1,\dots,q$ and put $\widehat{x}_0(t) := \widehat{x}_{0|i}(t)$. Assumption 12.4.3 (on p. 418) and (12.5.2) (on p. 419) yield

$$x_i(t) = \omega_i(t,\tau) + \lambda_i(t,\tau), \quad \text{where} \quad \lambda_i(t,\tau) := \sum_{r=\sigma_-(\tau)}^{t-1} \prod_{\nu=r+1}^{t-1} A_i(\nu) B_i(r) u_i(r)$$

for $i \ge 1$, $\lambda_0(t,\tau) := 0$, and $\tau := t - \theta_i(t)$. By (12.4.5) (on p. 417), $\lambda_i(t,\tau)$ is a deterministic function of $\mathfrak{Y}(\tau)$, $\mathfrak{S}(\tau)$, and $\overline{\theta}(0),\dots,\overline{\theta}(t)$. This and (12.10.1) give

$$\widehat{x}_i(t) = \boldsymbol{E}\left\{\omega_i[t,\tau]\middle|\mathfrak{Y}(\tau), \mathfrak{S}(\tau), \overline{\theta}(0),\dots,\overline{\theta}(t)\right\} + \lambda_i(t,\tau).$$

Due to (12.2.1), (12.2.3), (12.4.5), and (12.7.5) (on pp. 407, 408, 417, and 425),

$$\omega_i(t,\tau) = \Omega_i\left[t,\tau,\Xi,\mathfrak{Y}(\tau),\mathfrak{S}(\tau),\Upsilon(\tau)\right],$$

where $\Xi := \left[a, \{\xi(s)\}_{s=0}^{t-1}\right]$, the quantity $\Upsilon(\tau)$ is defined in (12.6.15) (on p. 424), and $\Omega_i(\cdot)$ is a deterministic function. Here the quantities $\{\overline{\theta}(s)\}$ are independent of Ξ and $\mathfrak{H}(s') := [\mathfrak{Y}(s'), \mathfrak{S}(s')]$ by Assumption 12.2.2 (on p. 410). So the following relation holds for the conditional distributions:

$$p\left\{d\Xi|\mathfrak{H}[\tau],\overline{\theta}(0),\dots,\overline{\theta}(t)\right\} = p\left\{d\Xi|\mathfrak{H}(s)\right\}\Big|_{s=t-\theta_i(t)}$$
$$= p\left\{d\Xi|\mathfrak{H}(s),\Upsilon(s)\right\}\Big|_{s=t-\theta_i(t)}.$$

Therefore

$$\begin{aligned}
&\boldsymbol{E}\left\{\omega_i\left[t,\tau\right]\Big|\mathfrak{H}(\tau),\overline{\theta}(0),\dots,\overline{\theta}(t)\right\} \\
&\quad= \boldsymbol{E}\left\{\Omega_i\left[t,\tau,\Xi,\mathfrak{H}(\tau),\Upsilon(\tau)\right]\Big|\mathfrak{H}(\tau),\overline{\theta}(0),\dots,\overline{\theta}(t)\right\} \\
&\quad= \int \Omega_i\left[t,\tau,\Xi,\mathfrak{H}(\tau),\Upsilon(\tau)\right] p[d\Xi|\mathfrak{H}(\tau),\overline{\theta}(0),\dots,\overline{\theta}(t)] \\
&\quad= \int \Omega_i\left[t,s,\Xi,\mathfrak{H}(s),\Upsilon(s)\right] p[d\Xi|\mathfrak{H}(s),\Upsilon(s)]\Big|_{s=t-\theta_i(t)} \\
&\quad= \boldsymbol{E}\left\{\omega_i\left[t,s\right]\Big|\mathfrak{Y}(s),\mathfrak{S}(s),\Upsilon(s)\right\}\Big|_{s=t-\theta_i(t)} \overset{(12.6.14)}{=\!=} \widehat{\omega}_i[t|t-\theta_i(t)].
\end{aligned}$$

Then (12.10.3) is immediate from (12.5.5), (12.5.6), (12.7.5) (on pp. 420 and 425), and Lemma 12.10.1. □

12.10.2 Proof of Theorem 12.7.1 on p. 426: Single Subsystem

In this case, the subsystem's serial number i can take only one value $i = 1$. However in the formulas to follow, we employ the general notation i instead of using merely 1. These formulas can actually be considered for any i in the general case of many subsystems and will be utilized in the next subsection, where this case is discussed.

We are going to show first that the problem can be reduced to the standard LQG optimal control problem. To this end, we denote $\widetilde{\sigma} := \sigma_i + \mu_i + \sigma - 1$, where σ, σ_i, and μ_i are taken from Assumptions 12.4.1 and 12.4.2 (on pp. 417 and 417), consider the linear space

$$\widehat{\mathfrak{Z}} := \left\{Z = \{z_{\nu,j}\}_{\nu=1}^{l}{}_{j=0}^{\widetilde{\sigma}} : z_{\nu,j} \in \mathbb{R}^{k_\nu}\ \forall \nu, j\right\},$$

where l is taken from (12.2.2) (on p. 407) and k_ν is the dimension of the sensor output y_ν, and put

$$x(-\widetilde{\sigma}) := x(-\widetilde{\sigma}+1) := \cdots := x(-1) := 0, s_i(-1) := 0;$$

$$X(t) := \textbf{col}\,\big[x(t), x(t-1), \ldots, x(t-\widetilde{\sigma})\big], \qquad \widetilde{W}(t) := \big\{\widetilde{w}_{\nu,j}\big\} \in \widehat{\mathfrak{Z}};$$

$$\widetilde{w}_{\nu,j} := \begin{cases} y_\nu(t-j) \text{ if } & (\nu, t-j) \in S[\tau_i(t)] \quad \text{and} \quad \tau_i(t) > \tau_i(t-1) \\ 0 & \text{otherwise} \end{cases}, \tag{12.10.4}$$

where $\tau_i(t) := t - \theta_i(t)$. In terms of these vectors, the primal problem shapes into

$$\text{minimize} \quad \boldsymbol{E} \sum_{t=0}^{T-1} \left[X(t+1)^{\mathrm{T}} \mathfrak{Q}(t+1) X(t+1) + u_i(t)^{\mathrm{T}} \Gamma_i(t) u_i(t)\right] \quad \text{subject to} \tag{12.10.5}$$

$$X(t+1) = \widetilde{\mathfrak{A}}(t) X(t) + \widetilde{\mathfrak{B}}(t) u_i(t) + \mathfrak{E}\,\xi(t) \qquad t = 0, \ldots, T-1; \tag{12.10.6}$$

$$X(0) = \mathfrak{a} := \textbf{col}\,(a, 0, \ldots, 0); \tag{12.10.7}$$

$$\widetilde{W}(t) = \widetilde{\mathfrak{C}}\Big\{t, S[\tau_i(t)], \theta_i(t), \theta_i(t-1)\Big\} X(t) + \widetilde{\Xi}\Big\{t, S[\tau_i(t)], \theta_i(t), \theta_i(t-1)\Big\}; \tag{12.10.8}$$

$$u_i(t) = \mathfrak{U}_i \left\{\widetilde{W}(0), \ldots, \widetilde{W}(t), S(0), S(1), \ldots, S[t-\theta_i(t)], \overline{\theta}(0), \ldots, \overline{\theta}(t)\right\}. \tag{12.10.9}$$

Here $\tau_i(t) := t - \theta_i(t)$ and $\mathfrak{E} := \textbf{col}\,(I, 0, \ldots, 0)$,

$$\widetilde{\mathfrak{A}}(t) := \begin{pmatrix} A(t) & 0 & \ldots & 0 & 0 \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \ldots & I & 0 \end{pmatrix}, \qquad \mathfrak{Q}(t) := \begin{pmatrix} \widetilde{Q}(t) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{12.10.10}$$

where $\widetilde{Q}(t) := Q(t)$ and $Q(t)$ is taken from (12.2.6) (on p. 408). Furthermore,

$$\widetilde{\mathfrak{B}}(t) := \textbf{col}\,\big[\widetilde{B}(t), 0, \ldots, 0\big],$$

where the matrix $\widetilde{B}(t) := \textbf{col}\,[0, B_1(t)]$ describes the influence of the control u_1 on the process $x = \textbf{col}\,[x_0, x_1]$ and $B_1(t)$ is taken from (12.4.7) (on p. 418). For any $t \leq T$,

$$S \subset \{(\nu, j) : \nu = 1, \ldots, l, j = 0, \ldots, \widetilde{\sigma}\}, \quad \theta' \geq \theta'',$$

and $X = \textbf{col}\,(x_0, x_1, \ldots, x_{\widetilde{\sigma}}) \in \mathbb{R}^{(\widetilde{\sigma}+1)n}$, the following relations hold:

$$\left.\begin{array}{r} \widetilde{\mathfrak{C}}[t, S, \theta', \theta'']Z := \mathfrak{C}[S]X \\ \widetilde{\Xi}[t, S, \theta', \theta''] := \Xi[t, S] \end{array}\right\} \quad \text{if} \quad t - \theta' > t - 1 - \theta'',$$

otherwise,

$$\begin{cases} \widetilde{\mathfrak{C}}[t, S, \theta', \theta'']Z := 0 \\ \widetilde{\Xi}[t, S, \theta', \theta''] := 0 \end{cases},$$

where $\mathfrak{C}[S]X$ and $\Xi[t, S]$ are defined by (12.9.10) and (12.9.11) (on p. 437) with $\sigma := \widetilde{\sigma}$. The random sets $\{S(j)\}$ and delays $\{\theta_i(j)\}$ are independent of $a, \{\xi(j)\}$, $\{\chi_\nu(j)\}$ by Assumption 12.2.2 (on p. 410). So by (12.2.3) and (12.2.5) (on p. 408), the vectors $a, \xi(t)$, and $\widehat{\Xi}\{t, S[\tau_i(t)], \theta_i(t), \theta_i(t-1)\}$ conditioned over $S(0), \ldots$, $S(T-1), \overline{\theta}(0), \ldots, \overline{\theta}(T-1)$ are Gaussian and independent. This yields that the problem resulting from (12.10.5)–(12.10.9) by expansion the class of admissible controls (12.10.9) as follows:

$$u_i(t) = \mathfrak{U}_i \left\{ \widetilde{W}(0), \ldots, \widetilde{W}(t), S(0), S(1), \ldots, S(T-1), \overline{\theta}(0), \ldots, \overline{\theta}(T-1) \right\}$$

in fact reduces to a particular case of the standard LQG optimal control problem (see Appendix C starting on p. 509). The solution of the latter problem is given by the linear feedback $u_i(t) = -\mathfrak{L}(t)X(t)$ that is optimal for the problem of minimizing the functional (12.10.5) subject to the constraints (12.10.6) in the case where $\xi(t) \equiv 0$ and the entire state X is measured. This feedback is fed by the minimum variance estimate $\widehat{X}(t|t)$ of $X(t)$ based on the observations $\widetilde{W}(0), \ldots, \widetilde{W}(t)$. This estimate is generated recursively by the Kalman filter (see Appendix C) so that only the sets $S[0], \ldots, S[t-\theta_i(t)]$ and the tuples $\overline{\theta}(0), \ldots, \overline{\theta}(t)$ from (12.10.8) are in fact required to compute $\widehat{X}(t|t)$. This in particular means that

$$\widehat{X}(t|t) = \boldsymbol{E}\left(X[t] \Big| \left\{ \widetilde{W}[j] \right\}_{j=0}^{t}, \left\{ S[j] \right\}_{j=0}^{t-\theta_i(t)}, \left\{ \overline{\theta}[j] \right\}_{j=0}^{t} \right). \tag{12.10.11}$$

Calculating the gain matrix $\mathfrak{L}(t)$ employs the coefficients of the functional (12.10.5) and the equation (12.10.6) for the time instants $t, t+1, \ldots, T$. Since these coefficients do not depend on $\{S(j)\}, \{\overline{\theta}(j)\}$ and are known a priori, the control appears to be of the form (12.10.9). So it furnishes the optimum over the strategies given by (12.10.9).

By invoking the formulas for the solution of the standard linear-qudratic control problem (i.e., formulas (C.18)–(C.21) on p. 512), it is easy to check that the above feedback $u_i(t) = -\mathfrak{L}(t)X(t)$ has the form $u_i(t) = -\mathfrak{L}_*(t)x(t)$, where $X(t)$ is the tuple from (12.10.4) and $u_i(t) = -\mathfrak{L}_*(t)x(t)$ is the optimal feedback for the problem of minimizing the functional (12.2.6) (on p. 408) subject to the constraints (12.2.1) (on p. 407) in the case where $\xi(t) \equiv 0$ in (12.2.1) and the entire state $x =$ **col** $[x_0, x_1]$ is measured. Likewise, it is straightforward to verify that the last feedback looks as follows:

$$u_i(t) = -L_i(t)x_i(t) - \overline{L}_i(t)x_0(t),$$

where $L_i(t), \overline{L}_i(t)$ are defined by (12.7.1)–(12.7.4) (on p. 424). Summarizing we see that the solution for the initial problem is given by

$$u_i(t) = -L_i(t)\hat{\hat{x}}_i(t) - \overline{L}_i(t)\hat{\hat{x}}_0(t),$$

where

$$\hat{\hat{x}}_j := \boldsymbol{E}\{x_j(t)|\widetilde{W}[0], \ldots, \widetilde{W}[t], S[0], \ldots, S[t-\theta_i(t)], \overline{\theta}(0), \ldots, \overline{\theta}(t)\}.$$

Invoking (12.2.3) (on p. 408) and (12.10.4) demonstrates that

$$\hat{\hat{x}}_j := \boldsymbol{E}\{x_j(t)|\mathfrak{Y}[\tau], \mathfrak{S}[\tau], \overline{\theta}(0), \ldots, \overline{\theta}(t)\},$$

where $\tau = t - \theta_i(t)$, and by (12.10.1) and (12.10.2), the optimal feedback takes the form

$$u_i(t) = -L_i(t)\widehat{x}_i(t) - \overline{L}_i(t)\widehat{x}_{0|i}(t).$$

Employing (12.10.3) shapes this formula into

$$u_i(t) = -\Big\{L_i[\tau+\theta_i(t)]\hat{\omega}_i[\tau+\theta_i(t)|\tau] + \overline{L}_i[\tau+\theta_i(t)]\hat{\omega}_0[\tau+\theta_i(t)|\tau]\Big\}$$
$$- L_i(t) \times \begin{cases} \overline{x}_i^{(0)}(t) & \text{if} \quad t < f \\ \overline{x}_i^{(2)}(t) & \text{if} \quad f \leq t - \theta_i(t) < \mathfrak{f}(t) \\ \overline{x}_i^{(1)}(t) & \text{otherwise} \end{cases}. \quad (12.10.12)$$

Now consider the control algorithm described in Sect. 12.7 (starting on p. 424). By the definition of $\theta_i(t)$, the control package currently used by the local controller was departed at the time τ. So by (12.7.7) (on p. 425), the summand in the curly brackets $\{\cdots\}$ from (12.10.12) equals the member $v^{(i)}_{\theta_i(t)}$ of this package. Here $\theta_i(t) = \tau_i + s_i$, where τ_i and s_i are the numbers generated by the ith local controller. It follows that (12.10.12) takes the form (12.7.8) (on p. 426). □

12.10.3 Proofs of Theorem 12.7.1 and Proposition 12.7.2: Many Subsystems

We start with a lemma that was in fact proved in the case of a single subsystem in the preceding subsection.

Lemma 12.10.3. *Let $i = 1, \ldots, q$ be given, and let any subsystem except for the ith one be controlled in accordance with some given strategy of the form* (12.4.5) *(on p.* 417*). Consider the problem* (12.7.9) *(on p.* 426*) of optimal control of the ith subsystem, where the control strategy is sought in the class* (12.4.5) *with i fixed. The solution for this problem is given by the feedback*

$$u_i(t) = -L_i(t)\widehat{x}_i(t) - \overline{L}_i(t)\widehat{x}_{0|i}(t). \quad (12.10.13)$$

Here $\widehat{x}_i(t)$, $\widehat{x}_{0|i}(t)$ are the estimates (12.10.1)*,* (12.10.2) *and $L_i(t), \overline{L}_i(t)$ are defined by* (12.7.1)–(12.7.4) *(on p.* 425*).*

Proof. For $j = 1, \ldots, q$, consider the matrix

$$\widetilde{B}_j(t) := \mathbf{col}\,[0, \ldots, 0, B_j(t), 0, \ldots, 0]$$

that describes the influence of the jth control u_j on the process $x = \mathbf{col}\,[x_0, \ldots, x_q]$. Here the matrix $B_j(t)$ is taken from (12.4.7) (on p. 418) and occupies the $(j+1)$th position. Introduce also the process $x_*(t)$ generated by the equations

$$x_*(t+1) = A(t)x_*(t) + \sum_{j\neq i} \widetilde{B}_j(t)u_j(t), \qquad x_*(0) = 0. \tag{12.10.14}$$

In terms of the difference $\mathfrak{x}(t) := x(t) - x_*(t)$, relations (12.2.1) and (12.2.2) (on p. 407) take the form:

$$\mathfrak{x}(t+1) = A(t)\mathfrak{x}(t) + \widetilde{B}_i(t)u_i(t) + \xi(t) \quad t = 0,\ldots,T-1, \quad \mathfrak{x}(0) = a; \tag{12.10.15}$$

$$\mathfrak{y}_\nu(t) := y_\nu(t) - C_\nu(t)x_*(t) = C_\nu(t)\mathfrak{x}(t) + \chi_\nu(t), \tag{12.10.16}$$

where $\nu = 1,\ldots,l$ and $t = 0,\ldots,T$. Put

$$\widetilde{Y}(t) := \{\mathfrak{y}_\nu[\theta]\}_{(\nu,\theta)\in S(t)}, \quad \widetilde{\mathfrak{Y}}(t) := [\widetilde{Y}(0),\ldots,\widetilde{Y}(t)].$$

We are going to show first that

$$\widetilde{\mathfrak{Y}}(t) = \widetilde{\mathcal{Y}}[\mathfrak{Y}(t),\mathfrak{F}(t)], \qquad \mathfrak{Y}(t) = \mathcal{Y}\left[\widetilde{\mathfrak{Y}}(t),\mathfrak{F}(t)\right], \tag{12.10.17}$$

where $\mathfrak{F}(t) := [\mathfrak{S}(t), \overline{\theta}(0),\ldots,\overline{\theta}(t-1), t]$ and $\mathfrak{Y}(t)$ is defined in (12.2.8) (on p. 409). (We recall that capital script letters denote deterministic functions.) For $t = 0$, (12.10.17) is evident since $x_*(0) = 0$ by (12.10.14). Let (12.10.17) be true for some t. If $S(t+1) = \emptyset$, then $Y(t+1) = \widetilde{Y}(t+1) = 0 \in \mathbb{R}$ and so (12.10.17) does hold for $t := t+1$. Suppose that $S(t+1) \neq \emptyset$. Due to (12.4.5) (on p. 417) and (12.10.14),

$$x_*(s) = \mathcal{X}_*[\mathfrak{Y}(s-1), \mathfrak{S}(s-1), \overline{\theta}(0),\ldots,\overline{\theta}(s-1), s].$$

Hence for $s \leq t+1$,

$$x_*(s) = \overline{\mathcal{X}}\left[\mathfrak{Y}(t), \mathfrak{S}(t), \overline{\theta}(0),\ldots,\overline{\theta}(t), s\right] \overset{(12.10.17)}{=\!=\!=} \widetilde{\mathcal{X}}\left[\widetilde{\mathfrak{Y}}(t), \mathfrak{S}(t), \overline{\theta}(0),\ldots,\overline{\theta}(t), s\right].$$

This and the definition of $\mathfrak{y}_\nu(t)$ from (12.10.16) clearly imply (12.10.17) for $t := t+1$.

Relations (12.10.17) shape (12.4.5) (on p. 417) into

$$u_i(t) = \mathcal{U}\left[t, \widetilde{\mathfrak{Y}}(\tau), \mathfrak{S}(\tau), \overline{\theta}(0),\ldots,\overline{\theta}(t)\right] \qquad \tau := t - \theta_i(t). \tag{12.10.18}$$

Let $\mathfrak{x} =$ **col** $(\mathfrak{x}_0,\ldots,\mathfrak{x}_q)$ be the partition corresponding to (12.4.1) (on p. 415). For the similar partition $x_*(t) =$ **col** $[x_{*,0}(t),\ldots,x_{*,q}(t)]$, Assumption 12.4.3 (on p. 418) and (12.10.14) imply that $x_{*,j}(t) = 0$ for $j = 0, i$. Hence

$$\mathfrak{x}_0(t) = x_0(t), \qquad \mathfrak{x}_i(t) = x_i(t) \qquad \forall t. \tag{12.10.19}$$

So in view of (12.4.8) (on p. 418), the problem (12.7.9) (on p. 426) can be rewritten in terms of the vector $\mathfrak{x}(t)$:

$$\text{minimize} \quad \boldsymbol{E}\sum_{t=0}^{T-1} \mathcal{G}_i[t, \mathfrak{x}(t+1), u_i(t)]$$

$$\text{subject to (12.10.15) and (12.10.16)} \tag{12.10.20}$$

over the control strategies (12.10.18). Here $\mathcal{G}_i(\cdot)$ is taken from (12.4.8) (on p. 418).

This problem can be reduced to a particular case of the standard LQG optimal control problem just as it was done in Subsect. 12.10.2. Now one, however, should replace (12.10.4) by

$$X(t) := \textbf{col}\left[\mathfrak{x}(t), \mathfrak{x}(t-1), \ldots, \mathfrak{x}(t-\widetilde{\sigma})\right], \qquad \widetilde{W}(t) := \left\{\widetilde{w}_{\nu,j}\right\} \in \mathfrak{Z};$$
$$\widetilde{w}_{\nu,j} := \begin{cases} \mathfrak{y}_\nu(t-j) \text{ if } & \begin{cases} (\nu, t-j) \in S[\tau_i(t)] \quad \text{and} \\ \tau_i(t) > \tau_i(t-1) \end{cases} \\ 0 & \text{otherwise} \end{cases}, \quad (12.10.21)$$

where $\tau_i(t) := t - \theta_i(t)$, put

$$\mathfrak{x}(-\widetilde{\sigma}) := \mathfrak{x}(-\widetilde{\sigma}+1) := \cdots := \mathfrak{x}(-1) := 0, \theta_i(-1) := 0,$$

and define the matrices $\widetilde{\mathfrak{B}}(t)$ in (12.10.6) and $\widetilde{Q}(t)$ in (12.10.10) by the formulas

$$\widetilde{Q}(t) := \begin{pmatrix} 0 & \ldots & \overline{Q}_i(t) & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \overline{Q}_i(t)^{\mathrm{T}} & \ldots & Q_i(t) & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \ldots & 0 & \ldots & 0 \end{pmatrix}, \qquad \widetilde{\mathfrak{B}}(t) := \begin{pmatrix} \widetilde{B}_i(t) \\ 0 \\ \cdot \\ 0 \end{pmatrix}.$$

Here the partition of the $n \times n$ matrix $\widetilde{Q}(t)$ corresponds to the partition of the state from (12.4.1), all three nonzero blocks of $\widetilde{Q}(t)$ belong to the ith row and column, and the enumeration of the rows and columns starts with 0.

By retracing the arguments from Subsect. 12.10.2, and invoking (12.10.19),we see that the solution for the problem under consideration is given by the feedback

$$u_i(t) = -L_i(t)\hat{\hat{x}}_i(t) - \overline{L}_i(t)\hat{\hat{x}}_0(t),$$

where

$$\hat{\hat{x}}_j := \boldsymbol{E}\{x_j(t)|\widetilde{W}[0], \ldots, \widetilde{W}[t], S[0], \ldots, S[t-\theta_i(t)], \overline{\theta}(0), \ldots, \overline{\theta}(t)\}.$$

Invoking (12.2.3) (on p. 408) and (12.10.21) demonstrates that

$$\hat{\hat{x}}_j := \boldsymbol{E}\{x_j(t)|\widetilde{\mathfrak{Y}}[\tau], \mathfrak{S}[\tau], \overline{\theta}(0), \ldots, \overline{\theta}(t)\},$$

where $\tau = t - \theta_i(t)$. Relations (12.10.1), (12.10.2), and (12.10.17) complete the proof. □

Remark 12.10.4. For each i, the auxiliary problem considered in Lemma 12.10.3 was introduced under the assumption that the control strategies for all subsystems except for the ith one are given. The minimum value of the cost functional in this problem does not depend on these strategies. This follows from, e.g., the fact that the problem can be rewritten in the form (12.10.5)–(12.10.9), which is invariant with respect to the above strategies.

Proofs of Theorem 12.7.1 (on p. 426) *and Proposition 12.7.2* (on p. 427). Lemma 12.10.3 and Remark 12.10.4 imply that the optimal way to control the ith subsystem is given by (12.10.13). This is true for not only the auxiliary problem (12.7.9) (on p. 426) associated with only this subsystem but also for the primal one concerning the entire set of subsystems. Denote $\tau := t - \theta_i(t)$. Then relations (12.10.3) shape (12.10.13) into (12.10.12). Now consider the control algorithm described in Sect. 12.7. By the definition of $\theta_i(t)$, the control package currently used by the ith local controller was departed at the time τ. So by (12.7.7) (on p. 425), the addend in the curly brackets from (12.10.12) equals the member $v^{(i)}_{\theta_i(t)}$ of this package. Here $\theta_i(t) = \tau_i + s_i$, where τ_i and s_i are the numbers generated by the ith local controller. It follows that (12.10.12) takes the form (12.7.8) (on p. 426). □

13

Linear-Quadratic Gaussian Optimal Control via Limited Capacity Communication Channels

13.1 Introduction

In this chapter, we proceed with studying optimal control problems in the case where controls are communicated via limited bandwidth channels. A situation of such a kind has been addressed in Sect. 12.4. However the issue of limited bandwidth was taken into account only implicitly to motivate distribution of control functions between the central and local controllers. The objective of this chapter is to address this issue explicitly. In other words, now examined is the performance best in the case where the bit-rate of the control channel is finite and given. Another distinction is that now we examine the case where the actuators are basically capable only to execute the currently received control signal, whereas actuators endowed with rather powerful computing modules were considered in Sect. 12.4. We also drop the system disintegration assumptions from Sect. 12.4 and deal with general linear plants.

Specifically, we study a finite-horizon linear-quadratic optimal control problem for a discrete-time partially observed system with Gaussian disturbances. The controller produces a control signal based on the prior observations. Unlike the classic linear-quadratic Gaussian (LQG) control problems (see Appendix C starting on p. 509), this signal is communicated to the actuators over a digital channel. This channel is capable of transferring no more than a given number of bits of information per unit time. So on its way to the actuator, the control signal must be first encoded, then transmitted, and finally decoded (see Fig. 13.1). The algorithms of encoding and decoding are part of the control strategy, which should be designed to achieve the best performance. The focus is on memoryless (static) decoders. This corresponds to the case where actuators are not equipped with computing and memory modules. The controller has access to the sensor data with arbitrarily high accuracy, which may be due to, e.g., the fact that it is colocated with the sensors. Unlike Chaps. 11 and 12, both observation and control channels provide no delays and dropouts.

We prove that the optimal strategy does exist (i.e., the minimum of the cost functional is attained) and demonstrate how this strategy can be designed. We also prove that the current optimal control codeword transmitted over the channel is determined on the basis of the minimum variance estimate of the current state. In this sense, this

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_13,

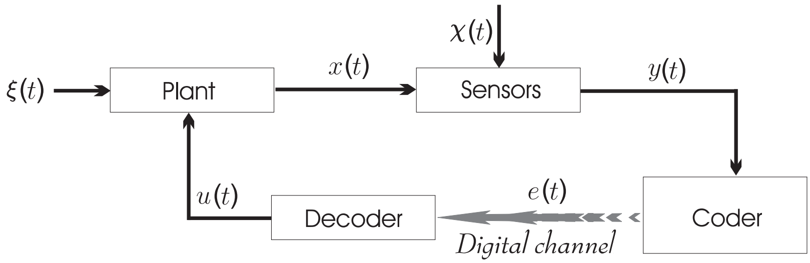


Fig. 13.1. Control system with a limited capacity channel.

estimate remains to be a "sufficient statistics," as in the classic LQG control theory. However the natural hypothesis that the optimal control strategy always consists in generating the standard LQG optimal control and its proper encoding (and decoding after transmission) fails to be true. This is proved by a counterexample. It shows that the best performance may be neither achieved nor approached if the codeword to be transmitted over the channel is generated from the knowledge of exclusively LQG optimal controls.

The main results of this chapter were originally presented in [110].

The body of the chapter is as follows. Section 13.2 presents the problem statement. Its solution is offered in Sect. 13.5. It is prefaced with Sects. 13.3 and 13.4. The former recalls some required classic results, whereas the latter shows that in general, the LQG optimal control and the minimum variance state estimate is not and is, respectively, a "sufficient statistics" for the coder. Sections 13.6 and 13.7 contain the proofs of the results of the chapter.

13.2 Problem Statement

The System

Consider the following discrete-time linear system:

$$x(t+1) = A(t)x(t) + B(t)u(t) + \xi(t) \quad t = 0, \dots, T-1, \qquad x(0) = a; \quad (13.2.1)$$

$$y(t) = C(t)x(t) + \chi(t) \qquad t = 0, \dots, T. \quad (13.2.2)$$

Here $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ are the state and control, respectively; $\xi(t) \in \mathbb{R}^n$ is a process disturbance; $y(t) \in \mathbb{R}^k$ is the measured output; and $\chi(t)$ is a measurement noise. The vectors $a, \xi(t)$, and $\chi(t)$ are random and satisfy the following.

Assumption 13.2.1. *The vectors $a, \xi(t)$, and $\chi(t)$ are Gaussian and independent with $\boldsymbol{E}\xi(t) = 0$ and $\boldsymbol{E}\chi(t) = 0$. The mean $\boldsymbol{E}a$ and the correlation matrices*

$$R_{aa} := \boldsymbol{E}[a - \boldsymbol{E}a][a - \boldsymbol{E}a]^{\mathrm{T}}, \quad R_{\xi\xi}(t) := \boldsymbol{E}\xi(t)\xi(t)^{\mathrm{T}}, \quad R_{\chi\chi}(t) := \boldsymbol{E}\chi(t)\chi(t)^{\mathrm{T}} \quad (13.2.3)$$

are known. So are the matrices $A(t), B(t)$, and $C(t)$.

The Control Loop

The actuators are situated at a remote location, where the control signals are sent over a digital communication channel. It is able to transfer no more than a given number b of bits per unit time. So the control must be in fact generated as an b-bit codeword $e(t)$. This is done by the *controller-coder* on the basis of the prior observations:

$$e(t) = \mathcal{E}[t, y(0), \dots, y(t)] \in \mathfrak{E}. \tag{13.2.4}$$

(We recall that capital script letters denote deterministic functions.) Here $\mathfrak{E} = \{e\}$ is the channel's *alphabet*. Since the "physical" nature of the elements e now is of no importance, we assume that $\mathfrak{E} = [1 : N]$, where $N := 2^b$.[1] The code-symbol $e(t)$ is transmitted and then translated by a *decoder* into the "final" control signal $u(t)$ that acts upon the plant (see Fig. 13.1). We consider the case of a memoryless decoder:

$$u(t) = \mathcal{U}[t, e(t)] \in \mathbb{R}^m. \tag{13.2.5}$$

Thus the control strategy consists of the coder (13.2.4) and the decoder (13.2.5).

LQG Optimal Control Problem with Communication Capacity Constraints

Introduce the functional

$$\mathcal{I}_T := \boldsymbol{E} \sum_{t=0}^{T-1} \Big[x(t+1)^{\mathrm{T}} Q(t+1) x(t+1) + u(t)^{\mathrm{T}} \Gamma(t) u(t) \Big]. \tag{13.2.6}$$

Here $Q(t+1) \geq 0$ and $\Gamma(t) > 0$ are symmetric $n \times n$ and $m \times m$ matrices, respectively. In this chapter, we consider the following problem.

Let the bit-rate of the channel b be given. We wish to minimize the functional (13.2.6) over the class of control strategies described by (13.2.4) and (13.2.5).

In doing so, we consider only measurable functions $\mathcal{E}(\cdot)$ in (13.2.4).

Remark 13.2.2. Putting (13.2.4) into (13.2.5) yields that

$$u(t) = \mathcal{V}[t, y(0), \dots, y(t)].$$

For given t, the function $\mathcal{V}[t, \cdot]$ can take any values from $\mathbb{R}^m$. However, their number must not exceed a given constant N. This makes the difference with the standard setup of the stochastic optimal control problem (see, e.g., [11, 19, 85, 123, 198]).

13.3 Preliminaries

The posed problem resembles the standard LQG optimal control problem[2] in many respects. So it does not come as a surprise that the solutions of these two problems have much in common. For the convenience of the reader, we repeat the relevant formulas concerning the standard LQG problem from Appendix C in this section.

---

[1] We do not suppose further that b is natural. So $N$is not necessarily a natural power of 2.

[2] See Subsect. C.1 in Appendix C.

Minimum Variance State Estimate

Let a control strategy (13.2.4), (13.2.5) be given. Then the state estimate

$$\widehat{x}(t|t) = \boldsymbol{E}\big[x(t)\big|y(0),\dots,y(t)\big] \tag{13.3.1}$$

can be generated recursively by the Kalman filter:

$$\widehat{x}(t+1|t+1) := \widehat{x}(t+1|t) + K(t+1)\Big[y(t+1) - C(t+1)\widehat{x}(t+1|t)\Big]; \tag{13.3.2}$$

$$P(t+1) := A(t)\overline{P}(t)A(t)^{\mathrm{T}} + R_{\xi\xi}(t), \quad \text{where}$$
$$\widehat{x}(t+1|t) := A(t)\widehat{x}(t|t) + B(t)u(t), \quad \overline{P}(t) := P(t) - K(t)C(t)P(t); \tag{13.3.3}$$

$$K(t) := P(t)C(t)^{\mathrm{T}}\overset{+}{\Lambda}(t), \qquad \Lambda(t) := \Lambda := C(t)P(t)C(t)^{\mathrm{T}} + R_{\chi\chi}(t), \tag{13.3.4}$$

and $\overset{+}{\Lambda}$ is the pseudoinverse of the matrix Λ (see Subsect. 11.3.1 starting on p. 374).

The estimation process is initialized by putting

$$\widehat{x}(-1|-1) := \boldsymbol{E}a, \quad P(0) = R_{aa},$$
$$u(-1) := 0, \quad B(-1) := 0, \quad A(-1) := I_n. \tag{13.3.5}$$

Solution for the Deterministic Linear-Quadratic Optimal Control Problem

We mean the problem of minimizing the functional (13.2.6) subject to the constraints (13.2.1) in the case where the process disturbances are removed ($\xi(t) \equiv 0$) and the entire state x is accessible for "on-line" measurements. The solution for this problem is given by the feedback $u(t) = -L(t)x(t)$, where

$$L(t) = F(t)^{-1}B(t)^{\mathrm{T}}\left[Q(t+1) + H(t+1)\right]A(t),$$
$$F(t) := \Gamma(t) + B(t)^{\mathrm{T}}\left[Q(t+1) + H(t+1)\right]B(t), \tag{13.3.6}$$

and the symmetric $n \times n$-matrices $H(T), H(T-1), \dots, H(0)$ are calculated recursively from $H(T) := 0$ in accordance with the following difference Ricatti equation:

$$H(t) := L(t)^{\mathrm{T}}\Gamma(t)L(t) + A_L(t)^{\mathrm{T}}\left[Q(t+1) + H(t+1)\right]A_L(t),$$
$$A_L(t) := A(t) - B(t)L(t). \tag{13.3.7}$$

13.4 Controller-Coder Separation Principle Does Not Hold

One of the major results of the classic LQG control theory is the controller-estimator separation principle. It states that LQG optimization problem always disintegrates into two independent problems. One of them is the deterministic linear-quadratic optimal control problem that results from dropping noises in the system model. Its solution is given by a linear feedback. The other one is the problem of minimum variance state estimation. The solution of the original stochastic problem results from merely putting such an estimate in place of the state in the feedback rule.[3]

A natural approach to control the plant via a limited capacity communication channel is to separate the control and encoding functions. In other words, the idea is to produce first the control that is optimal under neglecting the capacity constraints; i.e., the classic LQG optimal control

$$u_{\text{LQG}}(t) := -L(t)\widehat{x}(t|t).$$

Then this control must be coded, transmitted, and decoded with the resultant signal acting upon the plant (see Fig. 13.2). The encoding–decoding algorithm is optimized to achieve the best performance possible within this design scheme. A natural question is whether this approach permits to achieve the best performance. Were the answer always in the affirmative, it might be said that the controller-coder separation principle holds.

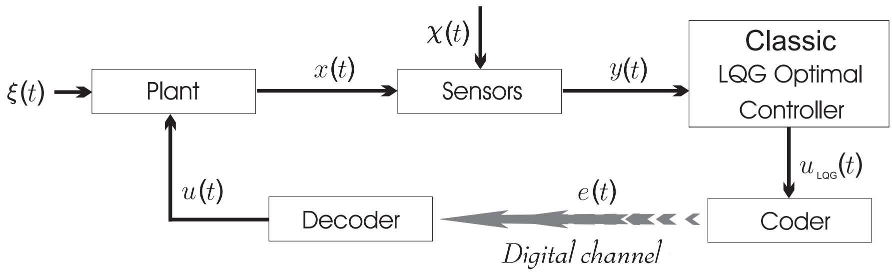


Fig. 13.2. A control system based on the controller-coder separation principle.

In this section, we show that this principle does not hold: Problems exist for which the answer is in the negative. To precise this statement, note that this principle in fact restricts the admissible control strategies to a certain subclass $\mathfrak{S}_{\text{sep}}$. Its critical feature is that the current codeword is generated from the knowledge of exclusively LQG optimal controls; i.e., (13.2.4) is replaced by

$$e(t) = \mathcal{E}[t, u_{\text{LQG}}(0), \dots, u_{\text{LQG}}(t)] \in \mathfrak{E}. \tag{13.4.1}$$

Thus the strategies from $\mathfrak{S}_{\text{sep}}$ are described by (13.2.5) and (13.4.1), whereas the primal class of strategies $\mathfrak{S}$ is given by (13.2.4) and (13.2.5). It is clear that $\mathfrak{S}_{\text{sep}} \subset \mathfrak{S}$. In this section, we show that problems exist for which the minimum of the cost functional over $\mathfrak{S}$ is strictly less than its infimum over $\mathfrak{S}_{\text{sep}}$.

Specifically, this holds for the following particular case of the primal problem (13.2.1), (13.2.2), and (13.2.6):

---

[3]See Appendix C starting on p. 509 for details.

$$x(t+1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} x(t) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} u(t) + \xi(t) \quad t = 0, 1, \quad x(0) = a;$$
$$y(t) = x(t) + \chi(t), \quad x \in \mathbb{R}^2, u \in \mathbb{R}, \quad T = 2;$$
$$\boldsymbol{E}a = 0, \; R_{aa} = R_{\xi\xi}(t) = R_{\chi\chi}(t) = I;$$
$$\mathfrak{I}_T = \boldsymbol{E}\Big\{ x_2(2)^2 + \frac{1}{2} x_2(1)^2 + u(1)^2 + u(0)^2 \Big\}. \quad (13.4.2)$$

Theorem 13.4.1. *Consider the particular case* (13.4.2) *of the LQG optimal control problem with communication capacity constraints posed in Sect.* 13.2. *Suppose that the channel transmits one bit of information per unit time; i.e.,* $b = 1$ *and* $N = 2$. *Then*

$$\inf_{\mathfrak{S}_{\text{sep}}} \mathfrak{I}_T > \inf_{\mathfrak{S}} \mathfrak{I}_T;$$

i.e., the infimum of the cost functional $\mathfrak{I}_T$ *from* (13.4.2) *over the control strategies based on the controller-coder separation principle is strictly greater than that over the primal class of control strategies given by* (13.2.4) *and* (13.2.5).

At the same time, a weaker estimator-coder separation principle is true. This claim is specified by the following.

Theorem 13.4.2. *Suppose that Assumption* 13.2.1 *holds. Consider the system* (13.2.1), (13.2.2)*, the cost functional* (13.2.6)*, and the LQG optimal control problem with communication capacity constraints, where optimization is over all control strategies* (13.2.4), (13.2.5). *Then the optimal control strategy exists. The optimal rule* (13.2.4) *to produce codewords has the form*

$$e(t) = \mathcal{E}\big[t, \widehat{x}(t|t)\big]. \quad (13.4.3)$$

Here $\widehat{x}(t|t)$ *is the minimum variance estimate* (13.3.1) *of the current state* $x(t)$ *and* $\mathcal{E}[t, x]$ *is a deterministic function.*

Thus like in the classic LQG optimal control problem, the minimum variance state estimate is a sufficient statistic. Furthermore, the Kalman filter producing this estimate appears to be a part of the solution of the optimal control problem with communication constraints.

Remark 13.4.3. Let a decoding rule (13.2.5) be given. Then formula (13.4.3) for the optimal encoding algorithm (13.2.4) can be obtained by the standard stochastic optimal control technique, e.g., by retracing the arguments from [200]. In doing so, it is convenient to put (13.2.5) into (13.2.1), thus excluding the variable $u(t)$, and to interpret $e(t)$ as a control ranging over a finite set $\mathfrak{E}$. The admissible "control strategies" are given by (13.2.4). (In [200], controls were taken from a finite-dimensional linear space, whereas now they are limited to a finite set. However this difference in fact influences neither the arguments nor the conclusion from [200].) The existence of the optimal encoding rule holds due to finiteness of the set $\mathfrak{E}$ of "controls" e and can be established via the standard dynamic programming arguments.

Modulo this remark, Theorem 13.4.2 mainly focuses on the existence of the optimal decoding rule. The proof of this theorem will be given in Sect. 13.6.

13.5 Solution of the Optimal Control Problem

To complete the description of the optimal control strategy, we note first that the decoding rule (13.2.5) is determined by the sequence $\omega_0, \dots, \omega_{T-1}$ of the reproduction (control) alphabets

$$\omega_t = \textbf{col}\left[u_t^{(1)}, \dots, u_t^{(N)}\right], \quad u_t^{(e)} := \mathfrak{U}[t,e] \in \mathbb{R}^m,$$

whose elements are enumerated by the symbols e of the channel alphabet $\mathfrak{E} = [1 : N]$. Indeed,

$$\mathfrak{U}[t,e] = u_t^{(e)} \qquad \forall t = 0, \dots, T-1, e \in \mathfrak{E}. \tag{13.5.1}$$

So the decoding scheme can be identified with the sequence $\{\omega_t\}$. Such a sequence related to the optimal strategy is said to be *optimal*.

In determining this strategy, the following dynamical programming procedure plays a key role:

$$\begin{aligned} q_T(x) := 0 \mapsto \cdots \mapsto q_t(x, \omega_{T-1}, \dots, \omega_t) \mapsto \cdots \mapsto \\ \mapsto \cdots \mapsto q_0(x, \omega_{T-1}, \dots, \omega_0) \mapsto q_{-1}(x, \omega_{T-1}, \dots, \omega_{-1}), \end{aligned} \tag{13.5.2}$$

where $x \in \mathbb{R}^n$. Its step is as follows:

$$q_t(x, \overbrace{\omega_{T-1}, \dots, \omega_t}^{w}) \tag{13.5.3}$$

$$\downarrow$$

$$q_t^{\mathcal{N}}(x, w) := \int_{\mathbb{R}^n} q_t(x + z, w)\, \mathcal{N}\left[d\, z \middle| 0, K(t)\Lambda(t)K(t)^{\mathrm{T}}\right] \tag{13.5.4}$$

$$\downarrow$$

$$\boxed{\begin{aligned} r_{t-1}(x, u, w) := \left\| F(t-1)^{1/2}\left[L(t-1)x + u\right]\right\|^2 \\ + q_t^{\mathcal{N}}\left[A(t-1)x + B(t-1)u, w\right] \end{aligned}} \tag{13.5.5}$$

$$\downarrow$$

$$q_{t-1}(x, \omega_{T-1}, \dots, \omega_t, \omega_{t-1}) := \min_{e=1,\dots,N} r_{t-1}(x, u_{t-1}^{(e)}, w). \tag{13.5.6}$$

Here $\omega_{t-1} = \textbf{col}\,(u^{(1)}, \dots, u^{(N)})$ and $\mathcal{N}(dx|m, M)$ is the Gaussian distribution with the mean m and the correlation matrix M. In (13.5.4) and (13.5.5), $K(t), \Lambda(t)$ and $L(t), F(t)$ are the matrices (13.3.4) and (13.3.6) generated by the Kalman filter and related to the solution of the deterministic linear-quadratic optimal control problem, respectively. To define $L(-1)$ and $F(-1)$, we set

$$Q(0) := 0 \quad \text{and} \quad \Gamma(-1) := I_m$$

[see also (13.3.5)]. Then

$$L(-1) = 0 \quad \text{and} \quad F(-1) = I_m. \tag{13.5.7}$$

For $t = T$, the argument w should be dropped everywhere.

Remark 13.5.1. The superfluous alphabet ω_{-1} is introduced for uniformity of notations. Only alphabets ω_{-1} containing 0 are considered.

Due to this remark, $q_{-1}(\cdot)$ does not depend on ω_{-1}. Specifically,

$$r_{-1}(x, u, w) = \|u\|^2 + q_0^{\mathcal{N}}(x, w)$$

and (13.5.6) implies that

$$q_{-1}(x, w, \omega_{-1}) = q_{-1}(x, w) = q_0^{\mathcal{N}}(x, w), \qquad w = \mathbf{col}\,(\omega_{T-1}, \dots, \omega_0).$$

Given $t = 1, \dots, T$ and $\omega_{T-1}, \dots, \omega_{t-1}$, we introduce the encoding rule (13.4.3) (for $t := t - 1$) by putting

$$\mathcal{E}(t-1, x) := \mathcal{E}[t-1, x|\omega_{T-1}, \dots, \omega_{t-1}] := e, \tag{13.5.8}$$

where e is the index furnishing the minimum on the right in (13.5.6). The function

$$x \mapsto \mathcal{E}[t-1, x|\omega_{T-1}, \dots, \omega_{t-1}]$$

can be chosen measurable [217, Sec.I.7].

Now we are in a position to complete the description of the optimal control.

Theorem 13.5.2. *Consider the LQG optimal control problem with communication capacity constraints posed in Sect.* 13.2. *Suppose that Assumption* 13.2.1 *(on p.* 448*) holds. Then the optimal reproduction alphabets*

$$\omega_t^{opt} = \mathbf{col}\left[\overset{o}{u}{}_t^{(1)}, \dots, \overset{o}{u}{}_t^{(N)}\right], \quad t = 0, \dots, T-1$$

are those furnishing the minimum of $q_{-1}(\boldsymbol{E}a, w)$ *over* $w = \mathbf{col}\,(\omega_{T-1}, \dots, \omega_0)$. *Here* a *is the initial vector from* (13.2.1) *(on p.* 448*). The optimal rules to produce codewords* (13.4.3) *(on p.* 452*) and to decode them* (13.2.5) *(on p.* 449*) are determined by these alphabets as follows:*

$$\mathcal{E}(t, x) := \mathcal{E}[t, x|\omega_{T-1}^{opt}, \dots, \omega_t^{opt}] \quad \text{in (13.4.3)} \quad \text{and} \quad \mathcal{U}(t, e) := \overset{o}{u}{}_t^{(e)} \quad \text{in (13.2.5)}.$$

Here $\mathcal{E}[t-1, x|\omega_{T-1}, \dots, \omega_{t-1}]$ *is defined by* (13.5.8).

The proof of this theorem will be given in Sect. 13.6.

Comments on the Basic Dynamical Programming Procedure (13.5.4)–(13.5.6)

Since the optimal encoding rule has the form (13.4.3), it is convenient to reformulate the primal problem as that of optimal control of the minimum variance state estimate generated by the Kalman filter. To this end, we note that elementary transformations shape (13.3.2) and (13.2.2) into

$$\begin{aligned}\widehat{x}(t+1|t+1) &= A(t)\widehat{x}(t|t) + B(t)u(t) + \beta(t);\\ y(t) &= C(t)\widehat{x}(t|t) + \gamma(t), \qquad t = -1, \dots, T-1;\\ \widehat{x}(-1|-1) &\overset{(13.3.5)}{=\!=} \boldsymbol{E}a, \qquad \text{where} \quad \gamma(t) := C(t)e(t) + \chi(t);\\ \beta(t) &:= K(t+1)C(t+1)A(t)e(t)\\ &\quad + K(t+1)C(t+1)\xi(t) + K(t+1)\chi(t+1); \quad (13.5.9)\end{aligned}$$

and

$$\begin{aligned}e(j) &:= x(j) - \widehat{x}(j|j) \quad \text{for} \quad j \geq 0, \qquad e(-1) := 0;\\ \xi(-1) &:= a - \boldsymbol{E}a, \quad C(-1) := 0, \quad \chi(-1) := 0. \quad (13.5.10)\end{aligned}$$

The following lemma summarizes the required technical facts about the system (13.5.9) and will be proved in Sect. 13.6.

Lemma 13.5.3. *The following statements are true:*

(i) The "noises" $\beta(t), \gamma(t)$ *in* (13.5.9) *do not depend on the control strategy;*
(ii) The random vector $\beta(t)$ *is zero-mean Gaussian and independent of the prior observations* $y(0), \dots, y(t)$*;*
(iii) The correlation matrix of the "noise" $\beta(t)$ *is given by*

$$\boldsymbol{E}\beta(t-1)\beta(t-1)^{\mathrm{T}} = K(t)\Lambda(t)K(t)^{\mathrm{T}}. \quad (13.5.11)$$

Here $K(t)$ *and* $\Lambda(t)$ *are the matrices* (13.3.4) *generated by the Kalman filter.*
(iv) The primal problem posed in Sect. 13.2 *is to minimize the functional*

$$\begin{aligned}\mathfrak{I}_T &:= \boldsymbol{E}\sum_{t=0}^{T-1} \mathcal{G}\left[\widehat{x}(t|t), u(t), t\right], \quad \text{where}\\ \mathcal{G}(x, u, t) &:= [u + L(t)x]^{\mathrm{T}} F(t)[u + L(t)x], \quad (13.5.12)\end{aligned}$$

over the processes in the system (13.5.9)*. Here* $L(t)$ *and* $F(t)$ *are given by* (13.3.6)*.*

By Theorem 13.4.2, attention can be restricted to the control strategy

$$\mathcal{W}_t := [\mathcal{E}(t, \cdot), \mathcal{U}(t, \cdot)], \qquad t = -1, \dots, T-1 \quad (13.5.13)$$

with encoding rules of the form (13.4.3). Formally, $\mathcal{W}_{-1}$ should be considered here although it does not influence the evolution of the system since $B(-1) = 0$ by

(13.3.5). We also recall that the decoding rule is determined (13.5.1) by the sequence $\{\omega_t\}_{t=-1}^{T-1}$ of reproduction alphabets each enumerated by the elements e of the channel alphabet. So the primal problem posed in Sect. 13.2 can be shaped into

$$\min_{\omega_{-1},\dots,\omega_{T-1}} \quad \min_{\mathcal{W}_{-1}\sqsubset\omega_{-1},\dots,\mathcal{W}_{T-1}\sqsubset\omega_{T-1}} \mathfrak{J}_T, \tag{13.5.14}$$

where $\mathcal{W}_t \sqsubset \omega_t \overset{\text{def}}{\Leftrightarrow} \mathcal{U}(t,e) \in \omega_t \ \forall e$ and $\mathfrak{J}_T$ is taken from (13.5.12).

Formulas (13.5.4)–(13.5.6) in fact result from applying the dynamical programming technique to the problem of finding the second $\min$ in (13.5.14). The function $q_t(\cdot)$ employed in these formulas is the standard "cost-to-go:"

$$q_t(x, \omega_{T-1}, \dots, \omega_t)$$
$$:= \min_{\mathcal{W}_t\sqsubset\omega_t,\dots,\mathcal{W}_{T-1}\sqsubset\omega_{T-1}} \boldsymbol{E}\left(\sum_{\theta=t}^{T-1} \mathcal{G}[\widehat{x}(\theta|\theta), u(\theta), \theta]\Big|\widehat{x}(t|t) = x\right). \tag{13.5.15}$$

In particular, $q_{-1}(\boldsymbol{E}a, \omega_{T-1}, \dots, \omega_0)$ equals the second $\min$ from (13.5.14). Relation (13.5.14) also illustrates the fact that minimizing $q_{-1}(\boldsymbol{E}a, w)$ over $w = (\omega_{T-1}, \dots, \omega_0)$, as is suggested by Theorem 13.5.2, gives rise to optimal alphabets.

13.6 Proofs of Lemma 13.5.3 and Theorems 13.4.2 and 13.5.2

Proof of Lemma 13.5.3. Statement (i) is straightforward from (13.3.4), (13.3.5), and (iii) of Theorem C.2 (on p. 511).

Statement (ii) follows from (ii) of Theorem C.2, along with Assumption 13.2.1 and (13.2.1), (13.2.2) (on p. 448).

Statement (iii). Due to (13.2.1), (13.2.2), (13.3.1), and (13.5.10) (on pp. 448, 450, and 455), the estimation error $e(t|t)$ is determined by a, $\xi(0)$, ..., $\xi(t-1)$, $\chi(0)$, ..., $\chi(t)$. So by Assumption 13.2.1 (on p. 448), it is independent of the zero-mean vectors $\xi(t)$ and $\chi(t+1)$. Hence the definition of $\beta(t)$ from (13.5.9) implies that

$$\begin{aligned}
\boldsymbol{E}\beta(t-1)\beta(t-1)^{\mathrm{T}} &= K(t)C(t)A(t-1)\left[\boldsymbol{E}e(t-1)e(t-1)^{\mathrm{T}}\right]A(t-1)^{\mathrm{T}}C(t)^{\mathrm{T}}K(t)^{\mathrm{T}} \\
&\quad + K(t)C(t)\left[\boldsymbol{E}\xi(t-1)\xi(t-1)^{\mathrm{T}}\right]C(t)^{\mathrm{T}}K(t)^{\mathrm{T}} + K(t)\left[\boldsymbol{E}\chi(t)\chi(t)^{\mathrm{T}}\right]K(t)^{\mathrm{T}} \\
&\overset{(13.2.3),\,(C.15)}{=\!=} K(t)C(t)\left[A(t-1)\overline{P}(t-1)A(t-1)^{\mathrm{T}} + R_{\xi\xi}(t-1)\right]C(t)^{\mathrm{T}}K(t)^{\mathrm{T}} \\
&\quad + K(t)R_{\chi\chi}(t)K(t)^{\mathrm{T}} \overset{(13.3.3)}{=\!=} K(t)\left[C(t)P(t)C(t)^{\mathrm{T}} + R_{\chi\chi}(t)\right]K(t)^{\mathrm{T}} \\
&\overset{(13.3.4)}{=\!=} K(t)\Lambda(t)K(t)^{\mathrm{T}}.
\end{aligned}$$

(We put $\overline{P}(-1) := \boldsymbol{E}e(-1)e(-1)^{\mathrm{T}} = 0$.)

Statement (iv) follows from Lemma 11.10.2 (on p. 397) applied to the system (11.2.1), (11.2.2) (on p. 372) with only one sensor $l = 1$ served by a perfect (i.e., providing no delays and data dropouts) channel. □

To prove the existence part of Theorem 13.4.2 (on p. 452), we employ the reformulation of the primal problem in the form described in (iv) of Lemma 13.5.3.

Remark 13.6.1. From now on, the sum in (13.5.12) is taken from $t = -1$.

This does not change the problem since the optimal value of $u(-1)$ equals 0 due to (13.5.7). Then the extra summand added to the sum from (13.5.12) vanishes.

The next lemma plays a key role in the proof and is based on the standard dynamic programming arguments.

Lemma 13.6.2. *Suppose that a time instant $t = -1, \dots, T$, reproduction alphabets $\omega_t, \dots, \omega_{T-1}$, and a "beginning" $\mathcal{W}_{-1}, \dots, \mathcal{W}_{t-1}$ of a control strategy* (13.5.13) *are given. Then*

$$\min_{\mathcal{W}_t \sqsubset \omega_t, \dots, \mathcal{W}_{T-1} \sqsubset \omega_{T-1}} \mathfrak{J}_T = \mathfrak{J}_t + \boldsymbol{E} q_t \big[\widehat{x}(t|t), \omega_{T-1}, \dots, \omega_t\big]. \tag{13.6.1}$$

Here $\mathfrak{J}_s$ with $s \in [-1 : T]$ is defined by (13.5.12) *(on p.* 455*) with s put in place of T. This minimum is attained at the control strategies $\overset{o}{\mathcal{W}}_t, \dots, \overset{o}{\mathcal{W}}_{T-1}$ given by* (13.4.3)*,* (13.5.1)*, and* (13.5.8) *(on pp.* 452, 453, and 454*).*

Remark 13.6.3. Since any sum from $j = s_-$ to $j = s_+ < s_-$ equals zero, $\mathfrak{J}_{-1} = 0$ in (13.6.1) with $t := -1$. If $t := T$, then $\min$ on the left and $\omega_{T-1}, \dots, \omega_t$ on the right should be dropped.

Corollary 13.6.4. *The minimum of the cost functional over the control strategies employing a given set of alphabets $\omega_0, \dots, \omega_{T-1}$ (i.e., such that $\mathcal{W}_j \sqsubset \omega_j \, \forall j$) equals*

$$q_{-1}(\boldsymbol{E}a, \omega_{T-1}, \dots, \omega_0)$$

and is attained at the strategies $\overset{o}{\mathcal{W}}_0, \dots, \overset{o}{\mathcal{W}}_{T-1}$.

Proof of Lemma 13.6.2. The proof will be by induction on $t = T, \dots, -1$. For $t := T$, the claim is obvious since $q_T(\cdot) \equiv 0$ by (13.5.2) (on p. 453). Suppose that it is true for some $t \geq 0$, and the following strategies and alphabets are given:

$$\mathcal{W}_{-1}, \dots, \mathcal{W}_{t-2}, \qquad \omega_{t-1}, \dots, \omega_{T-1}.$$

In the remainder of the proof, we consider strategies $\mathcal{W}_{t-1}, \dots, \mathcal{W}_{T-1}$ such that $\mathcal{W}_j \sqsubset \omega_j$. Now suppose that $\mathcal{W}_{t-1}$ is also given. By the induction hypothesis,

$$\min_{\mathcal{W}_t, \dots, \mathcal{W}_{T-1}} \mathfrak{J}_T = \mathfrak{J}_t + \boldsymbol{E} q_t \big[\widehat{x}(t|t), \omega_{T-1}, \dots, \omega_t\big]$$
$$\overset{(13.5.9),(13.5.12)}{=\!=\!=} \mathfrak{J}_{t-1} + \boldsymbol{E}\psi\big[\widehat{x}(t-1|t-1), u(t-1), \beta(t-1)\big],$$

where

$$\psi(x, u, \beta) := \big|F(t-1)^{1/2}[u + L(t-1)x]\big|^2 + q_t\big[A(t-1)x + B(t-1)u + \beta, \omega_{T-1}, \dots, \omega_t\big].$$

We denote by $p_Y(dY)$ and $p_\beta(d\beta)$ the probability distributions of the vectors $Y :=$ **col** $[y(0),\dots,y(t-1)]$ and $\beta(t-1)$, respectively. These vectors are independent, and $\beta(t-1)$ is zero-mean Gaussian with the correlation matrix (13.5.11) (on p. 455) by (ii) and (iii) of Lemma 13.5.3 (on p. 455), whereas $\widehat{x}(t-1|t-1)$ and $u(t-1)$ are functions of Y. Hence

$$p_\beta(d\beta) = \mathcal{N}\big[d\beta|0, K(t)\Lambda(t)K(t)^{\mathrm{T}}\big]$$

and

$$\begin{aligned}
&\boldsymbol{E}\psi\big[\widehat{x}(t-1|t-1), u(t-1), \beta(t-1)\big] \\
&\qquad = \int \psi\big[\widehat{x}(t-1|t-1), u(t-1), \beta\big] p_Y(dY) p_\beta(d\beta). \\
&\qquad \overset{(13.5.4),\,(13.5.5)}{=\!=\!=} \int r_{t-1}\big[\widehat{x}(t-1|t-1), u(t-1), \omega_{T-1},\dots,\omega_t\big] p_Y(dY);
\end{aligned}$$

$$\begin{aligned}
\min_{\mathcal{W}_t,\dots,\mathcal{W}_{T-1}} \mathfrak{J}_T &= \mathfrak{J}_{t-1} + \int r_{t-1}\big[\widehat{x}(t-1|t-1), u(t-1), \omega_{T-1},\dots,\omega_t\big] p_Y(dY) \\
&= \mathfrak{J}_{t-1} + \boldsymbol{E} r_{t-1}\big[\widehat{x}(t-1|t-1), u(t-1), \omega_{T-1},\dots,\omega_t\big]; \qquad (13.6.2)
\end{aligned}$$

$$\begin{aligned}
\min_{\mathcal{W}_{t-1},\dots,\mathcal{W}_{T-1}} \mathfrak{J}_T &= \min_{\mathcal{W}_{t-1}} \min_{\mathcal{W}_t,\dots,\mathcal{W}_{T-1}} \mathfrak{J}_T \\
&= \mathfrak{J}_{t-1} + \min_{\mathcal{W}_{t-1}} \int r_{t-1}\big[\widehat{x}(t-1|t-1), u(t-1), \omega_{T-1},\dots,\omega_t\big] p_Y(dY).
\end{aligned}$$

Now it becomes apparent that the optimal strategy $\mathcal{W}_{t-1}$ is given by (13.4.3), (13.5.1) (where $t := t-1$), and (13.5.8). For this strategy, the second summand in the last expression equals

$$\boldsymbol{E} q_{t-1}[\widehat{x}(t-1|t-1), \omega_{T-1},\dots,\omega_{t-1}]$$

due to (13.5.6) (on p. 453). Thus (13.6.1) does hold with $t := t-1$, which completes the proof. □

Thus to justify Theorems 13.4.2 and 13.5.2 (on pp. 452 and 454), it remains to show that the minimum $\min_w q_{-1}(\boldsymbol{E}a, w)$ is attained. To this end, some properties of the functions $q_t(\cdot)$ from (13.5.2) should be revealed. To this end, we denote

$$\lfloor \omega \rfloor := \min\Big\{\big\|u^{(1)}\big\|,\dots,\big\|u^{(N)}\big\|\Big\} \qquad (13.6.3)$$

whenever $\omega =$ **col** $\big(u^{(1)},\dots,u^{(N)}\big)$ and $u^{(i)} \in \mathbb{R}^m$.

Lemma 13.6.5. *Constants* $a_{-1},\dots,a_T, b_{-1},\dots,b_T > 0$ *exist such that*

$$0 \le q_t(x, \omega_{T-1},\dots,\omega_t) \le a_t\Big(\|x\|^2 + \lfloor \omega_{T-1} \rfloor^2 + \dots + \lfloor \omega_t \rfloor^2\Big) + b_t \qquad (13.6.4)$$

for all $t, x, \omega_{T-1},\dots,\omega_t$.

Proof. The proof will be by induction on $t = T, \dots, -1$. For $t = T$, (13.6.4) holds since $q_T(\cdot) \equiv 0$ by (13.5.2) (on p. 453). Let (13.6.4) be true for some $t = 0, \dots, T$. Due to the inequality $\|x+z\|^2 \le 2(\|x\|^2 + \|z\|^2)$ and (13.5.4) (on p. 453), we get

$$0 \le q_t(x+z, \omega_{T-1}, \dots, \omega_t) \\ \le a_t\Big(2\|x\|^2 + 2\|z\|^2 + \lfloor \omega_{T-1} \rfloor^2 + \dots + \lfloor \omega_t \rfloor^2\Big) + b_t; \quad (13.6.5)$$

$$0 \le q_t^{\mathcal{N}}(x, \omega_{T-1}, \dots, \omega_t) \le a_t\Big(2\|x\|^2 + \lfloor \omega_{T-1} \rfloor^2 + \dots + \lfloor \omega_t \rfloor^2\Big) + b_t \\ + 2\int_{\mathbb{R}^n} \|z\|^2\, \mathcal{N}\Big[dz\Big|0, K(t)\Lambda(t)K(t)^{\mathrm{T}}\Big].$$

Denote the sum of the last two summands by b_{t-1} and put

$$\varkappa := 2 \max_{t=-1,\dots,T} \Big[\|F(t)^{1/2}\|^2 + \|F(t)^{1/2}L(t)\|^2\Big], \\ \overline{a} := 2 \max_{t=-1,\dots,T} \Big[\|A(t)\|^2 + \|B(t)\|^2\Big]. \quad (13.6.6)$$

Then (13.5.5) (on p. 453) yields

$$0 \le r_{t-1}(x, u, \omega_{T-1}, \dots, \omega_t) \le \varkappa\big(\|x\|^2 + \|u\|^2\big) + b_{t-1} \\ + 2a_t \overbrace{\big|A(t-1)x + B(t-1)u\big|^2}^{\le \overline{a}\left(\|x\|^2 + \|u\|^2\right)} + a_t\Big(\lfloor \omega_{T-1} \rfloor^2 + \dots + \lfloor \omega_t \rfloor^2\Big). \quad (13.6.7)$$

This and (13.5.6) (on p. 453) result in (13.6.4) with $t := t-1$ and $a_{t-1} := \max\{\varkappa + 2a_t\overline{a}; a_t\}$. □

Corollary 13.6.6. *For any $t = -1, \dots, T$, the functions $q_t(\cdot)$, $q_t^{\mathcal{N}}(\cdot)$, $r_t(\cdot)$ are continuous. (We put $r_T(\cdot) := 0$.)*

Indeed, let us argue by induction on $t = T, \dots, -1$. For $t = T$, the claim is evident. Suppose that it holds for some $t = 0, \dots, T$. Formula (13.5.4) (on p. 453), along with the estimate (13.6.5) and the Lebesgue dominated convergence theorem, ensures that the function $q_t^{\mathcal{N}}(\cdot)$ is continuous. Then by (13.5.5) and (13.5.6) (on p. 453), so are $r_{t-1}(\cdot)$ and $q_{t-1}(\cdot)$.

Now we are going to show that the functions $q_t(\cdot)$ are actually continuous in a stronger sense. This fact plays a key role in proving that the minimum of $q_{-1}(\boldsymbol{E}a, \cdot)$ is attained. To this end, we introduce the following.

Definition 13.6.7. *An alphabet* $\omega =$ **col** $[u^{(1)}, \dots, u^{(N)}]$ *is called a* slight limit point *of a sequence of alphabets*

$$\{\omega(j)\}_{j=1}^{\infty}, \qquad \omega(j) = \textbf{col}\,[u^{(1)}(j), \dots, u^{(N)}(j)]$$

if and only if a nonempty set of indices $I \subset \{1, \dots, N\}$ exists such that the following statements are true:

(i) Whenever $i \in I$, the following limit exists:

$$\lim_{j\to\infty} u^{(i)}(j) =: u^{(i)}_\infty;$$

(ii) Whenever $i \notin I$, the following relation holds:

$$\|u^{(i)}(j)\| \to \infty \quad \text{as} \quad j \to \infty;$$

(iii) The alphabet $\{u^{(i)}\}_{i=1}^N$ consists of (maybe, repeating) limit points $u^{(i)}_\infty$, $i \in I$.

If (ii) holds for all $i = 1, \ldots, N$, the sequence is said to increase without limits.

Lemma 13.6.8. *Suppose that $t = -1, \ldots, T$ and sequences of alphabets*

$$\{\omega_t(j)\}_{j=1}^\infty, \ldots, \{\omega_{T-1}(j)\}_{j=1}^\infty, \qquad \omega_\theta(j) = \mathbf{col}\,[u^{(1)}_\theta(j), \ldots, u^{(N)}_\theta(j)]$$

are given. Then the following statements hold:

(i) If each of them $\{\omega_\theta(j)\}_{j=1}^\infty$ has a slight limit point ω_θ, then for any $R > 0$,

$$q_t\big[x, \omega_{T-1}(j), \ldots, \omega_t(j)\big] \to q_t\big[x, \omega_{T-1}, \ldots, \omega_t\big] \tag{13.6.8}$$

as $j \to \infty$ uniformly over $\|x\| \le R$;

(ii) If $t \le T-1$ and at least one of the sequences $\{\omega_\theta(j)\}_{j=1}^\infty$ with $\theta \ge \max\{0, t\}$ increases without limits, then

$$q_t\big[x, \omega_{T-1}(j), \ldots, \omega_t(j)\big] \to \infty \quad \text{as} \quad j \to \infty \quad \forall x.$$

Proof. The proof will be by induction on $t = T, \ldots, -1$. For $t = T$, the statements (i) and (ii) are obvious since $q_T(\cdot) \equiv 0$ by (13.5.2) (on p. 453). Let they hold for some $t \in [0 : T]$. It should be shown that these statements are valid for $t := t-1$.

Statement (i). Introduce the set I_θ associated with the sequence $\{\omega_\theta(j)\}_{j=1}^\infty$ by Definition 13.6.7. Thanks to (13.6.3), a constant $d > 0$ exists such that

$$\big\|u^{(i)}_\theta(j)\big\| \le d \quad \forall i \in I_\theta, \qquad \downharpoonleft \omega_\theta(j) \downharpoonright \le d$$

for all θ, j. So whenever $\|x\| \le R$, (13.6.4) gives

$$0 \le q_t[x + z, \omega_{T-1}(j), \ldots, \omega_t(j)] \le a_t\Big[2R^2 + 2\|z\|^2 + (T-t)d^2\Big] + b_t. \tag{13.6.9}$$

The function on the right is integrable with respect to

$$\mathcal{N}[dz|0, K(t)\Lambda(t)K(t)^{\mathrm{T}}].$$

Hence the Lebesgue dominated convergence theorem, the induction hypothesis, and (13.5.4) (on p. 453) imply that

$$q_t^{\mathcal{N}}\big[x, \omega_{T-1}(j), \ldots, \omega_t(j)\big] \to q_t^{\mathcal{N}}\big[x, \omega_{T-1}, \ldots, \omega_t\big] \quad \text{as} \quad j \to \infty \tag{13.6.10}$$

uniformly over $\|x\| \le R$. Whenever $i \in I_{t-1}$, we get by letting $j \to \infty$ and taking into account (13.5.5) (on p. 453),

$$r_{t-1}\big[x, u_{t-1}^{(i)}(j), \omega_{T-1}(j), \ldots, \omega_t(j)\big] \le \max_{\|x\| \le R, \|u\| \le d} \Big\{ \big\|F(t-1)^{1/2}[L(t-1)x+u]\big\|^2$$
$$+ q_t^{\mathcal{N}}\big[A(t-1)x + B(t-1)u, \omega_{T-1}(j), \ldots, \omega_t(j)\big] \Big\}$$
$$\to \max_{\|x\| \le R, \|u\| \le d} \Big\{ \big\|F(t-1)^{1/2}[L(t-1)x+u]\big\|^2$$
$$+ q_t^{\mathcal{N}}\big[A(t-1)x + B(t-1)u, \omega_{T-1}, \ldots, \omega_t\big] \Big\} < \infty.$$

Now let $s \notin I_{t-1}$. Thanks to Remark C.4 (on p. 512),

$$\|F(\tau)^{1/2}u\|^2 = u^{\mathsf{T}}F(\tau)u \ge 2\gamma\|u\|^2 \quad \forall \tau, u,$$

where $\gamma > 0$. By employing the evident inequality $\|u+v\|^2 \ge 1/2\|u\|^2 - \|v\|^2$ and (13.5.5) (on p. 453), (13.6.6), we see that

$$r_{t-1}\big[x, u_{t-1}^{(s)}(j), \omega_{T-1}(j), \ldots, \omega_t(j)\big] \ge \Big\|F(t-1)^{1/2}\big[L(t-1)x + u_{t-1}^{(s)}(j)\big]\Big\|^2$$
$$\ge 1/2\big\|F(t-1)^{1/2}u_{t-1}^{(s)}(j)\big\|^2 - \big\|F(t-1)^{1/2}L(t-1)x\big\|^2$$
$$\ge \gamma\big\|u_{t-1}^{(s)}(j)\big\|^2 - \varkappa R^2 \to \infty \quad (13.6.11)$$

as $j \to \infty$. So by letting $j \to \infty$, it follows from (13.5.6) (on p. 453) that whenever $\|x\| \le R$ and j is large enough,

$$q_{t-1}\big[x, \omega_{T-1}(j), \ldots, \omega_{t-1}(j)\big] = \min_{i \in I_{t-1}} r_{t-1}\big[x, u_{t-1}^{(i)}(j), \omega_{T-1}(j), \ldots, \omega_t(j)\big]$$
$$\xrightarrow{(13.5.5),(13.6.10)} \min_{i \in I_{t-1}} r_{t-1}\Big[x, \lim_{j\to\infty} u_{t-1}^{(i)}(j), \omega_{T-1}(j), \ldots, \omega_t(j)\Big]$$
$$= q_{t-1}\big[x, \omega_{T-1}, \ldots, \omega_{t-1}\big] \quad \text{uniformly over } \|x\| \le R.$$

The last relation holds by (iii) of Definition 13.6.7 and (13.5.6). Thus (i) of the lemma does hold for $t := t-1$.

Statement (ii). If one of the sequences

$$\{\omega_t(j)\}_{j=1}^{\infty}, \ldots, \{\omega_{T-1}(j)\}_{j=1}^{\infty}$$

increases without limits, then (13.5.4) (on p. 453), Fatou's lemma, and the induction hypothesis ensure that

$$q_t^{\mathcal{N}}[x, \omega_{T-1}(j), \ldots, \omega_t(j)] \to \infty \quad \text{as} \quad j \to \infty,$$

and so (ii) with $t := t-1$ is straightforward from (13.5.5) and (13.5.6) (on p. 453). If the sequence $\{\omega_{t-1}(j)\}_{j=1}^{\infty}$ increases without limits, (ii) with $t := t-1$ follows from (13.6.11). □

Proofs of Theorems 13.4.2 and 13.5.2 (on pp. 452 and 454). Modulo Corollary 13.6.4, it remains to show that the minimum of $q_{-1}(\boldsymbol{E}a, w)$ over $w = \mathbf{col}\,(\omega_{T-1}, \ldots, \omega_0)$ is achieved. To this end, we pick minimizing sequences $\{\omega_0(j)\}_{j=1}^{\infty}, \ldots, \{\omega_{T-1}(j)\}_{j=1}^{\infty}$:

$$q_{-1}\big[\boldsymbol{E}a, \omega_{T-1}(j), \ldots, \omega_0(j)\big] \to \inf_{\omega_0', \ldots, \omega_{T-1}'} q_{-1}\big[\boldsymbol{E}a, \omega_{T-1}', \ldots, \omega_0'\big] \quad (13.6.12)$$

as $j \to \infty$. By passing to subsequences, one can ensure that each sequence $\{\omega_\nu(j)\}_{j=1}^{\infty}$ either has a slight limit point ω_ν or increases without limits. However the second case in fact holds for no sequence due to (ii) of Lemma 13.6.8. Then (i) of this lemma ensures that the left-hand side of (13.6.12) converges to $q_{-1}\big[\boldsymbol{E}a, \omega_{T-1}, \ldots, \omega_0\big]$, which completes the proof. □

13.7 Proof of Theorem 13.4.1

We start with calculation of the function $r_0(\cdot)$ given by (13.5.5) (on p. 453) . Since the permutation of the components $u_t^{(1)}, u_t^{(2)}$ of ω_t does not affect $r_0(\cdot)$, it suffices to find $r_0(\cdot)$ assuming that $u_t^{(1)} \leq u_t^{(2)}$. In this section, we use the notation

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{1}{2}s^2\right)\, ds.$$

Lemma 13.7.1. *In the new variables*

$$z_1 = \frac{x_1 + x_2}{2}, \qquad z_2 = \frac{x_1 - x_2}{2}, \qquad \Delta_1^{\pm} := u_1^{(2)} \pm u_1^{(1)},$$

the function $r_0(\cdot)$ looks as follows:

$$r_0(x_1, x_2, u, \omega_1) = 2\Big(\frac{z_1}{2} + u\Big)^2 + 2\Big(\frac{z_1 + z_2 + u}{2} + u_1^{(1)}\Big)^2 - \sqrt{\frac{1.8}{\pi}}\Delta_1^- \exp\left[-\frac{(\Delta_1^+ + z_1 + z_2 + u)^2}{1.8}\right] + 2\Delta_1^-\big(z_1 + z_2 + u + \Delta_1^+\big)\Phi\Big(-\frac{\Delta_1^+ + z_1 + z_2 + u}{\sqrt{0.9}}\Big) + 0.45. \quad (13.7.1)$$

Proof. Elementary calculations in accordance with (13.3.3)–(13.3.7) (on p. 450) show that

$$P(0) = I, \;\; \Lambda(0) = 2I, \;\; K(0) = 1/2I, \;\; \overline{P}(0) = 1/2I;$$
$$K(0)\Lambda(0)K(0)^{\mathrm{T}} = 1/2I, \;\; P(1) = 3/2I, \;\; \Lambda(1) = 5/2I, \;\; K(1) = 3/5I;$$
$$\overline{P}(1) = 3/5I, \;\; K(1)\Lambda(1)K(1)^{\mathrm{T}} = 9/10I, \;\; F(1) = F(0) = 2;$$
$$L(1) = 1/2(0,1), \;\; L(0) = 1/4(1,1).$$

Relations (13.5.4)–(13.5.6) (on p. 453) imply

$$q_2(\cdot) \equiv 0, \; q_2^{\mathcal{N}}(\cdot) \equiv 0, \; r_1(x_1, x_2, u) = 2\big(x_2/2 + u\big)^2;$$
$$q_1(x_1, x_2, u_1^{(1)}, u_1^{(2)}) = 2\min\Big\{\big(x_2/2 + u_1^{(1)}\big)^2; \big(x_2/2 + u_1^{(2)}\big)^2\Big\}$$
$$= \begin{cases} 2\big(x_2/2 + u_1^{(2)}\big)^2 & \text{if } x_2 \le -\Delta_1^+ \\ 2\big(x_2/2 + u_1^{(1)}\big)^2 & \text{if } x_2 > -\Delta_1^+ \end{cases}.$$

By invoking (13.5.4) (on p. 453) once more, we have

$$q_1^{\mathcal{N}}\big(x_1, x_2, u_1^{(1)}, u_1^{(2)}\big) = 2\int_{-\infty}^{-\Delta_1^+ - x_2} \big(x_2/2 + z/2 + u_1^{(2)}\big)^2 \mathcal{N}(dz|0, 0.9)$$
$$+ 2\int_{-\Delta_1^+ - x_2}^{\infty} \big(x_2/2 + z/2 + u_1^{(1)}\big)^2 \mathcal{N}(dz|0, 0.9)$$
$$= 2\int_{-\infty}^{\infty} \big(x_2/2 + z/2 + u_1^{(1)}\big)^2 \mathcal{N}(dz|0, 0.9)$$
$$+ 2\int_{-\infty}^{-\Delta_1^+ - x_2} \Big[\big(x_2/2 + z/2 + u_1^{(2)}\big)^2 - \big(x_2/2 + z/2 + u_1^{(1)}\big)^2\Big] \mathcal{N}(dz|0, 0.9)$$
$$= 2\big(x_2/2 + u_1^{(1)}\big)^2 \int_{-\infty}^{\infty} \mathcal{N}(dz|0, 0.9) + 2\big(x_2/2 + u_1^{(1)}\big) \int_{-\infty}^{\infty} z\, \mathcal{N}(dz|0, 0.9)$$
$$+ 1/2\int_{-\infty}^{\infty} z^2\, \mathcal{N}(dz|0, 0.9) + 2\Delta_1^- \Bigg[\int_{-\infty}^{-\Delta_1^+ - x_2} \big(x_2 + \Delta_1^+\big)\, \mathcal{N}(dz|0, 0.9)$$
$$+ \int_{-\infty}^{-\Delta_1^+ - x_2} z\, \mathcal{N}(dz|0, 0.9)\Bigg] = 2\big(x_2/2 + u_1^{(1)}\big)^2 + 2\Delta_1^- \big(x_2 + \Delta_1^+\big)\Phi\Big(-\frac{\Delta_1^+ + x_2}{\sqrt{0.9}}\Big)$$
$$- \sqrt{\frac{1.8}{\pi}}\Delta_1^- \exp\left[-\frac{(\Delta_1^+ + x_2)^2}{1.8}\right] + 0.45. \quad (13.7.2)$$

Then (13.7.1) is straightforward from (13.5.5) (on p. 453) . □

By employing elementary calculus, formula (13.7.1) gives rise to the following.

Corollary 13.7.2. *The function $r_0(\cdot)$ can be represented in the form:*

$$r_0(z_1, z_2, u, \omega_1) = \frac{z_2^2}{2} + 2\Big(\frac{z_1 + u}{2} + u_1^{(1)}\Big)z_2$$
$$+ \begin{cases} \varphi_+(z_1, u, \omega_1) + \mathrm{o}_+(z_2, z_1, u, \omega_1) \\ 2\Delta_1^- z_2 + \varphi_-(z_1, u, \omega_1) + \mathrm{o}_-(z_2, z_1, u, \omega_1) \end{cases}.$$

Here $\varphi_\pm(\cdot)$, $\mathcal{O}_\pm(\cdot)$ *are continuous functions and*

$$\mathcal{O}_\pm(z_2, z_1, u, \omega_1) \to 0 \quad \text{as} \quad z_2 \to \pm\infty, \quad \text{respectively.}$$

Corollary 13.7.3. *Let* $\omega_t = \mathbf{col}\,[u_t^{(1)}, u_t^{(2)}],\ t = 0, 1$ *be given and* $u_0^{(1)} \le u_0^{(2)}$. *Then for* $\nu = 1, 2$, *the following inequality holds:*

$$r_0(z_1, z_2, u_0^{(\nu)}, \omega_1) \ge q_0(z_1, z_2, \omega_0, \omega_1) \\ + \max\Big\{\lambda_\nu(z_1) + (-1)^\nu\big[u_0^{(2)} - u_0^{(1)}\big]z_2 + \mathcal{O}_\nu(z_2, z_1); \mathbf{0}\Big\}. \quad (13.7.3)$$

Here $\lambda_\nu(\cdot)$, $\mathcal{O}_\nu(\cdot)$ *are continuous functions,* $\mathcal{O}_\nu(z_2, z_1) \to 0$ *as* $(-1)^\nu z_2 \to \infty$, *and* $q_0(\cdot)$ *is given by* (13.5.6) *(on p.* 453*).*

Proof. Let $\nu = 2$. (For $\nu = 1$, the arguments are similar.) By (13.5.6) (on p. 453),

$$r_0(z_1, z_2, u_0^{(j)}, \omega_1) \ge q_0(z_1, z_2, \omega_0, \omega_1)$$

for $j = 1, 2$. This and Corollary 13.7.2 yield

$$q_0(z_1, z_2, \omega_0, \omega_1) \le r_0(z_1, z_2, u_0^{(1)}, \omega_1) = \frac{z_2^2}{2} + 2\Big(\frac{z_1 + u_0^{(1)}}{2} + u_1^{(1)}\Big)z_2 \\ + \varphi_+(z_1, u_0^{(1)}, \omega_1) + \mathcal{O}_+(z_2, z_1, u_0^{(1)}, \omega_1) = r_0(z_1, z_2, u_0^{(2)}, \omega_1) + \big[u_0^{(1)} - u_0^{(2)}\big]z_2 \\ + \underbrace{\varphi_+(z_1, u_0^{(1)}, \omega_1) - \varphi_+(z_1, u_0^{(2)}, \omega_1)}_{-\lambda_2(z_1)} + \underbrace{\mathcal{O}_+(z_2, z_1, u_0^{(1)}, \omega_1) - \mathcal{O}_+(z_2, z_1, u_0^{(2)}, \omega_1)}_{-\mathcal{O}_2(z_2, z_1)}.$$

Invoking that

$$r_0(z_1, z_2, u_0^{(2)}, \omega_1) \ge q_0(z_1, z_2, \omega_0, \omega_1),$$

we arrive at (13.7.3) with $\nu = 2$. □

Another technical fact concerns the alphabets furnishing the minimum

$$\mathcal{Q} := \min_{w = \omega_1, \omega_0} q_{-1}(\boldsymbol{E}a, w) = \min_w q_0^{\mathcal{N}}(0, w) \\ = \min_w \int_{\mathbb{R}^2} q_0(x, w)\, p(dx), \quad p(dx) := \mathcal{N}\big[dx\big|0; 1/2I\big]. \quad (13.7.4)$$

Lemma 13.7.4. *If* $\overline{\omega}_t = \mathbf{col}\,\big[\overline{u}_t^{(1)}, \overline{u}_t^{(2)}\big],\ t = 0, 1$ *furnish the minimum* (13.7.4) *and* $\overline{u}_t^{(1)} \le \overline{u}_t^{(2)}$, *then in fact* $\overline{u}_0^{(1)} < \overline{u}_0^{(2)}$.

Proof. Suppose to the contrary that $\overline{u}_0^{(2)} = \overline{u}_0^{(1)} =: \overline{u}$. This and (13.5.6) (on p. 453) imply that the functional $\int_{\mathbb{R}} r_0\big(x, u, \omega_1\big)p(dx)$ attains the minimum value over u, ω_1 at $\overline{u}, \overline{\omega}_1$. The change of the variables

$$u \leftrightarrow u, u_1^{(1)} \leftrightarrow v_1^{(1)} := u_1^{(1)} + u/2, \qquad u_1^{(2)} \leftrightarrow v_1^{(2)} := u_1^{(2)} + u/2$$

makes all the summands in (13.7.1) except for the first one independent of the variable u. It follows that $\overline{u}$ furnishes the minimum of the integral of the first summand

$$2\int \Big(\frac{x_1+x_2}{4}+u\Big)^2 p(dx) = \frac{1}{8}\int \big(x_1+x_2\big)^2 p(dx) + u \underbrace{\int \big(x_1+x_2\big)p(dx)}_{=0} + 2u^2;$$

i.e., $\overline{u}=0$. Thus $(0,0), \overline{\omega}_1$ furnish the minimum in (13.7.4). So for any $u>0$,

$$\int r_0(x,0,\overline{\omega}_1)\, p(dx) \leq \int \min\big\{r_0(x,u,\overline{\omega}_1); r_0(x,-u,\overline{\omega}_1)\big\}\, p(dx),$$

$$0 \leq \int \min_{\alpha=1,2} u^{-1}\{r_0[x,(-1)^\alpha u,\overline{\omega}_1] - r_0[x,0,\overline{\omega}_1]\}\, p(dx)$$

$$= \int \min\left\{\frac{\partial r_0}{\partial u}\big[x,u\theta'(x),\overline{\omega}_1\big]; -\frac{\partial r_0}{\partial u}\big[x,-u\theta''(x),\overline{\omega}_1\big]\right\} p(dx), \quad (13.7.5)$$

where $\theta'(x), \theta''(x) \in [0,1]$. By putting $t := 1$ into (13.5.5) (on p. 453) and invoking the first relation from (13.7.2), we get

$$\frac{\partial r_0}{\partial u}\big[x,u,\overline{\omega}_1\big] = 2\int_{-\infty}^{-\Delta_1^+ - x_1 - u} \Big[\frac{x_1+u+z}{2} + \overline{u}_1^{(2)}\Big]\mathcal{N}(dz|0,0.9)$$

$$+\,2\int_{-\Delta_1^+ - x_1 - u}^{\infty} \Big[\frac{x_1+u+z}{2} + \overline{u}_1^{(1)}\Big]\mathcal{N}(dz|0,0.9) + 4u + x_1 + x_2. \quad (13.7.6)$$

Hence for $u \approx 0$, the function $\left|\frac{\partial r_0}{\partial u}(\cdot)\right|$ is estimated from above by a function integrable over $p(dx)$. So letting $u \to +0$ in (13.7.5) gives

$$0 \leq -\int \left|\frac{\partial r_0}{\partial u}\big[x,0,\overline{\omega}_1\big]\right| p(dx) \Rightarrow \frac{\partial r_0}{\partial u}\big[x,0,\overline{\omega}_1\big] = 0 \qquad \forall x$$

in violation of (13.7.6). The contradiction obtained proves that $\overline{u}_0^{(1)} < \overline{u}_0^{(2)}$. □

Now we in fact start the immediate proof of Theorem 13.4.1 (on p. 452). Suppose to the contrary to its conclusion that the infimum of the cost functional over $\mathfrak{S}$ equals that over $\mathfrak{S}_{\text{sep}}$. From now on, the cost functional is taken in the form $\mathfrak{J}_T$ given by (13.5.12) (on p. 455). Consider a sequence $\{\mathfrak{s}_j\}_{j=1}^\infty$ of control strategies that asymptotically minimizes the functional over $\mathfrak{S}_{\text{sep}}$. Corresponding to it are sequences of controls $\{u_j(t)\}_{j=1}^\infty$ and reproduction alphabets $\big\{\omega_t(j) = \textbf{col}\,\big[u_t^{(1)}(j), u_t^{(2)}(j)\big]\big\}_{j=1}^\infty$ with $u_t^{(1)}(j) \leq u_t^{(2)}(j)$. Here $t=0,1$.

Lemma 13.7.5. *The sequences* $\{\omega_t(j)\}_{j=1}^\infty$ *asymptotically furnish the minimum* $\mathcal{Q}$ *from* (13.7.4)*, and* $\mathfrak{J}_T[\mathfrak{s}_j] \xrightarrow{j\to\infty} \mathcal{Q}$*.*

Proof. Thanks to (13.5.15), $\mathcal{Q} = \inf_{\mathfrak{S}} \mathfrak{J}_T$, where $\inf_{\mathfrak{S}} \mathfrak{J}_T = \inf_{\mathfrak{S}_{\text{sep}}} \mathfrak{J}_T$ by assumption. This gives the second claim of the lemma. Then (13.5.15) (on p. 456) and Theorem 13.4.2 (on p. 452) complete the proof. □

This lemma and retracing the arguments from the last paragraph in Sect. 13.6 show that passing to a subsequence ensures existence of a slight limit point $\overline{\omega}_t = \mathbf{col}\left[\overline{u}_t^{(1)}, \overline{u}_t^{(2)}\right]$ of $\{\omega_t(j)\}_{j=1}^\infty$ and also the following fact.

Remark 13.7.6. The pair $\overline{\omega}_0, \overline{\omega}_1$ furnishes the minimum (13.7.4).

Lemma 13.7.7. *The asymptotically minimizing sequence* $\{\mathfrak{s}_j\}$ *can be chosen so that* $\omega_t(j) \to \overline{\omega}_t$ *as* $j \to \infty$ $\forall t$.

Proof. Note that

$$u_t^{(1)}(j) \leq u_t^{(2)}(j)\ \forall j \Rightarrow \overline{u}_t^{(1)} \leq \overline{u}_t^{(2)}.$$

Lemma 13.7.4 and Remark 13.7.6 yield $\overline{u}_0^{(1)} < \overline{u}_0^{(2)}$. Then Definition 13.6.7 (on p. 459) implies that $\omega_0(j) \to \overline{\omega}_0$ as $j \to \infty$. Suppose that $\omega_1(j) \not\to \overline{\omega}_1$. Then Definition 13.6.7 ensures that $\overline{u}_1^{(1)} = \overline{u}_1^{(2)} =: \overline{u}$. For any j, replacing the control $u_1(j)$ by the constant $\overline{u}$ results in a new strategy $\mathfrak{s}_j^* \in \mathfrak{S}_{\text{sep}}$ and changes the alphabet $\omega_1(j) := \overline{\omega}_1$. Thus it suffices the prove that the sequence $\{\mathfrak{s}_j^*\}$ is asymptotically minimizing.

Since the strategies $\mathfrak{s}_j$ and $\mathfrak{s}_j^*$ are identical for $t = 0$, so are the corresponding summands $\boldsymbol{E}\mathcal{G}\left[\widehat{x}(0|0), u(0), 0\right]$ in (13.5.12) (on p. 455), and these strategies give rise to a common estimate $\widehat{x}(1|1)$ with a probability distribution $p_1(dx)$. Then

$$\begin{aligned}\boldsymbol{E}\mathcal{G}\left[\widehat{x}(1|1), u_j(1), 1\right] &\overset{(13.5.15)}{\geq} \int_{\mathbb{R}^2} q_1\left[x, \omega_1(j)\right]\, p_1(dx) \\ &\xrightarrow{(13.6.4),(13.6.8)} \int_{\mathbb{R}^2} q_1\left[x, (\overline{u}, \overline{u})\right]\, p_1(dx) \\ &\overset{(13.5.5),(13.5.6)}{=\!=} \int_{\mathbb{R}^2} \left\| F(1)^{1/2}\left[L(1)\widehat{x}(1|1) + \overline{u}\right]\right\|^2 p_1(dx) = \boldsymbol{E}\mathcal{G}\left[\widehat{x}(1|1), \overline{u}, 1\right].\end{aligned}$$

The proof is completed as follows:

$$\inf_{\mathfrak{S}_{\text{sep}}} \mathfrak{J}_T = \lim_{j\to\infty} \mathfrak{J}_T\left[\mathfrak{s}_j\right] \geq \limsup_{j\to\infty} \mathfrak{J}_T\left[\mathfrak{s}_j^*\right] \geq \inf_{\mathfrak{S}_{\text{sep}}} \mathfrak{J}_T. \qquad \square$$

Proof of Theorem 13.4.1 (on p. 452). Now we complete the proof. We recall that we started from the assumption that the conclusion of the theorem violates. At $t = 0$, the LQG optimal control is given by

$$u_{\text{LQG}}(0) = -L(0)\widehat{x}(0|0) = -1/4\left[\widehat{x}_1(0|0) + \widehat{x}_2(0|0)\right] = -1/2\widehat{z}_1.$$

Here we employ the new variables

$$\widehat{z}_1 = 1/2\left[\widehat{x}_1(0|0) + \widehat{x}_2(0|0)\right], \qquad \widehat{z}_2 = 1/2\left[\widehat{x}_1(0|0) - \widehat{x}_2(0|0)\right],$$

as before. So far as $\mathfrak{s}_j \in \mathfrak{S}_{\text{sep}}$, the control $u_j(0)$ is a function of $u_{\text{LQG}}(0)$ and so $u_j(0) = \mathcal{V}[\widehat{z}_1]$. Since $u_j(0) \in \omega_0(j)$, we have $u_j(0) = u_0^{(\nu_j[\widehat{z}_1])}(j)$, where $\nu_j(\cdot) \in \{1, 2\}$ is a deterministic function. Thanks to the second claim of Lemma 13.7.5 and (13.5.5), (13.5.12), and (13.5.15) (on pp. 453 and 455), we have

$$\mathcal{Q} = \lim_{j\to\infty} \mathfrak{J}_T(\mathfrak{s}_j) \geq \limsup_{j\to\infty} \boldsymbol{E} r_0\big[\widehat{x}(0|0), u_j(0), \omega_1(j)\big]$$
$$= \limsup_{j\to\infty} \boldsymbol{E} r_0\big[\widehat{z}_1, \widehat{z}_2, u_0^{(\nu_j[\widehat{z}_1])}(j), \omega_1(j)\big].$$

Corollary 13.6.6 (on p. 459) and Lemma 13.7.7, along with the estimate (13.6.7) (on p. 459) and the Lebesgue dominated convergence theorem, ensure that

$$\boldsymbol{E} r_0\big[\widehat{z}_1, \widehat{z}_2, u_0^{(\nu_j[\widehat{z}_1])}(j), \omega_1(j)\big] - \boldsymbol{E} r_0\big[\widehat{z}_1, \widehat{z}_2, \overline{u}_0^{(\nu_j[\widehat{z}_1])}, \overline{\omega}_1\big] \to 0$$

as $j \to \infty$. Thus

$$\mathcal{Q} \geq \limsup_{j\to\infty} \boldsymbol{E} r_0\big[\widehat{z}_1, \widehat{z}_2, \overline{u}_0^{(\nu_j[\widehat{z}_1])}, \overline{\omega}_1\big].$$

Now we employ (13.7.3). As follows from Corollary 13.7.3 and Lemma 13.7.4, the set

$$\mathcal{M}^\nu(z_1) := \Big\{z_2 : \lambda_\nu(z_1) + (-1)^\nu\big[\overline{u}_0^{(2)} - \overline{u}_0^{(1)}\big]z_2 + \mathcal{O}_\nu(z_2, z_1) \geq 1\Big\}$$

contains an infinite subinterval of the real line for any z_1 and $\nu = 1, 2$. Denote by $I_{\mathcal{M}}(\cdot)$ the indicator function of the set $\mathcal{M}$. It follows from (13.7.3) that

$$\boldsymbol{E} r_0\big[\widehat{z}_1, \widehat{z}_2, \overline{u}_0^{(\nu_j[\widehat{z}_1])}, \overline{\omega}_1\big] \geq \boldsymbol{E} q_0\big[\widehat{z}_1, \widehat{z}_2, \overline{\omega}_0, \overline{\omega}_1\big] + \boldsymbol{E} I_{\mathcal{M}^{\nu_j[\widehat{z}_1]}(\widehat{z}_1)}\big[\widehat{z}_2\big].$$

Owing to (13.5.15) (with $t := 0, \omega_j := \overline{\omega}_j$), (13.5.12) (on p. 455) and Remark 13.7.6, the first summand on the right equals $\mathcal{Q}$. To estimate the second one, note that

$$\widehat{x}(0|0) = 1/2y(0) = 1/2[a + \chi(0)]$$

is a zero-mean Gaussian vector with the correlation matrix $1/2I$. Hence $\widehat{z}_1$ and $\widehat{z}_2$ are independent (nonsingular) Gaussian random quantities. It follows that the quantity

$$\mu(z_1) := \min_{\nu=1,2} \boldsymbol{E}\Big(I_{\mathcal{M}^\nu(\widehat{z}_1)}\big[\widehat{z}_2\big]\Big|\widehat{z}_1\Big)$$

is strictly positive function of $z_1 = \widehat{z}_1$. Thus

$$\varkappa := \boldsymbol{E} I_{\mathcal{M}^{\nu_j[\widehat{z}_1]}(\widehat{z}_1)}\big[\widehat{z}_2\big] = \boldsymbol{E}\boldsymbol{E}\Big(I_{\mathcal{M}^{\nu_j[\widehat{z}_1]}(\widehat{z}_1)}\big[\widehat{z}_2\big]\Big|\widehat{z}_1\Big) \geq \boldsymbol{E}\big[\mu(\widehat{z}_1)\big] > 0.$$

Summarizing, we arrive at a contradiction:

$$\mathcal{Q} \geq \limsup_{j\to\infty} \boldsymbol{E} r_0\big[\widehat{z}_1, \widehat{z}_2, \overline{u}_0^{(\nu_j[\widehat{z}_1])}, \overline{\omega}_1\big] \geq \mathcal{Q} + \varkappa.$$

This proves that the initial hypothesis $\inf_{\mathfrak{S}} \mathfrak{J}_T = \inf_{\mathfrak{S}_{\text{LQG}}} \mathfrak{J}_T$ fails to be true and so $\inf_{\mathfrak{S}} \mathfrak{J}_T < \inf_{\mathfrak{S}_{\text{LQG}}} \mathfrak{J}_T$. □

14

Kalman State Estimation in Networked Systems with Asynchronous Communication Channels and Switched Sensors

14.1 Introduction

In this chapter, we consider the sensor control problem that consists in estimating the state of an uncertain process based on measurements obtained over asynchronous communication channels from noisy controlled sensors.

The classic estimation theory deals with the problem of forming an estimate of a process given measurements produced by sensors observing the process. The standard assumption is that data transmission and information processing required by the algorithm can be performed instantaneously. However in a number of newly arisen engineering applications, only a limited number of sensors can be remotely linked with the estimator via low bandwidth communication channels during any measurement interval; and the estimator can dynamically select which sensors use the channels. This gives rise to the sensor control (scheduling) problem. Such a problem may also arise when a flexible or intelligent sensor is able to operate in several modes and the estimator can dynamically switch this mode. Finally, sensor control problems arise when measurements from a large number of sensors are available to the estimator, but the computational power is such that only data from a small selection of the sensors can be processed at any given time, hence, forcing the estimator to dynamically select which sensor data are important for the task at hand.

Sensor scheduling has been addressed for continuous-time stochastic systems in [17,126,155] under the assumption that the information exchange is instantaneous. It was assumed that the process is generated by a known linear system driven by a white noise. It is shown that the optimal sensor schedule can be computed before the experiment has commenced, and this schedule is independent of the sensor data.

In this chapter, a sensor control problem is studied for a networked system. We consider a discrete-time linear partially observed system with Gaussian disturbances. The observations are sent to the estimator over communication channels, which provide random transmission delays, may lose data, and do not keep the succession of messages. The estimator is given a dynamic control over the measurements: It administrates forming messages to be sent from sensors. The corresponding control

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_14,

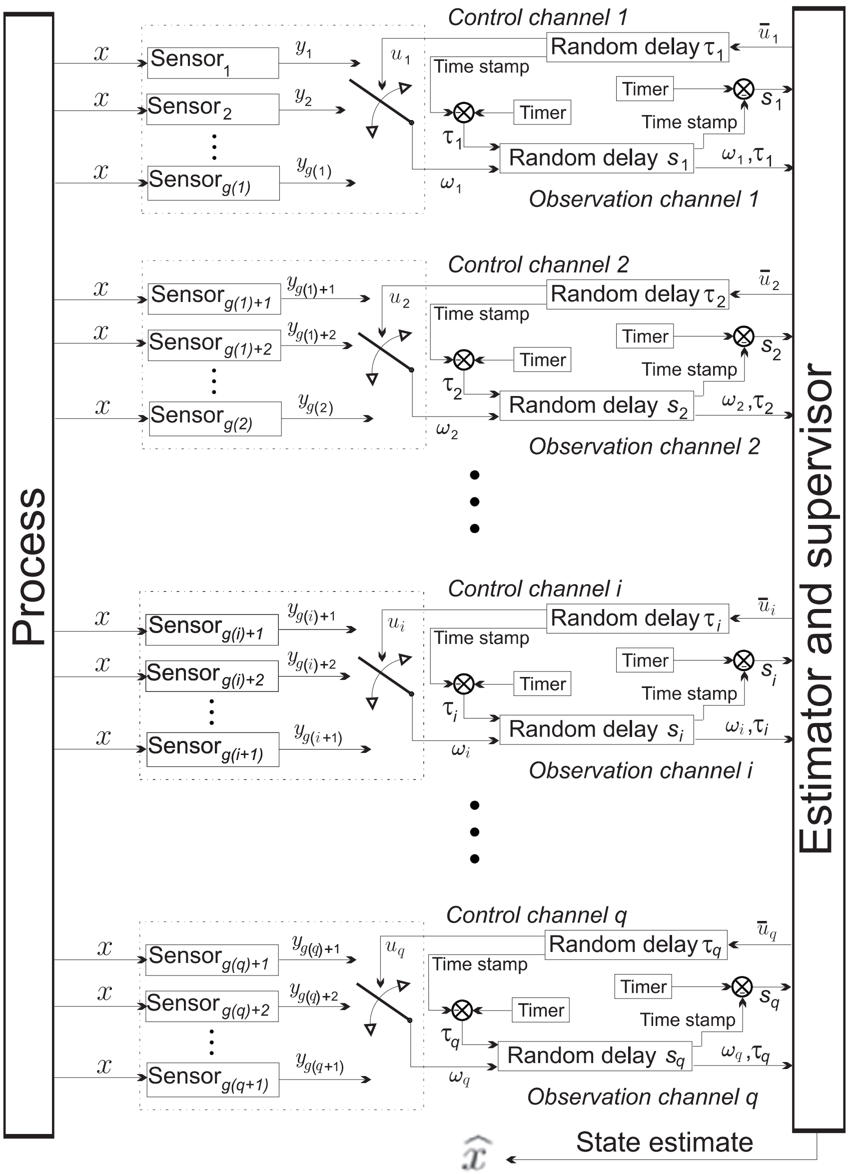


Fig. 14.1. Networked estimator with switched sensors.

is transmitted from the estimator to the remote sensor locations over communication channels with the features outlined (see Fig. 14.1). We consider the case where any message transferred via the channel is marked with a time stamp indicating the moment of the transfer beginning, which is characteristic of many channels. As a result, there is an awareness about the bygone states of the communication medium, whereas its future states are unknown. The statistics of the data delays in the communication channels is assumed to be known. The objective is to find a minimum variance state estimation. The minimum is over not only state estimates but also nonanticipating sensor control strategies.

The optimal state estimate and sensor control strategy are obtained. It is shown that the estimate is generated by an analog of the conventional Kalman filter provided the control strategy is given. We also demonstrate that the optimal sensor control can be computed by solving a difference Riccati equation and a dynamic programming procedure. It is proved that this control depends only on the history of the communication medium monitoring and does not employ the sensor data. Finally we apply ideas of model predictive control (see, e.g., [29]) to derive a nonoptimal but implementable real-time method for sensor control.

The main results of this chapter were originally presented in [118].

Its outline is as follows. Section 14.2 offers the problem statement. Section 14.3 contains the basic assumptions. The minimum variance state estimator and the optimal sensor control strategy are presented in Sects. 14.4 and 14.6, respectively. The proofs of the main results from these sections are given in Sects. 14.5 and 14.7, respectively. In Sect. 14.8, we apply the model predictive control approach.

14.2 Problem Statement

We consider the following discrete-time linear system:

$$x(t+1) = A(t)x(t) + \xi(t) \quad t = 0, \dots, T-1, \qquad x(0) = a; \tag{14.2.1}$$

$$y(t) = C(t)x(t) + \chi(t), \qquad t = 0, \dots, T. \tag{14.2.2}$$

Here $x(t) \in \mathbb{R}^n$ is the state; $\xi(t) \in \mathbb{R}^n$ is a process disturbance; $y(t) \in \mathbb{R}^k$ is the vector comprising the outputs of the sensors; and $\chi(t) \in \mathbb{R}^k$ is a noise.

Measurements transmission. The observations are sent to the estimator via communication channels. Due to the limited bandwidth of the channels, the entire information represented by y cannot be sent. Its part to be transmitted is chosen dynamically by the estimator by generating the *sensor control* $u \in U \subset \mathbb{R}^s$. As a result, transmitted is the vector

$$\omega(t) = D[t, u(t)]y(t) \in \mathbb{R}^l. \tag{14.2.3}$$

Here $D(t, u)$ is a given $l \times k$ matrix-function and the dimension of $\omega(t)$ is typically less than that of $y(t)$.

Remark 14.2.1. In this remark, we consider examples of sensor control and assume for the sake of simplicitness that any scalar coordinate y_i of the vector y from (14.2.2) represents the output of a particular sensor.

(i) The sensor control may consist in selecting a set of sensors to be linked with the estimator. Then $u = \textbf{col}\,(i[1], \ldots, i[l])$ indicates the serial numbers $i[\nu] = 1, \ldots, k$ of the corresponding sensors and $\omega := \textbf{col}\,\big(y_{i[1]}, \ldots, y_{i[l]}\big)$. (We suppose that as many as l sensors can be simultaneously connected with the estimator.)

(ii) Suppose that the sensors are organized into several groups

$$y_1, y_2, \ldots, y_k = \big|y_1, \ldots, y_{g[1]}\big|y_{g[1]+1}, \ldots, y_{g[2]}\big| \cdots \big|y_{g[l-1]+1}, \ldots, y_{g[l]}\big|,$$

each deputing one signal to be transmitted. Then in the previous example, the set of sensor controls U consists of the tuples $u = \textbf{col}\,(i[1], \ldots, i[l])$ such that $g[\nu-1]+1 \leq i[\nu] \leq g[\nu]$ for all ν, where $g[0] := 0$. (Fig. 14.1 concerns this case with $l = q$.[1])

(iii) Any group is equipped with a preprocessor converting linearly the measurements produced by this group into a single scalar signal

$$\omega_\nu := \alpha(\nu, 1) y_{g[\nu]+1} + \cdots + \alpha(\nu, p_\nu) y_{g[\nu]+p_\nu} \qquad (p_\nu := g[\nu] - g[\nu-1])$$

to be transmitted. Then

$$\begin{array}{l} \omega = \textbf{col}\,(\omega_1, \ldots, \omega_l) \\ u = \textbf{col}\,(u_1, \ldots, u_l) \end{array}, \quad \text{where} \quad u_\nu := \textbf{col}\,[\alpha(\nu, 1), \ldots, \alpha(\nu, p_\nu)]\,.$$

It is natural to normalize u_ν, e.g., as follows:

$$\alpha(\nu, i) \geq 0, \qquad \alpha(\nu, 1) + \cdots + \alpha(\nu, p_\nu) = 1.$$

The set U consists of all tuples u normalized in such a way. This means that all linear "conversions" are admitted.

The vector ω from (14.2.3) is partitioned into several portions to be communicated over parallel channels

$$\omega = \textbf{col}\,\big(\omega_1, \ldots, \omega_q\big), \qquad \omega_\nu \in \mathbb{R}^{l_\nu}. \tag{14.2.4}$$

These portions incur independent transmission delays. So the estimator receives

$$z(t) := \Big\{\omega_\nu[\theta]\Big\}_{(\nu,\theta)\in S(t)} \tag{14.2.5}$$

at the time t. Here $S(t)$ is some, maybe, empty set such that

$$(\nu, \theta) \in S(t) \Rightarrow \theta \leq t \quad \text{and} \quad S(t_1) \cap S(t_2) = \emptyset \quad \text{whenever} \quad t_1 \neq t_2. \tag{14.2.6}$$

Remark 14.2.2. The observation signals may arrive at the estimator out of order. They may be also lost due to, e.g., noise in the communication medium and protocol malfunctions.

[1] Fig. 14.1 depicts a particular and yet typical case of the system considered in this chapter.

Transmission of the sensor controls. The sensor control is also sent from the estimator to the sensors over communication channels. (They are called the *control* channels as distinct from those transferring the measurements and called the *observation* ones.) Several channels are employed since the sensors may be distributed over a variety of locations. Correspondingly, the control u is subjected to a fixed partition $u = \mathbf{col}\,(u_1, \ldots, u_r)$, with u_i transmitted via the ith control channel. These channels provide random transmission delays and may lose data. As a result, the sensors receive and employ in (14.2.3) the following control at the current time t:

$$u(t) = \mathbf{col}\left(\overline{u}_1\left[t - \tau_1(t)\right], \ldots, \overline{u}_r\left[t - \tau_r(t)\right]\right). \tag{14.2.7}$$

Here

$$\overline{u}(\theta) = \mathbf{col}\,[\overline{u}_1(\theta), \ldots, \overline{u}_r(\theta)]$$

is the control produced by the estimator at time θ.

Remark 14.2.3. If several messages arrive via the ith control channel out of order, accepted is the most updated of them (i.e., produced at the latest moment). If no message arrives at the current time, the sensor control u_i is kept unchanged. If no message has arrived until t, we put $u_i(t) := u_i^0$, where the control u^0 is chosen a priori. Then $\tau_i(t) := t + 1$ and $\overline{u}(-1) := u^0$ in (14.2.7). Hence

$$\tau_i(t) \geq 0 \quad \text{and} \quad t_1 \leq t_2 \Rightarrow t_1 - \tau_i(t_1) \leq t_2 - \tau_i(t_2). \tag{14.2.8}$$

Information about the past states of the communication network. Many communication channels equip the messages transferred with time stamps. This enables one to calculate the time taken to transmit the message at the moment of its arrival. We suppose that the estimator has access to this information. In particular, it is aware of the control channel transmission times: Their values are revealed at the sensor sites and sent to the estimator via the observation channels. The estimator can likewise receive information about channels competing with control and observation ones for network resources, which may be useful to predict future states of the channels serving the system at hand. All the above information concerns the bygone states of the communication medium.

Notation 14.2.4. *The information about the bygone states of the communication network currently received by the estimator is organized in a tuple denoted by* $\varkappa(t)$.

Although we do not specify the content of the message $\varkappa$, some assumptions about it will be given in the next section.

Problem statement. The class of admissible control strategies is as follows:

$$\overline{u}(t) = \mathfrak{U}\big[t, \mathfrak{Z}(t-1), \Delta(t)\big], \quad \text{where}$$
$$\mathfrak{Z}(t) := \mathbf{col}\left[z(0), \ldots, z(t)\right] \text{ for } t \geq 1,\ \mathfrak{Z}(-1) := 0 \in \mathbb{R},$$
$$\text{and} \quad \Delta(t) := \mathbf{col}\left[\varkappa(0), \ldots, \varkappa(t)\right]. \tag{14.2.9}$$

The problem is to find a minimum variance estimate of the current state $x(t)$ from (14.2.1). In other words, we look for a deterministic function of the observations

$$\widehat{x}(t|t) = \mathfrak{X}\big[t, \mathfrak{Z}(t), \Delta(t)\big] \tag{14.2.10}$$

that, along with the function $\mathfrak{U}(\cdot)$ from (14.2.9), minimizes the total estimation error

$$\mathfrak{I} := \sum_{t=0}^{T} \boldsymbol{E}\big\|e(t|t)\big\|^2, \quad e(t|t) := x(t) - \widehat{x}(t|t). \tag{14.2.11}$$

14.3 Assumptions

In this section, we state the assumptions adopted in this chapter.

14.3.1 Properties of the Sensors and the Process

Assumption 14.3.1. *The vectors $a, \xi(t)$, and $\chi(t)$ from* (14.2.1) *and* (14.2.2) *are random, Gaussian, and independent with $\boldsymbol{E}\xi(t) = 0$ and $\boldsymbol{E}\chi(t) = 0$. The mean $\boldsymbol{E}a$ and the correlation matrices*

$$R_{aa} := \boldsymbol{E}[a - \boldsymbol{E}a][a - \boldsymbol{E}a]^{\mathrm{T}}, \quad R_{\xi\xi}(t) := \boldsymbol{E}\xi(t)\xi(t)^{\mathrm{T}}, \quad R_{\chi\chi}(t) := \boldsymbol{E}\chi(t)\chi(t)^{\mathrm{T}}$$

are known. So are the matrices $A(t), B(t), C(t)$, and $D(t,u)$ from (14.2.1)–(14.2.3)*.*

Assumption 14.3.2. *The quantities $\{\varkappa(t)\}$ from Notation* 14.2.4 *are random and independent of a, $\{\xi(t)\}$, and $\{\chi(t)\}$.*

Assumption 14.3.3. *Given $t = 0, \ldots, T$, the matrix function $D(t, \cdot)$ from* (14.2.3) *is defined on a finite set $U \subset \mathbb{R}^s$. Consider the partition*

$$D(t, u) = \textbf{col}\left[D_1(t, u), \ldots, D_q(t, u)\right]$$

corresponding to the partition (14.2.4) *of the vector ω. Whenever $\nu \neq \eta$, the following relation holds:*

$$D_\nu(\theta, u) R_{\chi\chi}(\theta) D_\eta(\theta, u)^{\mathrm{T}} = 0 \qquad \forall \theta, u.$$

The last claim is true, e.g., in the cases from Remark 14.2.1 if the various scalar entries $\chi_1(\theta), \ldots, \chi_k(\theta)$ of the noise $\chi(\theta)$ are independent. The last condition is natural in the context of Remark 14.2.1 where these entries represent noises in different sensors.

Typically, any control channel serves a particular group of sensors, and these groups employ disjoint sets of observation channels. (See Fig. 14.1.) Then a given observation signal ω_ν from (14.2.4) is affected by only one control $\overline{u}_i$ from (14.2.7). A generalization of this property gives rise to the following assumption.

Assumption 14.3.4. *For any $\nu \in [1:q]$, a set $I_\nu \subset [1:r]$ exists such that the matrix $D_\nu(t, u)$ depends only on the entries u_i of $u =$ **col** $(u_1, \ldots, u_r)$ with $i \in I_\nu$.*

14.3.2 Information about the Past States of the Communication Network

We suppose that this information includes the observation channel time stamps. In view of Notation 14.2.4, this gives rise to the following.

Assumption 14.3.5. *The estimator is able to determine the set $S(t)$ from* (14.2.5) *at the current time t; i.e., $S(t) = \mathcal{S}[t, \varkappa(t)]$.*

We recall that capital script letters denote deterministic functions.

It is tacitly supposed in Assumption 14.3.4 that the observation ω_ν is produced by the sensors controlled by the signals u_i with $i \in I_\nu$. The corresponding delays $\tau_i(\theta)$ from (14.2.7) can be determined on the basis of the time stamps at the moment θ when the signal ω_ν departs. We suppose that this information is attached to ω_ν in the observation message. As a result, we arrive at the following assumption.

Assumption 14.3.6. *At the current time t, the estimator is able to determine the delays $\tau_i(\theta)$ from* (14.2.7) *for $i \in I_\nu$ and $(\nu, \theta) \in S(t)$; i.e.,*

$$\tau_i(\theta) = \mathcal{T}_i\big[\theta, \varkappa(t)\big]$$

whenever an index ν exists such that $(\nu, \theta) \in S(t) = \mathcal{S}[t, \varkappa(t)]$ and $i \in I_\nu$. Here the function $\mathcal{T}_i(\cdot)$ takes integer values and $0 \leq \mathcal{T}_i(\theta, \varkappa) \leq \theta + 1$.

14.3.3 Properties of the Communication Network

Assumption 14.3.7. *The delays in the communication channels are bounded by known constants: $t - \theta \leq \sigma$ and $\mathcal{T}_i(\theta, \varkappa) \leq \sigma_*$ whenever $(\nu, \theta) \in \mathcal{S}(t, \varkappa)$ and $i \in I_\nu$. Here $\mathcal{S}(\cdot)$ and $\mathcal{T}_i(\cdot)$ are taken from Assumptions* 14.3.5 *and* 14.3.6*, respectively.*

The following assumption is adopted only to simplify the formulations.

Assumption 14.3.8. *The set Υ of the values that can be taken by the tuple $\varkappa$ from Notation* 14.2.4 *is finite.*

The next assumption to follow means that the communication medium is a system with a finite aftereffect.

Assumption 14.3.9. *A known constant $\overline{\sigma} = 0, 1, \ldots$ exists such that for any $t = 0, 1, \ldots$ and $\varkappa_0, \ldots, \varkappa_{t+1} \in \Upsilon$, the following relation (where $\overline{\sigma}[t] := \min\{\overline{\sigma}, t\}$) holds for the conditional distribution of the tuple $\varkappa$ from Notation* 14.2.4*:*

$$\boldsymbol{P}\left[\varkappa(t+1) = \varkappa_{t+1} \middle| \varkappa(t) = \varkappa_t, \ldots, \varkappa(0) = \varkappa_0\right] = \boldsymbol{P}\left[\varkappa(t+1) = \varkappa_{t+1} \middle| \varkappa(t) = \varkappa_t, \ldots, \varkappa(t - \overline{\sigma}[t]) = \varkappa_{t-\overline{\sigma}(t)}\right]. \quad (14.3.12)$$

This distribution is known in advance for all $t = 0, 1, \ldots, T$.

Note that (14.3.12) is trivially true if $t \leq \overline{\sigma}$ since then $\overline{\sigma}(t) = t$.

Remark 14.3.10. We suppose that (14.2.6) and (14.2.8) hold almost surely.

Remark 14.3.11. We restrict ourselves to consideration of the sensor control strategies (14.2.9) and state estimates (14.2.10) with measurable functions $\mathcal{U}(\cdot)$ and $\mathcal{X}(\cdot)$, respectively.

14.4 Minimum Variance State Estimator for a Given Sensor Control

In this section, we assume that the sensor control (14.2.9) is chosen and fixed. Under this assumption, we find the minimum variance state estimate (14.2.10). As will be shown in Sect. 14.6, this estimate is a part of the solution of the primal problem stated in Sect. 14.2. More precisely, after determining the optimal control strategy (14.2.9), the complete solution results from supplementing it with the estimate found in this section.

Denote by $\widehat{x}(j|t)$ the minimum variance estimate of the state $x(j)$ based on

$$z(0), \ldots, z(t), \varkappa(0), \ldots, \varkappa(t).$$

Being coupled with certain matrices $P_{ij}(t), \overline{P}_{ij}(t), i, j = 0, \ldots, \sigma$, the tuple

$$\widehat{X}(t) = \textbf{col}\left[\widehat{x}(t|t), \widehat{x}(t-1|t), \ldots, \widehat{x}(t-\sigma|t)\right] \tag{14.4.1}$$

may be generated recursively by the following analog of the filter described in Subsect. 11.3.2 (starting on p. 375).

Explanation 14.4.1. Here σ is the constant from Assumption 14.3.7.

Recursive State Estimator

The next tuple $\widehat{X}(t+1)$ *is generated by equations*

$$\widehat{x}(j|t+1) = \widehat{x}(j|t) + \sum_{(\nu,\theta)\in S(t+1)} K_{t+1-j}^{(\nu,\theta)}(t+1)\left[\omega_\nu(\theta) - \widehat{\omega}_\nu(\theta|t)\right], \tag{14.4.2}$$

where $j = t+1, t, \ldots, t+1-\sigma$ *and*

$$\widehat{x}(t+1|t) := A(t)\widehat{x}(t|t), \quad \widehat{\omega}_\nu(\theta|t) := \mathfrak{C}_\nu(\theta)\widehat{x}(\theta|t),$$
$$\mathfrak{C}_\nu(\theta) := D_\nu\left[\theta, u(\theta)\right]C(\theta). \tag{14.4.3}$$

Here $C(\theta)$ *and* $D_\nu[\theta, u]$ *are the matrices from* (14.2.2) *and Assumption* 14.3.3*, respectively.*

The gain matrices $K_j^s(t)$ *are enumerated by the pairs* $[j, s]$ *with* $j = 0, \ldots, \sigma$ *and* $s = (\nu, \theta) \in S(t)$*, have the respective dimensions* $n \times l_\nu, l_\nu = \dim(\omega_\nu)$*, and are calculated as follows:*

$$K_j^s(t) = \sum_{(\nu,\theta)\in S(t)} P_{j,t-\theta}(t)\mathfrak{C}_\nu(\theta)^{\mathrm{T}}\overset{+}{\Lambda}{}^{s}_{(\nu,\theta)}. \tag{14.4.4}$$

Here $\overset{+}{\Lambda}$ *is the pseudoinverse of the square ensemble of matrices*[2] Λ *over the set* $S(t)$ *that is given by*

[2]See Subsect. 11.3.1 starting on p. 374.

$$\Lambda_{s_1}^{s_2} = \mathfrak{C}_{\nu_1}(\theta_1) P_{t-\theta_1, t-\theta_2}(t) \mathfrak{C}_{\nu_2}(\theta_2)^{\mathrm{T}} + \begin{cases} \mathfrak{D}_{\nu_1}(\theta_1) R_{\chi\chi}^{\nu_1}(\theta_1) \mathfrak{D}_{\nu_1}(\theta_1)^{\mathrm{T}} & \text{if} \quad s_1 = s_2 \\ 0 & \text{otherwise} \end{cases}, \quad (14.4.5)$$

where

$$s_i = (\nu_i, \theta_i) \in S(t), \qquad \mathfrak{D}_\nu(\theta) := D_\nu[\theta, u(\theta)].$$

The ensembles (over $[0:\sigma]$*) of matrices* $P_{ij}(t), \overline{P}_{ij}(t)$ *are generated recursively*

$$\mapsto \cdots \left\{\overline{P}_{ij}(t)\right\} \mapsto \left\{P_{ij}(t+1)\right\} \mapsto \left\{\overline{P}_{ij}(t+1)\right\} \cdots \mapsto \quad (14.4.6)$$

by equations

$$P_{ij}(t+1) = \begin{cases} A(t)\overline{P}_{ij}(t)A(t)^{\mathrm{T}} + R_{\xi\xi}(t) & \text{if} \quad i = j = 0 \\ A(t)\overline{P}_{i,j-1}(t) & \text{if} \quad i = 0, j \geq 1 \\ \overline{P}_{i-1,j}(t)A(t)^{\mathrm{T}} & \text{if} \quad i \geq 1, j = 0 \\ \overline{P}_{i-1,j-1}(t) & \text{if} \quad i, j \geq 1 \end{cases}; \quad (14.4.7)$$

$$\overline{P}_{ij}(t+1) = P_{ij}(t+1) - \sum_{(\nu,\theta)\in S(t+1)} K_i^{(\nu,\theta)}(t+1)\mathfrak{C}_\nu(\theta) P_{t+1-\theta,j}(t+1). \quad (14.4.8)$$

The recursion (14.4.2), (14.4.6) *is initialized by putting*

$$A(-1) := I, \ \widehat{x}(-1|-1) := \boldsymbol{E}a, \ \widehat{x}(-1-j|-1) := 0 \quad \forall j = 1, \ldots, \sigma, \quad R_{\xi\xi}(-1) := R_{aa}, \quad \overline{P}(-1) := 0. \quad (14.4.9)$$

Remark 14.4.2. Thanks to Assumptions 14.3.1, 14.3.4, and 14.3.6, the matrices (14.4.3) with $(\nu, \theta) \in S(t)$ become known to the decision maker at the time t. (We suppose that the values $\overline{u}[t - \widehat{\sigma}(t)], \ldots, \overline{u}[t]$ are kept in its memory. Here $\widehat{\sigma}(t) := \min\{\sigma + \sigma_*, t+1\}$ and σ, σ_* are taken from Assumption 14.3.7.)

Remark 14.4.3. It is easy to see that the proposed state estimator is basically that from Subsect. 11.3.2 (starting on p. 375) with the matrices $C_\nu(\theta)$ and $R_{\chi\chi}^\nu(\theta)$ altered as follows:

$$C_\nu(\theta) := \mathfrak{C}_\nu(\theta), \qquad R_{\chi\chi}^\nu(\theta) := \mathfrak{D}_\nu(\theta) R_{\chi\chi}^\nu(\theta) \mathfrak{D}_\nu(\theta)^{\mathrm{T}}.$$

The main result of this section is offered by the following.

Theorem 14.4.4. *Suppose that the control strategy* (14.2.9) *(on p.* 473*) is chosen and fixed. Then the above estimator generates the minimum variance estimations; i.e.,*

$$\widehat{x}(j|t) = \boldsymbol{E}\left[x(j)|\mathfrak{Z}(t), \Delta(t)\right]$$

whenever $t - \sigma \leq j \leq t$ *and* $j \geq 0$*. Here* $\mathfrak{Z}(t)$ *and* $\Delta(t)$ *are defined in* (14.2.9)*. The matrices* $P_{ij}(t), \overline{P}_{ij}(t)$ *generated by the estimator are the conditional covariance matrices of the estimation errors* $e(\theta|s) := x(\theta) - \widehat{x}(\theta|s)$*. More precisely, whenever* $i, j = 0, \ldots, \sigma$ *and* $t - i \geq 0, t - j \geq 0$*, the following relations hold:*

$$\overline{P}_{ij}(t) := \boldsymbol{E}\left[e(t-i|t)e(t-j|t)^{\mathrm{T}}\middle|\mathfrak{Z}(t-1), \Delta(t)\right];$$
$$P_{ij}(t) := \boldsymbol{E}\left[e(t-i|t-1)e(t-j|t-1)^{\mathrm{T}}\middle|\mathfrak{Z}(t-1), \Delta(t-1)\right].$$

14.5 Proof of Theorem 14.4.4

The critical point in the proof will be the following fact.

Lemma 14.5.1 ([55]). *Consider the linear discrete time system*

$$X(t+1) = \mathfrak{A}(t)X(t) + \Sigma(t), \quad X(0) = \mathfrak{a}, \quad Z(t) = \mathfrak{C}(t)X(t) + \Xi(t). \tag{14.5.1}$$

Here $X(t) \in \mathbb{R}^N$ *is the state,* $\Sigma(t) \in \mathbb{R}^N$ *is the process disturbance,* $Z(t) \in \mathbb{R}^K$ *is the sensor output, and* $\Xi(t) \in \mathbb{R}^K$ *is the noise. The matrix* $\mathfrak{C}(t)$ *is a deterministic function of the past measurements:*

$$\mathfrak{C}(t) = \mathcal{C}[t, Z(0), \ldots, Z(t-1)]. \tag{14.5.2}$$

This function $\mathcal{C}[\cdot]$ *is measurable and bounded and, along with the matrices*

$$\mathfrak{A}(0), \ldots, \mathfrak{A}(T-1),$$

is known in advance. For any t, the vector $\Sigma(t)$ *is independent of*

$$\mathfrak{a}, \Sigma(0), \ldots, \Sigma(t-1), \Xi(0), \ldots, \Xi(t),$$

Gaussian, and zero-mean. The vector $\mathfrak{a}$ *is Gaussian. Given*

$$t \quad \text{and} \quad Z(0), \ldots, Z(t-1),$$

the vector $\Xi(t)$ *is Gaussian, zero-mean, and independent of*

$$\mathfrak{a}, \Sigma(0), \ldots, \Sigma(t-1).$$

Then the minimum variance estimate

$$\widehat{X}(t) = \boldsymbol{E}\left[X(t)\middle|Z(0), \ldots, Z(t)\right] \tag{14.5.3}$$

is generated by the standard Kalman filter.

Remark 14.5.2. The formulas for the standard Kalman filter employ the correlation matrices $R_{\Xi\Xi}(t)$ of the noise $\Xi(t)$. Under the circumstances, utilized are the conditional covariance matrices

$$R_{\Xi\Xi}(t) := \boldsymbol{E}\left[\Xi(t)\Xi(t)^{\mathrm{T}}\middle|Z(0), \ldots, Z(t-1)\right].$$

Remark 14.5.3. Apart from the state estimate, the Kalman filter generates apriorial $\mathcal{P}(t)$ and aposteriorial $\overline{\mathcal{P}}(t)$ covariance matrices of the estimation error. Under the circumstances, the filter generates conditional covariance matrices, where conditioning is over $Z(0), \ldots, Z(t-1)$.

The proof of Theorem 14.4.4 follows the lines of the state augmenting technique considered in Sect. 11.7 (starting on p. 384).

State Space Augmentation

Now we are going to rewrite the primal system introduced in Sect. 14.2 in the form (14.5.1). To this end, we consider the linear space

$$\mathcal{Z} := \left\{ Z = \{z_{\nu,j}\}_{\nu=1}^{q} {}_{j=0}^{\sigma} : z_{\nu,j} \in \mathbb{R}^{l_\nu} \ \forall \nu, j \right\},$$

where q and l_ν are taken from (14.2.4) (on p. 472). Furthermore, we put

$$x(-\sigma) := x(-\sigma+1) := \cdots := x(-1) := 0$$

and introduce the vectors

$$X(t) := \textbf{col}\left[x(t), x(t-1), \ldots, x(t-\sigma)\right],$$

$$Z(t) := \left\{z_{\nu,j}\right\} \in \mathcal{Z}, \quad \text{where} \quad z_{\nu,j} := \begin{cases} \omega_\nu(t-j) & \text{if} \quad (\nu, t-j) \in S(t) \\ 0 & \text{otherwise} \end{cases}.$$

Here $Z(t)$ can be interpreted as the vector (14.2.5) (on p. 472) supplemented with several zeros. In terms of $X(t)$ and $Z(t)$, relations (14.2.1)–(14.2.5) (on pp. 471 and 472) take the form (14.5.1), where

$$\mathfrak{A}(t) := \begin{pmatrix} A(t) & 0 & 0 & \cdots & 0 & 0 \\ I & 0 & 0 & \cdots & 0 & 0 \\ 0 & I & 0 & \cdots & 0 & 0 \\ 0 & 0 & I & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & I & 0 \end{pmatrix}, \quad \Sigma(t) = \mathfrak{E}\xi(t), \quad \mathfrak{E} := \textbf{col}\left[I, 0, \ldots, 0\right];$$

$$\mathfrak{a} := \textbf{col}\,[a, 0, \ldots, 0]. \tag{14.5.4}$$

Furthermore for any $t \leq T$ and $X = \textbf{col}\,(x_0, x_1, \ldots, x_\sigma) \in \mathbb{R}^{(\sigma+1)n}$,

$$\mathfrak{C}[t]X := \left\{z_{\nu,j}\right\} \in \mathcal{Z}, \text{ where}$$
$$z_{\nu,j} := \begin{cases} D_\nu[t-j, u(t-j)]C(t-j)x_j & \text{if } (\nu, t-j) \in S(t) \\ 0 & \text{otherwise} \end{cases}; \tag{14.5.5}$$

$$\Xi[t] := \left\{\zeta_{\nu,j}\right\} \in \mathcal{Z}, \text{ where}$$
$$\zeta_{\nu,j} := \begin{cases} D_\nu[t-j, u(t-j)]\chi(t-j) & \text{if } (\nu, t-j) \in S(t) \\ 0 & \text{otherwise} \end{cases}. \tag{14.5.6}$$

Now choose and fix a realization $\overline{\Delta} = \{\varkappa_0, \ldots, \varkappa_T\}$ of the random sequence $\Delta = \{\varkappa(0), \ldots, \varkappa(T)\}$ that is taken with a positive probability. Until otherwise stated, we shall deal with conditional distributions given $\Delta = \overline{\Delta}$. Under this assumption, the sets $S(t) = \mathcal{S}[t, \varkappa_t]$ and the delays $\tau_i(\theta) = \mathcal{T}_i[\theta, \varkappa_t]$, $(\nu, \theta) \in S(t), i \in I_\nu$ from (14.2.7) (on p. 473) are fixed due to Assumptions 14.3.5 and 14.3.6 (on p. 475). Our next goal is to show that the assumptions of Lemma 14.5.1 hold under the circumstances. In so doing, we need more preliminaries.

A Preliminary Technical Fact

It concerns the following quantities (where $t \leq \tau$):

$$Y_{\nu,\theta} := D_\nu[\theta, u(\theta)]\chi(\theta),$$
$$Y^t(\tau) := \left\{Y_{\nu,\theta}\right\}_{(\nu,\theta)\in S(\tau),\theta\leq t}, \qquad Y(t) := Y^t(t). \quad (14.5.7)$$

Lemma 14.5.4. *Let* $t = 0, \ldots, T$. *Given* $z(0), \ldots, z(t-1)$, *the vectors*

$$Y(t), Y^t(t+1), \ldots, Y^t(T)$$

are mutually independent, zero-mean, Gaussian, and independent of

$$Y(0), \ldots, Y(t-1), a, \xi(0), \ldots, \xi(t-1).$$

Furthermore,

$$\boldsymbol{E}\left[Y_{\nu',\theta'}Y^{\mathrm{T}}_{\nu'',\theta''}\middle| z(0), \ldots, z(t-1)\right] = 0$$
$$\text{whenever} \quad (\nu',\theta'), (\nu'',\theta'') \in S(\tau), \quad (\nu',\theta') \neq (\nu'',\theta''), \quad \theta', \theta'' \leq t \leq \tau;$$
$$\boldsymbol{E}\left[Y_{\nu,\theta}Y^{\mathrm{T}}_{\nu,\theta}\middle| z(0), \ldots, z(t-1)\right] = D_{\nu,\theta}[\theta, u(\theta)]R_{\chi\chi}(\theta)D_{\nu,\theta}[\theta, u(\theta)]^{\mathrm{T}}$$
$$\text{whenever} \quad (\nu,\theta) \in S(\tau), \quad \theta \leq t \leq \tau. \quad (14.5.8)$$

Remark 14.5.5. Conditioning over $z(0), \ldots, z(t-1)$ is dropped everywhere if $t = 0$.

Remark 14.5.6. By (14.2.7) and (14.2.9) (on p. 473), $u(\theta)$ is a deterministic function of $z(0), \ldots, z(t-1)$ in (14.5.8).

Proof of Lemma 14.5.4. The proof will be by induction on $t = 0, \ldots, T$. For $t = 0$, the control $u(0)$ is deterministic thanks to Remark 14.2.3 and (14.2.9) (on p. 473). By Assumption 14.3.5 (on p. 475) and (14.5.7),

$$Y^0(t) = \{D_\nu[0, u(0)]\chi(0)\}_{\nu\in\mathcal{N}(t)},$$

where

$$\mathcal{N}(t) := \{\nu : (\nu, 0) \in \mathcal{S}[t, \varkappa_t]\}.$$

Here the sets $\mathcal{S}[0, \varkappa_0], \ldots, \mathcal{S}[T, \varkappa_T]$ are disjoint owing to (14.2.6) (on p. 472), and the vectors $D_\nu[0, u(0)]\chi(0)$ and $D_\eta[0, u(0)]\chi(0)$ with $\nu \neq \eta$ are uncorrelated thanks to Assumption 14.3.3 (on p. 474). So the statements to be proved follow from Assumption 14.3.1 (on p. 474).

Now suppose that these statements hold for some t. Owing to (14.5.7),

$$Y^{t+1}(\tau) = \left[Y^t(\tau), \mathcal{Y}^{t+1}(\tau)\right], \quad \text{where}$$
$$\mathcal{Y}^{t+1}(\tau) := \left\{D_\nu[t+1, u(t+1)]\chi(t+1)\right\}_{\nu:(\nu,t+1)\in S(\tau)} \quad (14.5.9)$$

for any $\tau \geq t+1$. By the induction hypothesis, the tuples

$$Y(t), Y^t(t+1), \dots, Y^t(T) \tag{14.5.10}$$

are mutually independent, zero-mean, Gaussian, and independent of

$$Y(0), \dots, Y(t-1), a, \xi(0), \dots, \xi(t-1) \quad \text{given} \quad z(0), \dots, z(t-1). \tag{14.5.11}$$

Due to (14.2.1)–(14.2.4), (14.2.6), (14.2.7), and Assumptions 14.3.1 and 14.3.2 (on pp. 471–474), $z(t)$ and all random quantities from (14.5.10) and (14.5.11) are independent of $\xi(t)$. Thanks to (14.2.1)–(14.2.7), (14.2.9), and (14.5.7),

$$z(t) = \mathcal{Z}\big[t, z(0), \dots, z(t-1), Y(t), a, \xi(0), \dots, \xi(t-1)\big].$$

It follows from the foregoing that given $z(0), \dots, z(t)$, the tuples $Y^t(t+1), \dots, Y^t(T)$ are mutually independent, zero-mean, Gaussian, and independent of

$$Y(0), \dots, Y(t), a, \xi(0), \dots, \xi(t).$$

Their distributions given $z(0), \dots, z(t)$ equal those given $z(0), \dots, z(t-1)$. Hence (14.5.8) is true provided the conditioning is over $z(0), \dots, z(t)$, whereas $\theta', \theta'', \theta \leq t, \tau \geq t+1$. Note also that $u(t+1)$ is a deterministic function of $z(0), \dots, z(t)$ by (14.2.7)–(14.2.9) and Assumptions 14.3.5 and 14.3.6 (on pp. 473–475). Furthermore the vector $\chi(t+1)$ is independent of $\xi(t), z(t)$ and all vectors from (14.5.10) and (14.5.11). By retracing the arguments from the first paragraph of the proof, we see that given $z(0), \dots, z(t)$, the tuples $\mathcal{Y}^{t+1}(t+1), \dots, \mathcal{Y}^{t+1}(T)$ are mutually independent, Gaussian, zero-mean, and independent of

$$Y(0), \dots, Y(t), Y^t(t+1), \dots, Y^t(T), a, \xi(0), \dots, \xi(t).$$

This and (14.5.9) imply that (14.5.8) holds with $t := t+1$, and given $z(0), \dots, z(t)$, the vectors $Y(t+1), Y^{t+1}(t+1), \dots, Y^{t+1}(T)$ are mutually independent, zero-mean, Gaussian, and independent of $Y(0), \dots, Y(t)$, a, $\xi(0), \dots, \xi(t)$. Thus the statements of the lemma are true for $t := t+1$, which completes the proof. $\square$

Completion of the Proof of Theorem 14.4.4

Now we recall that a realization of the random sequence $\{\varkappa(0), \dots, \varkappa(T)\}$ is fixed, and we deal with the corresponding conditional distributions. Then the assumptions of Lemma 14.5.1 hold for the augmented system. Indeed, formula (14.5.2) results from (14.2.7), (14.2.9) (on p. 473), and (14.5.5). It follows from (14.2.1)–(14.2.7), (14.2.9) (on pp. 471–473), (14.5.4), (14.5.6), and Assumptions 14.3.1 and 14.3.2 (on p. 474) that the vector $\Sigma(t)$ is independent of

$$\mathfrak{a}, \Sigma(0), \dots, \Sigma(t-1), \Xi(0), \dots, \mathit{\Xi}(t),$$

Gaussian, and zero-mean, whereas the vector $\mathfrak{a}$ is Gaussian. Lemma 14.5.4 justifies the last assumption of Lemma 14.5.1: Given t and $Z(0), \dots, Z(t-1)$, the vector $\mathit{\Xi}(t)$ is Gaussian, zero-mean, and independent of $\mathfrak{a}, \Sigma(0), \dots, \Sigma(t-1)$.

By Lemma 14.5.4, the standard Kalman filter (see Subsect. C.2 starting on p. 510 in Appendix C) gives the minimum variance estimate $\widehat{X}(t|t)$ of $X(t)$ based on $Z(0), \ldots, Z(t)$ and, due to the preliminary conditioning, $\varkappa(0), \ldots, \varkappa(T-1)$. This estimate is generated recursively so that only the sets $S(0), \ldots, S(t)$ and controls $u(0), \ldots, u(t)$ from (14.5.5)–(14.5.7) are required to compute $\widehat{X}(t|t)$. This, (14.2.7), (14.2.9), and Assumptions 14.3.5 and 14.3.6 (on p. 475) imply that $\widehat{X}(t|t)$ is in fact the minimum variance estimate of $X(t)$ based on $Z(0), \ldots, Z(t)$ and $\varkappa(0), \ldots, \varkappa(t)$, as is required.

Theorem 14.4.4 results from putting (14.5.4)–(14.5.6), and (14.5.8) into the formulas of the standard Kalman filter (i.e., (C.8)–(C.13) on pp. 510 and 511), along with elementary transformations of the resultant expressions.

14.6 Optimal Sensor Control

In this section, we obtain the optimal strategy to control the sensors. Along with the estimate from the previous section, it constitutes the complete solution of the primal problem.

Preliminaries

To introduce the optimal control, note that the recursion (14.4.6) (on p. 477) can be written in the form

$$\overline{P}(t) = \mathfrak{P}\Big\{\overline{P}[t-1], t, \varkappa[t], \overline{u}[t], \ldots, \overline{u}\big[t - \widehat{\sigma}(t)\big]\Big\}, \tag{14.6.1}$$

where

$$\overline{P} = \big\{\overline{P}_{ij}\big\}_{i,j=0}^{\sigma}, \qquad \widehat{\sigma}(t) := \min\{\sigma + \sigma_*, t+1\} \tag{14.6.2}$$

and σ, σ_* are taken from Assumption 14.3.7 (on p. 475). Indeed due to Assumptions 14.3.4–14.3.6, the matrix $D_\nu[\theta, u(\theta)]$ with $(\nu, \theta) \in S(t) = \mathcal{S}[t, \varkappa(t)]$ is determined by the tuple

$$\mathfrak{p} := \big\{t, \varkappa := \varkappa(t), u_0 := \overline{u}[t], \ldots, u_{\widehat{\sigma}(t)} := \overline{u}[t - \widehat{\sigma}(t)]\big\}.$$

Specifically,

$$D_\nu[\theta, u(\theta)] = \mathcal{D}_{\nu,\theta}(\mathfrak{p}) := D_\nu\left[\theta, \left\{u^{(j)}_{t-\theta+\mathcal{T}_j(\theta,\varkappa)}\right\}_{j\in I_\nu}\right] C(\theta). \tag{14.6.3}$$

Here **col** $(u_i^{(1)}, \ldots, u_i^{(r)})$ is the partition of u_i corresponding to (14.2.7) (on p. 473), and $t - \theta + \mathcal{T}_j(\theta, \varkappa) \leq \widehat{\sigma}(t)$ by Assumptions 14.3.6 and 14.3.7. The map $\mathfrak{P}(\cdot)$ in (14.6.1) acts as follows:

$$\alpha = \Big[\overline{P}, \mathfrak{p}\Big] \begin{array}{l} \nearrow \\ \xrightarrow{(14.6.3)} \end{array} \begin{array}{l} P = \big\{P_{ij}\big\} \\ \mathcal{D}_{\nu,\theta}(\mathfrak{p}) \end{array} \left| \begin{array}{l} \mapsto \Lambda \\ \quad\downarrow \\ \mapsto \mathfrak{P}(\alpha) := \Big\{\overline{P}^{+}_{ij}\Big\}_{i,j=0}^{\sigma} \end{array} \right. .$$

Here P_{ij} is the right-hand side of (14.4.7) (on p. 477) with $t := t-1, \overline{P}_{ij}(t-1) := \overline{P}_{ij}$, and the square ensemble of matrices Λ over $\mathcal{S}(t,\varkappa)$ is given by (14.4.5) (on p. 477), where

$$P_{ij}(t) := P_{ij}, \quad \mathfrak{D}_\nu(\theta) := \mathcal{D}_{\nu,\theta}(\mathfrak{p}), \quad \mathfrak{E}_\nu(\theta) := \mathfrak{D}_\nu(\theta)C(\theta),$$

and

$$\overline{P}_{ij}^{+} := P_{ij} - \sum_{s_1,s_2\in\mathcal{S}[t,\varkappa]} \mathcal{B}_{\nu_1,\theta_1}^{\mathrm{T}} \overset{+}{\Lambda}{}_{s_1}^{s_2} \mathcal{B}_{\nu_2,\theta_2},$$
$$\text{where } s_i = (\nu_i,\theta_i) \text{ and } \mathcal{B}_{\nu,\theta} := \mathcal{D}_{\nu,\theta}(\mathfrak{p})C(\theta)P_{t-\theta,j}. \quad (14.6.4)$$

Dynamic Programming Procedure

The key point in solution of the primal problem is the procedure:

$$\mathfrak{M}_T \mapsto \cdots \mapsto \mathfrak{M}_t := \left[W_t(\cdot) \curvearrowright \mathcal{W}_t(\cdot) \curvearrowright \mathcal{W}_t^m(\cdot)\right] \mapsto \cdots \mapsto \mathfrak{M}_0. \quad (14.6.5)$$

Here $W_t(\cdot), \mathcal{W}_t(\cdot), \mathcal{W}_t^m(\cdot)$ are real functions of the variables

$$\overline{P}, \quad \overline{\Delta} = \left[\varkappa_0,\ldots,\varkappa_{\overline{\sigma}(t)}\right], \quad \mathfrak{U} := \mathbf{col}\left[u_0,\ldots,u_{\widehat{\sigma}(t)}\right],$$

where $\overline{\sigma}(t)$ and $\widehat{\sigma}(t)$ are defined in Assumption 14.3.9 (on p. 475) and (14.6.2), respectively. The control variables on which the function $W_t(\cdot)$ depends in fact are $u_0,\ldots,u_{\widehat{\sigma}(t+1)-1}$, and $\mathcal{W}_t^m(\cdot)$ does not depend on u_0. The recursion (14.6.5) is initialized by putting $W_T(\cdot)\equiv 0$. Its step

$$W_t(\cdot) \to \mathcal{W}_t(\cdot) \to \mathcal{W}_t^m(\cdot) \to W_{t-1}(\cdot)$$

is as follows:

$$\mathcal{W}_t\left[\overline{P},\overline{\Delta},\mathfrak{U}\right] := \mathbf{tr}\,\mathfrak{P}_{00}\left[\overline{P},t,\varkappa_0,\mathfrak{U}\right] + W_t\Big\{\mathfrak{P}\left[\overline{P},t,\varkappa_0,\mathfrak{U}\right],\overline{\Delta},u_0,u_1,\ldots,u_{\widehat{\sigma}(t+1)-1}\Big\}, \quad (14.6.6)$$

where $\mathfrak{P}_{00}(\alpha)$ is the block of the matrix $\mathfrak{P}(\alpha)$;

$$\mathcal{W}_t^m\left[\overline{P},\overline{\Delta},u_1,\ldots,u_{\widehat{\sigma}(t)}\right] := \min_{u\in U}\mathcal{W}_t\Big\{\overline{P},\overline{\Delta},u,u_1,\ldots,u_{\widehat{\sigma}(t)}\Big\}; \quad (14.6.7)$$

$$W_{t-1}\left[\overline{P},\varkappa_0,\ldots,\varkappa_{\overline{\sigma}(t-1)},u_1,\ldots,u_{\widehat{\sigma}(t)}\right]$$
$$:= \sum_{\varkappa\in\Upsilon}\left\{\boldsymbol{P}\left[\varkappa|\varkappa_0,\ldots,\varkappa_{\overline{\sigma}(t-1)}\right]\times\mathcal{W}_t^m\left[\overline{P},\varkappa,\{\varkappa_\theta\}_{\theta=0}^{\overline{\sigma}(t)-1},\{u_\theta\}_{\theta=1}^{\widehat{\sigma}(t)}\right]\right\}. \quad (14.6.8)$$

Here $\boldsymbol{P}\left[\varkappa|\varkappa_0,\ldots,\varkappa_{\overline{\sigma}(t-1)}\right]$ is the conditional probability of the event $\varkappa(t) = \varkappa$ given $\varkappa(t-1) = \varkappa_0,\ldots,\varkappa(t-1-\overline{\sigma}[t-1]) = \varkappa_{\overline{\sigma}(t-1)}$, which is known in advance by Assumption 14.3.9 (on p. 475). Finally we put

$$\mathcal{U}_t\left[\overline{P},\overline{\Delta},u_1,\ldots,u_{\widehat{\sigma}(t)}\right] := u_*,$$

where u_* is an element furnishing the minimum in (14.6.7). Since the set U is finite by Assumption 14.3.3 (on p. 474), such an element does exist.

Remark 14.6.1. It is easy to see that the functions $W_t(\cdot)$, $\mathcal{W}_t(\cdot)$, and $\mathcal{W}_t^m(\cdot)$ are measurable. So the function $\mathcal{U}_t(\cdot)$ can be chosen measurable as well (see, e.g., [217, Sec. I.7]).

Main Result

The optimal sensor control strategy is described in the following theorem.

Theorem 14.6.2. *The following statements hold:*

(i) The infimum value of the error (14.2.11) *(on p.* 474*) over all sensor control strategies* (14.2.9) *and estimates* (14.2.10) *equals*

$$\boldsymbol{E}\mathcal{W}_0^m\left[0,\varkappa(0),u^0\right].$$

Here u^0 is taken from Remark 14.2.3 *(on p.* 473*).*

(ii) The infimum from (i) is attained at the control strategy

$$\overline{u}(t) = \mathcal{U}_t\Big\{\overline{P}\big[t-1\big],\varkappa\big[t\big],\ldots,\varkappa\big[t-\overline{\sigma}(t)\big],\overline{u}\big[t-1\big],\ldots,\overline{u}\big[t-\widehat{\sigma}(t)\big]\Big\}, \quad (14.6.9)$$

along with the state estimate described in Sect. 14.4. *We recall that*

$$\overline{\sigma}[t] := \min\{\overline{\sigma},t\} \quad \text{and} \quad \widehat{\sigma}(t) := \min\{\sigma+\sigma_*,t+1\},$$

where σ,σ_*, and $\overline{\sigma}$ are the constants from Assumptions 14.3.7 *and* 14.3.9 *(on p.* 475*), respectively. Furthermore, $\overline{u}(t)$ and $\overline{P}(t) = \{\overline{P}_{ij}(t)\}_{i,j=0}^{\sigma}$ are the sensor control and ensemble of matrices, respectively, produced by the estimator at time t.*

Remark 14.6.3. Formulas (14.6.1) and (14.6.9) imply that the current optimal sensor control is determined on the basis of the history of the communication medium observations: $\overline{u}(t) = \mathcal{F}[t,\varkappa(t),\ldots,\varkappa(0)]$. Thus the possibility to depend on $z(0),\ldots,z(t)$ given by (14.2.9) (on p. 473) is not utilized.

Remark 14.6.4. Let the distribution $\{p_\varkappa^0 := \boldsymbol{P}\left[\varkappa(0) = \varkappa\right]\}_{\varkappa\in\Upsilon}$ be known in advance. Then the mathematical expectation

$$\boldsymbol{E}\mathcal{W}_0^m\left[0,\varkappa(0),u^0\right] = \sum_{\varkappa\in\Upsilon} p_\varkappa^0\mathcal{W}_0^m\left[0,\varkappa,u^0\right]$$

is computable a priori as a deterministic function of the control u^0 from Remark 14.2.3 (on p. 473). Its minimizing over u^0 gives rise to an optimal value of u^0 thanks to (i) of Theorem 14.6.2.

14.7 Proof of Theorem 14.6.2 on p. 484

Lemma 14.7.1. *Consider a sensor control* (14.2.9) *(on p.* 473*) and denote*

$$\Omega(t) := \mathbf{col}\Big[a, \xi(0), \ldots, \xi(t), \chi(0), \ldots, \chi(t)\Big], \tag{14.7.1}$$

where $\xi(\theta)$*,* $\chi(\theta)$ *are the noises from* (14.2.1) *and* (14.2.2) *(on p.* 471*). Then* $\mathfrak{Z}(t)$ *and* $\overline{u}(t)$ *are deterministic functions of* $\Omega(t)$ *and* $\Delta(t)$*. Here* $\mathfrak{Z}(t)$ *and* $\Delta(t)$ *were introduced in* (14.2.9)*.*

Proof. The proof is by induction on t. For $t = 0$, the claim is immediate from (14.2.1)–(14.2.9) (on pp. 471–473). Let it be true for some $t = \theta \leq T - 1$. Then by (14.2.9), the statement of the lemma holds for $\overline{u}(t)$ with $t := \theta + 1$. Hence it follows from (14.2.1)–(14.2.5) and (14.2.7) that this statement is true for $\mathfrak{Z}(t), t := \theta + 1$ as well, which completes the proof. □

Lemma 14.7.2. *For any control strategy* (14.2.9) *(on p.* 473*), the state estimate described in Sect.* 14.4*, and* $\tau = 0, \ldots, T$*, the corresponding value of the cost functional* (14.2.11) *(on p.* 474*) obeys the bound*

$$\mathfrak{I} \geq \sum_{t=0}^{\tau} \boldsymbol{E}\big\|e(t|t)\big\|^2 + \boldsymbol{E}W_\tau\Big\{\overline{P}[\tau], \varkappa[\tau], \ldots, \varkappa\big[\tau - \overline{\sigma}(\tau)\big], \overline{u}[\tau], \ldots, \overline{u}\big[\tau - \widehat{\sigma}(\tau+1) + 1\big]\Big\}. \tag{14.7.2}$$

Here $\overline{P}(t), t = 0, \ldots, T$ *are the ensembles of matrices generated by the state estimation algorithm from Sect.* 14.4*.*

Proof. The proof will be by induction on $\tau = T, T-1, \ldots, 0$. For $\tau = T$, the claim is evident since $W_T(\cdot) \equiv 0$. Let (14.7.2) hold for some $\tau = 1, \ldots, T$. Due to Theorem 14.4.4 (on p. 477),

$$\mathfrak{I} \geq \sum_{t=0}^{\tau-1} \boldsymbol{E}\big\|e(t|t)\big\|^2 + \boldsymbol{E}\,\mathbf{tr}\,\overline{P}_{00}[\tau] + \boldsymbol{E}W_\tau\Big(\overline{P}[\tau], \{\varkappa[t]\}_{t=\tau-\overline{\sigma}(\tau)}^{\tau}, \{\overline{u}[t]\}_{t=\tau-\widehat{\sigma}(\tau+1)+1}^{\tau}\Big). \tag{14.7.3}$$

By (14.6.1), (14.6.6), and (14.6.7), we have for the sum μ of the last two addends on the right

$$\mu = \boldsymbol{E}\mathcal{W}_\tau\Big(\overline{P}[\tau-1], \{\varkappa[t]\}_{t=\tau-\overline{\sigma}(\tau)}^{\tau}, \{\overline{u}[t]\}_{t=\tau-\widehat{\sigma}(\tau)}^{\tau}\Big) \geq \boldsymbol{E}\,\overbrace{\mathcal{W}_\tau^m\Big(\overline{P}[\tau-1], \{\varkappa[t]\}_{t=\tau-\overline{\sigma}(\tau)}^{\tau}, \{\overline{u}[t]\}_{t=\tau-\widehat{\sigma}(\tau)}^{\tau-1}\Big)}^{\mathfrak{W}} \tag{14.7.4}$$

$$= \boldsymbol{E}\boldsymbol{E}\Big[\mathfrak{W}\Big|\Omega(\tau-1), \Delta(\tau-1)\Big]. \tag{14.7.5}$$

Owing to Lemma 14.7.1, (14.6.1), and (14.4.9) (on p. 477),

$$\overline{P}[\tau-1] \quad \text{and} \quad \{\overline{u}[t]\}_{t=\tau-\widehat{\sigma}(\tau)}^{\tau-1}$$

are deterministic functions of $\Omega(\tau-1)$ and $\Delta(\tau-1)$; so evidently are

$$\varkappa[\tau-1], \ldots, \varkappa\big[\tau-\overline{\sigma}(\tau)\big].$$

So the conditional mathematical expectation from (14.7.5) amounts to

$$\sum_{\varkappa\in\Upsilon}\Bigg[\mathcal{W}_\tau^m\Big(\overline{P}[\tau-1], \varkappa, \{\varkappa[t]\}_{t=\tau-\overline{\sigma}(\tau)}^{\tau-1}, \{\overline{u}[t]\}_{t=\tau-\widehat{\sigma}(\tau)}^{\tau-1}\Big)\times \\ \times \boldsymbol{P}\Big(\varkappa[\tau]=\varkappa\Big|\Omega[\tau-1], \Delta[\tau-1]\Big)\Bigg]. \tag{14.7.6}$$

Thanks to Assumption 14.3.2 (on p. 474), the random vector $\Omega(\tau-1)$ is independent of both $\varkappa[\tau]$ and $\Delta[\tau-1]$. So

$$\boldsymbol{P}\big(\varkappa[\tau]=\varkappa\big|\Omega[\tau-1], \Delta[\tau-1]\big) = \boldsymbol{P}\big(\varkappa[\tau]=\varkappa\big|\Delta[\tau-1]\big).$$

Then in view of (14.3.12) (where $t := \tau - 1$) and (14.6.8), the quantity (14.7.6) equals

$$W_{\tau-1}\Big(\overline{P}[\tau-1], \{\varkappa[t]\}_{t=\tau-1-\overline{\sigma}(\tau-1)}^{\tau-1}, \{\overline{u}[t]\}_{t=\tau-\widehat{\sigma}(\tau)}^{\tau-1}\Big).$$

This and (14.7.3) imply (14.7.2) with $\tau := \tau - 1$. □

Corollary 14.7.3. *For any control strategy* (14.2.9) *(on p.* 473*) and state estimate* (14.2.10) *(on p.* 474*), the following inequality holds:*

$$\mathfrak{I} \geq \boldsymbol{E}\mathcal{W}_0^m\Big[0, \varkappa(0), u^0\Big]. \tag{14.7.7}$$

Indeed by Theorem 14.4.4 (on p. 477), it suffices to consider the estimate described in Sect. 14.4. Then inequality (14.7.7) results from (14.7.2) with $\tau := 0$ by following the lines of (14.7.3)–(14.7.6), taking into account that now $\sum \boldsymbol{E}\|e(t|t)\|^2$ does not occur in (14.7.3), and invoking that $\overline{P}(-1) = 0$ by (14.4.9) (on p. 477).

Lemma 14.7.4. *Relation* (14.7.2) *holds with the equality sign for the control strategy* (14.6.9) *(on p.* 484*).*

Proof. The proof will be by induction on $t = T, T-1, \ldots, 0$. Since $W_T(\cdot) \equiv 0$, the statement of the lemma holds for $\tau := T$. If it is true for some $\tau = 1, \ldots, T$, then (14.7.3) is valid with $\geq$ replaced by $=$; so is (14.7.4) by the definition of $\mathcal{U}_\tau(\cdot)$. Hence retracing the arguments from the proof of Lemma 14.7.2 shows that the statement holds for $\tau := \tau - 1$. □

Corollary 14.7.5. *Relation* (14.7.7) *holds with the equality sign for the sensor control strategy* (14.6.9) *and the state estimate described in Sect.* 14.4.

This corollary follows from Lemma 14.7.4 just like Corollary 14.7.3 results from Lemma 14.7.2.

Proof of Theorem 14.6.2 (on p. 484*).* This theorem is straightforward from Corollaries 14.7.3 and 14.7.5. □

14.8 Model Predictive Sensor Control

The dynamic programming equations like those derived in Sect. 14.6 have been the subject of intensive research in the field of optimal control theory. In realistic situations, solution of such equations is often a hard task. In this section, we apply ideas of model predictive control (see, e.g., [29]) to give a nonoptimal but real-time implementable method for sensor control. To simplify the presentation, we restrict ourselves to consideration of a special case where minor alternations of the standard model predictive control approach are required.

Suppose that there are one observation and one control channels $q = r = 1$ in (14.2.4) and (14.2.7) (on pp. 472 and 473), and they keep the order of the messages. In other words, any set $\mathcal{S}(t, \varkappa)$ from Assumption 14.3.5 (on p. 475) is either empty or contains one element $s(t, \varkappa)$, and

$$s(t_1, \varkappa_1) < s(t_2, \varkappa_2) \qquad \text{whenever}$$
$$t_1 < t_2, \quad \mathcal{S}(t_i, \varkappa_i) \neq \emptyset, \quad \text{and} \quad \boldsymbol{P}[\varkappa(t_1) = \varkappa_1 \wedge \varkappa(t_2) = \varkappa_2] > 0.$$

We put $s(t, \varkappa) := \circledast$ whenever $\mathcal{S}(t, \varkappa) = \emptyset$, where $\circledast$ is a special "void" symbol.

Remark 14.8.1. From now on, we consider the state estimate described in Sect. 14.4.

One-Zone-Ahead Optimal Control

The standard idea of the model predictive control is to pick the current control so that the current state estimation error be minimal provided the previous controls are chosen and fixed. Under the circumstances, this scheme needs modification since the current control $\overline{u}(\theta)$ may not affect observations received until θ and hence the current error. So it is natural to minimize not the current but the summary error, where the sum is over the times when observations affected by $\overline{u}(\theta)$ arrive at the estimator. Due to (14.2.3), (14.2.7) (on pp. 471 and 473), and Assumption 14.3.6 (on p. 475), these times form the set

$$J_\theta := \Big\{ t : s := s[t, \varkappa(t)] \neq \circledast, s - \mathcal{T}_1[s, \varkappa(t)] = \theta \Big\}. \tag{14.8.8}$$

For

$$\theta \in \Theta := \{\theta = 0, \ldots, T : J_\theta \neq \emptyset\},$$

we denote

$$\tau_\theta := \min\{t : t \in J_\theta\}.$$

Our assumptions imply that $\tau_{\theta'} < \tau_{\theta''}$ if $\theta' < \theta'', \theta', \theta'' \in \Theta$. Furthermore we put

$$\mathfrak{J}_* := \big\{t : 0 \le t < \min_{\theta \in \Theta} \tau_\theta\big\}, \quad \mathfrak{J}_{\overline{\theta}} := \big\{t : \min_{\theta \in \Theta, \theta \ge \overline{\theta}} \tau_\theta \le t < \min_{\theta \in \Theta, \theta > \overline{\theta}} \tau_\theta\big\}, \tag{14.8.9}$$

where $\overline{\theta} = 0, \ldots, T$ and $\min_{\theta \in \emptyset} \tau_\theta := T + 1$. The interval $\mathfrak{J}_{\overline{\theta}}$ is the *zone of the influence* of the control $\overline{u}(\overline{\theta})$: All observations influenced by this control are received within this interval, and no other observation arrives within it. This interval may contain "void" times where no observation arrives. However it starts with a nonvoid time instant. Accordingly the time immediately following this interval is nonvoid and influenced by another control. The interval $\mathfrak{J}_*$ consists of times t such that no observation has arrived until t. It is easy to see that

$$\mathfrak{J}_* \cup \mathfrak{J}_0 \cup \ldots \cup \mathfrak{J}_T = [0, T],$$

and the sets $\mathfrak{J}_*, \mathfrak{J}_0, \ldots, \mathfrak{J}_T$ are pairwise disjoint; some of them may be empty. Hence the functional (14.2.11) (on p. 474) shapes into $\mathfrak{I} = \mathfrak{I}^T$, where

$$\mathfrak{I}^\theta := \boldsymbol{E} \sum_{t \in \mathfrak{J}_*} \big\|e(t|t)\big\|^2 + \boldsymbol{E} \sum_{\tau=0}^{\theta} \sum_{t \in \mathfrak{J}_\tau} \big\|e(t|t)\big\|^2. \tag{14.8.10}$$

Definition 14.8.2. *A sensor control strategy $\mathfrak{S}^0$ is said to be* one-zone-ahead optimal *if for any $\theta = 0, \ldots, T$ and any other sensor control strategy $\mathfrak{S}$ that coincides with $\mathfrak{S}^0$ on $[0, \theta - 1]$ provided $\theta \ge 1$, the following inequality holds:*

$$\mathfrak{I}^\theta\big[\mathfrak{S}^0\big] \le \mathfrak{I}^\theta\big[\mathfrak{S}\big].$$

Construction of the One-Zone-Ahead Optimal Control Strategy

By (14.8.8) and (14.8.10), the zone of the influence $\mathfrak{J}_\theta$ is determined uniquely

$$\mathfrak{J}_\theta = \mathfrak{J}_\theta(\Delta)$$

by the sequence $\Delta = \big\{\varkappa_t\big\}_{t=\theta}^{T}$, where $\varkappa_t := \varkappa(t)$. For any two sequences

$$\Delta' = \big\{\varkappa_t\big\}_{t=\theta-\overline{\sigma}(\theta)}^{\theta}, \qquad \Delta'' = \big\{\varkappa_t\big\}_{t=\theta+1}^{T} \subset \Upsilon,$$

we denote

$$\Delta' \curlywedge \Delta'' := \big\{\varkappa_\theta, \ldots, \varkappa_T\big\}$$

and

$$\mathfrak{Q}^\theta(\Delta') := \Bigg\{\Delta'' : \mathfrak{J}_\theta(\Delta' \curlywedge \Delta'') \neq \emptyset \quad \text{and} \quad \boldsymbol{P}\Big(\big\{\varkappa(t)\big\}_{t=\theta+1}^{T} = \Delta'' \Big| \big\{\varkappa(t)\big\}_{t=\theta-\overline{\sigma}(\theta)}^{\theta} = \Delta'\Big) > 0\Bigg\}. \tag{14.8.11}$$

Remark 14.8.3. Due to (14.3.12) (on p. 475) and the chain rule, the conditional probability from (14.8.11) amounts to

$$\boldsymbol{P}\Big(\varkappa_T\Big|\varkappa_{T-1},\ldots,\varkappa_{T-1-\overline{\sigma}(T-1)}\Big)\times\boldsymbol{P}\Big(\varkappa_{T-1}\Big|\varkappa_{T-2},\ldots,\varkappa_{T-2-\overline{\sigma}(T-2)}\Big)\times\cdots$$
$$\cdots\times\boldsymbol{P}\Big(\varkappa_{\theta+1}\Big|\varkappa_\theta,\ldots,\varkappa_{\theta-\overline{\sigma}(\theta)}\Big).$$

Here the quantity $\boldsymbol{P}\big[\varkappa|\varkappa_0,\ldots,\varkappa_{\overline{\sigma}(t-1)}\big]$ was introduced after formula (14.6.8) (on p. 483) and is known a priori by Assumption 14.3.9 (on p. 475). Hence so is the above probability from (14.8.11).

Now suppose that a sequence of controls $\mathfrak{U} = \textbf{col}\,\big[\overline{u}_1,\ldots,\overline{u}_{\widehat{\sigma}(\theta)}\big]$ and an element $\Delta'' \in \mathfrak{Q}^\theta(\Delta')$ are given. Consider a control variable u and equations (14.6.1) (on p. 482), where $t=\theta,\ldots,\tau$ and

$$\tau := \max\{t : t\in\mathfrak{J}_\theta(\Delta'\curlywedge\Delta'')\},\quad \varkappa(t) := \varkappa_t,$$
$$\overline{u}[\theta] := u, \overline{u}[\theta-1] := \overline{u}_1,\ldots,\overline{u}[\theta-\widehat{\sigma}(\theta)] := \overline{u}_{\widehat{\sigma}(\theta)}.$$

Remark 14.8.4. The undefined controls $\overline{u}[\theta+1],\ldots,\overline{u}[\tau]$ formally may but in fact do not occur in (14.6.1) (on p. 482).

Indeed by (14.6.3) and (14.6.4) (on pp. 482 and 483), occurring are controls that affect an observation arriving within the interval $[\theta,\tau]$. By the definition of the moment τ, the last such observation is affected by $\overline{u}(\theta)$. Since the communication channels keep the order of the messages, all controls in question are generated until θ.

Thanks to Remark 14.8.4, the recursion (14.6.1) with $t=\theta,\ldots,\tau$ is well defined. Along with the initial condition $\overline{P}(\theta-1) = \overline{P} = \overline{P}^{\mathrm{T}} \geq 0$, it gives rise to a sequence

$$\Big\{\overline{P}^\theta\big[t,\overline{P},u,\mathfrak{U},\Delta',\Delta''\big]\Big\}_{t=\theta-1}^{\tau}. \tag{14.8.12}$$

Now introduce the deterministic function

$$\mathcal{F}[\theta,\overline{P},u,\mathfrak{U},\Delta'] := \sum_{\Delta''\in\mathfrak{Q}^\theta(\Delta')}\boldsymbol{P}\Big[\big\{\varkappa(t)\big\}_{t=\theta+1}^{T} = \Delta''\Big|\big\{\varkappa(t)\big\}_{t=\theta-\overline{\sigma}(\theta)}^{\theta} = \Delta'\Big]\times$$
$$\times\sum_{t\in\mathfrak{J}_\theta(\Delta'\curlywedge\Delta'')}\mathbf{tr}\,\overline{P}_{00}^\theta[t,\overline{P},u,\mathfrak{U},\Delta',\Delta'']. \tag{14.8.13}$$

Here $\overline{P}_{00}^\theta(\cdot)$ is the block of $\overline{P}^\theta(\cdot)$. Finally we put

$$\widetilde{\mathfrak{U}}_\theta[\overline{P},\mathfrak{U},\Delta'] := u_m,$$

where u_m is an element furnishing the minimum

$$\min_{u\in U}\mathcal{F}[\theta,\overline{P},u,\mathfrak{U},\Delta'].$$

It is easy to see that the function $\mathcal{F}(\cdot)$ is measurable. So the function $\widetilde{\mathfrak{U}}_\theta(\cdot)$ can be chosen measurable as well (see, e.g., [217, Sec. I.7]).

Main Result

Theorem 14.8.5. *Assume that there are one control and one observation channels, and that these channels keep the order of the messages. Then the following sensor control strategy is one-zone-ahead optimal:*

$$\overline{u}(t) = \widetilde{\mathfrak{U}}_t\Big\{\overline{P}\big[t-1\big], \overline{u}\big[t-1\big], \ldots, \overline{u}\big[t-\widehat{\sigma}(t)\big], \varkappa\big[t\big], \ldots, \varkappa\big[t-\overline{\sigma}(t)\big]\Big\}. \quad (14.8.14)$$

Here

$$\widehat{\sigma}[t] := \min\{\sigma + \sigma_*, t\} \quad \text{and} \quad \overline{\sigma}(t) := \min\{\overline{\sigma}, t\},$$

where σ, σ_*, and $\overline{\sigma}$ are the constants from Assumptions 14.3.7 *and* 14.3.9 *(on p.* 475*), respectively. Furthermore, $\overline{P}(t), t = 0, \ldots, T$ is the ensemble of matrices generated by the state estimator described in Sect.* 14.4 *(starting on p.* 476*).*

Remark 14.8.6. For both optimal (14.6.9) (on p. 484) and one-zone-ahead optimal (14.8.14) strategies, the current control $\overline{u} = \overline{u}(t)$ is determined as an element minimizing a certain function $\mathcal{G}_t(\overline{u})$. Here

$$\mathcal{G}_t(\overline{u}) = \mathcal{W}_t\{\overline{P}[t-1], \varkappa[t], \ldots, \varkappa[t-\overline{\sigma}(t)], \overline{u}, \overline{u}[t-1], \ldots, \overline{u}[t-\widehat{\sigma}(t)]\}$$

in the case (14.6.9) (on p. 484) and

$$\mathcal{G}_t(\overline{u}) = \mathcal{F}\{t, \overline{P}[t-1], \overline{u}, \overline{u}[t-1], \ldots, \overline{u}[t-\widehat{\sigma}(t)], \varkappa[t], \ldots, \varkappa[t-\overline{\sigma}(t)]\}$$

in the case (14.8.14). In the case (14.6.9), calculation of $\mathcal{G}_t(\overline{u})$ for a given $\overline{u}$ requires certain functions $\mathcal{W}_T[\cdot], \ldots, \mathcal{W}_t[\cdot]$ to be determined in advance, which can be interpreted as computation of infinite arrays of reals. In contrast such a calculation requires computation of only a finite set of reals in the case (14.8.14).

Remark 14.8.7. It is straightforward to extend the approach considered in this subsection on a more general case where 1) the numbers of both observation and control channels are arbitrary; 2) all channels keep the order of the messages; 3) the observation channels are synchronous, i.e., provide a common transmission time and lose signals only simultaneously; and 4) the control channels are also synchronous.

Proof of Theorem 14.8.5

As long as the channels keep the order of the messages, the union $\mathcal{J}_* \cup \mathcal{J}_1 \cup \ldots \cup \mathcal{J}_t$ is an interval with the left end-point 0 for all t. Consider $\theta = 0, \ldots, T$ and two sensor control strategies $\mathfrak{S}$ and $\mathfrak{S}^0$ that coincide on $[0, \theta - 1]$. They produce two sequences of observations arriving at the estimator. Within the interval

$$\mathcal{J}_* \cup \mathcal{J}_1 \cup \ldots \cup \mathcal{J}_{\theta-1},$$

the entries of these sequences are affected by common controls and hence coincide. So do the related state estimates $\widehat{x}_{\mathfrak{S}}(t|t)$ and $\widehat{x}_{\mathfrak{S}^0}(t|t)$. This and (14.8.10) imply

$$\Delta\mathfrak{I}^\theta := \mathfrak{I}^\theta\big(\mathfrak{S}\big) - \mathfrak{I}^\theta\big(\mathfrak{S}^0\big) = \eta_{\mathfrak{S}} - \eta_{\mathfrak{S}^0}, \quad \text{where}$$
$$\eta_{\mathfrak{S}} := \boldsymbol{E}\sum_{t\in\mathfrak{J}_\theta}\big\|\underbrace{x(t)-\widehat{x}_{\mathfrak{S}}(t|t)}_{e_{\mathfrak{S}}(t|t)}\big\|^2. \quad (14.8.15)$$

We denote

$$Q := \{(t,\varkappa) : s := s(t,\varkappa) \neq \circledast \text{ and } s - \mathfrak{T}_1(s,\varkappa) = \theta\},$$

$j_\theta(t) := 1$ if $t \in \mathfrak{J}_\theta$ and $j_\theta(t) := 0$ otherwise. It follows from (14.8.8) and (14.8.9) that almost surely

$$\big\{j_\theta(t)=1\big\} \Leftrightarrow \bigvee_{\overline{t}=\theta}^{t}\Big\{[\overline{t},\varkappa(\overline{t})]\in Q\Big\}\bigwedge\bigwedge_{t'=\overline{t}}^{t}\Big\{s[t',\varkappa(t')]=\circledast\bigvee[t',\varkappa(t')]\in Q\Big\}.$$

Hence the event $\{j_\theta(t)=1\}$ and thus the random quantity $j_\theta(t)$ are measurable with respect to the σ-algebra generated by

$$\Delta(t) := [\varkappa(0),\dots,\varkappa(t)].$$

Then in view of (14.8.15),

$$\begin{aligned}
\eta_{\mathfrak{S}} &= \sum_{t=\theta}^{T}\boldsymbol{E}j_\theta(t)\big\|e_{\mathfrak{S}}(t|t)\big\|^2 \\
&= \sum_{t=\theta}^{T}\boldsymbol{E}\boldsymbol{E}\Big[j_\theta(t)\big\|e_{\mathfrak{S}}(t|t)\big\|^2\Big|\underbrace{z(0),\dots,z(t-1)}_{\mathfrak{z}(t-1)},\Delta(t)\Big] \\
&= \sum_{t=\theta}^{T}\boldsymbol{E}j_\theta(t)\boldsymbol{E}\Big[\big\|e_{\mathfrak{S}}(t|t)\big\|^2\Big|\mathfrak{z}(t-1),\Delta(t)\Big] \xlongequal{\text{Theorem 14.4.4}} \boldsymbol{E}\sum_{t=\theta}^{T}j_\theta(t)\,\mathbf{tr}\,\overline{P}_{00}^{\mathfrak{S}}(t) \\
&= \boldsymbol{E}\sum_{t\in\mathfrak{J}_\theta}\mathbf{tr}\,\overline{P}_{00}^{\mathfrak{S}}(t) = \boldsymbol{E}\,\overbrace{\boldsymbol{E}\Big[\sum_{t\in\mathfrak{J}_\theta}\mathbf{tr}\,\overline{P}_{00}^{\mathfrak{S}}(t)\Big|\Delta(\theta)\Big]}^{\mu}. \quad (14.8.16)
\end{aligned}$$

Here $\{\overline{P}_{00}^{\mathfrak{S}}(t)\}$ is the sequence of the matrices generated by the algorithm described in Sect. 14.4 for the strategy $\mathfrak{S}$. The quantity μ is the function of $\overline{\Delta} := \Delta(\theta)$:

$$\begin{aligned}
\mu = \mu(\overline{\Delta}) &= \boldsymbol{E}\Big[\sum_{t\in\mathfrak{J}_\theta}\mathbf{tr}\,\overline{P}_{00}^{\mathfrak{S}}(t)\Big|\Delta(\theta)=\overline{\Delta}\Big] \\
&= \sum_{\Delta''}\boldsymbol{E}\Big[\sum_{t\in\mathfrak{J}_\theta}\mathbf{tr}\,\overline{P}_{00}^{\mathfrak{S}}(t)\Big|\Delta(\theta)=\overline{\Delta},\{\varkappa(t)\}_{t=\theta+1}^{T}=\Delta''\Big]\times \\
&\qquad\times\boldsymbol{P}\Big[\big\{\varkappa(t)\big\}_{\theta+1}^{T}=\Delta''\Big|\Delta(\theta)=\overline{\Delta}\Big], \quad (14.8.17)
\end{aligned}$$

where $\Delta'' = \{\varkappa_t\}_{t=\theta+1}^{T}$. By following the lines of Remark 14.8.3, it can be shown that the above conditional probability equals

$$\boldsymbol{P}\Big[\{\varkappa(t)\}_{\theta+1}^{T} = \Delta'' \Big| \{\varkappa(t)\}_{\theta-\overline{\sigma}(\theta)}^{\theta} = \Delta'\Big],$$

where $\Delta' := \{\varkappa_t\}_{t=\theta-\overline{\sigma}(\theta)}^{\theta}$. By (14.8.11),

$$\mu = \sum_{\Delta''\in\mathfrak{Q}^\theta(\Delta')} \Bigg\{ \boldsymbol{P}\Big[\{\varkappa(t)\}_{\theta+1}^{T} = \Delta'' \Big| \{\varkappa(t)\}_{\theta-\overline{\sigma}(\theta)}^{\theta} = \Delta'\Big] \\ \times \sum_{t\in\mathfrak{J}_\theta(\Delta'\curlywedge\Delta'')} \boldsymbol{E}\Big[\mathbf{tr}\,\overline{P}_{00}^{\mathfrak{S}}(t)\Big|\varkappa(t') = \varkappa_{t'}\ \forall t'\Big]\Bigg\}. \quad (14.8.18)$$

Here by (14.6.1) (on p. 482) and the definition of the sequence (14.8.12),

$$\overline{P}_{00}^{\mathfrak{S}}(t) = \overline{P}_{00}^{\theta}\Big\{t, \overline{P}_{00}^{\mathfrak{S}}[\theta-1], \{\overline{u}^{\mathfrak{S}}[t']\}_{t'=\theta-\widehat{\sigma}(\theta)}^{\theta}, \Delta', \Delta''\Big\}, \quad (14.8.19)$$

where $\{\overline{u}^{\mathfrak{S}}[t']\}$ is the sequence of controls generated by the strategy $\mathfrak{S}$. By Lemma 14.7.1 (on p. 485), any control $\overline{u}^{\mathfrak{S}}[t'], t' \leq \theta$ is a deterministic function of $\overline{\Delta}$ and $\Omega(\theta)$. By (14.6.1) (on p. 482), so is $\overline{P}_{00}^{\mathfrak{S}}[\theta-1]$. Since $\Omega(\theta)$ is independent of $\{\varkappa(t)\}$ by Assumption 14.3.2 (on p. 474), we have for any deterministic function $\mathcal{G}(\cdot)$,

$$\boldsymbol{E}\Big(\mathcal{G}\big[\Omega(\theta), \overline{\Delta}\big]\Big|\varkappa(t') = \varkappa_{t'}\ \forall t'\Big) = \boldsymbol{E}\mathcal{G}\big[\Omega(\theta), \varkappa_0, \ldots, \varkappa_\theta\big] \\ = \boldsymbol{E}\Big(\mathcal{G}\big[\Omega(\theta), \overline{\Delta}\big]\Big|\varkappa(0) = \varkappa_0, \ldots, \varkappa(\theta) = \varkappa_\theta\Big).$$

This implies that in the mathematical expectation from (14.8.18), conditioning given $\varkappa(t') = \varkappa_{t'}\ \forall t'$ can be replaced by that given $\Delta(\theta) = \overline{\Delta}$. This and (14.8.13), (14.8.19) shape (14.8.18) into

$$\mu = \boldsymbol{E}\left(\mathcal{F}\Big\{\theta, \overline{P}^{\mathfrak{S}}[\theta-1], \{\overline{u}^{\mathfrak{S}}[t']\}_{t'=\theta-\widehat{\sigma}(\theta)}^{\theta}, \Delta'\Big\}\Big|\Delta(\theta) = \overline{\Delta}\right).$$

Since the control strategies $\mathfrak{S}$ and $\mathfrak{S}^0$ coincide on $[0, \theta-1]$, relations (14.2.9) and (14.6.1) (on pp. 473 and 482) yield

$$\overline{u}^{\mathfrak{S}}[t'] = \overline{u}^{\mathfrak{S}^0}[t'],\ t' = 0, \ldots \theta-1 \quad \text{and} \quad \overline{P}^{\mathfrak{S}}[\theta-1] = \overline{P}^{\mathfrak{S}^0}[\theta-1].$$

By (14.8.15) and (14.8.16),

$$\Delta\mathcal{I}^\theta = \boldsymbol{E}\Bigg(\mathcal{F}\Big\{\theta, \overline{P}^{\mathfrak{S}^0}[\theta-1], \overline{u}^{\mathfrak{S}}[\theta], \{\overline{u}^{\mathfrak{S}^0}[t']\}_{t'=\theta-\widehat{\sigma}(\theta)}^{\theta-1}, \Delta'\Big\} \\ - \mathcal{F}\Big\{\theta, \overline{P}^{\mathfrak{S}^0}[\theta-1], \overline{u}^{\mathfrak{S}^0}[\theta], \{\overline{u}^{\mathfrak{S}^0}[t']\}_{t'=\theta-\widehat{\sigma}(\theta)}^{\theta-1}, \Delta'\Big\}\Bigg).$$

By substituting here the control strategy $\mathfrak{S}^0$ given by (14.8.14), we arrive at the assertion of the theorem. □

15

Robust Kalman State Estimation with Switched Sensors

15.1 Introduction

This chapter proceeds with consideration of the sensor scheduling problem, which consists in estimating the state of an uncertain process based on measurements obtained by switching a given set of noisy sensors.

Classic estimation theory deals with the problem of forming an estimate of a process given measurements produced by sensors observing the process; e.g., see [8]. A standard solution is to compute the posterior density of the process state conditioned on all the available measurements. A more difficult class of estimation problem arises in applications such as robotics, command and control, and networked systems where an estimator is given dynamic control over the measurements. These sensor scheduling problems occur, for example, when a flexible or intelligent sensor is able to operate in one of several different measurement modes and the estimator can dynamically switch the sensor mode. Alternatively several sensors may be remotely linked to the estimator via a low bandwidth communication channel and only one sensor can send measurement data during any measurement interval. Again the estimator can dynamically select which sensor uses the channel. Finally, sensor scheduling problems arise when measurements from a large number of sensors are available to the estimator but the computational power is such that only data from a small selection of the sensors can be processed at any given time, hence, forcing the estimator to dynamically select which sensor data are important for the task at hand. Sensor scheduling has been addressed for stochastic systems in [17, 126, 155] where it is assumed that the process is generated by a known linear system with Gaussian input noise. It is shown that the optimal sensor schedule can be computed a priori and that this schedule is independent of the observed data. In particular a sufficient statistic for a linear zero-mean Gaussian processes with linear sensors and a minimum variance estimation objective is given by the estimation error covariance matrix, which can be determined by the solution to a Riccati differential equation. This matrix depends on the sequence of sensors used but is independent of the actual observed measurements. Hence for any given sequence of sensors, the estimation error covariance can be determined before the experiment has commenced. As a

A.S. Matveev and A.V. Savkin, *Estimation and Control over Communication Networks*,
doi: 10.1007/978-0-8176-4607-3_15,

consequence the optimal sensor sequence can be determined a priori and is given by the sequence which minimizes, under some measure (such as the trace of the error covariance matrix at the final time) the precomputable solution to a Riccati differential equation.

In practice, however, it often occurs that the system model is not precisely known and standard stochastic models cannot be readily applied. In the current chapter, we follow the approach of Chap. 5, where the uncertainties are modeled by unknown functions that satisfy an integral quadratic constraint. In this framework, the estimation problem is one of characterizing the set of possible states that could have given rise to the observed measurement. This makes the difference with Chap. 14, where a stochastic plant model was employed, and the objective was to find a minimum variance state estimate. Another distinction is that we do not assume any longer that the channels provide delays and may lose messages.

The results of this chapter were originally published in [175,176] (see also [174]).

The remainder of the chapter is organized as follows. In Sect. 15.2, we introduce a time-varying uncertain system in which the measurement process is defined by a collection of given sensors that we call basic sensors. Without loss of generality it is assumed that only one of the basic sensors can be used at any time. Hence, our sensor schedule is a rule for switching from one basic sensor to another. Furthermore, we introduce the concept of uniform robust observability for such systems. The objective is to ensure uniform robust observability and an optimal estimate of the system state. We show that the optimal switching rule can be computed by solving a set of Riccati differential equations of the game type and a dynamic programming procedure. It is shown that for the framework considered here the optimal sensor sequence depends on the history of measurements. This is unlike sensor scheduling problems with Gaussian noise models and mean-square estimation criteria where the sensor schedule is independent of the observed measurements and can be computed a priori using only the statistical structure of the measurement and process noise. In Sect. 15.3, we apply ideas of model predictive or finite horizon control (e.g., see [29]) to derive a nonoptimal but real-time implementable method for robust sensor scheduling. Finally, in Sect. 15.4 we present the proofs of the main results of the chapter.

15.2 Optimal Robust Sensor Scheduling

Consider the time-varying uncertain system defined on the finite interval $[0, T]$

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)w(t); \\ z(t) &= K(t)x(t); \\ y^*(t) &= C^*(t)x(t) + \nu^*(t), \end{aligned} \tag{15.2.1}$$

where $x(t) \in \mathbb{R}^n$ is the state; $w(t) \in \mathbb{R}^p$ and $\nu^*(t) \in \mathbb{R}^l$ are the uncertainty inputs; $z \in \mathbb{R}^q$ is the uncertainty output; and $y^*(\cdot)$ is the continuously measured output.

Here $A(\cdot)$, $B(\cdot)$, and $K(\cdot)$ are given piecewise continuous matrix functions. The matrix function $C^*(\cdot)$ is defined by a particular sensor schedule used to measure the system state. Note that the dimension of $C^*(\cdot)$ can change with time depending on the sensors used. Let N be a given positive integer, and let

$$0 = t_0 < t_1 < t_2 < \ldots < t_N = T$$

denote the permissible sensor switching times.

Suppose we have the following collection of measured outputs that are called *basic sensors*:

$$\begin{aligned} y^1(\cdot) &= C_1(\cdot)x(\cdot) + \nu^1(\cdot) \\ y^2(\cdot) &= C_2(\cdot)x(\cdot) + \nu^2(\cdot) \\ \vdots\ &\quad \vdots \\ y^k(\cdot) &= C_k(\cdot)x(\cdot) + \nu^k(\cdot) \end{aligned} \tag{15.2.2}$$

where $C_1(\cdot), C_2(\cdot), \ldots, C_k(\cdot)$ are given matrix functions. Let $I_j(\cdot)$ be a function that maps the set of past measurements $\{y^*(\cdot) \mid_{t_0}^{t_j}\}$ to the set of symbols $\{1, 2, \ldots, k\}$. Then for any sequence of functions $\{I_j\}_{j=0}^{N-1}$, we consider the following dynamic sensor schedule:

$$\forall j \in \{0, 1, \ldots, N\},\ \ y^*(t) = y^{i_j}(t)\ \ \forall t \in [t_j, t_{j+1}), \\ \text{where } i_j = I_j \left[y^*(\cdot) \mid_{t_0}^{t_j} \right]. \tag{15.2.3}$$

Hence the sensor schedule is a rule for sequencing the basic sensors and constructs a sequence of symbols $\{i_j\}_{j=0}^{N-1}$ from the past measurements. Let $\mathcal{L}$ denote the class of all sensor schedules of the form (15.2.2), (15.2.3).

As in Chap. 5, the uncertainty in the system (15.2.1) is required to satisfy the following integral quadratic constraint.

Given $X_0 = X_0^{\mathrm{T}} > 0$, $x_0 \in \mathbb{R}^n$, $d > 0$ and a finite time interval $[0, s]$, $s \leq T$, we consider the class of uncertainty inputs $\mathbf{col}\,[w(\cdot), \nu^*(\cdot)] \in \mathbf{L}_2[0, s]$ and initial conditions $x(0)$ such that

$$[x(0) - x_0]^{\mathrm{T}} X_0 [x(0) - x_0] + \int_0^s \left[\|w(t)\|^2 + \|\nu^*(t)\|^2 \right] dt \leq \\ d + \int_0^s \|z(t)\|^2 dt. \tag{15.2.4}$$

Notation 15.2.1. *Let $\mathcal{M} \in \mathcal{L}$ be a given sensor schedule and $y^*(\cdot)$ be the corresponding realized measured output. Then for the finite time interval $[0, s]$, $s \leq T$, $X_s\left[x_0, y^*(\cdot) \mid_0^s, d, \mathcal{M}\right]$ is the set of all possible states $x(s)$ at time s for the uncertain system* (15.2.1) *with the sensor schedule $\mathcal{M}$, uncertain inputs $w(\cdot)$ and $\nu^*(\cdot)$, and initial conditions satisfying the integral quadratic constraint* (15.2.4) *and compatible with the given output $y^*(\cdot)|_0^s$.*

The following definition is a natural extension of the concept of robust observability from Chap. 5.

Definition 15.2.2. *Let* $\mathcal{M} \in \mathcal{L}$ *be a given sensor schedule. The system* (15.2.1), (15.2.4) *is said to be* robustly observable with the sensor schedule $\mathcal{M}$ *on the interval* $[0, T]$, *if for any vector* $x_0 \in R^n$, *any time* $s \in [0, T]$, *any constant* $d > 0$, *and any realized measured output* $y^*(\cdot)$ *the set* $X_s \left[x_0, y^*(\cdot)\mid_0^s, d, \mathcal{M}\right]$ *is bounded.*

Notation 15.2.3. *Let* $\mathcal{A}(S)$ *be some measure of the size of a bounded convex set S. The hyperellipsoid centered about* $a \in \mathbb{R}^n$ *and determined by a square matrix* $M = M^{\mathrm{T}} > 0$ *and a real* $d > 0$ *is denoted by*

$$\mathcal{E}(M, a, d) := \left\{x \in \mathbb{R}^n : (x - a)^{\mathrm{T}} M (x - a) \leq d\right\}. \tag{15.2.5}$$

We suppose that the following assumptions hold.

Assumption 15.2.4. *For all* a_1, a_2, $\mathcal{A}(\mathcal{E}(M, a_1, d)) = \mathcal{A}(\mathcal{E}(M, a_2, d))$.

Assumption 15.2.5. *If* $d_1 > d_2$, *then* $\mathcal{A}(\mathcal{E}(M, a, d_1)) > \mathcal{A}(\mathcal{E}(M, a, d_2))$.

Assumption 15.2.6. $\mathcal{A}(\mathcal{E}(M, a, d)) \to \infty$ *as* $d \to \infty$.

Notation 15.2.7. *We will use the notation* $\mathcal{A}(M, d)$ *for the number* $\mathcal{A}(\mathcal{E}(M, a, d))$, *where* $\mathcal{E}(M, a, d)$ *is defined by* (15.2.5).

Explanation 15.2.8. This number does not depend on a by Assumption 15.2.4.

Definition 15.2.9. *Let* $\mathcal{M} \in \mathcal{L}$ *be a given sensor schedule. The uncertain system* (15.2.1), (15.2.4) *is said to be* uniformly and robustly observable with the sensor schedule $\mathcal{M}$ *on* $[0, T]$, *if it is robustly observable with this sensor schedule and for any vector* $x_0 \in \mathbb{R}^n$, *any constant* $d > 0$, *the following condition holds:*

$$c[x_0, d, \mathcal{M}] := \sup_{y^*(\cdot)} \mathcal{A}(X_T[x_0, y^*(\cdot)|_0^T, d, \mathcal{M}]) < \infty \tag{15.2.6}$$

where the supremum is taken over all fixed realized measured outputs $y^*(t)$.

Notation 15.2.10. *Let* $\mathcal{N}_0 \subset \mathcal{L}$ *denote the set of all sensor schedules such that the system* (15.2.1),(15.2.4) *is uniformly and robustly observable.*

Definition 15.2.11. *The uncertain system* (15.2.1), (15.2.4) *is said to be* uniformly and robustly observable via synchronous sensor switching with the basic sensors (15.2.2) *on* $[0, T]$ *if the set* $\mathcal{N}_0$ *is nonempty. In other words, if a sensor schedule exists such that the system* (15.2.1), (15.2.4) *is uniformly and robustly observable with this schedule.*

Definition 15.2.12. *Assume that the uncertain system* (15.2.1), (15.2.4) *is uniformly and robustly observable via synchronous sensor switching with the basic sensors* (15.2.2) *on* $[0, T]$. *Let* x_0 *be a given vector and* $d > 0$ *be a given number. A sensor schedule* $\mathcal{M}^0$ *is said to be* optimal *for the parameters* x_0 *and* d *if*

$$c[x_0, d, \mathcal{M}^0] = \inf_{\mathcal{M} \in \mathcal{N}_0} c[x_0, d, \mathcal{M}],$$

where $c[x_0, d, \mathcal{M}]$ *is defined by* (15.2.6).

Let $m := [i_0, i_1, \dots, i_{N-1}]$, where $1 \le i_j \le k$ be an index sequence defining a sensor schedule. Our solution to the optimal sensor scheduling problem with continuous time measurements involves the following Riccati differential equations associated with the sequence m:

$$\dot{P}^m(t) = A(t)P^m(t) + P^m(t)A^{\mathrm{T}}(t) + P^m(t)\Big[K^{\mathrm{T}}(t)K(t) - C_m^*(t)^{\mathrm{T}}C_m^*(t)\Big]P^m(t) + B(t)B^{\mathrm{T}}(t), \quad P^m(0) = X_0^{-1} \tag{15.2.7}$$

and the following set of state estimator equations:

$$\begin{aligned}\dot{\hat{x}}(t) = \big\{A(t) + P^m(t)\big[K^{\mathrm{T}}(t)K(t) - C_m^*(t)^{\mathrm{T}}C_m^*(t)\big]\big\}\,\hat{x}(t) \\ +P^m(t)C_m^{*'}(t)y^*(t); \\ \hat{x}(0) = x_0.\end{aligned} \tag{15.2.8}$$

Here

$$C_m^*(t) := C_{i_j}(t) \quad \text{for} \quad t \in [t_j, t_{j+1}), \ \ j = 0, 1, \dots, N-1. \tag{15.2.9}$$

Let $\hat{x}_0$ be a vector, and $y^*(\cdot)$ be a vector function. Introduce the following value:

$$F_j^i(\hat{x}_0, y^*(\cdot)) := \int_{t_j}^{t_{j+1}} \big[\|K(t)\hat{x}(t)\|^2 - \|(C^*(t)\hat{x}(t) - y^*(t))\|^2\big]\,dt, \tag{15.2.10}$$

where $\hat{x}(t)$ is the solution of (15.2.8) with $\hat{x}(t_j) = \hat{x}_0$ and $C_m^*(\cdot)$ defined by (15.2.9) with $i_j := i$.

For all $\hat{x}_0 \in \mathbb{R}^n$, $j = 1, 2, \dots, N$, and $1 \le i_j \le k$, introduce the functions

$$\hat{V}_j[\hat{x}_0, i_0, i_1, \dots, i_{j-1}] \in \mathbb{R}^{n\times n}, \quad v_j[\hat{x}_0, i_0, i_1, \dots, i_{j-1}] \in \mathbb{R}$$

as solutions of the following dynamic programming procedure. Firstly, we define

$$\begin{aligned}&v_N[\hat{x}_0, i_0, i_1, \dots, i_{N-1}] := 0 \quad \forall \hat{x}_0, i_0, i_1, \dots, i_{N-1}, \\ &\hat{V}_N[\hat{x}_0, i_0, i_1, \dots, i_{N-1}] := P^m(T)^{-1} \quad \forall \hat{x}_0, \ \ m = [i_0, i_1, \dots, i_{N-1}].\end{aligned} \tag{15.2.11}$$

Note that we do not assume that the solution of the Riccati equation (15.2.7) exists on $[0, T]$ for any m. If the solution does not exist for some m, we take $P^m(T)^{-1} := \infty$.

Furthermore, for all $\hat{x}_0 \in \mathbb{R}^n$ and $j = 0, 1, \dots, N-1$, let $i_j(\hat{x}_0)$ be an index for which the minimum in the following minimization problem is achieved:

$$\min_{i=1,2,\dots,k} \ \sup_{y^*(\cdot)\in \mathbf{L}_2[t_j, t_{j+1}]} \mathcal{A}(\hat{V}_{j+1}[\hat{x}(t_{j+1}), i_0, i_1, \dots, i_{j-1}, i], \hat{d}), \tag{15.2.12}$$

where

$$\hat{d} := F_j^i(\hat{x}_0, y^*(\cdot)) + v_{j+1}[\hat{x}(t_{j+1}), i_0, i_1, \ldots, i_{j-1}, i].$$

Note that this index may be nonunique. Moreover, if

$$\sup_{y^*(\cdot)\in \mathbf{L}_2[t_j,t_{j+1}]} \mathcal{A}(\hat{V}_{j+1}[\hat{x}(t_{j+1}), i_0, i_1, \ldots, i_{j-1}, i_j(\hat{x}_0)], \hat{d}) < \infty, \tag{15.2.13}$$

then a matrix $M_j(\hat{x}_0) > 0$ and a number $d_j(\hat{x}_0) > 0$ exist such that the supremum in (15.2.13) is equal to $\mathcal{A}(M_j(\hat{x}_0), d_j(\hat{x}_0))$. Now let

$$\begin{aligned} \hat{V}_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}] &:= M_j(\hat{x}_0); \\ v_j[\hat{x}(t_{j+1}), i_0, i_1, \ldots, i_{j-1}] &:= d_j(\hat{x}_0). \end{aligned} \tag{15.2.14}$$

The main result of this section is now given by the following theorem.

Theorem 15.2.13. *Consider the uncertain system* (15.2.1), (15.2.4) *with the basic sensors* (15.2.2)*. Then, the following two statements are equivalent:*

(i) The uncertain system (15.2.1)*,* (15.2.4) *is uniformly and robustly observable via synchronous sensor switching with the basic sensors* (15.2.2) *on* $[0, T]$*;*
(ii) The dynamic programming procedure defined by (15.2.11)*,* (15.2.12)*, and* (15.2.14) *has a finite solution*

$$\hat{V}_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}] > 0, \quad v_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}] \geq 0$$

for $j = 0, 1, \ldots, N-1$ *for all* $\hat{x}_0 \in \mathbb{R}^n$.

Furthermore, if condition (ii) holds and $i_j(\hat{x}_0)$ *is an index defined in the above dynamic programming procedure, then the sensor schedule defined by the sequence of indexes* $i_j(\hat{x}(t_j))$ *is optimal.*

The proof of Theorem 15.2.13 is given in Sect. 15.4.

15.3 Model Predictive Sensor Scheduling

The solution to the discrete-time dynamic equations derived in Sect. 15.2 has been the subject of much research in the field of optimal control theory. It is difficult to solve dynamic programming equations in realistic situations. In this section, we apply ideas of model predictive control (e.g., see [29]) to give a nonoptimal but real-time implementable method for sensor switching.

Definition 15.3.1. *Assume that the uncertain system* (15.2.1), (15.2.4) *is uniformly and robustly observable via synchronous sensor switching with the basic sensors* (15.2.2) *on* $[0, T]$*. Let* x_0 *be a given vector and* $d > 0$ *be a given number. A sensor schedule* $\mathcal{M}^0 \in \mathcal{N}_0$ *is said to be* one-step-ahead optimal *for the parameters* x_0 *and* d *if for any* $j = 0, 1, \ldots, N-1$*, any realized measured output* $y^*(\cdot)|_0^{t_j}$ *and any schedule* $\mathcal{M}$ *such that* $\mathcal{M}$ *coincides with* $\mathcal{M}^0$ *on* $[0, t_j]$*, the following condition holds:*

$$\sup_{h(\cdot)\in\mathcal{H}} \mathcal{A}(X_{t_{j+1}}[x_0, h(\cdot)|_0^{t_{j+1}}, d, \mathcal{M}]) \geq$$
$$\sup_{h(\cdot)\in\mathcal{H}} \mathcal{A}(X_{t_{j+1}}[x_0, h(\cdot)|_0^{t_{j+1}}, d, \mathcal{M}^0]),$$

where $\mathcal{H} := \{h(\cdot) \in \mathbf{L}_2[0, t_{j+1}] : h(t) = y^*(t) \quad \forall t \in [0, t_j]\}$.

Remark 15.3.2. The idea of Definition 15.3.1 is very straightforward: We wish to design a schedule such that at any sensor switching time t_j, the upper bound of the size of the set of all possible states $X_{t_{j+1}}[x_0, y^*(\cdot)|_0^{t_{j+1}}, d, \mathcal{M}^0]$ is minimal.

Let $j \leq N-1$ and $i_0, i_1, \ldots, i_{j-1}$ be a fixed sequence of indexes ($1 \leq i_r \leq k$), and let $i = 1, 2, \ldots, k$. The result of this section involves the following k pairs of Riccati differential equations associated with the sequence $[i_0, i_1, \ldots, i_{j-1}, i]$ and defined over the time interval $[t_j, t_{j+1}]$:

$$\begin{aligned} \dot{P}^i(t) = A(t)P^i(t) + P^i(t)A^{\mathrm{T}}(t) + \\ P^i(t)[K^{\mathrm{T}}(t)K(t) - C_i^{\mathrm{T}}(t)C_i(t)]P^i(t) + B(t)B^{\mathrm{T}}(t); \\ P^i(t_j) = P^m(t_j); \end{aligned} \quad (15.3.15)$$

$$\begin{aligned} -\dot{Y}^i(t) = [A(t) + P^i(t)K(t)^{\mathrm{T}}K(t)]^{\mathrm{T}}Y^i(t) + \\ Y^i(t)[A(t) + P^i(t)K(t)^{\mathrm{T}}K(t)] + K(t)^{\mathrm{T}}K(t) \\ +Y^i(t)P^i(t)K(t)^{\mathrm{T}}K(t)K(t)^{\mathrm{T}}K(t)P^i(t)Y^i(t); \\ Y^i(t_{j+1}) = 0. \end{aligned} \quad (15.3.16)$$

Here $P^m(\cdot)$ is the solution to the Riccati equation (15.2.7) with

$$C_m^*(t) := C_{i_r}(t) \quad \text{for} \quad t \in [t_r, t_{r+1}), \;\; r = 0, 1, \ldots, j-1. \quad (15.3.17)$$

Furthermore, introduce the following values:

$$c_i[x_0, d] := \mathcal{A}\left[P^i(t_{j+1})^{-1}, d + \hat{x}(t_j)^{\mathrm{T}}Y^i(t_j)\hat{x}(t_j)\right],$$

where $\hat{x}(t)$ is defined by (15.2.8). If for some i the solution to at least one of the Riccati equations does not exist on the time interval $[t_j, t_{j+1}]$, we take $c_i[x_0, d] := \infty$.

Now we are in a position to present a method to design a one-step-ahead-optimal sensor switching strategy.

Theorem 15.3.3. *Consider the uncertain system* (15.2.1)*,* (15.2.4) *with the basic sensors* (15.2.2)*. A schedule* $\mathcal{M}^0$ *is one-step-ahead optimal if and only if for any* $j = 0, 1, \ldots, N-1$*, and any sensor index sequence* $i_0, i_1, \ldots, i_j$ *associated with some realized measured output* $y^*(\cdot)|_0^{t_j}$*, the following two statements hold:*

(i) For $i = i_j$*, the solution* $P^i(\cdot)$ *to the Riccati equation* (15.3.15) *is defined and positive-definite on the interval* $[t_j, t_{j+1}]$*, and the solution* $Y^i(\cdot)$ *to the Riccati equation* (15.3.16) *is defined and nonnegative-definite on the interval* $[t_j, t_{j+1}]$*;*

(ii) The following minimum:

$$\min_{i=1,2,\ldots,k} c_i[x_0, d]$$

is achieved at $i = i_j$*.*

Remark 15.3.4. Note that our method to design a one-step-ahead-optimal sensor switching rule requires at each step an on-line solution to k pairs of Riccati differential equations and a simple look-up procedure to determine which of the basic sensors to use.

15.4 Proof of Theorems 15.2.13 and 15.3.3

To prove the main result of this chapter, we consider the time-varying uncertain system:

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)w(t); \\ z(t) &= K(t)x(t); \\ y(t) &= C(t)x(t) + v(t), \end{aligned} \tag{15.4.18}$$

where $x(t) \in \mathbb{R}^n$ is the *state*; $w(t) \in \mathbb{R}^p$ and $v(t) \in \mathbb{R}^l$ are the *uncertainty inputs*; $z(t) \in \mathbb{R}^q$ is the *uncertainty output*; and $y(t) \in \mathbb{R}^l$ is the *measured output*; and $A(\cdot), B(\cdot), K(\cdot)$, and $C(\cdot)$ are bounded piecewise continuous matrix functions.

The uncertainty in the above system is required to satisfy the following integral quadratic constraint. Let $X_0 = X_0^{\mathrm{T}} > 0$ be a given matrix, $x_0 \in \mathbb{R}^n$ be a given vector, and $d > 0$ be a given constant. For a given finite time interval $[0, s]$, we will consider the uncertainty inputs $w(\cdot)$ and $v(\cdot)$ and initial conditions $x(0)$ such that

$$\begin{aligned} [x(0) - x_0]^{\mathrm{T}} X_0 [x(0) - x_0] + \int_0^s (\|w(t)\|^2 + \|v(t)\|^2) dt \\ \leq d + \int_0^s \|z(t)\|^2 dt. \end{aligned} \tag{15.4.19}$$

The following definition is a natural simplification of Definition 15.2.9 for the uncertain system (15.4.18) and (15.4.19).

Definition 15.4.1. *The uncertain system* (15.4.18)*,* (15.4.19) *is said to be* uniformly and robustly observable *on* $[0, T]$*, if it is robustly observable and for any vector* $x_0 \in \mathbb{R}^n$*, any constant* $d > 0$*, the following condition holds:*

$$c[x_0, d] := \sup_{y_0(\cdot)} \mathcal{A}(X_T[x_0, y_0(\cdot)|_0^T, d]) < \infty, \tag{15.4.20}$$

where the supremum is taken over all fixed measured outputs $y_0(t)$*.*

Our necessary and sufficient condition for uniform and robust observability involves the following Riccati differential equations:

$$\dot{P}(t) = A(t)P(t) + P(t)A(t)^{\mathrm{T}} + \\ P(t)\left[K(t)^{\mathrm{T}}K(t) - C(t)^{\mathrm{T}}C(t)\right]P(t) + B(t)B(t)^{\mathrm{T}}. \quad (15.4.21)$$

Also, we consider a set of state equations of the form

$$\dot{\hat{x}}(t) = \left\{A(t) + P(t)\left[K(t)^{\mathrm{T}}K(t) - C(t)^{\mathrm{T}}C(t)\right]\right\}\hat{x}(t) \\ +P(t)C(t)^{\mathrm{T}}y_0(t); \quad (15.4.22)$$

$$-\dot{Y}(t) = \left[A(t) + P(t)K(t)^{\mathrm{T}}K(t)\right]^{\mathrm{T}}Y(t) \\ +Y(t)\left[A(t) + P(t)K(t)^{\mathrm{T}}K(t)\right] + K(t)^{\mathrm{T}}K(t) \\ +Y(t)P(t)K(t)^{\mathrm{T}}K(t)K(t)^{\mathrm{T}}K(t)P(t)Y(t). \quad (15.4.23)$$

The following lemma gives a necessary and sufficient condition for uniform robust observability.

Lemma 15.4.2. *Let* $X_0 = X_0^{\mathrm{T}} > 0$ *be a given matrix. Consider the uncertain system* (15.4.18), (15.4.19)*. Then the system* (15.4.18), (15.4.19) *is uniformly and robustly observable on* $[0, T]$ *if and only if the following two statements hold:*

(i) The solution $P(\cdot)$ *to the Riccati equation* (15.4.21) *with initial condition* $P(0) = X_0^{-1}$ *is defined and positive-definite on the interval* $[0, T]$*;*
(ii) The solution $Y(\cdot)$ *to the Riccati equation* (15.4.23) *with boundary condition* $Y(T) = 0$ *is defined and nonnegative-definite on the interval* $[0, T]$.

Furthermore, if (i) and (ii) hold, then the upper bound (15.4.20) *is given by*

$$c[x_0, d] = \mathcal{A}\left[P(T)^{-1}, d + x_0^{\mathrm{T}}Y(0)x_0\right]. \quad (15.4.24)$$

Proof of Lemma 15.4.2. Necessity. The necessity of (i) immediately follows from Definition 15.4.1 and Lemma 5.3.2 (on p. 120). Now we prove the necessity of (ii). Indeed, Lemma 5.3.2 implies that the set $X_T[x_0, y_0(\cdot)|_0^T, d]$ is defined by

$$X_s[x_0, y_0(\cdot)|_0^s, d] = \\ \left\{x_s \in \mathbb{R}^n : \begin{array}{l}[x_s - \hat{x}(s)]^{\mathrm{T}}P(s)^{-1}[x_s - \hat{x}(s)] \leq \\ d + \rho_s[y_0(\cdot)]\end{array}\right\}, \quad (15.4.25)$$

where

$$\rho_s[y_0(\cdot)] := \int_0^s \left[\|K(t)\hat{x}(t)\|^2 - \|(C(t)\hat{x}(t) - y_0(t))\|^2\right] dt \quad (15.4.26)$$

and $\hat{x}(\cdot)$ is defined by the equation (15.4.22) with initial condition $\hat{x}(0) = x_0$.

This and the requirement (15.4.20) imply that

$$\sup_{y_0(\cdot)\in \mathbf{L}_2[0,T]} \int_0^T \left[\|K(t)\hat{x}(t)\|^2 - \|(C(t)\hat{x}(t) - y_0(t))\|^2\right] dt < \infty, \quad (15.4.27)$$

where the supremum is taken over all solutions to the linear system (15.4.22) with $\hat{x}(0) = x_0$. Using the linear substitution

$$\hat{y}(t) = C(t)\hat{x}(t) - y_0(t), \quad (15.4.28)$$

the requirement (15.4.27) can be rewritten as

$$\sup_{\hat{y}(\cdot)\in \mathbf{L}_2[0,T]} \int_0^T \left[\|K(t)\hat{x}(t)\|^2 - \|\hat{y}(t)\|^2\right] dt < \infty, \quad (15.4.29)$$

where the supremum is taken over all solutions to the linear system

$$\dot{\hat{x}}(t) = \left[A(t) + P(t)K(t)^{\mathrm{T}}K(t)\right]\hat{x}(t) + P(t)C(t)^{\mathrm{T}}\hat{y}(t) \quad (15.4.30)$$

with $\hat{x}(0) = x_0$. Furthermore, (15.4.29) immediately implies that

$$\sup_{\hat{y}(\cdot)\in \mathbf{L}_2[t_0,T]} \int_{t_0}^T \left[\|K(t)\hat{x}(t)\|^2 - \|\hat{y}(t)\|^2\right] dt < \infty, \quad (15.4.31)$$

where the supremum is taken over all solutions to the linear system (15.4.30) with $\hat{x}(t_0) = x_0$. Moreover, it is easy to see that

$$\sup_{\hat{y}(\cdot)\in \mathbf{L}_2[t_0,T]} \int_{t_0}^T \left[\|K(t)\hat{x}(t)\|^2 - \|\hat{y}(t)\|^2\right] dt \geq 0$$

for any x_0. This and (15.4.31) immediately imply that condition (ii) holds; e.g., see [34]. This completes the proof of this part of the theorem.

Sufficiency. Assume that conditions (i) and (ii) hold. Then, it follows from Lemma 5.3.2 that the set $X_s[x_0, y_0(\cdot)|_0^s, d]$ is given by (15.4.25). Furthermore, using the linear substitution (15.4.28) and standard results of the theory of linear quadratic optimal control (e.g., see [34]), we obtain that

$$\sup_{y_0(\cdot)\in \mathbf{L}_2[0,T]} \int_0^T \left[\|K(t)\hat{x}(t)\|^2 - \|(C(t)\hat{x}(t) - y_0(t))\|^2\right] dt = x_0^{\mathrm{T}}Y(0)x_0, \quad (15.4.32)$$

where the supremum is taken over all solutions to the linear system (15.4.22) with $\hat{x}(0) = x_0$. Finally, (15.4.25), (15.4.26) and (15.4.32) imply uniform and robust observability and the equation (15.4.24). This completes the proof of the lemma. □

Proof of Theorem 15.2.13 (on p. 498*)*. For all $j = 0, 1, \ldots, N-1$, $\hat{x}_0 \in \mathbb{R}^n$, $1 \leq i_0 \leq k$, …, $1 \leq i_{j-1} \leq k$, there exists a number $s_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}] \leq \infty$ defined as

$$s_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}] := \sup_{y^*(\cdot) \in \mathbf{L}_2[t_j, T]} \mathcal{A}(X_T[x_0, y^*(\cdot)|_0^T, d]),$$

where the supremum is taken over all solutions of the system (15.2.1), (15.2.4) with the sensor switching sequence $i_0, i_1, \ldots, i_{j-1}$ on $[0, t_j)$ and $\hat{x}(t_j) = \hat{x}_0$.

According to Lemma 15.4.2, the set $X_T[x_0, y^*(\cdot)|_0^T, d]$ is always an ellipsoid of the form

$$\mathcal{E}(M, a, d) := \left\{ x \in \mathbb{R}^n : (x - a)^{\mathrm{T}} M (x - a) \leq d \right\},$$

where M is a matrix from the finite set of matrices $P^m(T)^{-1}$. Hence if

$$s_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}] < \infty,$$

then a matrix $\hat{V}_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}]$ and a number $v_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}]$ exist such that

$$s_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}] = \mathcal{A}(\mathcal{E}(\hat{V}_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}], a, v_j[\hat{x}_0, i_0, i_1, \ldots, i_{j-1}]))$$

for some vector a.

The dynamic programming procedure (15.2.11), (15.2.12), (15.2.14) can be derived from Bellman's principle of optimality (e.g., see [18]):

An optimal policy has the property that no matter what the previous decisions (i.e., controls) have been, the remaining decisions must constitute an optimal policy with regard to the state resulting from those previous decisions.

The optimality of the sensor switching policy defined in the theorem follows immediately from the relationships (15.2.11), (15.2.12), and (15.2.14), and from Bellman's principle of optimality. This completes the proof of this theorem. □

Proof of Theorem 15.3.3 (on p. 499*).* The statement of this theorem immediately follows from Lemma 15.4.2. □

Appendix A: Proof of Proposition 7.6.13 on p. 215

We consider a nondecreasing flow of σ-algebras

$$\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \mathfrak{F}_2 \subset \dots$$

in a probability space. Let τ be a Markov time[1] with respect to this flow.

Notation A.1. *Let $\mathfrak{F}_\tau$ denote the σ-algebra of random events R such that*

$$R \wedge \{\tau = k\} \in \mathfrak{F}_k \quad \text{for all} \quad k < \infty.$$

In this appendix, the symbols τ and V (with possible indices) stand for a Markov time and a random variable, respectively. The following properties are justified in, e.g., Lemmas 1.5, 1.8, and 1.9 [94, Ch.1]:

$$\tau \text{ is } \mathfrak{F}_\tau \text{ -measurable;} \tag{A.1}$$

$$\boxed{V_k \text{ is } \mathfrak{F}_k \text{ -measurable for all } k < \infty} \Rightarrow \boxed{V_\tau \text{ is } \mathfrak{F}_\tau \text{ -measurable}}; \tag{A.2}$$

$$\tau' \prec \tau'' \Rightarrow \mathfrak{F}_{\tau'} \subset \mathfrak{F}_{\tau''}; \tag{A.3}$$

$$\boldsymbol{P}\big[V \in \mathcal{V} \big| \mathfrak{F}_\tau\big] = \sum_k \boldsymbol{P}\big[V \in \mathcal{V} \big| \mathfrak{F}_k\big] \cdot I_{\tau=k} + I_{\tau=\infty \wedge V \in \mathcal{V}}. \tag{A.4}$$

Explanation A.2. It is supposed that V_k is defined for $k = 0, 1, \dots$ and $k = \infty$. So in (A.2), V_τ is well defined even if $\tau = \infty$.

Explanation A.3. The relation $\prec$ is defined in Notation 7.6.12 (on p. 215); and I_R is the indicator of the random event R.

Proof of Proposition 7.6.13 (on p. 215). We are going to apply Theorem 6.9.10 (on p. 171) to the flow of σ-algebras $\mathfrak{F}_i^* := \mathfrak{F}_{\tau_i+1}, i = 1, 2, \dots$ and

$$J_i := I_{V_{\tau_i} \in \mathcal{V} \wedge \tau_i < \infty}, \qquad b_i := i.$$

[1] See Definition 7.6.11 on p. 215.

This flow is nondecreasing, and J_i is $\mathfrak{F}_i^*$-measurable by (A.1)–(A.3). Relation (6.9.4) (on p. 171) is true since $0 \le J_i \le 1$ a.s. Thus (6.9.5) (on p. 171) holds a.s. by Theorem 6.9.10 (on p. 171). Now in (6.9.5),

$$\begin{aligned}
\boldsymbol{E}[J_i|\mathfrak{F}_{i-1}^*] &= \boldsymbol{P}[V_{\tau_i} \in \mathcal{V} \wedge \tau_i < \infty | \mathfrak{F}_{\tau_{i-1}+1}] \\
&\overset{\text{(A.4)}}{=} \sum_{k=0}^{\infty} \boldsymbol{P}[V_{\tau_i} \in \mathcal{V} \wedge \tau_i < \infty | \mathfrak{F}_{k+1}] I_{\tau_{i-1}=k} + I_{V_{\tau_i} \in \mathcal{V} \wedge \tau_i < \infty \wedge \tau_{i-1} = \infty} \\
&\overset{\text{a)}}{=} \sum_{k=0}^{\infty} \boldsymbol{P}[V_{\tau_i} \in \mathcal{V} \wedge \tau_i < \infty \wedge \tau_{i-1} = k | \mathfrak{F}_{k+1}] \\
&\overset{\tau_{i-1} \prec \tau_i}{=} \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \boldsymbol{P}[V_{\tau_i} \in \mathcal{V} \wedge \tau_i = j \wedge \tau_{i-1} = k | \mathfrak{F}_{k+1}] \\
&= \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \boldsymbol{P}[V_j \in \mathcal{V} \wedge \tau_i = j | \mathfrak{F}_{k+1}] I_{\tau_{i-1}=k} \\
&\overset{\text{b)}}{=} \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \boldsymbol{P}\big[V_j \in \mathcal{V}\big] \boldsymbol{P}[\tau_i = j | \mathfrak{F}_{k+1}] I_{\tau_{i-1}=k} \\
&\overset{\tau_{i-1} \prec \tau_i}{=} \boldsymbol{P}(\mathcal{V}) \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \boldsymbol{P}[\tau_i = j | \mathfrak{F}_{k+1}] I_{\tau_{i-1}=k} \\
&= \boldsymbol{P}(\mathcal{V}) \sum_{k=0}^{\infty} \boldsymbol{P}[\tau_i < \infty | \mathfrak{F}_{k+1}] I_{\tau_{i-1}=k} + \boldsymbol{P}(\mathcal{V}) \underbrace{I_{\tau_i < \infty \wedge \tau_{i-1} = \infty}}_{=0 \text{ a.s.}} \\
&\overset{\text{(A.4)}}{=} \boldsymbol{P}(\mathcal{V}) \boldsymbol{P}\big[\tau_i < \infty \big| \mathfrak{F}_{\tau_{i-1}+1}\big] = \boldsymbol{P}(\mathcal{V}) \boldsymbol{P}\big[\tau_i < \infty \big| \mathfrak{F}_{i-1}^*\big].
\end{aligned}$$

Here a) holds since $\{\tau_{i-1} = k\} \in \mathfrak{F}_k \subset \mathfrak{F}_{k+1}$ and

$$\tau_{i-1} \prec \tau_i \Rightarrow \boldsymbol{P}\big[\tau_i < \infty \wedge \tau_{i-1} = \infty\big] = 0;$$

b) holds since V_j is independent of $\mathfrak{F}_j$, where $\mathfrak{F}_j \ni \{\tau_i = j\}$ and $\mathfrak{F}_j \supset \mathfrak{F}_{k+1}$ for $j \ge k+1$. Thus (6.9.5) (on p. 171) takes the form

$$\frac{1}{k} \sum_{i=1}^{k} I_{V_{\tau_i} \in \mathcal{V} \wedge \tau_i < \infty} - \boldsymbol{P}(\mathcal{V}) \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{P}\big[\tau_i < \infty \big| \mathfrak{F}_{i-1}^*\big] \xrightarrow{k \to \infty} 0 \quad \text{a.s.}$$

By taking here $\mathcal{V} := \mathfrak{V}$ (where $\mathfrak{V}$ is the image space of V_i; i.e., $V_i \in \mathfrak{V}$ a.s.), we get

$$\frac{1}{k} \sum_{i=1}^{k} I_{\tau_i < \infty} - \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{P}\big[\tau_i < \infty \big| \mathfrak{F}_{i-1}^*\big] \xrightarrow{k \to \infty} 0 \quad \text{a.s.}$$

By combining the last two displayed formulas, we arrive at (7.6.5) (on p. 215). □

Appendix B: Some Properties of Square Ensembles of Matrices

In this appendix, we consider square ensembles of matrices over a finite set $S = \{s\}$.[1] The entries of the ensemble Λ are denoted by $\Lambda_{s_1}^{s_2}$ and are matrices of the respective sizes $k(s_1) \times k(s_2)$, where the integers $k(s), s \in S$ are given. Similarly to matrices, we introduce the following natural operations:

The sum

$$\Lambda' + \Lambda'' = \Lambda \Leftrightarrow \Lambda_{s_1}^{s_2} = (\Lambda')_{s_1}^{s_2} + (\Lambda'')_{s_1}^{s_2} \quad \forall s_1, s_2;$$

The product

$$\Lambda' \star \Lambda'' = \Lambda \Leftrightarrow \Lambda_{s_1}^{s_2} = \sum_{s \in S} (\Lambda')_{s_1}^{s} (\Lambda'')_{s}^{s_2} \quad \forall s_1, s_2;$$

The adjoint

$$\Lambda' = \Lambda^{\mathrm{T}} \Leftrightarrow (\Lambda')_{s_1}^{s_2} = \left(\Lambda_{s_2}^{s_1}\right)^{\mathrm{T}} \quad \forall s_1, s_2.$$

It is easy to see that endowed with the first two operations, the variety $\mathfrak{L}(S)$ of ensembles constitute a ring [124] with unity: $(\Lambda_{\text{unit}})_{s_1}^{s_2} := I_{k(s_1)}$ whenever $s_1 = s_2$ and $(\Lambda_{\text{unit}})_{s_1}^{s_2} := 0$ otherwise. The transformation (11.3.3) (on p. 375) establishes an isomorphism $\Lambda \leftrightarrow M$ between this ring and that of $k \times k$ matrices M, where $k := \sum_{s \in S} k(s)$. In particular, this implies that the inverse Λ^{-1} introduced in Subsect. 11.3.1 (starting on p. 374) is the inverse in the ring $\mathfrak{L}(S)$. Furthermore, $\Lambda \leftrightarrow M \Rightarrow \Lambda^{\mathrm{T}} \leftrightarrow M^{\mathrm{T}}; (\Lambda^{\mathrm{T}})^{\mathrm{T}} = \Lambda$.

The symbol Λ_π denotes the ensemble corresponding to π_M under the isomorphism (11.3.3) (on p. 375). Here $M \leftrightarrow \Lambda$ and π_M is the matrix corresponding to the orthogonal projection of $\mathbb{R}^k$ onto **Im** M.

Lemma B.1. *The ensemble Λ_π does not depend on the enumeration of the set S underlying the isomorphism* (11.3.3) *(on p.* 375*).*

Proof. Given two enumerations $\{s'(1), \ldots, s'(q)\}$ and $\{s''(1), \ldots, s''(q)\}$, the matrices M' and M'' related to a given Λ by the corresponding isomorphisms (11.3.3),

[1] See Definition 11.3.1 on p. 375.

respectively, are clearly similar: $M' = UM''U^{-1}$, where U is the matrix corresponding to the permutation of the coordinates that transforms $s'(\nu) = s''(j_\nu)\,\forall\nu$ the second enumeration into the first one. In other words, U is the block matrix whose νth raw $(0,\dots,0,I,0,\dots,0)$ is constituted by the matrices of respective sizes $k[s'(\nu)]\times k[s''(1)],\dots,k[s'(\nu)]\times k[s''(q)]]$ with the unit matrix on the j_νth position. Since the matrix U is orthogonal, it establishes an automorphism of the Euclidean space $\mathbb{R}^k$ onto itself. It follows that the matrix $U\pi_{M''}U^{-1}$ is the orthogonal projection from $\mathbb{R}^k$ onto

$$U\,\mathbf{Im}\,M'' = \mathbf{Im}\,\left(UM''U^{-1}\right) = \mathbf{Im}\,M'.$$

Hence $U\pi_{M''}U^{-1} = \pi_{M'}$. This implies that the matrices $\pi_{M''}$ and $\pi_{M'}$ are transformed into a common ensemble Λ_π by the isomorphisms (11.3.3) (on p. 375) corresponding to the enumerations $\{s'(1),\dots,s'(q)\}$ and $\{s''(1),\dots,s''(q)\}$, respectively. $\square$

The next lemma gives some properties of the pseudoinverse $\overset{+}{\Lambda}$ introduced in Subsect. 11.3.1 (starting on p. 374) and the ensemble Λ_π.

Lemma B.2. *The following relations hold for any square self-adjoint* $\Lambda = \Lambda^{\mathrm{T}}$ *ensemble of matrices* Λ*:*

$$\overset{+}{\Lambda}\Lambda = \Lambda_\pi, \quad \Lambda_\pi^{\mathrm{T}} = \Lambda_\pi, \quad \overset{+}{\Lambda}\Lambda_\pi = \overset{+}{\Lambda}. \tag{B.1}$$

Proof. Under the isomorphism (11.3.3) (on p. 375), relations (B.1) shape into

$$\overset{+}{M}M = \pi_M, \quad \pi_M^{\mathrm{T}} = \pi_M, \quad \overset{+}{M}\pi_M = \overset{+}{M}, \tag{B.2}$$

where $M \leftrightarrow \Lambda$ and so $M^{\mathrm{T}} = M$. By the definition on p. 374, $\overset{+}{M} = M_0^{-1}\pi_M$, where $M_0 : (\ker M)^\perp \to \mathbf{Im}\,M$ is obtained by restricting M on $(\ker M)^\perp$ and $^\perp$ stands for the orthogonal complement. So we have

$$\overset{+}{M}M = \left(M_0^{-1}\pi_M\right)M = M_0^{-1}\left(\pi_M M\right) \overset{(a)}{=} M_0^{-1}M \overset{(b)}{=} \left(M|_{\ker M^\perp}\right)^{-1}M.$$

Here (a) holds since π_M is a projection onto $\mathbf{Im}\,M$ and so $\pi_M Mx = Mx\,\forall x$, and (b) holds by the definition of M_0. Now we observe that $\ker M^\perp = \mathbf{Im}\,M$ so far as $M^{\mathrm{T}} = M$ [124]. Thus

$$\overset{+}{M}M = \left(M|_{\mathbf{Im}\,M}\right)^{-1}M.$$

It follows that $\overset{+}{M}Mx = x$ for $x \in \mathbf{Im}\,M$, whereas $\overset{+}{M}Mx = 0$ for $x \in (\mathbf{Im}\,M)^\perp = \ker M$. This means that $\overset{+}{M}M$ is the orthogonal projection of $\mathbb{R}^k$ onto $\mathbf{Im}\,M$; i.e., the first relation from (B.2) does hold.

Well-known properties of orthogonal projections [124] are given by the second relation from (B.2) and the relation $\pi_M\pi_M = \pi_M$. Hence

$$\overset{+}{M}\pi_M = M_0^{-1}\pi_M\pi_M = M_0^{-1}\pi_M = \overset{+}{M};$$

i.e., the last relation from (B.2) does hold as well. $\square$

Appendix C: Discrete Kalman Filter and Linear-Quadratic Gaussian Optimal Control Problem

In this appendix, we recall some classic results concerning minimum variance state estimation and optimal control for discrete-time linear systems perturbed by Gaussian white noises. The proofs of these results and their various extensions can be found in many textbooks and monographs in the control theory; see, e.g., [8, 30, 85].

C.1 Problem Statement

The System

We consider discrete-time linear systems of the form:

$$x(t+1) = A(t)x(t) + B(t)u(t) + \xi(t) \quad t = 0, \ldots, T-1, \qquad x(0) = a; \quad \text{(C.1)}$$

$$y(t) = C(t)x(t) + \chi(t) \qquad t = 0, \ldots, T. \quad \text{(C.2)}$$

Here $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ are the state and control, respectively; $\xi(t) \in \mathbb{R}^n$ is a process disturbance; $y(t) \in \mathbb{R}^k$ is the measured output; and $\chi(t)$ is a measurement noise. It is supposed that the system meets the following requirements.

Assumption C.1. *The vectors $a, \xi(t)$, and $\chi(t)$ are random, Gaussian, and independent with $\boldsymbol{E}\xi(t) = 0$ and $\boldsymbol{E}\chi(t) = 0$. The mean $\boldsymbol{E}a$ and the correlation matrices*

$$R_{aa} := \boldsymbol{E}[a - \boldsymbol{E}a][a - \boldsymbol{E}a]^{\mathrm{T}}, \quad R_{\xi\xi}(t) := \boldsymbol{E}\xi(t)\xi(t)^{\mathrm{T}},$$
$$R_{\chi\chi}(t) := \boldsymbol{E}\chi(t)\chi(t)^{\mathrm{T}} \quad \text{(C.3)}$$

are known. So are the matrices $A(t), B(t)$, and $C(t)$.

The Control

The control should be generated on the basis of the available data, i.e., in the form

$$u(t) = \mathcal{U}[t, y(0), \ldots, y(t)]. \quad \text{(C.4)}$$

The function $\mathcal{U}[\cdot]$ can also depend on the known matrices and vector mentioned in Assumption C.1.

State Estimation Problem

In this problem, the control strategy (C.4) (identified with the function $\mathcal{U}[\cdot]$) is given. The problem is to determine the best (minimum variance)

$$\boldsymbol{E}\|\widehat{x}(t|t) - x(t)\|^2 \to \min \tag{C.5}$$

estimate $\widehat{x}(t|t) \in \mathbb{R}^n$ of the current state $x(t)$ based on the currently available data $y(0), \ldots, y(t)$. In other words, the estimate should be of the form

$$\widehat{x}(t|t) = \mathcal{X}[t, y(0), \ldots, y(t)],$$

where the function $\mathcal{X}[\cdot]$ can also depend on the known matrices and vector mentioned in Assumption C.1. The $\min$ in (C.5) is over all such functions $\mathcal{X}[\cdot]$. A well-known fact from the probability theory (see, e.g., [94, Ch. 1]) states that the minimum variance estimate exists and equals the conditional mathematical expectation

$$\widehat{x}(t|t) = \boldsymbol{E}\left[x(t)\big|y(0), \ldots, y(t)\right]. \tag{C.6}$$

Of interest are explicit formulas for $\widehat{x}(t|t)$ that fit to generate it on-line, i.e., so that the estimate $\widehat{x}(t|t)$ of the current state $x(t)$ is produced at the current time t.

Linear-Quadratic Gaussian (LQG) Optimal Control Problem

This problem is to minimize the quadratic cost functional

$$\mathcal{I}_T := \boldsymbol{E}\sum_{t=0}^{T-1}\Big[x(t+1)^{\mathrm{T}}Q(t+1)x(t+1) + u(t)^{\mathrm{T}}\Gamma(t)u(t)\Big] \tag{C.7}$$

over the class of control strategies described by (C.4). Here $Q(t+1) \geq 0$ and $\Gamma(t) > 0$ are symmetric $n \times n$ and $m \times m$ matrices, respectively.

C.2 Solution of the State Estimation Problem: The Kalman Filter

Let a control strategy (C.4) be given. Being coupled with certain $n \times n$ symmetric matrices $P(t)$, the minimum variance estimate (C.6) of the state $x(t)$ can be generated recursively in the following way.

Kalman Filter

The next estimate is given by

$$\widehat{x}(t+1|t+1) := \widehat{x}(t+1|t) + K(t+1)\Big[y(t+1) - C(t+1)\widehat{x}(t+1|t)\Big]. \tag{C.8}$$

Here

$$\widehat{x}(t+1|t) := A(t)\widehat{x}(t|t) + B(t)u(t); \tag{C.9}$$

$$K(t) := P(t)C(t)^{\mathrm{T}}\overset{+}{\Lambda}(t); \tag{C.10}$$

$$\Lambda(t) := \Lambda := C(t)P(t)C(t)^{\mathrm{T}} + R_{\chi\chi}(t); \tag{C.11}$$

and $\overset{+}{\Lambda}$ is the pseudoinverse of the matrix Λ (see Subsect. 11.3.1 starting on p. 374). The matrices $P(t)$ are generated recursively via a scheme of the form

$$\ldots \mapsto P(t) \mapsto \overline{P}(t) \mapsto P(t+1) \mapsto \ldots$$

as follows:

$$\overline{P}(t) := P(t) - K(t)C(t)P(t), \qquad P(t+1) = A(t)\overline{P}(t)A(t)^{\mathrm{T}} + R_{\xi\xi}(t). \tag{C.12}$$

The recursion is initialized by putting

$$\widehat{x}(-1|-1) := \boldsymbol{E}a, \quad P(0) = R_{aa},$$
$$u(-1) := 0, \quad B(-1) := 0, \quad A(-1) := I_n. \tag{C.13}$$

The *Kalman filter* is a device that produces the state estimate $\widehat{x}(t|t)$ along with the matrices $P(t)$ in accordance with formulas (C.8)–(C.13).

Properties of the Kalman Filter

Theorem C.2. *Suppose that Assumption* C.1 *is valid. Then the following statements hold:*

(i) The output $\widehat{x}(t|t)$ of the Kalman filter is the minimum variance estimate (C.6) *of the current state $x(t)$;*

(ii) The estimation error

$$e(t|t) := x(t) - \widehat{x}(t|t) \tag{C.14}$$

is a zero-mean Gaussian random vector independent of the prior observations $y(0), \ldots, y(t)$;

(iii) The control strategy (C.4) *does not effect the estimation error* (C.14)*: Any two strategies give rise to common errors;*

(iv) The matrices $P(t), \overline{P}(t)$ generated by the Kalman filter are the covariance matrices of the estimation errors:

$$\overline{P}(t) = \boldsymbol{E}e(t|t)e(t|t)^{\mathrm{T}}, \qquad P(t) = \boldsymbol{E}e(t|t-1)e(t|t-1)^{\mathrm{T}}, \tag{C.15}$$

where $e(t|t)$ is defined by (C.14)*,*

$$e(t|t-1) := x(t) - \widehat{x}(t|t-1), \tag{C.16}$$

and $\widehat{x}(t|t-1)$ is introduced by (C.9).

C.3 Solution of the LQG Optimal Control Problem

This solution is given by the Kalman filter and the feedback optimal in the problem of minimizing the cost functional (C.7) in the case where the noise is removed from (C.1) and the entire state $x(t)$ is accessible for on-line measurements. So we start with recalling the construction of this feedback.

Deterministic Linear-Quadratic Optimal Control Problem

This is the problem of minimizing the cost functional (C.7) subject to the constraints

$$x(t+1) = A(t)x(t) + B(t)u(t), \qquad x(0) = a. \tag{C.17}$$

Here $a \in \mathbb{R}^n$ is a given deterministic vector and minimization is over all control programs $u(0), \ldots, u(T-1)$.

Theorem C.3. *The solution of the deterministic linear-quadratic optimal control problem is generated by the feedback*

$$u(t) = -L(t)x(t), \tag{C.18}$$

where

$$L(t) = F(t)^{-1}B(t)^{\mathsf{T}}\left[Q(t+1) + H(t+1)\right]A(t); \tag{C.19}$$

$$F(t) := \Gamma(t) + B(t)^{\mathsf{T}}\left[Q(t+1) + H(t+1)\right]B(t), \tag{C.20}$$

and the symmetric $n \times n$-matrices

$$H(T), H(T-1), \ldots, H(0)$$

are calculated recursively in accordance with the following difference Ricatti equation, where $H(T) = 0$, $A_L(t) := A(t) - B(t)L(t)$, and $t = T-1, T-2, \ldots, 0$,

$$H(t) = L(t)^{\mathsf{T}}\Gamma(t)L(t) + A_L(t)^{\mathsf{T}}\left[Q(t+1) + H(t+1)\right]A_L(t). \tag{C.21}$$

It should be remarked that the optimal gain matrix $L(t)$ does not depend on the initial state a. For any a, the corresponding process generated by the feedback (C.18) in the system (C.17) furnishes the minimum of the cost functional under the initial condition $x(0) = a$.

Remark C.4. It easily follows from (C.20), (C.21), and the inequalities $Q(t) \geq 0, \Gamma(t) > 0$ that $H(t) \geq 0$ and $F(t) > 0$.

LQG Optimal Control Problem

Theorem C.5. *Suppose that Assumption* C.1 *is valid. Then the solution of the LQG optimal control problem results from replacing the state* $x(t)$ *by its minimum variance estimate* $\widehat{x}(t|t)$ *in the feedback* (C.18) *providing the solution for the corresponding deterministic optimal control problem.*

In other words, the solution of the LQG optimal control problem is given by

$$u(t) = -L(t)\widehat{x}(t|t).$$

Here $L(t)$ *is determined in accordance with* (C.19)–(C.21)*, and* $\widehat{x}(t|t)$ *is the output of the Kalman filter.*

Appendix D: Some Properties of the Joint Entropy of a Random Vector and Discrete Quantity

In this appendix, we formally justify the properties (6.7.12) and (6.7.14) (on p. 151) of the joint entropy of a random vector and discrete quantity. The only reason for this special consideration is that the joint entropy is not as conventional and well-studied a tool as the differential entropy of a random vector and the entropy of a random discrete quantity. For the last two kinds of entropy, the above properties are well known; see, e.g., [38, 40, 50, 60, 152, 190]. The arguments underlying these properties basically remain in use for the case of the joint entropy.

D.1 Preliminary technical fact.

We start with showing that $h(V|F), H(V,F)$, and $I(V,F)$ are finite if $h(V) \in \mathbb{R}$.

Lemma D.1. *Suppose that $V \in \mathbb{R}^s$ and $F \in \mathfrak{F} = \{f\}$ are the random vector and quantity, respectively; the set $\mathfrak{F}$ is finite; and $h(V) \in \mathbb{R}$. Then $H(V,F) \in \mathbb{R}, I(V,F) \in \mathbb{R}, h(V|F) \in \mathbb{R}$, and $h_f(V) \in \mathbb{R}$ whenever $\boldsymbol{P}(F=f) > 0$.*

Proof. By the arguments from the footnote on p. 150, there exists the conditional probability density $p(\cdot|f)$ of the random vector V given that $F = f$. We denote by $p(f) := \boldsymbol{P}(F = f)$ the probability mass function, and by df the counting measure: $\mathfrak{F}' \subset \mathfrak{F} \mapsto |\mathfrak{F}'|$, where $|\mathfrak{F}'|$ is the size of the set $\mathfrak{F}'$. It is easy to see that $p(v|f)p(f)$ and $\frac{p(v|f)p(f)}{p(v)p(f)}$ are the densities of the joint distribution of V and F with respect to the product measures $dv \otimes df$ and $\boldsymbol{P}_V(dv) \otimes \boldsymbol{P}_F(df)$, respectively. The standard arguments concerning the mutual information $I(V,F)$ defined by formula (6.3.1) (on p. 137) and based on the Jensen inequality show that the function $p(v|f)p(f)\log_2 \frac{p(v|f)p(f)}{p(v)p(f)}$ is integrable with respect to $dv \otimes df$ and

$$I(V,F) = \int p(v|f)p(f)\log_2 \frac{p(v|f)}{p(v)}\, dvdf \geq 0. \tag{D.1}$$

Since for almost all v,

$$\sum_{f'} p(v|f')p(f') = p(v) \Rightarrow p(v|f)p(f) \leq p(v), \tag{D.2}$$

the function $p(v|f)p(f)\log_2 p(v)$ is summable for any f and

$$h(V) = -\sum_f \int p(v|f)p(f)\log_2 p(v)\;dv.$$

At the same time, the integrand in (D.1) is equal to

$$p(v|f)p(f)\log_2 p(v|f) - p(v|f)p(f)\log_2 p(v).$$

Hence the function $p(v|f)p(f)\log_2 p(v|f)$ is integrable with respect to $dv \otimes df$. It follows that whenever $p(f) > 0$, the function $p(v|f)\log_2 p(v|f)$ is integrable and the conditional differential entropy

$$h_f(V) = -\int p(v|f)\log_2 p(v|f)\;dv < \infty.$$

Moreover,

$$\begin{aligned} h(V|F) = \sum_f p(f)h_f(V) = -\int p(v|f)p(f)\log_2 p(v|f)\;dvdf \\ = h(V) - I(V,F) \le h(V) < \infty. \quad \text{(D.3)} \end{aligned}$$

On the other hand, whenever $p(f) > 0$,

$$\begin{aligned} &h_f(V) = -\int p(v|f)\log_2 p(v|f)\;dv \\ &= -\int_{\{v:p(v|f)<1\}} p(v|f)\log_2 p(v|f)\;dv - \int_{\{v:p(v|f)\ge 1\}} p(v|f)\log_2 p(v|f)\;dv \\ &\ge -\int_{\{v:p(v|f)\ge 1\}} p(v|f)\log_2 p(v|f)\;dv \overset{\text{(D.2)}}{\ge} -\int_{\{v:p(v|f)\ge 1\}} \frac{p(v)}{p(f)}\log_2 \frac{p(v)}{p(f)}\;dv \\ &= -\frac{1}{p(f)}\int_{\{v:p(v|f)\ge 1\}} p(v)\log_2 p(v)\;dv + \frac{\log_2 p(f)}{p(f)}\int_{\{v:p(v|f)\ge 1\}} p(v)\;dv > -\infty. \end{aligned}$$

Thus we see that $h_f(V) \in \mathbb{R}$. It follows that $h(V|F) \in \mathbb{R}$ and so $I(V,F) \in \mathbb{R}$, thanks to (D.3). □

Remark D.2. Formula (D.3) established in the proof of Lemma D.1 is identical to the first part of (6.7.13) (on p. 151). In the particular case where $\widehat{V} = V$, formula (6.7.12) (on p. 151) follows from (D.1), (D.3), and Lemma D.1.

D.2 Proof of (6.7.14) on p. 151.

$$H(V, F_1|F) \overset{(a)}{=} \sum_f \boldsymbol{P}(F = f) H_{F=f}(V, F_1)$$
$$\overset{(b)}{=} \sum_f \boldsymbol{P}(F = f) \Big[h_{F=f}(V|F_1) + H_{F=f}(F_1) \Big]$$
$$\overset{(a)}{=} \sum_f \boldsymbol{P}(F = f) \sum_{f_1} \boldsymbol{P}(F_1 = f_1 | F = f) h_{F=f}(V|F_1 = f_1) + H(F_1|F)$$
$$= \sum_{f, f_1} \boldsymbol{P}(F_1 = f_1 \wedge F = f) h_{F=f \wedge F_1 = f_1}(V) + H(F_1|F) \overset{(a)}{=} h(V|F_1, F) + H(F_1|F).$$

Here (a) holds by the definition of the averaged conditional entropy and (b) is valid by (6.7.11) (on p. 151). Thus the first relation in (6.7.14) (on p. 151) does hold.

To prove the inequality from (6.7.14), we first note that due to (D.1),

$$I(V, F) = -\int p(v|f)p(f) \log_2 p(f)\, dvdf + \int p(v|f)p(f) \log_2 \frac{p(v|f)p(f)}{p(v)}\, dvdf$$
$$= H(F) - \int p(v) dv \sum_f -\frac{p(v|f)p(f)}{p(v)} \log_2 \frac{p(v|f)p(f)}{p(v)}.$$

Here for almost all v, the sum is the entropy of a probability distribution over $\mathfrak{F}$. Hence this sum is nonnegative, and we see that

$$I(V, F) \leq H(F).$$

It follows that

$$H(V, F) \overset{(a)}{=} h(V|F) + H(F) \overset{(b)}{=} h(V) - I(V, F) + H(F) \geq h(V).$$

Here (a) holds by (6.7.11) (on p. 151) and (b) is due to (D.3). Thus

$$H(V, F_1|F) = \sum_f p(f) H_{F=f}(V, F_1) \geq \sum_f p(f) h_{F=f}(V) = h(V|F);$$

i.e., the inequality from (6.7.14) (on p. 151) does hold. □

D.3 Proof of (6.7.12) on p. 151.

The strict inequalities from (6.7.12) are apparent from (6.7.11) (on p. 151) and Lemma D.1. To prove the nonstrict one $H(\widehat{V}|F) \leq H(\widehat{V})$, we invoke the corresponding property of the entropy of the discrete random quantities $H(F_1|F) \leq H(F_1)$ [38, 40, 50, 60, 152, 190] and observe that

$$
\begin{aligned}
H(\widehat{V}|F) = H(V, F_1|F) &\overset{(6.7.14)}{=\!=} h(V|F_1, F) + H(F_1|F) \\
&= \sum_{f, f_1} \boldsymbol{P}(F = f, F_1 = f_1) h_{F=f, F_1=f_1}(V) + H(F_1|F) \\
&= \sum_{f_1} \boldsymbol{P}(F_1 = f_1) \sum_{f} \boldsymbol{P}(F = f | F_1 = f_1) h_{F_1=f_1}(V|F = f) + H(F_1|F) \\
&= \sum_{f_1} \boldsymbol{P}(F_1 = f_1) h_{F_1=f_1}(V|F) + H(F_1|F) \\
&\overset{(D.3)}{\leq} \sum_{f_1} \boldsymbol{P}(F_1 = f_1) h_{F_1=f_1}(V) + H(F_1|F) = h(V|F_1) + H(F_1|F) \\
&\leq h(V|F_1) + H(F_1) \overset{(6.7.11)}{=\!=} H(V, F_1) = H(\widehat{V}). \qquad \square
\end{aligned}
$$

References

1. H. Abou-Kandil, G. Freiling, and G. Jank. On the solution of discrete-time Markovian jump linear-quadratic control problems. *Automatica*, 31:765–768, 1995.
2. R. L. Adler, A. G. Konheim, and M. H. McAndrew. Topological entropy. *Transactions of the American Mathematical Society*, 114(2):309–319, 1965.
3. R. L. Adler and B. Weiss. Entropy a complete metric invariant for automorphisms of torus. *Proceedings of The National Academy of Sciences, USA*, 57(6):1573–1577, 1967.
4. R. Ahlswede. Network information flow. *IEEE Transactions on Information Theory*, 46(6):1204–1216, 2000.
5. R. Ahlswede and N. Cai. Seminoisy deterministic multiple-access channels: coding theorems for list codes and codes with feedback. *IEEE Transactions on Information Theory*, 48(8):2153–2162, 2002.
6. I. F. Akyilidiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Communications Magazine*, 40(8):102–114, 2002.
7. I.F. Akyilidiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: Survey. *Computer Networks*, 38:393–422, 2002.
8. B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, N.J., 1979.
9. B. D. O. Anderson, C. Yu, B. Fidan, and J. M. Hendrickx. Control and information architectures for formations. In *Proceedings of the IEEE Conference on Control Applications*, Munich, Germany, October 2006.
10. K. Aström and B. Wittenmark. *Computer-Controlled Systems: Theory and Design*. Prentice Hall, New York, 1997.
11. K. J. Astrom. *Introduction to Stochastic Control Theory*. Academic Press, New York, 1970.
12. K. J. Astrom and B. Wittenmark. *Computer-Controlled Systems*. Prentice Hall, Upper Saddle River, N.J., 1997.
13. B. Azimi-Sadjadi. Stability of networked control systems in the presence of packet losses. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 676–681, Atlantis, Bahamas, December 2004.
14. T. Başar and P. Bernhard. H^{∞}*-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Birkhäuser, Boston, 1991.
15. J. Baillieul. Feedback designs for controlling device arrays with communication channel bandwidth constraints. In *Proceedings of the ARO Workshop on Smart Structures*, Penns. State Uni., University Park, P.A., August 1999.

16. C. R. Baker. Coding capacity for a class of additive channels. *IEEE Transactions on Information Theory*, 33(2):233–243, 1991.
17. J.S. Baras and A. Bensoussan. Optimal sensor scheduling in nonlinear filtering of diffusion processes. *SIAM Journal on Control and Optimization*, 27(4):786–813, 1989.
18. R.E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.
19. A. Bensoussan. *Stochastic control of partially observable systems*. Cambridge University Press, New York, N.Y., 1992.
20. T. Berger. *Rate Distortion Theory*. Prentice-Hall, Englewood Cliffs, N.J., 1971.
21. D. P. Bertsekas and I. B. Rhodes. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, 16(2):117–128, 1971.
22. E. Bierstone and P. Milman. Semialgebraic and subanalytical sets. *IHES Pub. Math*, 67:5–42, 1988.
23. J.-D. Boissonnat and M.A. Yvinec. *Algorithmic Geometry*. Cambridge University Press, Cambridge, M.A., 1998.
24. V. Borkar, S. Mitter, and S. Tatikonda. Optimal sequential vector quantization of Markov sources. *SIAM Journal on Control and Optimization*, 40(1):135–148, 2001.
25. Th. L. Boullion and P. L. Odell. *Generalized Inverse Matrices*. Wiley-Interscience, New York, 1971.
26. R. Bowen. *Equilibrium States and Ergodic Theory of Anosov Diffeomorphisms*. Springer-Verlag, Berlin, 1975.
27. R. W. Brockett. Stabilization of motor networks. In *Proceedings of the 34th IEEE Conference on Decision and Control*, New Orleans, L.A., December 1995.
28. R. W. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, 45(7):1279–1289, 2000.
29. E.F. Camacho and C. Bordons. *Model Predictive Control in the Process Industry*. Springer-Verlag, London, 1995.
30. D.E. Catlin. *Estimation, Control, and the Discrete Kalman Filter*. Springer, New York, 1989.
31. S. C. Chang and E. J. Weldon. Coding for T-user multiple access channels. *IEEE Transactions on Information Theory*, 25:684–691, 1979.
32. T. M. Cheng and A. V. Savkin. Output feedback stabilization of nonlinear networked control systems with non-decreasing nonlinearities: matrix inequalities approach. *International Journal of Robust and Nonlinear Control*, 17(5-6):387–404, 2007.
33. T.M. Cheng, V. Malyavej, and A. V. Savkin. Set-valued state estimation for uncertain continuous-time systems via limited capacity communication channels. In *Proceedings of the 17th IFAC World Congress*, Seoul, South Korea, July 2008.
34. D. J. Clements and B. D. O Anderson. *Singular Optimal Control: The Linear-Quadratic Problem*. Springer-Verlag, Berlin, 1978.
35. O. L. V. Costa and M. D. Fragozo. Discrete-time LQ-optimal control problems with infinite Markov jump parameter systems. *IEEE Transactions on Automatic Control*, 40:2076–2088, 1995.
36. O. L. V. Costa and R. P. Marques. Robust H_2-control for discrete-time Markovian jump linear systems. *International Journal of Control*, 73(1):11–21, 2000.
37. I. Csisár. The method of types. *IEEE Transactions on Information Theory*, 44(6):2505–2523, 1998.
38. I. Csiszar and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Channels*. Academic Press, New York, 1982.

39. S. Dasgupta. Control over bandlimited communication channels: limitation to stabilizability. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 176–181, Maui, H.I., December 2003.
40. A. Dembo, Th. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991.
41. C. De Persis. Results on stabilization of nonlinear systems under data-rate constraints. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 3986–3991, Paradise Island, Bahamas, 2004.
42. C. De Persis and A. Isidori. Stabilization by state feedback implies stabilization by encoded state feedback. *Systems and Control Letters*, 53(3-4):249–258, 2004.
43. J. J. Deyst and Ch. F. Price. Conditions for asymptotic stability of the discrete minimum-variance linear estimator. *IEEE Transactions on Automatic Control*, AC-13(6):702–705, 1968.
44. R. L. Dobrushin. Information transmission in a channel with feedback. *Theory of Probability and Applications*, 34:367–383, 1958.
45. N. G. Dokuchaev and A. V. Savkin. Recursive state estimation via limited capacity communication channels. In *Proceedings of the 38th IEEE Conference on Decision and Control*, pages 4929–4932, Phoenix, A.Z., December 1999.
46. P. Dorato, R. Tempo, and G. Muscato. Bibliography on robust control. *Automatica*, 29(1):201–214, 1993.
47. N. Elia. When Bode meets Shannon: control-oriented feedback communication schemes. *IEEE Transactions on Automatic Control*, 49(9):1477–1492, 2004.
48. N. Elia and S. K. Mitter. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 44(9):1384–1400, 2001.
49. A. Ephremides and B. Hajek. Information theory and communication networks: an unconsummated union. *IEEE Transactions on Information Theory*, 44(6):2416–2434, 1998.
50. R. M. Fano. *Transmission of Information, A Statistical Theory of Communication*. Wiley, New York, 1961.
51. J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formation. *IEEE Transactions on Automatic Control*, 49(9):1465–1476, 2004.
52. A. Feinstein. Error bounds in noisy channels with memory. *IRE Transactions on Information Theory*, PGIT-1:13–14, 1955.
53. W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. J. Wiley & Sons, New York, 1971.
54. A. Feuer and G.C. Goodwin. *Sampling in Digital Signal Processing and Control*. Birkhauser, Boston, M.A., 1996.
55. V. N. Fomin. *Operator Approach in the Linear Filtering Theory*. St. Petersburg University Publishers, Saint Petersburg, 1996.
56. M. D. Fragozo. On a partially observable LQG problem for systems with Markovian jumping parameters. *Systems and Control Letters*, 10:349–356, 1988.
57. D. H. Fremlin. *Measure Theory*. Torres Fremlin, UK, 2000.
58. J. Freudenberg, J. Braslavsky, and R. Middleton. Control over signal-to-noise ratio constrained channels: stabilization and performance. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 191–196, Seville, Spain, December 2005.
59. R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, IT-11:3–18, 1965.
60. R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

61. F. R. Gantmacher. *The Theory of Matrices*, volume 1. Chelsea Publ. Com., New York, 1959.
62. A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Dordrecht, 1992.
63. G. C. Goodwin, S. F. Graebe, and M. E. Salgado. *Control System Design*. Prentice Hall International, Inc., Upper Saddle River, N.J., 2001.
64. G. C. Goodwin, H. Haimovich, D. E. Quevedo, and J. S. Welsh. A moving horizon approach to networked control system design. *IEEE Transactions on Automatic Control*, 49(9):1427–1445, 2004.
65. R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, New York, 1990.
66. M. Green and D. J. N. Limebeer. *Linear Robust Control*. Prentice-Hall, Englewood Cliffs, N.J., 1994.
67. B. E. Griffits and K. A. Loparo. Optimal control of a jump-linear Gaussian systems. *International Journal of Control*, 42(4):791–819, 1985.
68. S. Guiasu. *Information Theory with Applications*. McGraw-Hill Int., New York, 1977.
69. T. S. Han. Slepian-Wolf-Cover theorem for networks of channels. *Information and Control*, 47:67–83, 1980.
70. J. Hespanha, A. Ortega, and L. Vasudevan. Towards the control of linear systems with minimum bit-rate. In *Proceedings of the International Simposium on the Mathematical Theory of Networks and Systems*, Notre Dame, I.N., 2002.
71. D. Hristu and K. Morgansen. Limited communication control. *Systems and Control Letters*, 37:193–205, 1999.
72. O. C. Imer, S. Yüksel, and T. Başar. Optimal control of LTI systems over unreliable communication links. *Automatica*, 42:1429–1439, 2006.
73. H. Ishii and T. Başar. Remote control of LTI systems over networks with state quantization. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 830–835, Las Vegas, N.V., USA, December 2002.
74. H. Ishii and B. A. Francis. Stabilization with control networks. In *Proceedings of the IEE Control Confrence*, Cambridge, UK, July 2000.
75. H. Ishii and B. A. Francis. *Limited Data Rate in Control Systems with Networks*. Springer-Verlag, Berlin, 2002.
76. A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.
77. R. Jain, T. Simsek, and P. Varaiya. Control under communication constraints. In *Proceedings of the 41th IEEE Conference on Decision and Control*, pages 3209–3216, Las Vegas, N.V., December 2002.
78. Y. Ji and H. J. Chizeck. Jump linear quadratic gaussian control in continuous time. *IEEE Transactions on Automatic Control*, 37(12):1884–1892, 1992.
79. Y. Ji, H. J. Chizeck, X. Feng, and K. Loparo. Stability and control of discrete-time jump linear systems. *Control Theory and Technology Journal*, 7(2):247–270, 1991.
80. R. E. Kalman and J. E. Bertram. A unified approach to the theory of sampling systems. *J. Franklin Institute*, 26(7):405–436, 1959.
81. P. P. Khargonekar, I. R. Petersen, and K. Zhou. Robust stabilization of uncertain systems and H^∞ optimal control. *IEEE Transactions on Automatic Control*, AC-35(3):356–361, 1990.
82. A. N. Kolmogorov. A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Doklady Akademii Nauk SSSR*, 119(5):861–864, 1958.
83. A. N. Kolmogorov. Entropy per unit time as a metric invariant of automorphisms. *Doklady Akademii Nauk SSSR*, 124(4):754–755, 1959.

84. J. Körner and A. Orlitsky. Zero-error information theory. *IEEE Transactions on Information Theory*, 44(6):2207–2229, 1998.
85. H. J. Kushner. *Introduction to Stochastic Control*. Holt, Rinehart, and Winston, New York, 1971.
86. B. F. La Scala and R. J. Evans. Minimum necessary data rates for accurate track fusion. In *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference ECC 2005*, pages 6966–6971, Seville, Spain, December 2005.
87. F. L. Lewis. *Optimal Control*. Wiley, New York, 1986.
88. D. Liberzon. Hybrid feedback stabilization of systems with quantized signals. *Automatica*, 39:1543–1554, 2003.
89. D. Liberzon. Hybrid stabilization of systems with quantized signals. *Automatica*, 39(9):1543–1554, 2003.
90. D. Liberzon. On stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 48(2):304–307, 2003.
91. D. J. N. Limebeer, B. D. O. Anderson, P. P. Khargonekar, and M. Green. A game theoretic approach to H^∞ control for time-varying systems. *SIAM Journal on Control and Optimization*, 30:262–283, 1992.
92. Q. Ling and M. Lemmon. Optimal dropout compensation in networked control systems. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, Maui, H.I., December 2003.
93. L. W. Liou and A. Ray. A stochastic regulator for integrated communication and control systems: part 1 - formulation of control law. *ASME Journal of Dynamic Systems, Measurement, and Control*, 113:604–611, 1991.
94. R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes I. General Theory*. Springer, New York, 2nd edition, 2001.
95. M. Loéve. *Probability Theory*, volume 2. Springer, New York, 4th edition, 1978.
96. L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, IT-25:1–7, 1979.
97. R. Luck and A. Ray. An observer-based compensator for distributed delays. *Automatica*, 26(5):903–908, 1990.
98. R. Luck, A. Ray, and Y. Halevi. Observability under recurrent loss of data. *AIAA Journal of Guidance, Control, and Dynamics*, 15(1):284–287, 1992.
99. V. Malyavej, I. R. Manchester, and A. V. Savkin. Precision missile guidance using radar/multiple-video sensor fusion via communication channels with bit-rate constraints. *Automatica*, 42(5):763–769, 2006.
100. V. Malyavej and A. V. Savkin. The problem of robust Kalman state estimation via limited capacity digital communication channels. *Systems and Control Letters*, 54(3):283–292, 2005.
101. N. C. Martins. *Information theoretic aspects of the control and mode estimation of stochastic systems*. PhD thesis, Massachusetts Institute of Technology, 2004.
102. N. C. Martins, M. A. Dahleh, and N. Elia. Feedback stabilization of uncertain systems using a stochastic digital link. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 1889–1895, Atlantis, Bahamas, December 2004.
103. N. C. Martins, M. A. Dahleh, and N. Elia. Feedback stabilization of uncertain systems in the presence of a direct link. *IEEE Transactions on Automatic Control*, 51(3):438–447, 2006.
104. A. S. Matveev and A. V. Savkin. *Qualitative Theory of Hybrid Dynamical Systems*. Birkhauser, Boston, M.A., 2000.

105. A. S. Matveev and A. V. Savkin. Optimal computer control via communication channels with irregular transmission times. *International Journal of Control*, 76(2):165–177, 2003.
106. A. S. Matveev and A. V. Savkin. The problem of state estimation via asynchronous communication channels with irregular transmission times. *IEEE Transactions on Automatic Control*, 48(4):670–676, 2003.
107. A. S. Matveev and A. V. Savkin. An analogue of Shannon information theory for networked control systems. Stabilization via a noisy discrete channel. In *Proceedings of the 43th IEEE Conference on Decision and Control*, pages 4491–4496, Atlantis, Paradise Island, Bahamas, December 2004.
108. A. S. Matveev and A. V. Savkin. An analogue of Shannon information theory for networked control systems. State estimation via a noisy discrete channel. In *Proceedings of the 43th IEEE Conference on Decision and Control*, pages 4485–4490, Atlantis, Paradise Island, Bahamas, December 2004.
109. A. S. Matveev and A. V. Savkin. Optimal control via asynchronous communication channels. *Journal of Optimization Theory and Applications*, 122(3):539–572, 2004.
110. A. S. Matveev and A. V. Savkin. The problem of LQG optimal control via a limited capacity communication channel. *Systems and Control Letters*, 53:51–64, 2004.
111. A. S. Matveev and A. V. Savkin. Stabilization of multisensor networked control systems with communication constraints. In *Proceedings of the 5th Asian Control Conference*, pages 1916–1923, Melbourne, Australia, July 2004.
112. A. S. Matveev and A. V. Savkin. Almost sure nonobservability and nontabilizability of unstable noisy plants via communication channels with packet losses. In *the 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005*, pages 7338–7341, Seville, Spain, December 2005.
113. A. S. Matveev and A. V. Savkin. Comments on "Control over Noisy Channels" and relevant negative results. *IEEE Transactions on Automatic Control*, 50(12):2105–2110, December 2005.
114. A. S. Matveev and A. V. Savkin. Decentralized stabilization of linear systems via limited capacity communication networks. In *the 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005*, pages 1155–1161, Seville, Spain, December 2005.
115. A. S. Matveev and A. V. Savkin. Multi-rate stabilization of linear multiple sensor systems via limited capacity communication channels. *SIAM Journal on Control and Optimization*, 44(2):584–618, 2005.
116. A. S. Matveev and A. V. Savkin. Zero error capacity as the border of the domain of almost sure observability over noisy channels. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 3219–3224, Seville, Spain, 2005.
117. A. S. Matveev and A. V. Savkin. On a problem related to application of digital networked communication technology to stabilization of noisy plants over noisy channels. In *Proceedings of the IEEE Conference on Control Applications*, pages 2072–2077, Munich, Germany, October 2006.
118. A. S. Matveev and A. V. Savkin. Optimal state estimation in networked systems with asynchronous communication channels and switched sensors. *Journal of Optimization Theory and Applications*, 128(1):1–27, 2006.
119. A. S. Matveev and A. V. Savkin. Stabilization of stochastic linear plants via limited capacity stochastic communication channels. In *Proceedings of the 45th IEEE Conference on Decision and Control*, San Diego, C.A., December 2006.

120. A. S. Matveev and A. V. Savkin. An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels. *SIAM Journal on Control and Optimization*, 46(4):1323–1361, 2007.
121. A. S. Matveev and A. V. Savkin. Shannon zero error capacity in the problems of state estimation and stabilization via noisy communication channels. *International Journal of Control*, 80(2):241–255, 2007.
122. A. S. Matveev and A. V. Savkin. Decentralized stabilization of networked systems under data-rate constraints. In *Proceedings of the 17th IFAC World Congress*, Seoul, South Korea, July 2008.
123. P. S. Maybeck. *Stochastic Models, Estimation, and Control*, volume 3. Academic Press, New York, 1982.
124. N. H. McCoy. *Fundamentals of Abstract Algebra*. Allyn and Bacon Inc., Boston, 1974.
125. J. S. Meditch. *Stochastic Optimal Linear Estimation and Control*. McGraw-Hill, New York, 1969.
126. B. M. Miller and W. J. Runggaldier. Optimization of observations: a stochastic control approach. *SIAM Journal on Control and Optimization*, 35(5):1030–1052, 1997.
127. S. K. Mitter. Control with limited information: the role of systems theory and information theory. In *ISIT 2000 Plenary Talk, IEEE Information Theory Society Newsletter*, December 2000.
128. L. A. Montestruque and P. Antsaklis. Stability of model-based networked control systems with time-varying transmission times. *IEEE Transactions on Automatic Control*, 49(9):1562–1572, 2004.
129. J. B. Moore. Fixed-lag smoothing results for linear dynamical systems. *Australian Telecommunication Research*, 7(2):16–21, 1973.
130. K. M. Nagpal and P. P. Khargonekar. Filtering and smoothing in an H^∞ setting. *IEEE Transactions on Automatic Control*, AC-36(2):152–166, 1991.
131. N. E. Nahi. Optimal recursive estimation with uncertain observations. *IEEE Transactions Information Theory*, 15:457–462, 1969.
132. G. N. Nair and R. J. Evans. State estimation via a capacity-limited communication channel. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 866–871, San Diego, C.A., December 1997.
133. G. N. Nair and R. J. Evans. State estimation via a limited communication channel. In *Proceedings of the 37th Conference on Decision and Control*, pages 251–256, Tampa, F.L., 1998.
134. G. N. Nair and R. J. Evans. A finite-dimensional coder-estimator for rate-constrained state estimation. In *Proceedings of the 14th IFAC World Congress*, pages 19–24, Beijing, China, July 1999.
135. G. N. Nair and R. J. Evans. Stabilization with data-rate-limited feedback: tightest attainable bounds. *Systems and Control Letters*, 41(1):49–56, 2000.
136. G. N. Nair and R. J. Evans. Mean square stabilisability of stochastic linear systems with data rate constraints. In *Proceedings of the 41th IEEE Conference on Decision and Control*, pages 1632–1637, Las Vegas, N.V., December 2002.
137. G. N. Nair and R. J. Evans. Exponential stabilisability of finite-dimensional linear systems with limited data rates. *Automatica*, 39(4):585–593, 2003.
138. G. N. Nair and R. J. Evans. Stabilisability of stochastic linear systems with finite feedback data rates. *SIAM Journal on Control and Optimization*, 43(2):413–436, 2004.
139. G. N. Nair, R. J. Evans, and P. E. Caines. Stabilizing decentralized linear systems under data rate constraints. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 3992–3997, Atlantis, Bahamas, December 2004.

140. G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran. Topological feedback entropy and nonlinear stabilization. *IEEE Transactions on Automatic Control*, 49(9):1585–1597, 2004.
141. G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran. Topological feedback entropy for nonlinear systems. In *Proceedings of the 5th Asian Control Conference*, Melbourne, Australia, July 2004.
142. G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans. Feedback control under data rate constraints: an overview. *Proceedings of the IEEE*, 95(1):108–137, 2007.
143. G. N. Nair, M. Huang, and R. J. Evans. Optimal infinite horizon control under a low data rate. In *Proceedings of the 14th IFAC Symposium on System Identification*, Newcastle, Australia, March 2006.
144. K. S. Narendra and J. H. Taylor. *Frequency Domain Criteria for Absolute Stability*. Academic Press, New York, 1973.
145. J. Neveu. *Mathematical Foundations of the Calculus of Probabilities*. Holden-Day, San Francisco, C.A., 1965.
146. J. Nilsson. *Real-time control systems with delays*. PhD thesis, Department of Automatic Control, Lund Institute of Technology, 1998.
147. M. J. Panik. *Fundamentals of Convex Analysis: Duality, Separation, Representation, and Resolution*. Kluwer Academic Publishers, Boston, M.A., 1993.
148. I. R. Petersen and A. V. Savkin. *Robust Kalman Filtering for Signals and Systems with Large Uncertainties*. Birkhauser, Boston, M.A., 1999.
149. I. R. Petersen and A. V. Savkin. Multi-rate stabilization of multivariable discrete-time linear systems via a limited capacity comunication channel. In *Proceedings of the 40th IEEE Conference on Decision and Control*, pages 304–309, Orlando, F.L., USA, 2001.
150. I. R. Petersen and A. V. Savkin. State feedback stabilization of uncertain discrete-time linear systems via a limited capacity communication channel. In *Proceedings of the 2002 Conference on Information, Decision and Control*, Adelaide, Australia, February 2002.
151. I. R. Petersen, V. A. Ugrinovskii, and A. V. Savkin. *Robust Control Design Using H^∞ Methods*. Springer-Verlag, London, 2000.
152. M. S. Pinsker. Information and information stability of random variables and processes. In G.M. Jenkins and E. Parzen, editors, *Holden-Day Series in Time Series Analysis*. Holden-Day, INC., San Francisco, C.A., 1964.
153. P. Pleinevaux and J.-D. Decotignie. Time critical communication networks: fieldbuses. *IEEE Network*, 2(3):55–63, 1988.
154. M. Pollicott and M. Yuri. *Dynamical Systems and Ergodic Theory*. Cambridge University Press, Cambridge, M.A., 1998.
155. C. Rago, P. Willet, and Y. Bar-Shalom. Censoring sensor: a low communication rate scheme for distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, 28(2):554–568, 1996.
156. R. Rajamani and Y. M. Cho. Existence and design of observers for nonlinear systems: Relation to distance to unobservability. *International Journal on Control*, 69(5):717–731, 1998.
157. C. R. Rao. *Generalized Inverse of Matrices and its Applications*. J. Wiley & Sons, New York, 1971.
158. N. Ratnakar and G. Kramer. The multicast capacity of deterministic relay networks with no interference. *IEEE Transactions on Information Theory*, 52(6):2425–2432, 2006.
159. J. H. Reif and H. Wang. Social potential fields: A distributed behavioral control for autonomous robots. *Robotics and Autonomous Systems*, 27:171–194, 1999.

160. W. Ren and R. W. Beard. *Distributed Consensus in Multi-vehicle Cooperative Control*. Springer-Verlag, London, 2008.
161. H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Optimization Methods in Statistics*, pages 233–257. Academic Press, New York, 1971.
162. V. A. Rokhlin. Entropy of metric authomorphism. *Doklady Akademii Nauk SSSR*, 124(5):980–983, 1959.
163. B. A. Sadjadi. Stability of networked control systems in the presence of packet losses. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, Maui, H.I., December 2003.
164. A. Sahai. *Anytime Information Theory*. PhD thesis, Massachusetts Institute of Technology, February 2001.
165. A. Sahai. The necessity and sufficiency of anytime capacity for control over a noisy communication channel. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 1896–1901, Atlantis, Bahamas, December 2004.
166. A. Sahai and S. Mitter. The necessity and sufficiency of anytime capacity for stabilization of a linear system over noisy communication links—Part I: Scalar systems. *IEEE Transactions on the Information Theory*, 52(8):3369–3395, 2006.
167. A. Sahai, S. Tatikonda, and S.Mitter. Control of LQG systems under communication constraints. In *Proc. of the 1999 American Control Conference*, volume 4, pages 2778–2782, 1999.
168. A. V. Savkin. A new approach to design of networked control systems: H^∞ state estimation via limited capacity communication channels. *International Journal of Hybrid Systems*, 3(1):1–8, 2003.
169. A. V. Savkin. Coordinated collective motion of groups of autonomous mobile robots: Analysis of Vicsek's model. *IEEE Transactions on Automatic Control*, 49(6):981–983, 2004.
170. A. V. Savkin. Detectability and output feedback stabilizability of nonlinear networked control systems. In *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference*, pages 8174–8178, Seville, Spain, 2005.
171. A. V. Savkin. Analysis and synthesis of networked control systems: topological entropy, observability, robustness, and optimal control. *Automatica*, 42(1):51–62, 2006.
172. A. V. Savkin and T. M. Cheng. Detectability and output feedback stabilizability of nonlinear networked control systems. *IEEE Transactions on Automatic Control*, 52(4):730–735, 2007.
173. A. V. Savkin and R. J. Evans. A new approach to robust control of hybrid systems over infinite time. *IEEE Transactions on Automatic Control*, 43(9):1292–1296, 1998.
174. A. V. Savkin and R. J. Evans. *Hybrid Dynamical Systems. Controller and Sensor Switching Problems*. Birkhauser, Boston, M.A., 2002.
175. A. V. Savkin, R. J. Evans, and E. Skafidas. The problem of optimal robust sensor scheduling. In *Proceedings of the 39th IEEE Conference on Decision and Control*, pages 3791–3796, Sydney, Australia, December 2000.
176. A. V. Savkin, R. J. Evans, and E. Skafidas. The problem of optimal robust sensor scheduling. *Systems and Control Letters*, 43(2):149–157, 2001.
177. A. V. Savkin and A. S. Matveev. Cyclic linear differential automata: a simple class of hybrid dynamical systems. *Automatica*, 36(5):727–734, 2000.
178. A. V. Savkin and I. R. Petersen. A connection between H^∞ control and the absolute stabilizability of uncertain systems. *Systems and Control Letters*, 23(3):197–203, 1994.

179. A. V. Savkin and I. R. Petersen. Minimax optimal control of uncertain systems with structured uncertainty. *International Journal of Robust and Nonlinear Control*, 5(2):119–137, 1995.
180. A. V. Savkin and I. R. Petersen. Nonlinear versus linear control in the absolute stabilizability of uncertain linear systems with structured uncertainty. *IEEE Transactions on Automatic Control*, 40(1):122–127, 1995.
181. A. V. Savkin and I. R. Petersen. Recursive state estimation for uncertain systems with an integral quadratic constraint. *IEEE Transactions on Automatic Control*, 40(6):1080, 1995.
182. A. V. Savkin and I. R. Petersen. An uncertainty averaging approach to optimal guaranteed cost control of uncertain systems with structured uncertainty. *Automatica*, 31(11):1649–1654, 1995.
183. A. V. Savkin and I. R. Petersen. Model validation for robust control of uncertain systems with an integral quadratic constraint. *Automatica*, 32(4):603–606, 1996.
184. A. V. Savkin and I. R. Petersen. Set-valued state estimation via a limited capacity communication channel. *IEEE Transactions on Automatic Control*, 48(4):676–680, 2003.
185. A. V. Savkin, I. R. Petersen, E. Skafidas, and R.J. Evans. Hybrid dynamical systems: Robust control synthesis problems. *Systems and Control Letters*, 29(2):81–90, 1996.
186. Y. Sawaragi, T. Katayama, and S. Fujishige. Sequential state estimation with interrupted observation. *Information and Control*, 21:56–71, 1972.
187. S. Sengupta. Analysis of communication losses in vehicle control problems. In *Proceedings of the 2001 American Control Conference*, pages 1491–1496, Arlington, V.A., 2001.
188. C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 623–656, 1948.
189. C. E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, IT-2:8–19, 1956.
190. C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, I.L., 1949.
191. M. Simonnet. *Measures and Probability*. Springer, New York, 1996.
192. T. Simsek, R. Jain, and P. Varaiya. Scalar estimation and control with noisy binary observation. *IEEE Transactions on Automatic Control*, 49(9):1598–1603, 2004.
193. T. Simsek and P. Varaiya. Noisy data-rate estimation: renewal codes. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 3148–3154, Maui, H.I., December 2003.
194. B. Sinopoli, L. Schenato, M. Franceschetti, K. Poola, and Sh. Sastry. Kalman filtering with intermittent observations. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, Maui, H.I., December 2003.
195. B. Sinopoli, L. Schenato, M. Franceschetti, K. Poola, and Sh. Sastry. Time varying optimal control with packet losses. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 1938–1943, Atlantis, Bahamas, December 2004.
196. B. Sinopoli, C. Sharp, S. Schaffert, L. Schenato, and Sh. Sastry. Distributed control applications within sensor networks. *IEEE Proceedings. Special Issue on Distributed Sensor Networks*, 2003.
197. M. W. Spong. Modeling and control of elastic joint robots. *Transactions of the ASME*, 109:310–319, 1987.
198. R. F. Stengel. *Stochastic Optimal Control: Theory and Application*. J. Wiley & Sons, New York, 1986.
199. D. J. Stilwell and B. E. Bishop. Platoons of underwater vehicles. *IEEE Control Systems Magazine*, 20(6):45–52, 2000.

200. C. Striebel. Sufficient statistics in the optimum control of stochastic systems. *Journal of Mathematical Analysis and Applications*, 12:576–592, 1965.
201. S. Tatikonda and S. Mitter. Control over noisy channels. *IEEE Transactions on Automatic Control*, 49(7):1196–1201, 2004.
202. S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7):1056–1068, 2004.
203. S. Tatikonda, A. Sahai, and S. Mitter. Stochastic linear control over a communication channel. *IEEE Transactions on Automatic Control*, 49(9):1549–1561, 2004.
204. S. C. Tatikonda. *Control Under Communication Constraints*. PhD thesis, Massachusetts Institute of Technology, 2000.
205. S. C. Tatikonda. Some scaling properties of large distributed control systems. In *Proceedings of the 42nd Conf. on Decision and Control*, pages 3142–3147, Maui, H.I., December 2003.
206. S. C. Tatikonda, A. Sahai, and S. K. Mitter. Control of LQG systems under communication constraints. In *Proceedings of the 37th IEEE Conference on Decision and Control*, pages 1165–1170, Tampa, F.L., 1998.
207. V. P. Telang and M. A. Herro. Error control coding for the n-user $\mod -2$ mulpitle-access channel. *IEEE Transactions on Information Theory*, 44:1632–1641, 1998.
208. C. Tepedelenlioglu and G. Giannakis. Transmitter redundancy for blind estimation and equalization of time- and frequency - selective channels. *IEEE Transactions on Signal Processing*, 48(7):2029–2042, 2000.
209. N. Tsai and A. Ray. Stochastic optimal control under randomly varying delays. *International Journal of Control*, 69:1179–1202, 1997.
210. N. Tsai and A. Ray. Compensatability and optimal compensation under randomly varying distributed delays. *International Journal of Control*, 72(9):826–832, 1999.
211. A. van der Schaft and H. Schumacher. *An Introduction to Hybrid Dynamical Systems*. Springer-Verlag, London, 1999.
212. P. Vanroose. Code construction for the noiseless binary switching multiple-access channel. *IEEE Transactions on Information Theory*, 34(5):1100–1106, 1988.
213. R. S. Varga. *Matrix Iterative Analysis*. Prentice Hall, Englewood Cliffs, N.J., 1962.
214. S. Verdú. Fifty years of Shannon theory. *IEEE Transactions on Information Theory*, 44(6):2057–2078, 1998.
215. E. Verriest and M. Egerstedt. Control with delayed and limited information: a first look. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 1231–1236, Las Vegas, N.V., December 2002.
216. T. Ward. *Entropy of Compact Group Automorphisms*. Available online at www.mth.uea.ac.uk/h720/lecture_notes.
217. J. Warga. *Optimal Control of Differential and Functional Equations*. Academic Press, New York and London, 1970.
218. N. Wiener. *Cybernetics*. MIT Press, Cambridge, M.A., 1948.
219. H. S. Witsenhausen. Separation of estimation and control for discrete time systems. *Proceedings of IEEE*, 59(11):1557–1566, 1971.
220. W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints - Part 1. State estimation problems. *IEEE Transactions on Automatic Control*, 42(9):1294–1299, 1997.
221. W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints - part 2: stabilization with limited information feedback. *IEEE Transactions on Automatic Control*, 44:1049–1053, 1999.
222. V. A. Yakubovich. *The Methods of Research of Nonlinear Control Systems*, chapter 2-3. Nauka, Moscow, 1975.

223. V. A. Yakubovich. Absolute stability of nonlinear systems with a periodically nonstationary linear part. *Soviet Physics Doklady*, 32(1):5–7, 1988.
224. V. A. Yakubovich, G. A. Leonov, and A. Kh. Gelig. *Stability of Stationary Sets in Control Systems with Discontinuous Nonlinearities*. World Scientific, Singapore, 2004.
225. M. Yu, L. Wang, T. Chu, and F. Hao. An LMI approach to networked control systems with data packet dropout and transmission delays. *International Journal of Hybrid Systems*, 3:291–303, 2003.
226. M. Yu, L. Wang, T. Chu, and G. Xie. Stabilization of networked control systems with data packet dropout and network delay via switching system approach. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 3539–3544, Atlantis, Bahamas, December 2004.
227. S. Yüksel and T. Başar. Quantization and coding for decentralized LTI systems. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 2847–2852, Maui, H.I., December 2003.
228. S. Yüksel and T. Başar. Minimum rate coding for state estimation over noiseless channels. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 4503–4508, Atlantis, Bahamas, December 2004.
229. S. Yüksel and T. Başar. Coding and control over discrete noisy forward and feedback channels. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 2517–2522, Seville, Spain, December 2005.
230. C. Zhang, R. Middleton, and R. J. Evans. An algorithm for multirate sampling adaptive control. *IEEE Transactions on Automatic Control*, 37(7):792–796, 1989.
231. W. Zhang, M. Branicky, and S. Phillips. Stability of networked control systems. *IEEE Control Systems Magazine*, 21(2):84–99, 2001.
232. R. Ziegler and J. Cioffi. Estimation of time-varying digital radio channels. *IEEE Transactions on Vehicular Technology*, 41:134–151, 1992.
233. K. S. Zigangirov. Upper bounds on the error probability for channels with feedback. *Problems of Information Transmission*, 6(2):87–92, 1970.

Index