KTH Electrical Engineering

Control over Low-Rate Noisy Channels

LEI BAO

Doctoral Thesis in Telecommunications
Stockholm, Sweden 2009

TRITA-EE 2009:032
ISSN 1653-5146
ISBN 978-91-7415-364-4

KTH, School of Electrical Engineering
Communication Theory Laboratory
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i telekommunikation den 16 June, 2009 klockan 13:15 i hörsal Q1, Osquldasvägen 4, Stockholm.

Tryck: Universitetsservice US AB

Abstract

Networked embedded control systems are present almost everywhere. A recent trend is to introduce radio communication in these systems to increase mobility and flexibility. Network nodes, such as the sensors, are often simple devices with limited computing and transmission power and low storage capacity, so an important problem concerns how to optimize the use of resources to provide sustained overall system performance. The approach to this problem taken in the thesis is to analyze and design the communication and control application layers in an integrated manner. We focus in particular on cross-layer design techniques for closed-loop control over non-ideal communication channels, motivated by future control systems with very low-rate and highly quantized sensor communication over noisy links. Several fundamental problems in the design of source–channel coding and optimal control for these systems are discussed.

The thesis consists of three parts. The first and main part is devoted to the joint design of the coding and control for linear plants, whose state feedback is transmitted over a finite-rate noisy channel. The system performance is measured by a finite-horizon linear quadratic cost. We discuss equivalence and separation properties of the system, and conclude that although certainty equivalence does not hold in general it can still be utilized, under certain conditions, to simplify the overall design by separating the estimation and the control problems. An iterative optimization algorithm for training the encoder–controller pairs, taking channel errors into account in the quantizer design, is proposed. Monte Carlo simulations demonstrate promising improvements in performance compared to traditional approaches.

In the second part of the thesis, we study the rate allocation problem for state feedback control of a linear plant over a noisy channel. Optimizing a time-varying communication rate, subject to a maximum average-rate constraint, can be viewed as a method to overcome the limited bandwidth and energy resources and to achieve better overall performance. The basic idea is to allow the sensor and the controller to communicate with a higher data rate when it is required. One general obstacle of optimal rate allocation is that it often leads to a non-convex and non-linear problem. We deal with this challenge by using high-rate theory and Lagrange duality. It is shown that the proposed method gives a good performance compared to some other rate allocation schemes.

In the third part, encoder–controller design for Gaussian channels is addressed. Optimizing for the Gaussian channel increases the controller complexity substan-

tially because the channel output alphabet is now infinite. We show that an efficient controller can be implemented using Hadamard techniques. Thereafter, we propose a practical controller that makes use of both soft and hard channel outputs.

Keywords: sensor networks, linear quadratic cost, stochastic control, joint source–channel coding, joint coding and control, rate allocation, soft source decoding.

Acknowledgments

I take the opportunity to acknowledge all those who have supported me in the development of this thesis.

First and foremost, I would like to express my sincere gratitude to my supervisors, Professor Mikael Skoglund and Professor Karl Henrik Johansson. I own many thanks to Mikael for your guidance and your support throughout the years. I very much appreciate for introducing me to this joint field of communication and control, and for the freedom and time to explore and experiment different research ideas. I'm grateful to Kalle for your belief in me and your continuous encouragement. I am truly grateful for many insightful remarks on this thesis. Your passion for science and research work has always been so inspiring to me. Special thanks are also extended to Dr. Carlo Fischione for all the valuable discussions, and for the collaboration leading to joint publications.

I want to thank all my current and former colleagues at the laboratories Communication Theory, Signal Processing and Automatic Control, for the wonderful working environment, for many interesting discussions, and for all the joyful trips together. I would like to extend my thanks to all of you who have proof read parts of the manuscript of this thesis; especially, Hieu Do, Jinfeng Du, Shengqian Han, Johannes Karlsson, Yi Liu, Zhanyu Ma, Tobias Oechtering, Zhongwei Si, Ming Xiao, Sha Yao, Guoqiang Zhang, and Xi Zhang. I thank Annika Augustsson, Karin Demin, Karin Karlsson, and Tetiana Viekhova for helping me with administrative issues, the computer support term for the aid of computer issues, and especially, Johannes and Xi for making MATLAB and LATEX run smoothly. I'm in debt to all my fellow friends in the Q-building for many happy lunch and dinner hours.

I wish to thank Professor Sandro Zampieri who acts as the opponent on this thesis, and Professor Thomas Eriksson, Professor Magnus Jansson, and Professor Erik Larsson for participating in the committee.

Last but not least, I want to express my greatest gratitude to my dearest mom and Ling for your love throughout the years.

Lei Bao
Stockholm, May 2009

Contents

Chapter 1

Introduction

1.1 Background

Networked embedded control systems are present almost everywhere. Application areas include industrial automation, aerospace and medical systems, as well as consumer electronics such as home electronics and mobile phones. The systems are often connected through either wired (e.g., wired local area networks) or wireless (e.g., bluetooth) communication technologies. A consequence of the rapidly growing number of connected components (nodes) is the increasing demand for efficient sharing of resources. Integrating technological advancements in sensing, communication, computation and control has brought up many engineering challenges, such as finding efficient ways of processing available information at each distributed node, as well as exchanging useful information among the nodes.

Traditional communication theory has been mainly focused on optimal strategies for transmitting information, while traditional control theory provides methodologies for designing controllers to interact with the environment. Until recently, the research work in these two disciplines has largely been carried out separately. A traditional control system is based on an underlying assumption of perfect communication links between the plant and the controller, e.g., [Aok67, Ber76]. The controller is assumed to have perfect access to the sensor observations, and the decision of the controller is available directly at the input of the actuator. Under these ideal assumptions, there is no limitation on how much data it is possible to transmit at each time instant, and there are neither delays nor transmission errors in the links between the plant and the controller. Advanced mathematical tools are developed to govern the interplay among the plant, sensor, and controller under these ideal assumptions.

A recent trend is to perform control using wireless sensor networks, which takes advantage of the mobility and the flexibility offered by wireless solutions, e.g., [LG04, PKA04, SLT04, DLGH05, WPJ+07]. In such networks, the sensor observations are typically quantized and transmitted over noisy links. Challenges, such as data delays and data drops, are encountered. Concerning control over non-ideal

communication links, relatively little work has been performed so far. To develop methods and tools for the analysis and synthesis of feedback control over imperfect communication links is therefore of great importance.

The constraints imposed by the imperfect communication links are complex. As discussed above, quantization and transmission errors are examples of crucial obstacles. The quantization deteriorates the signals transmitted between the plant and the controller. This can potentially degrade the overall system performance substantially. Although quantization in feedback control systems was studied since the dawn of control engineering, the results have mainly been restricted to treating quantization errors as additive white noise. Moreover, in almost all applications, simple quantizers, such as uniform quantizers, are employed because of practical reasons. However, for applications with extremely low data rate requirements and high communication costs, it is natural to study closer-to-optimal solutions.

Transmission errors are unavoidable in communications over unreliable media, for example in wireless networks. Therefore, robustness to transmission errors is one of the fundamental requirements of all modern communication systems. Concerning control applications, relatively little has been done to take into account imperfect communications in the overall system design. However, due to the delay sensitivity, it is not suitable to use long block codes to reduce the uncertainties, as commonly done in traditional communication systems. When facing the constraint on the codeword length, a joint design which combines the source compression and the channel protection is expected to achieve satisfactory performance. One of the main objectives of this thesis is to study the joint design of coding and control for an efficient use of the limited communication resources.

1.2 Motivating Examples

Before presenting the specific control problem studied in this thesis, let us first consider two examples of sensor networks in control applications: (i) motion control of non-contact objects in Example 1.2.1, and (ii) environmental control of greenhouses in Example 1.2.2.

Example 1.2.1. ***Motion Control of Non-Contact Objectives***

In industrial production, there are situations when fragile materials need to be transported by using non-contact methods. Figure 1.1 illustrates an example, inspired by [BBCea00], where a planar object is transported by a number of air jets. To eliminate the potential disturbances, such as external forces, the air jets can provide alternative air beams of varied angles and forces. The position of the object is monitored by spatially distributed sensor nodes. As long as the system is working under the normal condition, that is to say, there is no disturbance detected, the same set of air beams are applied on the object. On the other hand, once a disturbance is detected, a new set of air beams will be selected, based on the sensor measurements. The control objective is often to achieve a satisfactory performance on average, over

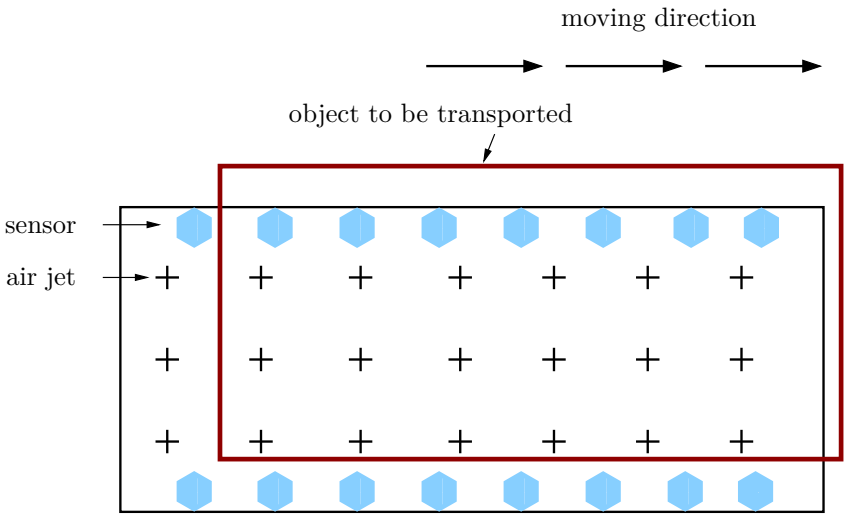


Figure 1.1: An example of a sensor network in motion control of a non-contact object in industrial manufacturing.

a finite time horizon. The questions, such as how to deploy the sensor network and how to utilize the sensor measurements, will be part of the controller design.

Example 1.2.2. ***Environmental Control of Greenhouses***

The following example, inspired by [Jon01], illustrates the future use of control over sensor networks in agricultural applications. The new generation of greenhouses will support automatic environmental regulations. Figure 1.2 depicts such a greenhouse, which exploits a feedback control system to maintain a perfect growing environment for plants. More precisely, the "perfect" environment is specified by a number of primary variables, such as temperature, humidity, light, CO_2 levels etc. The control task is to keep the primary variables within certain ranges provided by the grower. Typically, these values are altered continuously by complicated biological and chemical processes, both inside and outside the greenhouse. As an example, the solar radiation directly and indirectly affects all the primary variables. The variables are monitored by the sensor nodes deployed over a large area. As soon as any value exceeds the predefined limit, the actuators, such as heaters, fans, illumination and irrigation equipments, will act automatically to regulate the environment into favorable conditions.

Above, we have given two specific examples of using sensor networks in control applications. To have a general picture of the problems particularly studied in this thesis, let us consider the wireless networked control system depicted in Figure 1.3. The system consists of a number of sensor nodes that are connected

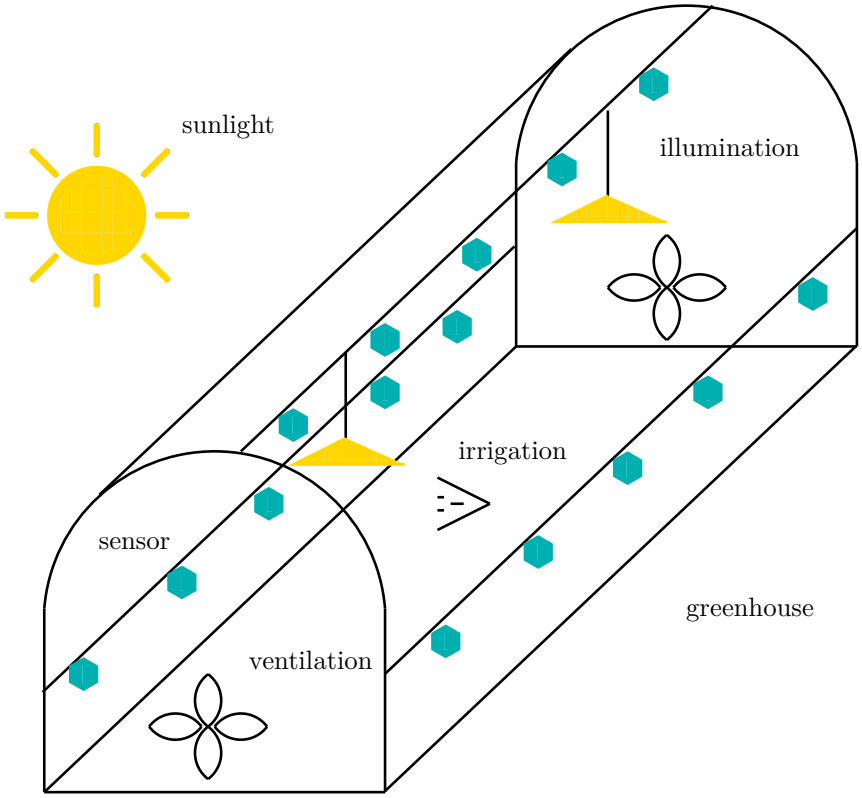


Figure 1.2: An example of a sensor network in environmental control of greenhouse vegetable production.

through a shared wireless medium to a central control node. The sensors are spatially distributed over a large area and they measure the state of a control object, which is affected by local disturbances. The control commands for keeping the states around the equilibrium working points are executed through either a common actuator as shown in Figure 1.3, or distributed actuators. The described system is quite representative for many emerging applications, with control using wireless sensor networks, as can be found in industrial automation, environmental monitoring, surveillance etc.

An important application of the work in this thesis is that wireless networked control systems in future can operate at inaccessible places. They could be deployed in remote locations, such as underwater, outer space; or deployed in extremely harsh environments, such as close to explosive or radioactive sources, or imbedded in the walls of skyscrapers, or even imbedded in human bodies. The monitoring and control tasks are performed by a large set of sensor nodes, even tens of thousands

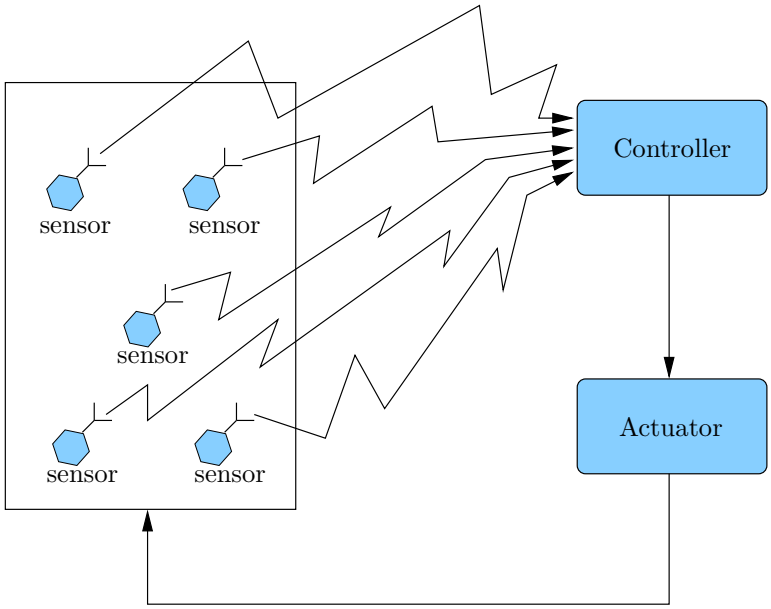


Figure 1.3: A control system utilizes data from wireless sensor network. The results of this thesis can be applied to optimize the use of the wireless medium.

[ZL08], over a widely spread area (in proportion to the sensor size). After the deployment, it may not be possible to maintain or recharge the network and its nodes manually for reasons such as the large number of nodes or the physical location of the network. A major challenge is to design a sustainable network which is highly energy efficient, since the life time of such a network is determined by battery-powered devices [GW02]. In many applications, the sensor nodes will be simple, inexpensive and small, which means that the computing power and the storage capacity are highly limited. To utilize the limited communication resources efficiently, it is especially interesting to study the case where the transmission from each sensor is limited to to a few possible symbols and each symbol consists of a few bits. Such simple transmission schemes are motivated in particular from the point of view of energy efficiency and delay requirements. Constraining the resolution per transmission allows for low-energy transmission, and avoiding coding over multiple channel uses results in low latency. This requires new types of protocols which work directly on the bit-level and perform joint low-delay coding (compression and/or quantization and error control) and control.

Motivated by the above-mentioned challenges this thesis studies serval fundamental problems in the area of control with highly limited information over imperfect communication channels. In particular, a stochastic control problem is formulated where the sensor measurements are conveyed to the controller through a

low-rate noisy channel. The optimization of the joint encoder–controller and the optimization of rate allocation over time are the two main topics studied in this thesis. Before proceeding, in the remaining part of this chapter we give a short introduction to related research topics in stochastic control and coding over noisy channels.

1.3 Stochastic Control

Stochastic control theory deals with the analysis and synthesis of controllers for dynamic systems subject to stochastic disturbances. Solutions to stochastic control problems rely heavily on *dynamic programming*. A brief introduction to the concept and techniques of dynamic programming is given in Section 1.3.1. Besides dynamic programming, we will see later that estimation theory also plays an important role in stochastic control. We will then in Section 1.3.2 present the *certainty equivalence property*, which describes the separation of state estimation and control. There exists a rich body of introductory literature in stochastic control theory, e.g., [Aok67, Åst70, Ath71, Ber95, BS96, Söd02].

1.3.1 Dynamic Programming

It is well-known that a decision which is optimal for the current time instant in most cases is not necessarily the best one for the future evolution. Dynamic programming captures this fact and provides decisions based on a combined cost of the current state and the expected future states.

A general stochastic control problem is formulated below. The state space model of the plant is given by

$$\begin{aligned} x_{t+1} &= F_t(x_t, u_t, v_t), \\ y_t &= G_t(x_t, e_t), \end{aligned}$$

where F_t is the system function and G_t is the measurement function at time t. The variables x_t, u_t, y_t represent the state, the control and the measurement, respectively. Finally, v_t and e_t denote the process noise and the measurement noise. The subscript t is a time index.

The design goal is to find the optimal control sequence, which minimizes an objective function $\mathbf{E}\{J_{tot}\}$, with J_{tot} given by

$$J_{tot} = \sum_{t=0}^{T} J_t(x_t, y_t, u_t).$$

Here, T denotes a finite time horizon and J_t denotes a function measuring the instantaneous system performance. Finally, $\mathbf{E}\{\cdot\}$ is the expectation operator.

Let us first review some results [Åst70, Söd02] that are useful for the studied optimization problem. For brevity, the time index t will be ignored for the moment. When having the complete state information, i.e., both x and y are available, the

following result has been proved: assuming the function $J(x, y, u)$ has a unique minimum as a function of the control input, at $u^\star(x, y)$ (throughout the thesis starred entities, *, denote optima), the minimization and the expectation are commutative, i.e.,

$$\min_{u(x,y)} \mathbf{E}\{J(x, y, u)\} = \mathbf{E}\{J(x, y, u^\star(x, y))\} = \mathbf{E}\left\{\min_u J(x, y, u)\right\}.$$

On the other hand, when only y is available, i.e., the incomplete state information scenario, one can show that

$$\min_{u(y)} \mathbf{E}\{J(x, y, u)\} = \mathbf{E}\{J(x, y, u^\star(y))\} = \mathbf{E}_y\left\{\min_u \mathbf{E}\{J(x, y, u)|\, y\}\right\},$$

where $\mathbf{E}\{J(x, y, u)|\, y\}$ is assumed having a unique minimum at $u^\star(y)$. Obviously, the complete state information scenario can be considered as a special case of the incomplete state information scenario.

Dynamic programming is based on the *principle of optimality*. The intuitive idea is that a truncation of the optimal control sequence, $\{u_t^\star, \ldots, u_T^\star\}$, is also the optimal policy for the truncated problem whose cost is a summation from time t to T. Let the boldface notation $\mathbf{x}_s^t$ describe a sequence, i.e., $\mathbf{x}_s^t = \{x_s, \ldots, x_t\}$, $s, t \in \mathbb{Z}^+$, $s \leq t$, with $\mathbb{Z}^+$ denoting the set of non-negative integers. Based on the principle of optimality, the optimal u_t is the one that minimizes a sum of the future costs:

$$u_t^\star = \arg\min_{u_t} \mathbf{E}\left\{\sum_{s=t}^{T} J_s(x_s, y_s, u_s)\Big|\mathbf{y}_0^t\right\},$$

where $\mathbf{y}_0^t$ represents all past measurements. Introduce the optimal *"cost-to-go"* function at time t, such as

$$\gamma_t^\star \triangleq \min_{\mathbf{u}_t^T} \mathbf{E}\left\{\sum_{s=t}^{T} J_s(x_s, y_s, u_s)\Big|\mathbf{y}_0^t\right\},$$

which is associated to the optimal control sequence $\{u_t^\star, \ldots, u_T^\star\}$. The optimal control function $u_t^\star$ at time t can then be derived by solving

$$\gamma_t^\star = \min_{u_t} \mathbf{E}\left\{ J_t(x_t, y_t, u_t) + \gamma_{t+1}^\star \big|\, \mathbf{y}_0^t\right\}.$$

1.3.2 Certainty Equivalence

A certainty equivalence (CE) controller [The57, BST74, TBS75, WW81] is obtained by replacing the full state observation x_t in the optimal deterministic solution, where process noise is absent and the perfect state observations are available, with a state estimate, more specifically,

$$u_t^{OpD} = g_t^{det}(x_t), \tag{1.1}$$

$$u_t^{CE} = g_t^{det}(\hat{x}_t), \tag{1.2}$$

where $g_t^{det}(\cdot)$ denotes the optimal deterministic solution and

$$\hat{x}_t \triangleq \mathbf{E}\left\{ x_t | \, \mathbf{y}_0^t, \mathbf{u}_0^{t-1} \right\},$$

is the conditional mean estimate based on all past measurements and control inputs. In the case that u_t is completely determined by $\mathbf{y}_0^t$, we also write $\hat{x}_t = \mathbf{E}\{ x_t | \, \mathbf{y}_0^t \}$. The equation (1.2) exposes a clear separation between the estimation and the control, since the estimate $\hat{x}_t$ is computed separately and then applied to derive the control input.

Sometimes, a CE controller is also the optimal control policy, which is often termed *certainty equivalence property*. However, a CE controller is in general only a suboptimum solution, with a few exceptions. The most well-known exception is the linear quadratic Gaussian (LQG) control problem [Aok67, Ath71, Ber76], a classical example with a linear plant, a quadratic cost, and Gaussian distributed uncertainties, i.e., initial-state, process noise, and measurement noise. Furthermore, a so-called *classical information pattern* is required, which means that all past measurements are known to the controller. For this example, the separation in (1.2) applies and the optimal control is a linear function of the conditional mean estimate,

$$u_t^\star = \ell_t \mathbf{E}\left\{ x_t | \, \mathbf{y}_0^t, \mathbf{u}_0^{t-1} \right\} = \ell_t \hat{x}_t,$$

where the *linear feedback control law* ℓ_t is specified by the system parameters. Some applications of the LQG controller can be found in e.g., [KL99, PBC$^+$07, Loo09].

Finally, we introduce the *separation property*, by which $\hat{x}_t$ is a *sufficient statistic* [Aok67] to derive an optimal control, such as

$$u_t^\star = g_t^{sep}(\hat{x}_t).$$

Separation property is a weak notion of the CE property, since the optimal control $g_t^{sep}(\cdot)$ is not necessary the optimal deterministic solution $g_t^{det}(\cdot)$.

1.4 Elements of Source and Channel Coding

Recall the example in Figure 1.3. The signal path from a sensor to the controller can be modeled as a point-to-point communication link. In Figure 1.4, a block diagram of the elementary building blocks in a traditional communication system is given, where the *source coding* and *channel coding* units are presented briefly in Section 1.4.1 and Section 1.4.2.

1.4.1 Source Coding

Source coding deals with the compression of source data by removing the redundancy in the data sequence. Based on whether or not the original data sequence can be reconstructed, source coding techniques are divided into two categories. The first, referred to as *lossless coding*, commonly associated with discrete sources, requires the data processing to be reversible. For example, consider a discrete random

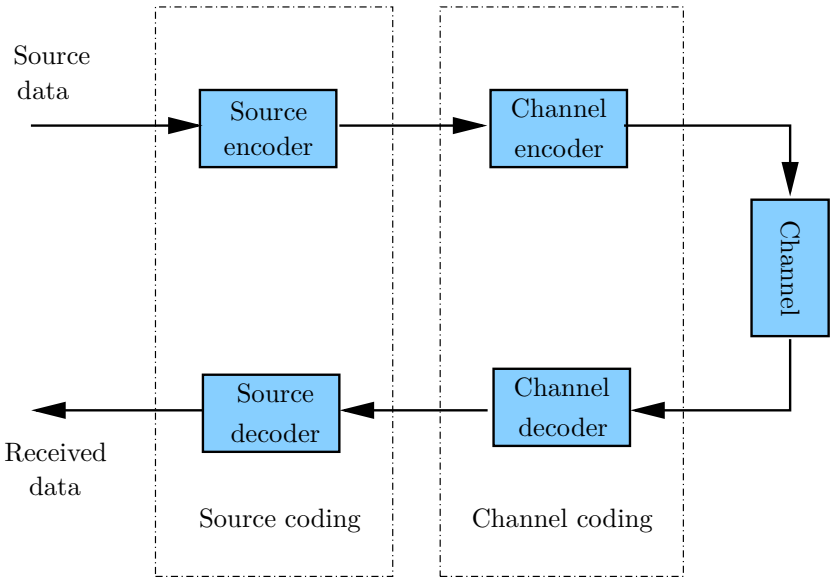


Figure 1.4: The functional diagram of a digital communication system.

variable x with alphabet $\mathcal{X}$. The *entropy* of x, which is a measure of uncertainty, is defined as

$$\mathcal{H}(x) = -\sum_{\mathcal{X}} \mathcal{P}(x) \log_2 \mathcal{P}(x),$$

with $\mathcal{P}(x)$ denoting the probability mass function. By using base 2 logarithm, the entropy $\mathcal{H}(x)$ is measured in bits. For a discrete stationary random process $\{x_n\}$, $x_n \in \mathcal{X}$, the average length of a uniquely decodable code is bounded by the *entropy rate* [CT91] of the source, which is defined by

$$\mathcal{H}_\infty = \lim_{m\to\infty} \frac{1}{m} \mathcal{H}(x_1, x_2, \ldots, x_m),$$

when the limit exists; and the joint entropy $\mathcal{H}(x_1, x_2, \ldots, x_m)$ is defined by

$$\mathcal{H}(x_1, \ldots, x_m) = -\sum_{x_1 \in \mathcal{X}} \cdots \sum_{x_m \in \mathcal{X}} \mathcal{P}(x_1, \ldots, x_m) \log \mathcal{P}(x_1, \ldots, x_m),$$

with $\mathcal{P}(x_1, \ldots, x_m)$ denoting a joint distribution.

In *lossy coding*, commonly associated with analog sources, the reversibility is not required. Assume now the source $\{x_n\}$, $x_n \in \mathbb{R}$, is a stochastic process of analog data. A *block source encoder* maps a sequence of source sample, $\mathbf{x}_1^m \in \mathbb{R}^m$, into an integer index $i \in \mathcal{L} = \{0, \ldots, 2^R - 1\}$, usually in its binary representation, with R denoting the *rate*. In order to discuss the quality of a source code, a distortion

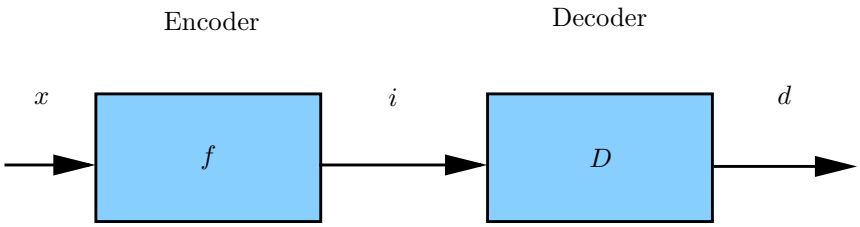


Figure 1.5: A source coding diagram, where x is the random source symbol, i is the integer index, and d is an estimate of x.

between the source sequence $\mathbf{x}_1^m$ and the reconstructed sequence $\mathbf{d}_1^m$, $J(\mathbf{x}_1^m, \mathbf{d}_1^m)$, should be specified. The distortion function J varies from application to application. In general, the main features of a proper distortion function are non-negativity, physical meaningfulness and ease of calculation. The conflicting relation between a given distortion and a minimum rate is stated in a rate distortion function, see e.g., [Gal68, CT91, Kle04], which is one of the fundamentals in lossy source coding.

Lossy source coding and quantization are two closely related terms. Quantization describes a process of approximating a large set of values into a smaller set of discrete symbols. Scalar quantizers operate on scalar data ($m = 1$), and vector quantizers operate on multidimensional vectors ($m > 1$). Figure 1.5 depicts the source coding over an error-free channel. Here, x is a scalar analog source sample. The encoder function $f(\cdot)$ performs the mapping of a source sample to an integer index: $i = f(x)$. The integer set $\mathcal{L} = \{0, 1, \ldots, 2^R - 1\}$, with R denoting the *rate*, contains all possible indices. (Throughout the thesis the transmitted and received indices share the same integer set $\mathcal{L}$, if nothing else is stated.) The *encoding cell* $\mathcal{S}(k)$, for the index value k, contains all source symbols assigned the index value k:

$$\mathcal{S}(k) \triangleq \{x : f(x) = k\}, \quad k \in \mathcal{L},$$

i.e., $x \in \mathcal{S}(k) \Leftrightarrow i = k$. The task of a decoding function $D(\cdot)$ is to reconstruct an estimate d of the source symbol x, based on the index i, i.e., $d = D(i)$. We define the *reconstruction* $d(k)$, which is associated with index value $i = k$ through

$$i = k \Rightarrow d = D(k) = d(k).$$

The set of all reconstructions is termed a *codebook*.

A source-optimized quantizer provides a set of reconstructions and their corresponding encoding cells based on the statistics of the source. The optimality is typically measured by minimizing a distortion J (or a cost), e.g., the *mean squared error* (MSE) distortion for a scalar source x with the pdf $p(x)$:

$$J(x, d) = \sum_{k=0}^{2^R-1} \int_{\mathcal{S}(k)} (x - d(k))^2 p(x) \mathrm{d}x.$$

1.4.2 Channel Coding

When quantized symbols are transmitted over a noisy channel, transmission errors are unavoidable. Channel coding deals with protecting information bits against channel errors by carefully adding redundant bits. *Channel capacity* is an upper bound of the achievable rate above which error-free transmission is not possible. For example, consider a discrete memoryless channel which can be described by a conditional probability function $\mathcal{P}(j|\,i)$, where $i \in \mathcal{L}_I$ and $j \in \mathcal{L}_J$ are the channel input and output, respectively. The channel capacity for this class of channels is

$$\begin{aligned}\mathcal{C} &= \max_{\mathcal{P}(i)} \mathcal{I}(i;j) \\ &= \max_{\mathcal{P}(i)} \left\{ \sum_{i\in\mathcal{L}_I} \sum_{j\in\mathcal{L}_J} \mathcal{P}(j|\,i)\mathcal{P}(i) \log \frac{\mathcal{P}(j|\,i)}{\sum\limits_{i\in\mathcal{L}_I} \mathcal{P}(j|\,i)\mathcal{P}(i)} \right\},\end{aligned}$$

where $\mathcal{I}(i;j)$ denotes the *mutual information* between the input i and the output j. Intuitively, mutual information describes the information about i shared by j. The references [Wic95, MS98, RU02] provide useful introductions to the subject of channel coding.

1.4.3 Source–Channel Separation Theorem

The source–channel separation theorem states that under certain conditions combining separately designed source and channel codes still achieve the optimal performance for transmission over noisy channels (e.g., [Sha48, Sha59, CT91]). For lossless coding, it particularly states that there exists a source–channel code which can be used to transmit a stationary and ergodic source with arbitrary low probability of error if the entropy rate is lower than the channel capacity. Conversely, the source can not be transmitted reliably, i.e., the probability of error is bounded away from zero, if the source entropy rate exceeds the channel capacity. First, because the source satisfies the *asymptotic equipartition property* [CT91], it is possible to reconstruct the source symbol with arbitrary low probability of error when it is compressed to a rate arbitrarily close to the entropy rate (per source symbol). Thereafter, by using channel codes of very long block lengths, the probability of error in the transmission approaches to zero, since it is most likely that the channel output is *jointly typical* [CT91] with the channel input. We see that when the separation theorem applies, reliable transmission is possible even though the source code has not taken into account the channel statistics, and the channel code is designed without the consideration of the source statistics. More results for a wide range of classes of sources and channels can be found in e.g., [CT91, VV95].

There are obvious advantages to separate the source coding and channel coding problems. The separation theorem has laid ground for many practical designs. However, it is worth noting that the source–channel coding theorem relies on the

assumption of infinitely long block codes. For applications with strict delay constraints, a joint design of source–channel codes appears to outperform a separate design in many cases. The control problem formulated in this thesis is such an example.

The remaining part of this section is devoted to two specific channel coding related topics: (*i*) non-redundant channel coding in Section 1.4.4, and (*ii*) channel optimized quantization in Section 1.4.5.

1.4.4 Non-redundant Channel Coding

The main objective of channel coding is to combat channel errors by means of adding redundant bits. Apparently, the more properly added redundant bits, the more reliable the transmission will be. Non-redundant coding is the special case with no extra bits added, where the objective is to label the codewords carefully, referred to as the *index assignment* (IA). The index assignment deals with the combinatorial optimization problem of assigning non-redundant codewords to integer indices. In Figure 1.6 we illustrate the signal path from the input of the source encoder to the output of the channel decoder, where IA is used. A binary symmetric channel is considered in the example, where $b(i)$ and $b(j)$ denote the binary transmitted and received codewords, associated with the transmitted and received integer indices, i.e., i and j. Finally, $d(k)$ is the real-valued reconstruction associated with the index value k. Index assignment has been shown to be important for transmission over noisy channels. It is well recognized that a carelessly designed IA will reduce the system performance seriously. The varied ability to combat the channel errors is attributed to the different conditional transition probabilities among the binary codewords. More results on this topic can be found in e.g., [RS76, ZG90, Kna93, SH94, KA96].

Below, we discussed one special method, the so-called *simulated annealing* (SA), which is often used in practice to optimize IA. Index assignment is a combinatorial optimization problem, which belongs to the family of non-convex optimization problems. An exhaustive search for a global optimum is often exceedingly computationally expensive. A fairly good local optimum can, however, be approached with lower complexity by using an optimization algorithm termed *simulated annealing*. Simulated annealing, originated in metallurgy, is a controlled heating and cooling technique. It is used to freeze the material into a minimum energy structure that the defects in the material can be reduced. By observing the similarities between this metallurgic process and the optimization problem with multiple local optima, [KGV83] has developed an efficient algorithm to solve non-convex optimization problems. Details about the simulated annealing used in the index assignment problems can be read in e.g., [EHSW87, SH87, Far90]. Here we give a brief description of the simulated annealing procedure by using a binary example. Let $b(k)$ be the binary representation of the integer value k, e.g., $b(k)$ is the natural binary code. We define a state Π as a permutation vector that specifies the mapping from the index l (as well as the reconstruction $d(l)$) to the binary codeword $b(k)$ through

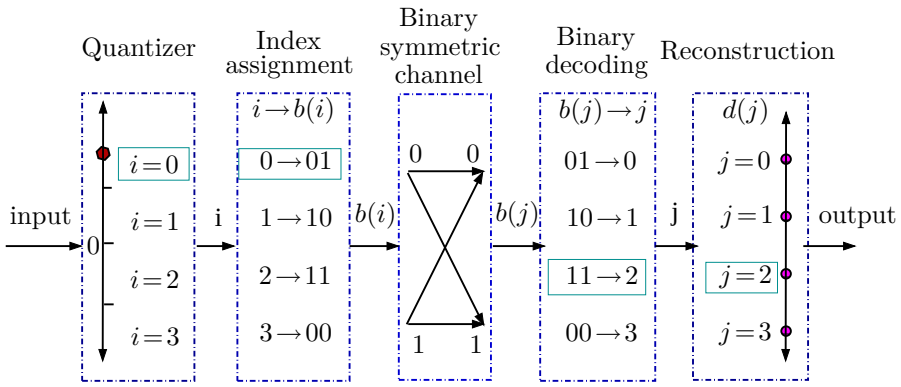


Figure 1.6: A diagram of data transmission over a binary symmetric channel. As an example, the index $i=0$ is mapped to the binary codeword 01, and the codeword 11 is received due to the error imposed by the channel. The decoded index is $j=2$.

$k=\Pi(l)$. The goal of the optimization is to find the best state $\Pi^\star$ which minimizes a certain objective function. In the beginning of the process, an initial-state and an initial temperature are selected. At each iteration a new state is obtained by randomly changing two entries in the previous state. The new state replaces the previous state if a lower cost is obtained; otherwise, the replacement takes place with a probability determined by the temperature, which decreases along with the rounds of the iteration. At a high temperature, the probability to accept a new IA is high, while when the temperature decreases, it is more and more unlikely to accept a new IA. The iteration terminates if one of the following conditions is fulfilled:

1. The cost drop exceeds a predefined number.
2. The temperature is below certain predefined freezing temperature.
3. There are too many unsuccessful trials.

The quality of the results is of course closely related to the number of iterations.

1.4.5 Channel Optimized Scalar Quantization

In traditional source–channel separate design, the source code does not take into account the channel properties. Near-optimal performance can still be achieved if the codeword length goes to infinity. On the other hand, for short codewords, it is well recognized that a quantizer design should take the channel properties into consideration [FV87, Far90, FV91, Sko97, Lin98]. When the channel $\mathcal{P}(j|i)$ and the source pdf $p(x)$ are perfectly known, a design method referred to as *channel optimized quantization* (COQ) [FV87, Far90, FV91, Sko97, Lin98] can be used. The terminology "quantizer" is used here since the joint source–channel coder performs a mapping from an infinite set of continuous values to a finite set of discrete symbols.

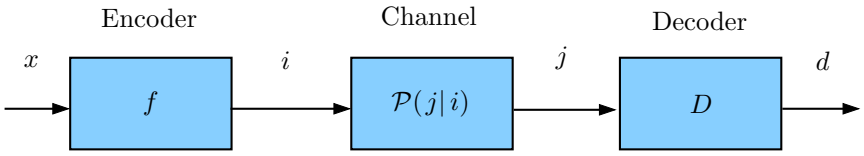


Figure 1.7: Quantization over a noisy channel with a finite number of possible outputs, where x and d are the source symbol and its reconstruction, and i and j are the transmitted and received indices.

In Figure 1.7 quantization over a noisy channel is depicted. The encoder function $f(\cdot)$ maps the source sample x to an integer index $i \in \mathcal{L} = \{0, \ldots, 2^R - 1\}$, with R denoting the transmission rate. The encoding cell associated with the index value k is defined as $\mathcal{S}(k) \triangleq \{x : f(x) = k\}$. Comparing with Figure 1.5, the encoded index i is now fed into a channel described by the transition probability function $\mathcal{P}(j|\,i)$, with $j \in \mathcal{L}$ denoting the received index. The channel output j is then used by the decoder to produce the estimate d, $j = l \Rightarrow d = d(l)$, where $d(l)$ denotes the reconstruction associated with the index value l.

In short, a channel optimized quantization should find the optimal encoding cells and the reconstructions that together minimize a certain distortion function, e.g., the MSE distortion below,

$$\mathbf{E}\left\{(x-d)^2\right\} = \sum_{k=0}^{2^R-1} \int_{\mathcal{S}(k)} p(x) \left(\sum_{l=0}^{2^R-1} \mathcal{P}(j=l|\,i=k)(x-d(l))^2 \right) \mathrm{d}x.$$

Unfortunately this type of quantizer is difficult to design in practice. Up to now, the solution is mostly obtained by using a method similar to the so-called Lloyd–Max algorithm, e.g., [Kie82, GG92]. The basic idea is to alternate between optimizing the encoder and decoder: to find the optimal decoder mapping for the fixed encoder mapping, and then to find the optimal encoder mapping for the fixed decoder mapping, back and forth. This iterative method can usually provide a locally optimal solution. As an example, we show the necessary conditions of a solution to the above-mentioned minimum mean squared error (MMSE) quantizer (e.g., [GG92]). For a fixed encoder, the best reconstruction $d^\star(l)$ in terms of MMSE is

$$d^\star(l) = \mathbf{E}\left\{x|\,j=l\right\} = \frac{\displaystyle\sum_{k=0}^{2^R-1} \mathcal{P}(j=l|\,i=k) \int_{\mathcal{S}(k)} xp(x)dx}{\displaystyle\sum_{k=0}^{2^R-1} \mathcal{P}(j=l|\,i=k) \int_{\mathcal{S}(k)} p(x)\mathrm{d}x}.$$

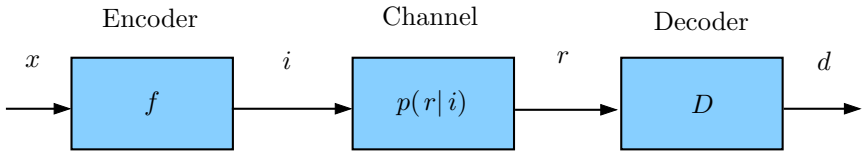


Figure 1.8: Quantization over a noisy channel with an infinite number of possible outputs. The source symbol and its reconstruction are x and d. The channel input and the real-valued output are i and r.

Furthermore, the optimal encoding cells for fixed reconstructions are

$$\mathcal{S}^{\star}(l) = \Bigg\{ x : \sum_{k=0}^{2^R-1} \mathcal{P}(j=k|\, i=l)(x-d(k))^2 \leq \sum_{k=0}^{2^R-1} \mathcal{P}(j=k|\, i=m)(x-d(k))^2, \ \forall m \in \mathcal{L} \Bigg\}, \quad l \in \mathcal{L}.$$

For further detail of the design of scalar and vector quantizers for noisy channels, the interested reader is referred to [FM84, FV87, Far90, FV91, GG92, Sko97].

1.4.6 Hadamard-Based Soft Decoding

Hadamard-based soft decoding techniques were developed for channels whose output $r \in \mathbb{R}^R$, with R denoting the transmission rate, is a real-valued vector. The Hadamard transform has been shown to be very useful in the implementation of estimator-based decoders. Here, a brief review of some useful results on Hadamard-based soft decoding is provided. For further detail on this topic the reader is referred to e.g., [KA96, Sko99a, Sko99b].

A Hadamard matrix, H_n, is a $2^n \times 2^n$ square matrix with binary elements of $\{-1, 1\}$, defined as

$$H_n = H_1 \otimes H_{n-1}, \quad H_n = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

where $\otimes$ denotes the Kronecker product. For brevity, the suffix n, used to specify the matrix dimension, is dropped in the subsequent analysis. The l^{th} column of the Hadamard matrix H, denoted by $h(l)$, is related to the integer value l and its natural binary representation, $b(l) = [\, b^{[1]}(l) \quad b^{[2]}(l) \quad \ldots \quad b^{[R]}(l) \,]$, as

$$h(l) = \begin{bmatrix} 1 \\ b^{[R]}(l) \end{bmatrix} \otimes \begin{bmatrix} 1 \\ b^{[R-1]}(l) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ b^{[1]}(l) \end{bmatrix}.$$

Here, $b^{[k]}(l), k \in \{1, 2, \ldots, R\}$ denotes the k^{th} bit of the binary codeword $b(l)$.

Consider a communication system in Figure 1.8 where $f(\cdot)$ and $D(\cdot)$ denote the encoding and decoding functions, respectively. A source sample x is transmitted to the receiver side through a noisy channel, described by a conditional density function $p(r|i)$, where $i \in \mathcal{L}_I$ and $r \in \mathbb{R}^R$ are the channel input and output, respectively. Note that the channel input alphabet is in this case different from the channel output alphabet. It is well-known that the minimum mean squared error decoder based on the received value r, is the conditional mean estimate,

$$D(r) = \mathbf{E}\{x|\, r\} = \sum_l \mathcal{P}(i = l|\, r)\mathbf{E}\{x|\, i = l\}.$$

Here, the decoder $D(r)$, in terms of the Hadamard representation, can be expressed as a product of two matrices, as

$$D(r) = \bar{T}\hat{h}(r). \tag{1.3}$$

First, the matrix $\bar{T}$ is obtained by solving the following matrix equation

$$\left[\ \mathbf{E}\{x|\, i=0\} \quad \mathbf{E}\{x|\, i=1\} \quad \ldots \quad \mathbf{E}\{x|\, i=2^n-1\}\ \right] = \bar{T}H,$$

where H is the Hadamard matrix, and $\mathbf{E}\{x|\, i=l\}$, $l \in \{0, \ldots, 2^n - 1\}$ is the conditional centroid given the channel input i. Second, the vector $\hat{h}(r)$ can be computed according to

$$\hat{h}(r) = \frac{R_{hh} \cdot \hat{p}(r)}{m_h' \cdot \hat{p}(r)}, \tag{1.4}$$

where $(\cdot)'$ denotes the transpose. The terms R_{hh} and m_h above are defined as

$$R_{hh} \triangleq \sum_{l=0}^{2^n-1} \mathcal{P}(i = l)\ h(l)h(l)',$$
$$m_h \triangleq \sum_{l=0}^{2^n-1} \mathcal{P}(i = l)\ h(l).$$

Recall that $h(l)$ denotes the l^{th} column of the Hadamard matrix H. Consider again (1.4). It is interesting to note that R_{hh} and m_h are independent of the channel output r, while

$$\hat{p}(r) \triangleq \mathbf{E}\left\{h \middle| r, \mathcal{P}(i) = \frac{1}{2^n}\right\},$$

the a posteriori expectation given evenly distributed channel input, is a function of r. Finally, it is worth mentioning that (1.3) reveals a clear separation of the channel decoding and the source decoding. In particular, $\bar{T}$ can be viewed as performing the source decoding, while $\hat{h}(r)$ operates as the channel decoding.

1.4.7 Optimal Bit Allocation

Classical bit allocation problems in communications are formulated for multiple sources, commonly in speech and video applications, to achieve a better performance for a fix total number of bits for date transmission or data storage, e.g., [Seg76, GG92, BV04]. Let us show the basic concept by using a simple example where the sources $\{x_1, x_2, \ldots, x_n\}$ are random variables with zero-mean and variances $\{\sigma_1^2,\ \sigma_2^2, \ldots,\ \sigma_n^2\}$. The task is to distribute a total of R_{tot} bits to the sources such that an overall objective function $\mathbf{E}\{J_{tot}(\mathbf{R}_1^n)\}$ with $J_{tot}(\mathbf{R}_1^n)$ given by

$$J_{tot}(\mathbf{R}_1^n) = \sum_{k=1}^{n} J_k(R_k),$$

is optimized. The cost J_k, associated with the source x_k, is a function of R_k, the number of bits allocated to x_k. The optimization problem is typically expressed in the following form,

$$\begin{aligned} \min_{\mathbf{R}_1^n} \quad & \mathbf{E}\{J_{tot}(\mathbf{R}_1^n)\}, \\ \text{s.t.} \quad & \sum_{k=1}^{n} R_k \le R_{tot}, \quad R_k \in \mathbb{Z}^+,\ k = 1, \ldots, n. \end{aligned} \tag{1.5}$$

Without imposing the requirement that R_k is a positive and integer number, the optimization problem (1.5) is usually solved by using Lagrange duality [BV04]. The basic idea is to introduce a so-called Lagrange multiplier θ, and then find a solution $\{R_1, R_2, \ldots, R_n, \theta\}$ that minimizes the Lagrangian η

$$\eta(\mathbf{R}_1^n, \theta) = \mathbf{E}\{J_{tot}(\mathbf{R}_1^n)\} + \theta\left(\sum_{k=1}^{n} R_k - R_{tot}\right).$$

In general the result is given by solving the system of equations

$$\begin{aligned} \frac{\partial}{\partial R_1}\eta(\mathbf{R}_1^n, \theta) &= 0, \\ &\vdots \\ \frac{\partial}{\partial R_n}\eta(\mathbf{R}_1^n, \theta) &= 0, \\ \sum_{k=1}^{n} R_k &= R_{tot}. \end{aligned}$$

The major challenge of this approach lies often in the derivation of the objective function in terms of the rates $\mathbf{R}_1^n$. In many cases we have to resort to approximations and simplifications to formulate a useful objective function. In addition, how to evaluate the quality of the solution is another challenging research topic which can be read in many textbooks (e.g., [BV04]) for an introductory study.

1.5 Control with Quantized Feedback

In this section, a brief introduction to the emerging research area of control with feedback over communication channels is given, by discussing a number of influential publications in this field. Before moving to an introductory survey, commonly used nomenclature is first presented.

Regarding the encoder, decoder, and controller, the term *static* refers to a time-invariant function, and *dynamic* refers to a time-varying function. A static quantizer is memoryless [FX06], while for a dynamic quantizer, the memory access pattern varies from *finite memory* to *infinite memory*. The importance of *information patterns* in the analysis of system behavior and design of control policies has been well recognized, e.g., [Wit71, Ho80]. The concept of the information pattern is introduced to specify what information is available at each unit in a connected system. Several authors have in their work, e.g., [Tat00], shown that the information pattern is instrumental to the achievable performance for quantized control systems.

The effect of quantized feedback on the overall performance of a control system is not a new topic in the literature. Traditionally, quantization errors have mainly been modeled as additive white noise [Wid61, GG92, WKL96], uncorrelated with the signal being quantized. The advantage of this model is that many tools and methods from traditional stochastic control theory can be readily applied. The model is valid at high rate, but fails to work at low rate. [Cur72] is one of the first works which studied coarsely quantized measurements. The author treated the quantized signal as non-linear partial observations and used non-linear estimation theory to solve the estimation problem. In the beginning of the 1990's, research interest in the stability analysis of quantized control systems was triggered by [Del90]. In that paper the author showed that, even for the simplest dynamic system (linear, scalar and noiseless), it is not possible to stabilize an unstable plant asymptotically with a static controller of a finite data rate. The author has also pointed out that treating quantization errors as white noise is no longer useful when the quantizer resolution is coarse, especially for unstable plants. The work of Delchamps has encouraged a rigorous study of the impact of quantization effects on control systems. Since then, controller design for data rate limited systems has continuously attracted researchers from different disciplines, e.g., applied mathematics, automatic control, communication and computer science. A wide range of interesting problems have been formulated.

The study of asymptotic behaviors of a dynamic system has theoretical importance. In the scope of automatic control, interesting features are for example *asymptotic observability*, which describes the asymptotic property of the state estimate; and *asymptotic stability* which describes the asymptotic behavior of the controlled state signals, e.g., [Tat00]. Since asymptotic stability cannot be achieved by using a static (memoryless) quantizer, *practical stability* [WB97] is formulated, where the state trajectories are only required to be bounded within a certain region.

Information theory has been the mathematical foundation for the designs of all modern communication systems. It provides fundamental limits of reliable data

compression and data transmission. Ongoing research has shown that several fundamental concepts in information theory, such as channel capacity and rate distortion theory, are in their traditional forms not useful for closed-loop control systems [Sah00, Tat00, Sah04, TM04b, TSM04]. The reasons are multiple. First, it is a common fact that control applications are much more time critical than conventional communication applications. Second, in communication applications the performance of the current transmission has usually no impact on what information to transmit in the future, while the objective of a feedback control is to affect the future states.

In the remaining part of this section, we give a short introduction to several important research problems in control over finite-rate noisy channels. First in Section 1.5.1 we address the so-called *minimum rate* control problem, which is closely related to stability properties of closed-loop systems. In Section 1.5.2, an introduction to research on static (memoryless) quantizers is provided. From the implementation point of view, static quantizers are an important class of quantizers. The study of their asymptotic behavior involves advanced mathematics. Even for a time-varying uniform quantizer which is easy to build, it is not transparent to understand. Finally, a number of important works on control over noisy channels are introduced, where we can see that channel errors have brought many new challenges to the encoder and controller design.

1.5.1 Minimum Rate and Stability

Since the main objective of control is to cope with uncertainty, the majority of control problems deal with various stability issues. Regarding control with quantized feedback, a fundamental problem that has triggered considerably attention during the last decades is to find the smallest feedback data rate necessary to stabilize an unstable dynamic system asymptotically. In the sequel, we refer to this rate as the *minimum rate* [WB99, Bai02, HOV02, NE03, YB04, TM04b]. An important overview of the influential publications on feedback control under rate constraints can be found in [NFZE07].

Regarding error-free communication links, the solutions to the minimum-rate problems have mostly exploited a "volume" based analysis. The intuitive idea is that a growth of the signal space for the state signal (often addressed as the uncertainty "volume"), due to unstable poles, should be counteracted by the "volume" reduction along the coding/quantization. One fundamental result of the limited rate is that, in order to attain asymptotic stability, a quantizer must be dynamic and has infinite reconstruction levels [Del90]. However, asymptotic stability can refer to different things, e.g., a bounded asymptotic worst-case state norm [Tat00] or a bounded asymptotic average state norm [NE02]. The asymptotic worst-case state norm provides conservative solutions and the analysis is more or less straightforward, while the asymptotic average state norm has appealed to information theoretic advances, such as different entropy power [NE03].

The minimum rate of a discrete time linear scalar plant ($x_{t+1} = ax_t + u_t$, where a

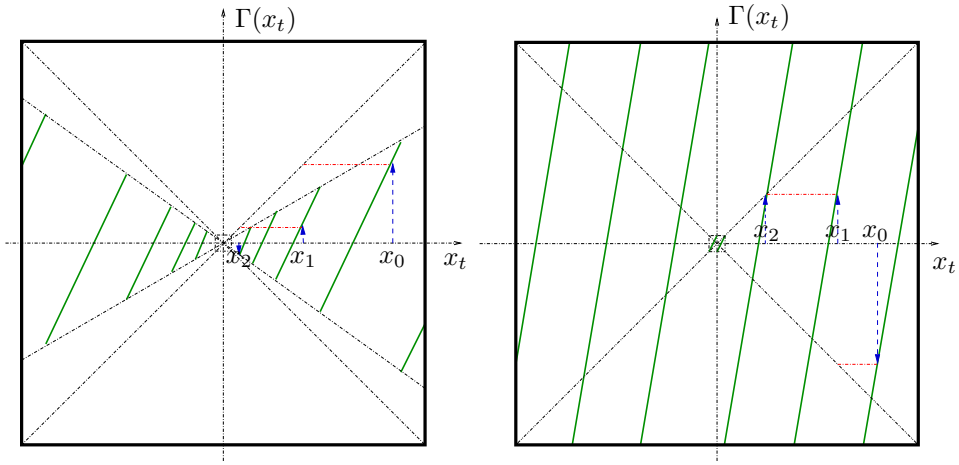
describes the linear dynamics) was investigated in e.g., [Tat00]. For an unstable discrete system the minimum data rate is $R > \log_2 |a|$. The corresponding result for a continuous time linear plant is $R > a \log_2 e$ [Bai02]. A generalization to multidimensional systems is pursued in e.g., [Tat00, HOV02, NE02]. A common way to tackle a multidimensional system is to transform the system matrix A (the counterpart of a in the multidimensional case) into its Jordan canonical form. Under the assumption that the system can be decoupled into several independent one-dimensional systems, the tools developed for scalar systems can be applied. A lower bound of the minimum rate for multidimensional systems is given by $R > \sum_k \log_2 |\lambda_k^u|$, where λ_k^u denotes the k^{th} unstable eigenvalue of the system matrix A. A realistic data rate should take on non-negative integer values. In [LL05a, LL05b], the authors have proposed a practical coding scheme that uses integer rates. In their approach, a transformation to the Jordan canonical form is not required. Instead, at each time t, the quantization is pursued only along the most critical direction. Regarding systems perturbed by stochastic disturbances, under certain assumptions on the process noise, a lower bound of minimum rate has been derived in [NE03], based on differential entropy power. Related recent work can be found in [GN08] where the authors have shown the mean squared internal stability for systems of unbounded noise, random initialization and arbitrarily initialized encoder–controller states.

There are a variety of stabilizable controllers which are based on the zooming technique [BL00, Lib02a, Lib02b, Lib03, BL06, LN07]. The basic principle behind these solutions is that when the trajectory is close to an equilibrium point a zooming-in operation will increase the quantizer resolution, while when the trajectory is far from the equilibrium point a zooming-out operation will reduce the quantizer resolution. It is worth mentioning that a memoryless dynamic quantizer has infinite quantization levels asymptotically. Adaptive quantizers based on the zooming idea can also be read in [NE03, EWRF07].

Extensions to non-linear systems can be found in [Lib02b, Lib03, Bai04, LE04, NEMM04, Per04, LH05]. For example in [Lib02b], the author has applied the zooming strategy to a non-linear time-invariant system with an unknown initial-state. The author has studied two quantized systems. In the first system, the quantizer is located at the observation link, while in the second system, the quantizer is located at the actuation link. The conditions for the global asymptotic stability are derived for the both cases.

1.5.2 Static Quantizer

From the implementation point of view, it is useful to study static quantizers. Recent research has shown an increasing interest in the fundamental properties of this class of quantizers. For example, attention has been given to issues such as useful performance measures for a static quantizer [FZ05a], and the impact of the strict memory restriction on the long-term behavior [FZ05a, PB07]. Although recent research in this field has been mainly addressed to the simplest system model (linear, scalar and noiseless), a lot of notable results have been reported, e.g., [Del88,



(a) Logarithmic quantizer (b) Chaotic quantizer

Figure 1.9: Examples of piecewise affine maps for quantized control systems. The dashed arrows illustrate the system evolutions from x_0 to x_2. (a) A piecewise affine mapping of a logarithmic quantizer. (b) A piecewise affine mapping of a chaotic quantizer.

Del89, Del90, PGB02, FZ03, PPBJ04, FZ05a, PB05, Del06, PB07, PB08]. They together have established a variety of theoretical fundamentals that lead to practical quantized feedback control systems.

In the remaining part of this section, a couple of important properties of static (memoryless) quantizers are introduced. First of all, let $q(\cdot)$ denote a static quantization mapping, and $F(\cdot)$ a static feedback control. Thus, we can describe the system evolution as

$$x_{t+1} = ax_t + u_t = ax_t + F(q(x_t)) = \Gamma(x_t),$$

where x_{t+1} depends only on x_t, since u_t is completely determined by x_t. As a matter of fact, the system evolution can be described by a piecewise affine mapping $\Gamma(\cdot)$, as illustrated in Figure 1.9. In the figure, the current state x_t is presented on the x-axis, and the state one-step-ahead $\Gamma(x_t)$ is presented on the y-axis. In particular, the mapping illustrates the two most important properties of a static quantizer. The first property is the *attractivity* [FZ05a], referring to as the attraction from a (large) initial region to a (small) target region. The second property is the practical stability, related to the ability of keeping the trajectories within a target region. As stated previously, for quantizers with finite reconstruction levels, the asymptotic stability is not achievable. Instead, the practical stability is an applicable stability measure.

A good control strategy should both give satisfactory steady state and transient properties. In [FZ03, FZ05a] the authors have shown that there is, however, a conflicting relation between the steady state and the transient behavior for quantized systems. The authors were particularly interested in the quantitative analysis and comparison among different quantizers. They have suggested the *contraction rate* for the steady-state performance and the *expected time* for the transient behavior. Generally speaking, a large contraction rate is a desired feature, since it means a smaller target region for a given initial region. At the same time, the expected time is strived to be small, which indicates an efficient control and a quick entrance to the target region. Unfortunately, there is a trade-off between the contraction rate and the expected time. The authors have evaluated this conflicting relation for uniform quantizers, logarithmic quantizers and chaotic quantizers. Other works regarding controlled invariance can be found in e.g., [PB07, PB08] and the references therein. Here, we mention a special property of the logarithmic quantizer. Shown by e.g., [EM99, EM01], a logarithmic quantizer is special in the way that when the number of the reconstruction level is a prior entity, the logarithmic quantizer is shown to be the most efficient quantization scheme, in the sense of quadratic Lyapunov stability [Kha96]. Lyapunov stability is a stability measurement commonly used in non-linear control. The intuitive idea behind this measure can be simply explained as, a stabilizable control will make the state follow a path along which the "energy" is continuously decreasing. The work of [EM01] is further developed in e.g., [EF02, FX05, FX07].

For many researchers, the quantizer design has been part of the stabilization problems. Unlike them, in [PGB02, PPBJ04, PB05, PB07] the authors have considered the quantizer as a given building block and investigated its capabilities. In particular, they worked with the construction of attractive and invariant sets for stabilizing unstable plants using uniform quantizers. Due to the simplicity in the implementation, the uniform quantizer is the most widely used quantization scheme in practice. Besides uniform quantizers and logarithmic quantizers, a special class of quantizers called chaotic quantizers have also received increasing attention, e.g., [FZ03, FZ05a]. The chaotic behavior is exploited to mitigate the conflicting relations between attractivity and invariance, done by combining two feedback policies. The attraction from the initial region to the target region is accomplished by one of the control strategies, which takes the advantage of the chaotic behavior of the affine mapping. All the trajectories will eventually enter the target region. When the trajectory is within the target region, it switches to the second controller,whose task is to keep the trajectories within the target region ever after, cf., Figure 1.9b.

1.5.3 Control over Noisy Channels

In many of the aforementioned works, the communication channel is essentially error-free, and the only limitation imposed by the channel is the data rate. In the last decade, the extension to control over noisy channels has brought many new challenges which invoked a great deal of research interests. Some influential works

include [SJV04, Sim04, TM04a, MS05, Xu05, MT06a, MT06b, MS06a, MS06b, NB06, SM06, Ten06, MS07, MFDN07, TM09]. Regarding stabilizing unstable plants over noisy channels, some of the early contributions have been e.g., [Sah04, Sim04, Xu05, MS06a]. In those works, the solutions are typically based on a kind of control–channel separation principle. The basic concept is to design the stabilizing control assuming the channel is error-free, and then, require the control command to be reliably transmitted over the noisy channel. Particularly for unstable systems, where delay becomes a critical issue, the validity of the classical Shannon capacity on problems of stabilizing through noisy channels is challenged. In [Sah04], the author has shown that the Shannon capacity is not a proper entity for characterizing feedback systems, instead, the novel concept of anytime capacity is introduced to show the necessary rate to achieve anytime reliability over noisy channels. This new concept is related to Gallager's error exponent [Gal68]. However, unlike in Gallager's problem, the decoder has to produce new estimates for each new received channel output, and cannot wait until it has received a long codeword. Further results on the anytime capacity can be found in e.g., [Sim04, DS06, SM06]. Recent results on the characterization of the capacity of channels with feedback can be read in e.g., [MS07, YKT07, TM09].

In early work where the communication channel is error-free, less attention was paid to the encoder design. Quantizers are often considered as fixed system components, e.g., [BL00, NE03], and simple quantizers are used, for example static quantizers as discussed in Section 1.5.2. However, in the presence of channel errors where the encoder does not have perfect knowledge about what message the controller will receiver, the encoder design becomes important. How to design the encoder for noisy channels can be found in e.g., [BMT01, TSM04, MT06a, Tat08]. In [TSM04], sequential rate–distortion theory is developed to handle the delay and causality restrictions typically encountered in control systems. The author has emphasized the difference between statistical dependence in traditional information theory and causality in the control context. The quantity of *directed mutual information* [Mas90] is pointed out to be instrumental in dealing with sequential rate distortion problems. In the same paper, the sequential rate–distortion functions for Gauss-Markov processes are computed. In [SV03] state estimation over a binary symmetric channel is studied, where the encoder–decoder adopts the zooming idea to capture the state trajectory and bound estimation errors. More precisely, the quantizer is uniform and its range is adaptively adjusted according to the state evolution. In [MT06a] some results on real-time (finite decoder memory) encoder–decoders for tracking the outputs of a Markov process are presented. In [Tat08], the author has introduced the concept of *joint sufficient* of the encoder and decoder to deal with the uncertainty in communications.

Up till now, most work on control with limited information has been devoted to analyzing stability properties, while optimal designs for general criteria are much less explored in the literature. However, the problem of optimal stochastic control over communication channels is addressed in e.g., [BM95, Tat00, MS02, MS04a, MS04b, TM04b, NFZE07, CF08]. In [BM95], the considered system is stable and the

measurement noise is absent. An innovation coding scheme is proposed to convey the information about the process noise to the controller periodically. Since the time between two transmissions is increased when longer codewords are used, there is a trade-off between the resolution and delay. In [Tat00], the research on the separation principle and CE controllers has been carried forward. An analysis of systems with different information patterns is pursued. The authors have separated the total distortion into two parts. The first is obtained by assuming that a full state knowledge is available and the second comes from the use of a sequential quantizer, referring to as the procedure of successively quantizing the outputs of a dynamic system. Unlike [BM95], where the state measurement is quantized and transmitted over the noisy channel, in [MS04a] the control input is quantized and transmitted over a rate-limited channel. In that work, the existence of the optimal controller and its properties are explicitly studied.

1.6 Contributions and Outline

In the previous section, a brief overview of recent advances in control using quantized feedback was given. We noticed that the existing work has been mainly devoted to various stability analyses. Moreover, static quantizers, such as uniform quantizers and logarithmic quantizers, are commonly considered in the literature, since they are easy to implement in practice. However, for applications with extremely low data rates and high communication costs, it would be reasonable to study optimal encoder–controllers. This is an important problem in networked control systems in which a large number of sensor nodes need to limit their individual access to the communication medium. However, in many situations, there is a challenging conflict between the amount of transmitted data and the response time. In particular for emerging control applications, not necessarily constrained by today's communication protocols, limits imposed on available signaling bandwidth from communication channels can severely restrict the closed-loop performance and even destabilize the system. To develop methods and tools for the analysis and synthesis of feedback control over imperfect communication links is therefore of great importance.

The common theme of the thesis is the design and analysis of encoder–controllers for control over low-rate noisy channels, motivated by wireless networked control systems. In particular, we focus on the performance of a single plant (node) in the network under resource constraints. In the case that there is no congestion problems, each plant can of course be treated individually. The remaining part of this thesis is organized as follows. The first part, Chapters 2–4, is devoted to the optimization of encoder–controller mappings over finite-input finite-output noisy channels. We focus explicitly on *iterative design* of encoder–controller pairs. The result is a practical synthesis technique for joint optimization of the quantization, error protection and control over a low-rate noisy channel. More specifically, in Chapter 2, we describe the general model of the closed-loop control system with quantized feedback, using a state space model. A stochastic optimization problem

with a quadratic linear cost is formulated. In Chapter 3, a training method is developed to optimize encoder–controller mappings iteratively. Considerations on the complexity of the proposed training process can be found in Chapter 4. In the same chapter, the problem of control over longer horizons is studied by comparing a number of coding–control strategies. The second part, Chapter 5, studies the problem of allocating the instantaneous rates over time, for systems described in Chapter 2. The rate allocation is important since it offers an alternative way to overcome the highly limited communication resources and to achieve a better overall control performance. In the third part, Chapter 6, we extend the result in Chapter 3 to design encoder-controllers for channels with infinite-outputs. Finally, in Chapter 7, some topics for future research are suggested. Below, the contributions are described in more detail for each chapter.

Chapter 2

In this chapter, a general model of a closed-loop control system for a single plant is described which embodies process noise, measurement noise and a noisy channel with finite input and output alphabets. To cope with the imperfect communication between the transmitter and the receiver, an encoder unit is enclosed in the system to protect information messages against channel distortions. Because of the potential transmission errors, side-information to the encoder is introduced which can inform the encoder about the channel outputs. By varying the information pattern at each component, the model can represent a wide range of practical scenarios.

As mentioned previously, this thesis deals with several aspects of the design of encoder–controllers in a networked control system, where the communication between the sensor and the actuator is modeled as a finite-rate noisy channel. Unlike the works on minimum data rate control, our main concern is an optimal average performance over a finite horizon, given a fixed data rate. Therefore, a stochastic optimization problem with a quadratic linear cost is formulated. The optimization problem can be viewed as an extension of the classical LQG problem, where we study the impact of a low-rate noisy channel on overall system performance. However, so far, there is hardly any result on the minimum data rate for this general system, since when both measurement noise and channel errors are present, the characterization of such a rate is extremely difficult.

Chapter 3

In this chapter, we study a closed-loop control system with feedback transmitted over a finite-rate noisy channel. With the objective to minimize the expected linear quadratic cost over a finite horizon, we propose a joint design of the sensor measurement quantization, channel error protection, and optimal controller actuation. This encoder–controller optimization problem is known to be hard in general, an iterative design procedure can be derived inspired by traditional quantizer designs in which the controller is optimized for a fixed encoder, then the encoder is optimized

for a fixed controller etc. Several properties of such a scheme are presented. For a fixed encoder, expressions for the optimal controller is derived for various amount of side-information available in the encoder about the symbols received at the controller. It is shown that the certainty equivalence controller is optimal when the encoder is optimal and has full side-information. For a fixed controller, expressions for the optimal encoder is given and implications are discussed for the special cases when process, sensor, or channel noise is not present. Numerical experiments are carried out to demonstrate the performance obtained by employing the proposed iterative design procedure and to compare it with other relevant schemes.

It is worth to point out that our approach treats transmission uncertainties differently from many aforementioned works by not employing the coding–control separation concept. A design approach based on the separation of coding and control, e.g., [SM06, Sav06] relies significantly on the information theoretical results of coding schemes which can ensure reliable communication over noisy channels. However, how to find such coding schemes is still an open question, which is the main obstacle to the implementation of the separation-based controllers in practice. In this thesis, we consider practically implementable encoder–controllers. As a matter of fact, we design for situations where the communication between the sensors and controller is highly expensive that information is conveyed by only a few bits. At low data rates, it is especially beneficial to employ a joint coding–control design that the codeword accomplishes source coding, protection against channel error and control simultaneously.

The present chapter is partly published in the papers [BSJ06a, BSJ06b, BSJ07a, BSJ07b, BSJ08, BSJ09].

Chapter 4

In this chapter, several practical considerations on the implementation of the training algorithm proposed in Chapter 3 are addressed. A major part of this chapter is studying the complexity of the encoder which is revealed to be critical to the implementation of the training algorithm. In general, the proposed training algorithm is complicated to implement. Certainly, overcoming the complexity barrier is one of the major challenges from a practical point of view. In the first part, two special encoder properties, namely *sufficient statistics* and *regularity*, are described. These two properties appear to be very useful in the implementation of the training algorithm. Thereafter, a closed-form approximation of the expected future cost is proposed, which can be used in assessing the regularity of the encoder. In the second part, we deal with the complexity problem of longer horizon by an empirical study of several systems that exploit complexity reduced approximations.

This chapter is partly based on the papers [BSJ08, BSJ09].

Chapter 5

In this chapter we study the problem of optimizing the rate allocation over time, which is an alternative method to improve the efficiency of the overall control performance, given highly limited communication resources. Owing to the non-stationarity of the state observations, it is natural to expect considerable gains by employing a non-uniform allocation of transmission rates over time. Hence, an even distribution of bits to all sensor measurements is often not efficient for control systems. The optimal rate allocation aims at an efficient way of exploiting the limited communication resources to achieve a better overall performance. How to achieve the optimal rate allocation in control systems is a challenging task. One of the main obstacle to optimize the instantaneous rate is the lack of tractable distortion functions, which we need to use as objective functions for the rate optimization problem. Furthermore, such an optimization problem is often non-convex and non-linear, which implies that it is difficult to compute the optimal solution in practice.

Our main contribution is a novel method for rate allocation for state feedback control of a linear system over a noisy channel. By resorting to an approximation based on high-rate quantization theory, we are able to derive a computationally feasible scheme that seeks to minimize the overall distortion over a finite time horizon. The resulting rate allocation is not necessarily evenly distributed. Practical considerations concerning the integer-rate constraint and the accuracy of the high-rate approximation are discussed and illustrated through numerical examples.

This chapter is partly published in the papers [BSFJ09a, BSFJ09b, BSFJ09c, BSFJ09d].

Chapter 6

This chapter studies the generalization to infinite-output channels. It is not difficult to realize that the system complexity is substantially increased if the channel output has an infinite alphabet. In practice, the extension from a finite alphabet to an infinite alphabet affects not only the training stage, but also the implementation of the trained encoder–controller pair. It no longer holds that the trained encoder–controller can be implemented as a simple look-up table. To get more insight into the optimal controller, we first introduce a Hadamard-based soft controller which fully exploits the information in the channel outputs. However, because of the computational and memory demands it is difficult to implement such a controller in practice. Alternatively, we can apply the results from Chapter 3 to employ a controller using only quantized channel outputs. Of course, if the channel is heavily noisy the degradation in system performance could be serious by only using quantized channel outputs. Given the challenges discussed above, we propose a combined encoder–controller which exploits both the hard and soft channel outputs.

Chapter 7

The concluding chapter summarizes the thesis. Examples for future work are suggested, where a broad range of research topics are presented.

1.7 Notation

Throughout the thesis, the following notations are used.

a	:	The linear dynamics of a scalar dynamic system.
b_t	:	In general, denotes binary codeword at time t.
c_t	:	In general, denotes centroid at time t.
$d_H(\cdot,\cdot)$	:	The Hamming distance of two binary codewords.
d_t	:	Decoded symbol at time t.
$D_t(\cdot)$	:	Decoding mapping at time t.
e_t	:	Measurement noise at time t.
$\mathbf{E}\{\cdot\}$	:	Mathematical expectation.
$\mathbf{E}\{\cdot\|\cdot\}$	:	Conditional expectation.
$E_t(\cdot)$	:	Channel mapping at time t.
$f_t(\cdot)$	:	Encoding mapping at time t.
$g_t(\cdot)$	:	Control mapping at time t.
H	:	Hadamard matrix.
i_t	:	Index-valued channel input at time t.
j_t	:	Index-valued channel output at time t ("hard" channel output).
$J(\cdot)$	:	Objective function.
$\ln\{\cdot\}$	:	The natural logarithm.
$\mathcal{L}$	:	In general, denotes a finite set of integer indices.
ℓ_t	:	Optimal linear feedback control law at time t.
M	:	In general, denotes memory in the sense of the context.
$\mathcal{N}(\bar{m},\sigma^2)$	:	Gaussian distribution function with mean $\bar{m}$ and variance σ^2.
$\mathbb{N}$	:	The set of natural numbers, $\{1,2,3,\dots\}$.
$\mathcal{O}$	:	The Ordo operator.
P_t	:	The weighting matrix in the linear quadratic objective function, associated to control input u_t.
$p(\cdot)$	:	Probability density function.
$p(\cdot\|\cdot)$	:	Conditional probability density function.
$\mathcal{P}(\cdot)$	:	Probability.

$\mathcal{P}(\cdot\|\cdot)$	:	Conditional probability.
r_t	:	Real-valued channel output at time t ("soft" channel output).
R_t	:	Rate at time t.
$\mathbb{R}$	:	The set of real numbers.
$\mathcal{S}$	:	In general, denotes an encoding cell.
T	:	Time horizon.
$\mathbf{T_r}\{\cdot\}$	:	The trace of a matrix.
u_t	:	Control variable at time t.
v_t	:	Process noise at time t.
V_t	:	The weighting matrix in the linear quadratic objective function, associated to state x_t.
w_t	:	Channel noise at time t.
x_t	:	State variable at time t.
y_t	:	Measurement variable at time t.
z_t	:	Side-information at time t.
$Z_t(\cdot)$	:	Side-information mapping at time t.
$\mathbb{Z}$	:	The set of integer numbers.
$\mathbb{Z}^+$	:	The set of non-negative integer numbers, including 0.
γ	:	Cost-to-go.
ϵ	:	The crossover probability of a binary symmetric channel.
ρ	:	The weighting parameter in a scalar control objective.
σ^2	:	In general, denotes signal variance.
Δ	:	In general, denotes difference in the sense of the context.
$(\cdot)^\star$	:	In general, indicates optimality in the sense of the context.
$(\cdot)'$	:	The transpose of a matrix.
$(\cdot)!$	:	The factorial of a non-negative integer.
$(\cdot)^\dagger$	:	The Moore-Penrose pseudoinverse of a matrix.
$\triangleq$	:	Equality by definition.
$\otimes$	:	Kronecker product.
$*$	:	Convolution.

Below are some common rules for indexing, if not stated otherwise.

1. The variables t and s are reserved to the time indices. The variables k, l, m, n are also commonly used to denote an index, but not necessary a time index.

2. The boldface notation $\mathbf{x}_s^t$, s, $t \in \mathbb{Z}^+$, $s \leq t$, is taken to describe the time series $\{x_s, x_{s+1}, \ldots, x_t\}$. Especially, x_t, $t < 0$, is not defined, and $\{\mathbf{x}_s^t : t < s\}$ is an empty set.
3. The notation $x^{[k]}$, $k \in \mathbb{N}$, is taken to specify the k^{th} element of x.
4. The notation $x_{[k]}$, $k \in \mathbb{N}$, is taken to specify the iteration result of the k^{th} round.

1.8 Acronyms

Acronyms will be defined at their first occurrence in the thesis. For convenience, these acronyms are listed below.

BSC Binary symmetric channel.

BGC Binary Gaussian channel.

CE Certainty equivalence.

DMC Discrete memoryless channel.

i.i.d. Independent identically-distributed.

IA Index assignment.

LQ Linear quadratic.

LQG Linear quadratic Gaussian.

MMSE Minimum mean squared error.

MSE Mean squared error.

pdf Probability density function.

pmf Probability mass function.

RA Rate Allocation.

SI Side-information.

Chapter 2

Coding and Control for Discrete Memoryless Channels

2.1 Introduction

This chapter describes the closed-loop control system studied in Chapters 3–5, where a linear plant is controlled using quantized feedback over a memoryless noisy channel. The channel has finite-input and finite-output alphabets. The main objective is to find the optimal coding–control strategy which solves a stochastic optimization problem with a quadratic linear cost.

It is widely-recognized in stochastic control that for most systems the optimal performance is achievable only when the coding and control are designed jointly. A closed-form solution to the optimal control is in general difficult to characterize. However, for traditional control systems without any channel imperfections, there are cases where the separation principle applies and the closed-form solution is known, e.g., the LQG problem described in Chapter 1. In this thesis, we focus on the impact of an imperfect channel on the encoder–controller design. We restrict our discussion to the linear plant and quadratic objective functions that the closed-form solution to optimal stochastic control for ideal channels under Gaussian assumptions (initial-state and noises pdf's), is well-known.

The remaining part of this chapter is organized as follows. In Section 2.2 we define the control system with encoder, controller, and communication channel. Thereafter, Section 2.3 is devoted to a discussion of properties of the side-information from the controller to the encoder. The problem statement which concerns a linear quadratic (LQ) objective over a finite horizon is formulated in Section 2.4. Finally, the operation of the encoder–controller is illustrated in Section 2.5, as well as the concept of optimizing the encoder–controller mappings by training.

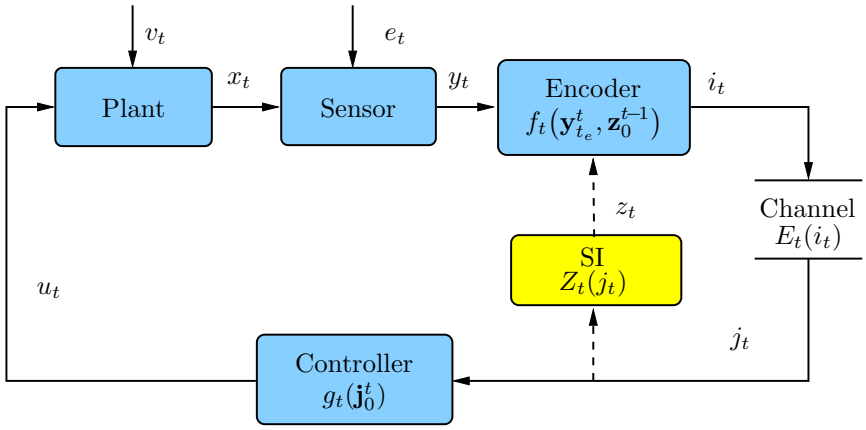


Figure 2.1: A general model of a system with feedback control over a discrete memoryless channel. The dashed line indicates potential side-information signaling from the controller to the encoder. At time t, the oldest measurement the encoder can use is y_{t_e}, $t_e = t - M_e$, where M_e specifies the total number of past measurements the encoder can use.

2.2 System Model

In the most general form, we consider the control system with a communication channel depicted in Figure 2.1. The sensor observation is encoded and transmitted to the controller over an unreliable communication channel, and the control command is derived based on the received data. Information available at each building block in the closed-loop system is specified by the so-called information pattern [Tat00]. It is well recognized that the information pattern is important in characterizing the achievable system performance. Let us set to describe each building block by specifying the information pattern.

The multi-variable linear plant is governed by the following equations:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + v_t, \\ y_t &= Cx_t + e_t, \end{aligned} \tag{2.1}$$

where $x_t \in \mathbb{R}^{\mathtt{n}}, u_t \in \mathbb{R}^{\mathtt{m}}, y_t \in \mathbb{R}^{\mathtt{p}}$, are the state, the control, and the measurement, respectively. The matrices $A \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$, $B \in \mathbb{R}^{\mathtt{n}\times\mathtt{m}}$, $C \in \mathbb{R}^{\mathtt{p}\times\mathtt{n}}$, are known; where (A, C) is state observable and (A, B) is state controllable. The variable $v_t \in \mathbb{R}^{\mathtt{n}}$ denotes the process noise, and $e_t \in \mathbb{R}^{\mathtt{p}}$ denotes the measurement noise. The noise signals are i.i.d. and mutually independent. They are also independent of the system initial-state x_0. We assume that the initial-state and the noise processes are zero-mean, and their probability density functions are known.

We consider an encoder that causally utilizes the encoder information. By the *encoder information*, we mean the set of variables whose values are known to the

encoder. The encoder is then a mapping from the set of the encoder information to a discrete set of symbols. We take each symbol to be represented by an integer index. At time t, the index is $i_t \in \mathcal{L}_I = \{0, \ldots, L_I - 1\}$, $L_I \in \mathbb{N}$. We are interested in the class of encoder mappings described by the function

$$i_t = f_t\left(\mathbf{y}_{t_e}^t, \mathbf{z}_0^{t-1}\right), \quad t_e = t - M_e, \;\; M_e \in \{0, \ldots, t\}. \tag{2.2}$$

Here, M_e specifies how many of the past measurements can be used by the encoder. For example, $M_e = 0$, is the special case where only the most recent measurement is available, while when $M_e = t$, the encoder can use the full sequence $\mathbf{y}_0^t$.

Given the sequence of the past side-information, $\mathbf{z}_0^{t-1}$, and measurements, $\mathbf{y}_{t_e}^t$, the encoder produces an index i_t, and the transmitted index is then received as $j_t \in \mathcal{L}_J = \{0, \ldots, L_J - 1\}$, where $L_J \geq L_I$. The side-information z_t represents available feedback to the encoder about the value of the symbol j_t received at the controller. In this thesis, we define the *side-information* (SI) at the encoder to be produced as

$$z_t = Z_t(j_t) \in \mathcal{L}_Z = \{0, \ldots, L_Z - 1\}, \quad L_Z \in \mathbb{N}, \;\; 1 \leq L_Z \leq L_J, \tag{2.3}$$

where $Z_t : \mathcal{L}_J \mapsto \mathcal{L}_Z$ is deterministic and memoryless. The side-information is further discussed in Section 2.3.

Non-ideal communication channels have a variety of negative impacts on the overall performance of control systems. We restrict our attention to two types of channel imperfections: the limited channel rate and the potential bit errors. Let the encoder output index, i_t, be transmitted over a *discrete memoryless channel* (DMC), with input and output alphabets $\mathcal{L}_I$ and $\mathcal{L}_J$, respectively. One use of the channel is defined as

$$j_t = E_t(i_t), \tag{2.4}$$

where $E_t : \mathcal{L}_I \mapsto \mathcal{L}_J$ is a memoryless random mapping. Since we assume $L_J \geq L_I$, the output alphabet is potentially larger than the input alphabet. Conditioned on the transmitted symbol i_t, the mapping to j_t is independent of other parameters in the system, e.g., the process and measurement noise. The fact that the channel is bandlimited is captured by the finite size of the input-alphabet $\mathcal{L}_I$. We define the *rate* of the transmission (in bits per channel use) as $R = \log_2 L_I$. Our transmission model encompasses all the common memoryless models; for example, the binary symmetric channel (BSC) and the binary erasure channel (BEC) [CT91].

At the receiver side, we consider a controller that causally utilizes all the available *controller information* $\mathbf{j}_0^t$, to produce the control command

$$u_t = g_t(\mathbf{j}_0^t) \in \mathbb{R}^{\mathrm{m}}. \tag{2.5}$$

According to (2.5), when the entire sequence $\mathbf{j}_0^t$ is available at the controller, all controls $\mathbf{u}_0^t$ are completely specified. Therefore, the controller $g_t(\mathbf{j}_0^t, \mathbf{u}_0^{t-1})$ belongs to the class of controllers (2.5).

Here we mention a special type of controller that consists of a separate decoder and a controller. The decoder is the deterministic mapping

$$d_t = D_t(\mathbf{j}_0^t),$$

and the separate controller is defined by the mapping

$$u_t = g_t(\mathbf{d}_0^t),$$

given the past decoded symbols $\mathbf{d}_0^t$. Note that, since u_t is fully determined by $\mathbf{d}_0^t$, the controller $g_t(\mathbf{d}_0^t, \mathbf{u}_0^{t-1})$ also belongs to the class of controllers (2.5).

Throughout the thesis, we use "encoder" and "controller" in quite general terms ("the corresponding boxes in Figure 2.1"), as well as specific terms ("the mappings f_t and g_t at time t"). When there is a need to be specific, we will adopt the following notation and terminology. Let the notation $\mathbf{f}_0^{T-1} = \{f_t\}_0^{T-1}$ mean the entire sequence of encoder mappings, from time 0 to $T-1$. We will call $\mathbf{f}_0^{T-1}$ the "encoder" of the system (since the sequence specifies the operation of the encoder for all relevant time-instants), and we will say that f_t is the "encoder mapping at time t" or "the component of $\mathbf{f}_0^{T-1}$ at time t" when there is a need to specify a specific time-instant and the corresponding mapping. For subsequences of the full sequence of encoder mappings, we will use similar notation, and say, e.g., "the encoder components $\mathbf{f}_s^t$". This notation will also be used for the controller, its individual mappings or components, and subsequences, namely, $\mathbf{g}_0^{T-1}$, g_t and $\mathbf{g}_s^t$, respectively.

In the remaining part of this section, two examples, the hydraulic network and the binary symmetric channel, are used to illustrate the utility of the linear plant (2.1) and the discrete memoryless channel (2.4).

Example 2.2.1. ***Example of a Plant: Hydraulic Network***

This example is used to show how a physical process can be modeled using the linear state space model (2.1)*. Figure 2.2 depicts a large-scale hydraulic network which can be found in many industrial and domestic applications. The network consists of a number of connected nodes. Each node in the network is modeled as a single tank, and the reader is referred to e.g., [ÅL92, Joh00] for further discussion of higher dimensional tank processes. In each single tank system, the water level h is controlled by regulating the voltage ϑ to the pump, based on sensor measurements. According to Bernoulli's law [Joh00], the system equation is*

$$\frac{\mathrm{d}h}{\mathrm{d}t} = -\frac{\alpha}{\mathcal{A}}\sqrt{2gh} + \frac{k\vartheta}{\mathcal{A}}, \tag{2.6}$$

where $\mathcal{A}$ and α are the cross-sections of the tank and the outlet hole, respectively. Meanwhile, g is the gravitational acceleration constant and k is a proportional constant relating the flow to the voltage. By sampling with a sampling interval ΔT, the differential equation (2.6) *can be written as a difference equation,*

$$h_{t+1} = h_t - \frac{\alpha \Delta T}{\mathcal{A}}\sqrt{2gh_t} + \frac{k\Delta T}{\mathcal{A}}\vartheta_t. \tag{2.7}$$

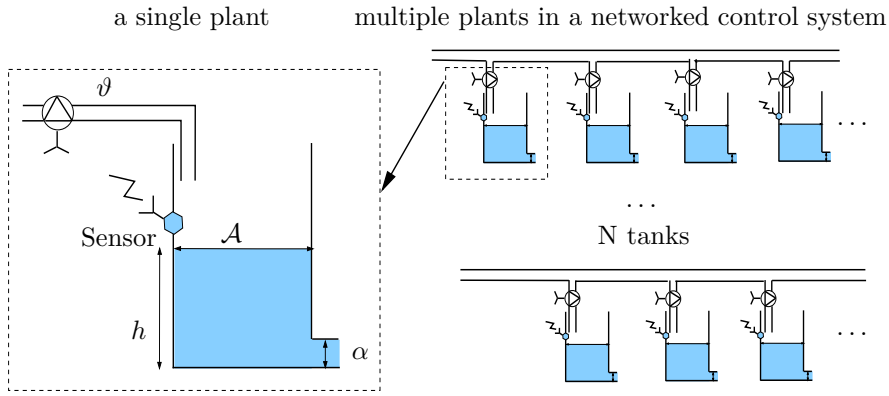


Figure 2.2: A schematic diagram of a networked control system with multiple plants, where each plant can be modeled as a a single tank process. The sensor measurement of the water level is transmitted to the controller by wireless communication.

Next, we linearize (2.7) *around the steady state* (h_{ss}, ϑ_{ss})*, which are related to* (h_t, ϑ_t) *as*

$$h_t = h_{ss} + \Delta h_t$$
$$\vartheta_t = \vartheta_{ss} + \Delta\vartheta_t.$$

Rewriting (2.7) *in terms of* h_{ss} *and* ϑ_{ss}*, gives*

$$h_{ss}+\Delta h_{t+1}=h_{ss}+\Delta h_t-\frac{\alpha\Delta T}{\mathcal{A}}\sqrt{2gh_{ss}}\left(1+\frac{\Delta h_t}{h_{ss}}\right)^{1/2}+\frac{k\Delta T}{\mathcal{A}}(\vartheta_{ss}+\Delta\vartheta_t).$$

The linearization of (2.7) *becomes*

$$\Delta h_{t+1} = \left(1-\frac{\alpha\Delta T}{\mathcal{A}}\sqrt{\frac{g}{2h_{ss}}}\right)\Delta h_t + \frac{k\Delta T}{\mathcal{A}}\Delta\vartheta_t + \mathcal{O}(\Delta h_t^2),$$

because of the following observations. First, at the steady state, it follows that $\mathrm{d}h/\mathrm{d}t=0$*, and* ϑ_{ss} *is related to* h_{ss} *as*

$$\vartheta_{ss} = \frac{\alpha}{k}\sqrt{2gh_{ss}}.$$

Second, we use the Taylor expansion

$$\left(1+\frac{\Delta h_t}{h_{ss}}\right)^{1/2} = 1+\frac{1}{2}\frac{\Delta h_t}{h_{ss}} + \mathcal{O}(\Delta h_t^2),$$

where $\mathcal{O}(\Delta h_t^2)$ *denotes the higher order non-linear terms. By denoting* $x_t = \Delta h_t$ *and* $u_t=\Delta\vartheta_t$*, the single tank process can be described by a scalar version of* (2.1)*, with*

$$A = 1-\frac{\alpha\Delta T}{\mathcal{A}}\sqrt{\frac{g}{2h_{ss}}}, \quad B = \frac{k\Delta T}{\mathcal{A}}, \quad C = 1. \tag{2.8}$$

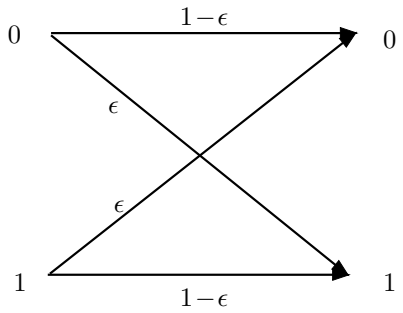


Figure 2.3: A binary symmetric channel, where ϵ is the crossover probability.

Note that, A in (2.8) *is often less than* 1*, which indicates that the plant has stable open-loop dynamics. We have thus shown that the tank system can be modeled as the linear process* (2.1) *where* v_t *could be the modeling error, and* e_t *could be the sensor measurement error.*

Example 2.2.2. ***Binary Symmetric Channel***

A class of discrete memoryless channels, which is one of the most widely studied channel models in coding theory and information theory, is the so-called binary symmetric channel. A binary symmetric channel has binary input and binary output, and it can be simply described by a crossover probability, $\epsilon = \mathcal{P}(0|1) = \mathcal{P}(1|0)$*, see Figure 2.3. Let* b *and* $\hat{b}$ *denote the transmitted and received binary codeword of length* R*, associated to the transmitted index* i *and received index* j*, respectively. Independent transmission for each binary bit leads to the following conditional probability function* $\mathcal{P}(\hat{b}\,|b)$*, in terms of the crossover probability* ϵ*,*

$$\mathcal{P}(\hat{b}\,|b) = (1-\epsilon)^{R - d_H(\hat{b},b)} \epsilon^{d_H(\hat{b},b)},$$

where $d_H(\hat{b}, b)$ *is the* Hamming distance *between the binary codewords* b *and* $\hat{b}$*, i.e., the number of bits by which they differ. Clearly, the binary symmetric channel can be represented by the channel model* (2.4)*.*

2.3 Encoder Side-Information

Since the transmission rate R is limited, the main motivation for using memory-based encoder–controllers is to increase the resolution of quantized observations. For memory-based schemes, the system performance relies heavily on the encoder's knowledge about the controller state and the controller's belief in the encoder state. In the presence of a noisy channel, care has to be taken in specifying how to synchronize the states of the encoder and controller.

Channel errors can give rise to a serious synchronization problem for memory-based encoder–controller pairs. In the case of error-free communications, that is, $\mathbf{i}_0^t = \mathbf{j}_0^t$, there is no uncertainty in the information the controller receives. Hence, the encoder and the controller have identical information about the previous events. Such a synchronization between the encoder and the controller is violated when the channel contains transmission errors. The encoder is not able to make a perfect prediction in advance about the index j_t that the controller will receive. More importantly, the controller has lost the ability to deduce the exact transmitted indices $\mathbf{i}_0^t$ by examining the received $\mathbf{j}_0^t$.

We use the term *encoder side-information* to specify the explicit feedback from the controller to the encoder about the channel outputs $\mathbf{j}_0^{t-1}$. Consequently, *no* SI is the extreme case when there is no feedback information at all about j_t, and *full* SI denotes the situation that the encoder knows exactly the previously received symbols $\mathbf{j}_0^{t-1}$. The latter is the case when the channel is error-free, so that $j_t = i_t$, or when there is an error-free SI channel of the same rate from the output of the forward channel to the encoder.

According to our definition of the SI mapping $z_t = Z_t(j_t)$ in (2.3), we have $z_t = j_t$ and $L_Z = L_J$, when full SI is available; while $z_t = 0$ and $L_Z = 1$, when there is no SI at the encoder. Between the extremes, there are a variety of cases with incomplete SI, for which $1 < L_Z < L_J$. Note that in Figure 2.1 we illustrate the mapping from j_t to z_t as an explicit SI channel, even though this information can be obtained by other means. For example, full SI can be obtained if the encoder knows the previous control commands, $\mathbf{u}_0^{t-1}$, and the controller is an invertible mapping, since then $\mathbf{j}_0^{t-1}$ can be deduced from $\mathbf{u}_0^{t-1}$. In general, when the SI is transmitted back to the encoder via a separate noisy channel, assuming the SI mapping to be deterministic may be optimistic; results based on this assumption then serve as bounds on the achievable performance. However, as we also include the possibility of very low-rate SI feedback (e.g. $L_Z = 2$), noiseless transmission may be motivated by the application of heavy error protection in the SI link.

Via the SI link the encoder will be informed about the potential transmission errors. It is interesting to note that conventional *automatic repeat request* (ARQ) protocols can also be considered as a special case of our model. However, if no retransmission will take place, the controller must be designed to maintain robustness to transmission errors, as the scenarios studied in this thesis.

2.4 Problem Statement

In this section we formulate the optimization problem studied in Chapter 3. Generally speaking, the optimality of a system depends on the design criterion. Our goal is to solve an optimal encoding–control problem and thereby to find the suitable encoder and controller mappings. The adopted performance measure is the following

LQ cost with a finite horizon $T>0$,

$$J_{tot} = \sum_{t=1}^{T} J_t = \sum_{t=1}^{T} x_t' V_t x_t + u_{t-1}' P_{t-1} u_{t-1}, \tag{2.9}$$

where J_t is taken to denote the instantaneous cost. The matrices V_t and P_t are symmetric and positive definite. The design objective is to find the encoder–controller mappings which minimize the expected value of the cost J_{tot}, namely,

$$\mathbf{E}\{J_{tot}\} = \mathbf{E}\left\{\sum_{t=1}^{T} x_t' V_t x_t + u_{t-1}' P_{t-1} u_{t-1}\right\}. \tag{2.10}$$

The LQ cost (2.10) is well established in the scope of stochastic control [Aok67]. There is often a practical relevance in this objective function since the quadratic terms $x_t' x_t$ and $u_t' u_t$ are related to the energy of the state and the control. The matrices V_t and P_t specify the importance factor of the control input with respect to the state. The cost can be interpreted as minimizing the state variance at all time instances, with a power constraint on the control input.

For ease of reference, we define Problem 2.4.1 below as the main optimization problem of the chapter.

Problem 2.4.1. *Consider the system in Figure 2.1. Given the linear plant* (2.1) *and the memoryless channel* (2.4)*, find the encoder* (2.2) *and controller* (2.5) *that minimize the LQ cost* (2.10)*.*

We use the notation $\{f_t^\star(\mathbf{y}_{t_e}^t, \mathbf{z}_0^{t-1})\}_{t=0}^{T-1}$ and $\{g_t^\star(\mathbf{j}_0^t)\}_{t=0}^{T-1}$ for the optimal[1] mappings that solve Problem 2.4.1. Of course, Problem 2.4.1 can be viewed as an extension of the traditional LQ problem, where now the optimal encoder–controller minimizes a cost function influenced by the initial-state, process noise, measurement noise, and the noisy channel.

2.5 Encoder–Controller Operation and Iterative Design

The encoder–controller pair presented in Section 2.2 are functions with memory. They map a continuous range of real values (sensor measurements) into a small set of discrete symbols (control inputs). The small set of symbols is referred to as a codebook, whose entities are control actions labeled with memory-based index sequences. In Figure 2.4, we illustrate the encoder and controller operation at $t=0$ (for simplicity), where $\mathfrak{m}=\mathfrak{p}=2$. The encoder maps a measurement $y_0 \in \mathbb{R}^2$ into an index i_0. For example, the measurement y_0 marked with a square is fed into the encoder, and the corresponding output is $i_0=3$. The index $i_0=3$ is then fed into the channel. After passing the channel, the controller receives $j_0=4$, which is used

[1] Note that the solution is not necessarily unique.

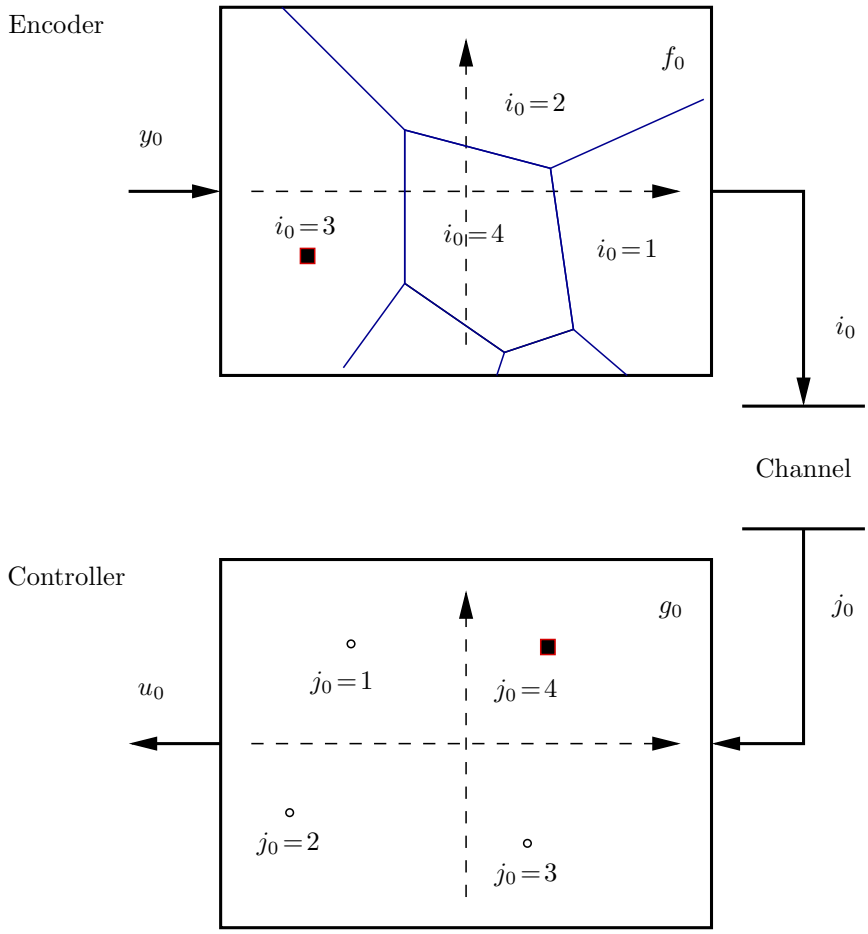


Figure 2.4: The operation of the encoder–controller mappings f_0 and g_0 for $y_0 \in \mathbb{R}^2$. For example, the measurement y_0, marked by a square, is mapped to $i_0 = f_0(y_0) = 3$. The controller receives $j_0 = 4$ and produces the control input $u_0 = g_0(4)$.

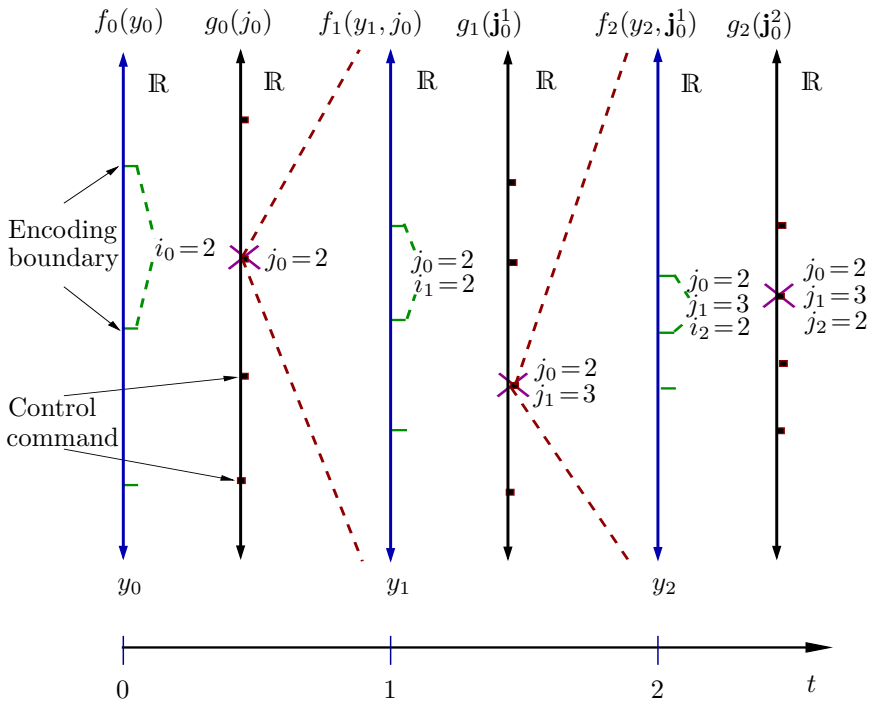


Figure 2.5: The operation of the encoder–controller mappings $\mathbf{f}_0^2$ and $\mathbf{g}_0^2$ for $y_t \in \mathbb{R}$, i.e, $M_e = 0$. In particular, $f_0(y_0)$, $g_0(j_0)$, $f_1(y_1, j_0)$, $g_1(\mathbf{j}_0^1)$, $f_2(y_2, \mathbf{j}_0^1)$ and $g_2(\mathbf{j}_0^2)$ for $\mathbf{j}_0^2 = (2, 3, 2)$ are depicted. For example, y_0 within the interval marked by the dashed lines is mapped by the encoder to $i_0 = 2$. Then, for all $j_0 = 2$, the control input $u_0 = g_0(j_0)$ is marked by a cross. Likewise, given $j_0 = 2$, y_1 within the interval marked by the dashed lines is mapped to $i_1 = 2$; and given $j_0 = 2$ and $j_1 = 3$, the control command $u_1 = g_1(\mathbf{j}_0^1)$ is marked with another cross. Finally, given $j_0 = 2$, $j_1 = 3$, y_2 within the interval marked by the dashed lines is mapped to $i_2 = 2$; and given $j_0 = 2$, $j_1 = 3$ and $j_2 = 2$, the control command $u_2 = g_2(\mathbf{j}_0^2)$ is marked with the third cross.

to produce the control input marked with another square. Note that at $t>0$, the encoder will also use the sequence of past SI, and the controller mapping utilizes the full sequence $\mathbf{j}_0^t$. A simple illustration of the operations of $\mathbf{f}_0^2$ and $\mathbf{g}_0^2$ is given in Figure 2.5. In the example, we assume that $M_e=0$, $y_t\in\mathbb{R}$, and full SI is available at the encoder. By the encoder mapping $f_0(y_0)$, all y_0 in the interval marked by the dashed lines are mapped to $i_0=2$. Then, for all $j_0=2$, the control command $u_0=g_0(j_0)$ is marked with a cross. Similarly, given $j_0=2$, all y_1 within the interval marked of the dashed lines are mapped to $i_1=2$; and given $j_0=2$ and $j_1=3$, the control command $u_1=g_1(\mathbf{j}_0^1)$ is marked with another cross. Finally, given $j_0=2$ and $j_1=3$, the measurement y_2 within the interval marked of the dashed lines will mapped to $i_2=2$; and given $j_0=2$, $j_1=3$ and $j_2=2$, the control command $u_2=g_2(\mathbf{j}_0^2)$ is marked with the third cross.

In general, finding an exact solution to Problem 2.4.1 is a hard problem, because the optimization problem is non-linear and non-convex. Therefore, we propose a method to optimize the encoder–controller pair iteratively, with the goal of finding locally optimal solutions. Roughly speaking, the problem of specifying an encoder mapping is equivalent to characterizing a partition of the measurement space, as illustrated in Figure 2.4 and Figure 2.5. At the controller side, the problem of specifying a controller is to find the values of the admissible controls as a finite set of points in $\mathbb{R}^{\mathsf{m}}$. At time t, there are at most L_J^{t+1} different control inputs, because of the size of the channel output alphabet $\mathcal{L}_J$ and the number of transmissions so far, i.e., $t+1$.

To implement the training-based approach, we define two stages. In the first stage, referred to as the training stage, computations are pursued in a simulated environment to derive an optimized encoder–control pair. While in the second stage, referred to as the operation stage, the trained encoder–controller pair, in form of a look-up table is applied to real data. Illustrated in Figure 2.6 is an example of the operation stage with the trained codebook at the encoder and controller. The training stage is thoroughly studied in Chapters 3–4, leading to an iterative optimization algorithm. The idea of the iterative training is inspired by traditional quantizer design [Far90, GG92], where the encoder is updated by fixing the controller, and the controller is updated by fixing the encoder, back and forth. The iteration terminates when a certain convergence is reached. Unfortunately, this method can not guarantee convergence to a global optimum, but by influencing the initial conditions of the design it is possible to search for good locally optimal designs.

2.6 Summary

In this chapter, we first introduced the general system model for control of a linear plant over a finite-rate noisy channel. The model embodies process noise, measurement noise and a noisy channel with finite input and finite output alphabets. By varying the information pattern at each component, this model can describe a wide range of practical scenarios. Second, the overall design problem is formulated to

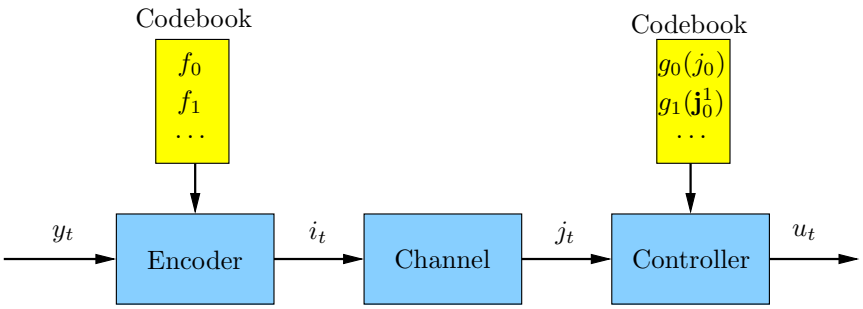


Figure 2.6: The encoder, the channel and the controller. There is a trained codebook at the encoder and the controller based on received symbols $\mathbf{j}_0^t$.

minimize a quadratic cost by optimizing the encoder–controller mappings. Finally, how the encoder–controller pair operate is described by illustrative examples. In the next chapter, criteria for updating the encoder and controller mappings are developed.

Chapter 3

Iterative Design

3.1 Introduction

In this chapter, we propose a joint design of the sensor measurement quantization, channel error protection, and optimal controller actuation, with the objective to minimize the expected linear quadratic cost over a finite horizon. As argued in Chapter 2, this encoder–controller optimization problem is known to be hard in general, and an *iterative design* procedure can be derived inspired by traditional quantizer designs in which the controller is optimized for a fixed encoder, then the encoder is optimized for a fixed controller etc. Unlike in most previous work on *minimum data rate control*, our main concern is optimal average performance over a finite horizon, given a fixed transmission rate. We derive and analyze optimality criteria that we employ to suggest an iterative design algorithm. The performance of the optimized system is then investigated through numerical simulations.

As a significant step in motivating our optimality criteria, we comment on the validity of decoder–controller separation in iterative encoder–controller design. Similar discussions have previously appeared in, for example, [TM04b, TSM04, NFZE07]. However, these works focused on *existence* of optimal systems, while we need a separation criterion that can be used at each step in our design algorithm. As will be discussed below, this leads to some subtle but important differences in the validity of the corresponding results. We also emphasize that in contrast to [TSM04, NFZE07] we consider control over a noisy channel, and in contrast to [MS04b] we allow for general initial-state and noise distributions[1]. In addition, we introduce a general model, not present in previous work, that captures the potential presence of partial SI at the encoder regarding the received symbols at the controller. The details of the system model can be found in Chapter 2.

The remaining part of this section is organized as follows. First, the problem of finding the optimal control policy for a fixed encoder is addressed in Section 3.2. In

[1]That is, not necessarily Gaussian; meaning that sufficient statistics at the encoder can in general not be computed based on a Kalman filter, even in the presence of perfect encoder SI.

Section 3.3, we consider the problem of optimizing one single encoder component, assuming that the controller and other encoder mappings are fixed. The encoding problem for certain special cases of the general settings from Section 2.2 is discussed in more detail to illustrate the impact of the process noise, measurement noise and transmission errors. The overall training algorithm is presented in Section 3.4. Numerical simulations are performed in Section 3.5 to demonstrate the performance of the training method. Finally, a brief summary of the chapter can be found in Section 3.6.

Throughout this chapter, we denote the conditional mean estimator of the state x_s, based on the history of the received indices $\mathbf{j}_0^t$, as

$$\breve{x}_{s|t} \triangleq \mathbf{E}\left\{ x_s | \, \mathbf{j}_0^t \right\}, \quad s \leq t,$$

to distinguish from the previously mentioned conditional mean estimator $\hat{x}_{s|t} = \mathbf{E}\{x_s|\mathbf{y}_0^t\}$. With an ideal communication channel between the sensor and controller, the state measurements are directly available at the controller, and the controller can compute the estimate $\hat{x}_{s|t}$. However, when there is a non-ideal channel between the sensor and controller, the controller has only access to the received data $\mathbf{j}_0^t$ to estimate x_t, which gives $\breve{x}_{s|t}$. For brevity, we will use $\breve{x}_t$ as a short notation for $\breve{x}_{t|t} = \mathbf{E}\{ x_t | \, \mathbf{j}_0^t\}$. Then, let $\tilde{x}_t$ be the estimation error in estimating the state at time t, as

$$\tilde{x}_t = x_t - \breve{x}_t = x_t - \mathbf{E}\left\{ x_t | \, \mathbf{j}_0^t \right\}.$$

3.2 Optimal Controller for Fixed Encoder

This section presents the main results on the problem of characterizing optimal controllers. In particular, we investigate the optimal controller mapping g_t, assuming the encoder $\mathbf{f}_0^{T-1}$ is fixed. We begin with a study of the general case in Section 3.2.1, leading to an optimal control equation which is hard to solve. Thereafter, in Section 3.2.2, we investigate the full SI scenario by introducing a modified system using an "open-loop encoder". The optimal control equation is solved for this type of system. Finally, in Section 3.2.3, we discuss how to use the results in Section 3.2.2 in designing the system for the general case of partial SI.

3.2.1 General Case

The problem of finding the optimal control assuming the encoder is fixed fits well into the setting of stochastic optimal control, e.g., [Aok67]. We apply dynamic programming to derive the optimal control policy recursively. In this case, the observations available at the controller are the integer indices $\mathbf{j}_0^t$. Resembling a classical result in LQ control, we present the following proposition.

Proposition 3.2.1. *Consider a fixed encoder* $\mathbf{f}_0^{T-1}$*. Given the plant* (2.1) *and the memoryless channel* (2.4)*, a controller mapping* (2.5) *that minimizes the LQ*

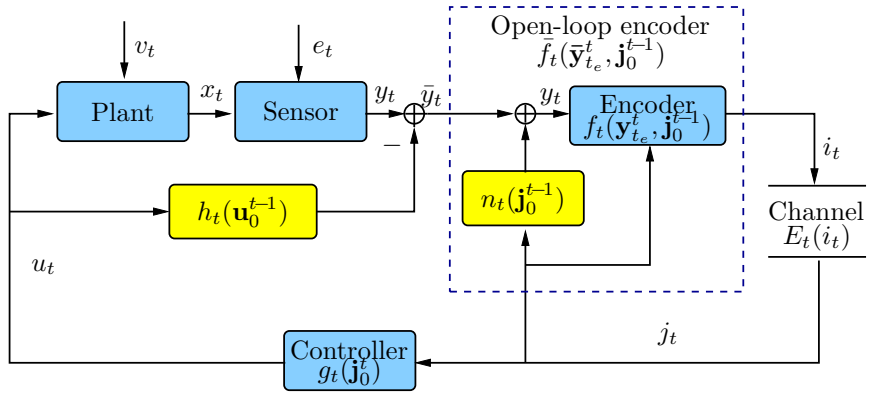


Figure 3.1: An equivalent implementation of the system when a special class of encoders $f_t(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1})$ is employed.

cost (2.10) *fulfills the following recursive relation*

$$\begin{aligned} u_{t-1}^\star &= \arg\min_{u_{t-1}} \gamma_t, \\ \gamma_t &= \lambda_t + \mathbf{E}\left\{ \gamma_{t+1}^\star \middle| \mathbf{j}_0^{t-1} \right\}, \\ \lambda_t &\triangleq \mathbf{E}\left\{ (Ax_{t-1} + Bu_{t-1} + v_{t-1})' V_t (Ax_{t-1} + Bu_{t-1} + v_{t-1}) \right. \\ &\qquad \left. + u_{t-1}' P_{t-1} u_{t-1} \middle| \mathbf{j}_0^{t-1} \right\}, \end{aligned} \tag{3.1}$$

for $t = 1, \ldots, T$, *where the cost-to-go* γ_t *is initialized at* $t = T+1$ *with the optimal cost-to-go* $\gamma_{T+1}^\star = 0$.

The proof of Proposition 3.2.1 is given in Appendix 3.A.

Unfortunately, it is in general not possible to solve (3.1) efficiently. One main obstruction lies in how the term $\mathbf{E}\left\{\gamma_{t+1}^\star | \mathbf{j}_0^{t-1}\right\}$ is affected by past controls. This term is difficult to analyze, since the received indices $\mathbf{j}_0^{t-1}$ are themselves functions of past controls via encoding and transmission. Hence, obtaining an explicit solution to (3.1) is typically not feasible. In the following two subsections we will first investigate the case of full SI and demonstrate that this assumption significantly simplifies the problem. Then we will discuss how to apply the conclusions derived assuming full SI in the general case.

3.2.2 Full Side-Information

An explicit solution to the optimal control problem (3.1) can essentially be obtained only in a few special cases. In this section we look at the special case when the encoder has full SI, $z_t = j_t$. In this case we are able to provide a characterization of the optimal system.

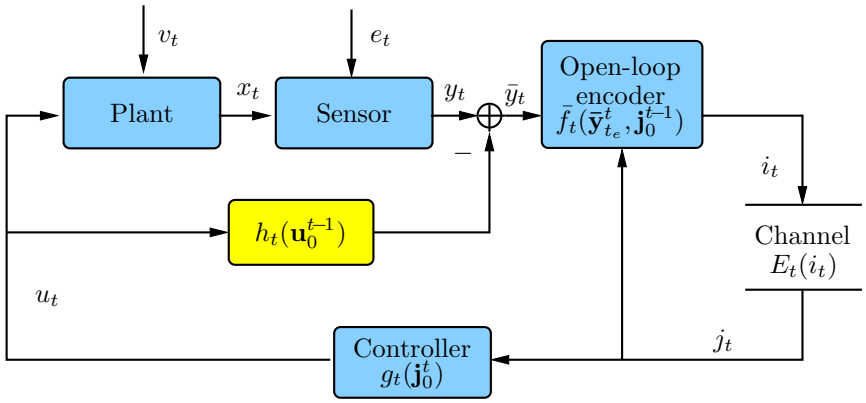


Figure 3.2: A closed-loop control system that employs an open-loop encoder.

Assume that full SI is available at the encoder, $z_t = j_t$. The encoder mapping at t is given as $f_t(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1})$. Now, let us look at the system in Figure 3.1, where the function $h_t(\mathbf{u}_0^{t-1})$ is the mapping that gives

$$\bar{y}_t = y_t - h_t(\mathbf{u}_0^{t-1}) = y_t - \sum_{s=0}^{t-1} CA^{t-1-s}Bu_s.$$

Note that $\bar{y}_t$ is the part of y_t remaining after removing the effect of all previous control commands, so $\bar{y}_t$ depends only on x_0, $\mathbf{v}_0^{t-1}$ and e_t,

$$\bar{y}_t = CA^t x_0 + \sum_{s=0}^{t-1} CA^{t-1-s}v_s + e_t.$$

That is, $\bar{y}_t$ is the equivalent *open-loop measurement*. Furthermore, in Figure 3.1 the mapping $n_t : \mathcal{L}_J^t \mapsto \mathbb{R}^{\mathfrak{p}}$ is defined as

$$n_t(\mathbf{j}_0^{t-1}) \triangleq \sum_{s=0}^{t-1} CA^{t-1-s}Bg_s(\mathbf{j}_0^s),$$

where the output is used to reproduce the measurement y_t. Observe that the only parameters needed to specify h_t are the matrices A, B and C, while n_t requires knowing the controller mappings $\mathbf{g}_0^{t-1}$. That is, if the controller is changed, the mapping n_t also changes. Since the operation carried out by h_t and n_t cancel out, it is obvious that the system in Figure 3.1 gives exactly the same value for the cost (2.9) as the one shown in Figure 2.1, provided the same encoder $\mathbf{f}_0^{T-1}$ and controller $\mathbf{g}_0^{T-1}$ are used.

Now consider the mapping from the open-loop measurement $\bar{y}_t$ and the SI $\mathbf{j}_0^{t-1}$ to the index i_t, as enclosed within the dashed line in Figure 3.1. We call this the *open-loop encoder mapping*. Motivated by Figure 3.1, we define the system in Figure 3.2,

assuming the same controller but a completely general mapping $\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1})$ (that is, one not necessarily related to f_t and n_t as in Figure 3.1). We call the system in Figure 3.2 the *open-loop encoder system*. For a given open-loop encoder system as in Figure 3.2, with the open-loop encoder $\bar{\mathbf{f}}_0^{T-1}$ and controller $\mathbf{g}_0^{T-1}$ specified, one can construct a system with the original structure of Figure 2.1 which has exactly the same performance. The corresponding system in Figure 2.1 utilizes the same controller $\mathbf{g}_0^{T-1}$, together with an encoder whose components are determined by $\bar{\mathbf{f}}_0^{T-1}$ and $\mathbf{g}_0^{T-1}$ as

$$\begin{aligned} &f_t(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}) \\ =&\bar{f}_t\left(y_{t_e} - \sum_{s=0}^{t_e-1} CA^{t_e-1-s}Bg_s(\mathbf{j}_0^s), \ldots, y_t - \sum_{s=0}^{t-1} CA^{t-1-s}Bg_s(\mathbf{j}_0^s), \mathbf{j}_0^{t-1}\right), \ t=0,\ldots,T-1. \end{aligned} \tag{3.2}$$

Note that, $\{\mathbf{j}_0^{-1}\}$ is an empty set, as explained in Chapter 1. For the open-loop encoder system, we formulate the following problem.

Problem 3.2.2. *Consider an open-loop encoder system as illustrated in Figure 3.2. Given the linear plant* (2.1) *and the memoryless channel* (2.4)*, find the open-loop encoder and controller mappings* $\{\bar{f}_t^\star(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1})\}_{t=0}^{T-1}$ *and* $\{g_t^\star(\mathbf{j}_0^t)\}_{t=0}^{T-1}$*, that minimize the LQ cost* (2.10)*.*

It is worthwhile here to emphasize our line-of-thought: Neither of the systems in Figures 3.1–3.2 fits our original model in Figure 2.1, in particular since in Figure 2.1 there is no connection from the controls to the encoder (via the mapping h_t). Hence it is not obvious why formulating Problem 3.2.2 is relevant. In the following propositions we will however demonstrate how the optimal control problem can be solved for the open-loop encoder system in Figure 3.2, while we are not able to solve it in the case of the original system in Figure 2.1. We therefore emphasize here that the purpose of introducing the open-loop encoder system in Figure 3.2 is to have access to a "*virtual help-system*" in the sense that we will optimize the design for this system and argue that the solution is useful also in the original system.

Open-loop systems similar to the one shown in Figure 3.2 were considered also in [TSM04, NFZE07] (for noiseless channels). However, there is a very important, albeit quite subtle, difference in that the equivalent systems in [TSM04, NFZE07] need to use the controller mapping inside the encoder to "open the loop." This is in contrast to our open-loop encoder system in Figure 3.2 where the encoder is a completely arbitrary mapping, and where the loop is opened by subtracting the controls via h_t. This is crucial for the approach in this chapter, since in each step of the iterative design the controller is fixed while updating the encoder, and vice versa. In particular, a controller that perfectly fits the encoder is not known when updating the encoder. Hence, as will be demonstrated, our open-loop encoder system is useful in iterative design, while this is not the case with previously proposed equivalent systems. To our knowledge, this issue has not been settled in previous work.

Now, given the plant, the memoryless channel and the design criterion, the solutions to the original Problem 2.4.1 and the corresponding Problem 3.2.2 for the open-loop encoder system are closely related, as revealed by the following proposition.

Proposition 3.2.3.

1. *Consider a solution* $\{f_t^\star(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$ *to Problem 2.4.1. The same controller* ${\mathbf{g}_0^\star}^{T-1}$ *and the open-loop encoder specified by* $\{f_t^\star(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$ *according to,*

$$\begin{aligned}&\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1})\\ =&f_t^\star\left(\bar{y}_{t_e}+\sum_{s=0}^{t_e-1} CA^{t_e-1-s}Bg_s^\star(\mathbf{j}_0^s),\ldots,\bar{y}_t+\sum_{s=0}^{t-1} CA^{t-1-s}Bg_s^\star(\mathbf{j}_0^s),\mathbf{j}_0^{t-1}\right), t=0,\ldots,T-1,\end{aligned} \tag{3.3}$$

   jointly solve Problem 3.2.2.

2. *Consider a solution* $\{\bar{f}_t^\star(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$ *to Problem 3.2.2. The same controller* ${\mathbf{g}_0^\star}^{T-1}$ *and the encoder* $\{f_t(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1})\}_{t=0}^{T-1}$ *which are specified by the encoder–controller* $\{\bar{f}_t^\star(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$ *according to*

$$\begin{aligned}&f_t(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1})\\ =&\bar{f}_t^\star\left(y_{t_e}-\sum_{s=0}^{t_e-1} CA^{t_e-1-s}Bg_s^\star(\mathbf{j}_0^s),\ldots,y_t-\sum_{s=0}^{t-1} CA^{t-1-s}Bg_s^\star(\mathbf{j}_0^s),\mathbf{j}_0^{t-1}\right), t=0,\ldots,T-1,\end{aligned} \tag{3.4}$$

   jointly solve Problem 2.4.1.

Proof. The proof is based on straightforward observations.

1. If the open-loop encoder $\{\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1})\}_{t=0}^{T-1}$ derived according to (3.3), and ${\mathbf{g}_0^\star}^{T-1}$ do not jointly solve Problem 3.2.2, another solution to Problem 3.2.2 provides a cost lower than the one given by $\{\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$. If this is the case, using the encoder specified by the solution to Problem 3.2.2 according to (3.2), jointly with the controller of the same solution, must lead to a lower cost than the one resulting from $\{f_t^\star(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$. This contradicts the statement that $\{f_t^\star(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$ is a solution to Problem 2.4.1. Hence, the statement in Proposition 3.2.3 must be true.

2. The proof is similar to part 1.

□

Proposition 3.2.3 indicates that a solution to Problem 3.2.2 specifies a solution to Problem 2.4.1, and vice versa. As we will discuss below, when using the iterative

design approach, Problem 3.2.2 is in general easier to solve than Problem 2.4.1. Hence, in the special case of full SI we will focus on finding a solution to Problem 3.2.2, and derive a corresponding solution to Problem 2.4.1 according to (3.4). Again, we note that Problem 3.2.2 is not an equivalent problem (as the system in Figure 2.1 cannot be transformed into the system in Figure 3.2). However, we will argue that striving to solve Problem 3.2.2 will result in encoder–controller pairs that can be converted to fit the original scenario in Figure 2.1. Recall that the iterative design approach alternates between specifying an encoder for a fixed controller and updating the controller for a fixed encoder. In the remaining part of this subsection, we will therefore study the optimal control problem for the open-loop encoder system in Figure 3.2, assuming a *fixed and general open-loop encoder* $\bar{\mathbf{f}}_0^{T-1}$ (this encoder may be completely unrelated to any encoder $\mathbf{f}_0^{T-1}$ in the original system).

Consider now a fixed sequence of open-loop encoder mappings $\bar{\mathbf{f}}_0^{T-1}$ in the open-loop encoder system (that is, $\bar{f}_t$ is fixed and does not change when the controller mappings change). Note that for any such fixed mappings, the transmitted indices $\mathbf{i}_0^t$ do not depend on the controls $\mathbf{u}_0^{t-1}$, since the open-loop measurements $\bar{\mathbf{y}}_{t_e}^t$ does not depend on $\mathbf{u}_0^{t-1}$ and since $i_0 = \bar{f}_0(\bar{y}_0)$, $i_1 = \bar{f}_1(\bar{\mathbf{y}}_{t_e}^1, j_0)$, $i_2 = \bar{f}_2(\bar{\mathbf{y}}_{t_e}^2, j_0, j_1)$, etc., and j_t depends only on i_t and potential channel errors. By virtue of this fact, we will be able to solve (3.1), as revealed by Proposition 3.2.4.

Proposition 3.2.4. *Consider the open-loop encoder system in Figure 3.2, assuming a fixed open-loop encoder* $\bar{\mathbf{f}}_0^{T-1} = \{\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t,\ \mathbf{j}_0^{t-1})\}_{t=0}^{T-1}$. *Given the plant* (2.1) *and the memoryless channel* (2.4), *the controller* $u_t = g_t(\mathbf{j}_0^t)$ *that minimizes the LQ cost* (2.10) *is given by*

$$u_t = \ell_t \breve{x}_t, \tag{3.5}$$

where $\breve{x}_t = \mathbf{E}\{x_t | \mathbf{j}_0^t\}$. *The linear control law* ℓ_t *is recursively computed as*

$$\begin{aligned}
\ell_t &= -(P_t + B'(V_{t+1} + I_{T-t-1})B)^\dagger B'(V_{t+1} + I_{T-t-1})A, \\
I_{T-t-1} &\triangleq A'(V_t + I_{T-t-2})A - \pi_{T-t-1}, \\
\pi_{T-t-1} &\triangleq A'(V_t + I_{T-t-2})B(P_{t-1} + B'(V_t + I_{T-t-2})B)^\dagger \\
&\qquad \times B'(V_t + I_{T-t-2})A,
\end{aligned} \tag{3.6}$$

where I_t *is initialized with* $I_1 = A'V_TA - A'V_TB(P_{T-1} + B'V_TB)^\dagger B'V_TA$, *and* $(\cdot)^\dagger$ *denotes the Moore-Penrose pseudoinverse. The resulting optimal cost-to-go* $\gamma_{t+1}^\star$ *is*

$$\begin{aligned}
\gamma_{t+1}^\star &= \mathbf{E}\left\{ x_t' I_{T-t} x_t + \varpi_{T-t} | \mathbf{j}_0^t \right\}, \\
I_{T-t} &= A'(V_{t+1} + I_{T-t-1})A - \pi_{T-t}, \\
\pi_{T-t} &= A'(V_{t+1} + I_{T-t-1})B(P_t + B'(V_{t+1} + I_{T-t-1})B)^\dagger \\
&\qquad \times B'(V_{t+1} + I_{T-t-1})A, \\
\varpi_{T-t} &= \varpi_{T-t-1} + \mathbf{T_r}\{(V_{t+1} + I_{T-t-1})Q_t\} + \mathbf{E}\left\{ \tilde{x}_t' \pi_{T-t} \tilde{x}_t | \mathbf{j}_0^t \right\}.
\end{aligned} \tag{3.7}$$

The proof of Proposition 3.2.4 is given in Appendix 3.B.

The results (3.5) and (3.6) illustrate that given a fixed open-loop encoder $\overline{\mathbf{f}}_0^{T-1}$, it is possible to characterize the optimal control policy (3.1) explicitly. Observe that the optimal control policy (3.5) is decomposed into a separate estimator/decoder and a controller. Hence, the *separation property* holds [Aok67]. Additionally, one can show that the derived optimal controller (3.5) is a *certainty equivalence* (CE) controller. As explained in Chapter 1, the CE controller in general does not provide optimum performance. In our case, we are able to show that the resulting CE controller in (3.5) is optimal for the open-loop encoder system in Figure 3.2, assuming a fixed open-loop encoder. However, since the open-loop encoder system is not an instance of the original system in Figure 2.1, we cannot claim that the CE controller structure is optimal given a fixed encoder $\mathbf{f}_0^{T-1}$ in the original system.

The CE controller is optimal if the second moment of the estimation error $\tilde{x}_t$ does not depend on past controls [BST74]. This is obviously the case in the open-loop system since the dependence on u_t is removed before encoding. (This is not needed in the classical linear quadratic Gaussian (LQG) problem, where the CE controller is optimal because the estimation error is Gaussian distributed with a fixed variance.) As mentioned, similar approaches are also exploited in, e.g., [TSM04, NFZE07], assuming fully observed plants and noiseless transmission. The problem gets more involved when measurement noise and transmission errors are present. And, again, the corresponding open-loop system in [NFZE07] requires the encoder to have instantaneous access to the controller mapping, while our result is directly applicable to iterative encoder–controller design (as described in Section 3.4).

Given a fixed encoder $\mathbf{f}_0^{T-1}$, used in the original system in Figure 2.1, the CE controller is not necessarily the corresponding optimal control strategy. Still, in the *jointly optimal pair* $\{\mathbf{f}_0^{\star\,T-1}, \mathbf{g}_0^{\star\,T-1}\}$ that solves Problem 2.4.1, the controller $\mathbf{g}_0^{\star\,T-1}$ is a CE controller, as concluded in the following proposition.

Proposition 3.2.5. *If* $\{f_t^\star\left(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}\right), g_t^\star\}_{t=0}^{T-1}$ *solves Problem 2.4.1, the controller* $\mathbf{g}_0^{\star\,T-1}$ *is the CE controller given by* (3.5)–(3.6) *for* $\mathbf{f}_0^{T-1} = \mathbf{f}_0^{\star\,T-1}$.

Proof. Given the linear plant (2.1), the memoryless channel (2.4) and the LQ cost (2.10), one can find a solution $\{\bar{f}_t^\star(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}), g_t^\star\}_{t=0}^{T-1}$ to Problem 3.2.2. According to Proposition 3.2.3, the following encoder

$$\begin{aligned} &f_t^\star(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1}) \\ =&\bar{f}_t^\star\left(y_{t_e} - \sum_{s=0}^{t_e-1} CA^{t_e-1-s}Bg_s^\star(\mathbf{j}_0^s), \ldots, y_t - \sum_{s=0}^{t-1} CA^{t-1-s}Bg_s^\star(\mathbf{j}_0^s), \mathbf{j}_0^{t-1}\right), t=0,\ldots,T-1, \end{aligned}$$

and the controller $\mathbf{g}_0^{\star\,T-1}$ jointly specify a solution to Problem 2.4.1. As shown by Proposition 3.2.4, $\mathbf{g}_0^{\star\,T-1}$ given by (3.5)–(3.6), is a CE controller in the open-loop encoder system. Observe that given $\overline{\mathbf{y}}_{t_e}^t$ and $\mathbf{j}_0^{t-1}$, the encoders $\bar{f}_t^\star(\overline{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1})$ and $f_t^\star(\mathbf{y}_{t_e}^t, \mathbf{j}_0^{t-1})$ produce exactly the same i_t. Therefore, $\breve{x}_t$ and consequently the CE controller are identical for both systems in Problem 2.4.1 and Problem 3.2.2. □

Again, Proposition 3.2.5 states that the optimal controller corresponding to the optimal encoder $\mathbf{f}_0^{\star\, T-1}$ is the CE controller. While, in the general case, the optimal controller corresponding to any encoder $\mathbf{f}_0^{T-1}$ does not necessarily satisfy the separation principle. This is in contrast to the open-loop encoder system, where the optimal controller for any given open-loop encoder is a CE controller. Thus, in designing for the "virtual help-system" defined by the open-loop encoder system, we can use a CE controller in each step of the design. Then, since we know that the controller in the *optimal* pair for the original system can be separated, we employ the design we get for the open-loop system in the original system, via translation. Proposition 3.2.5 motivates this last step.

3.2.3 Partial Side-Information

When only partial SI ($L_Z < L_J$) is available at the encoder in the case of the original system in Figure 2.1, there is no result corresponding to Proposition 3.2.5 that motivates using the open-loop encoder system with a CE controller as a basis for the design. In the optimal encoder–controller pair for the original system, the controller may not be separated without loss in the case of partial SI. Since we are not able to solve (3.1) in the general case, we resort to using the CE controller as a sub-optimal alternative to solving (3.1). That is, in the case of partial SI at the encoder, we propose an iterative design for the original system in Figure 2.1 based on *constraining the controller to be a CE controller*. For completeness, and for later reference, we state the corresponding expressions as follows: Consider the original system Figure 2.1, assuming a fixed encoder $\mathbf{f}_0^{T-1}$. By the "CE controller" for this system, we mean the corresponding mapping (3.5),

$$u_t = \ell_t \breve{x}_t,$$

where $\breve{x}_t = \mathbf{E}\{x_t | \mathbf{j}_0^t\}$, and where the linear control law ℓ_t is computed as

$$\begin{aligned}
\ell_t &= -(P_t + B'(V_{t+1} + I_{T-t-1})B)^\dagger B'(V_{t+1} + I_{T-t-1})A, \\
I_{T-t-1} &= A'(V_t + I_{T-t-2})A - \pi_{T-t-1}, \\
\pi_{T-t-1} &= A'(V_t + I_{T-t-2})B(P_{t-1} + B'(V_t + I_{T-t-2})B)^\dagger B'(V_t + I_{T-t-2})A,
\end{aligned}$$

where I_t is initialized as $I_1 = A'V_TA - A'V_TB(P_{T-1} + B'V_TB)^\dagger B'V_TA$.

3.3 Optimal Encoder for Fixed Controller

In this section, we address the problem of optimizing the encoder component f_t, for a fixed controller $\mathbf{g}_0^{T-1}$ and fixed encoder components $\mathbf{f}_0^{t-1}$ and $\mathbf{f}_{t+1}^{T-1}$. The optimal encoder mapping needs to take the impact of the predicted future state evolutions into account. The following results are a straightforward consequence of the system assumptions and the design criterion.

Proposition 3.3.1. *Consider a fixed controller* $\mathbf{g}_0^{T-1}$ *and fixed encoder components* $\mathbf{f}_0^{t-1}$, $\mathbf{f}_{t+1}^{T-1}$. *Given the linear plant* (2.1) *and the memoryless channel* (2.4), *the encoder component* $f_t(\mathbf{y}_{t_e}^t, \mathbf{z}_0^{t-1})$ *that minimizes the LQ cost* (2.10) *is given by*

$$i_t = \arg\min_{k\in\mathcal{L}_I} \mathbf{E}\left\{ \sum_{s=t+1}^{T} (x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1}) \Big| \mathbf{y}_{t_e}^t, i_t = k, \mathbf{z}_0^{t-1} \right\}. \tag{3.8}$$

Proof. The proof follows the principle of optimality. Recall the LQ cost (2.10), i.e.,

$$\mathbf{E}\{J_{tot}\} = \mathbf{E}\left\{ \sum_{t=1}^{T} x_t' V_t x_t + u_{t-1}' P_{t-1} u_{t-1} \right\}.$$

The mapping f_t influences $\mathbf{E}\{J_{tot}\}$ by producing i_t based on $\mathbf{y}_{t_e}^t$ and $\mathbf{z}_0^{t-1}$. Hence, it influences the states and controls that depend on i_t, i.e., $\mathbf{x}_{t+1}^T$ and $\mathbf{u}_t^{T-1}$.

Let $\mathcal{S}_t(k, \mathbf{z}_0^{t-1})$ denote the set of all $\mathbf{y}_{t_e}^t$ such that $i_t = k \in \mathcal{L}_I$ given $\mathbf{z}_0^{t-1}$. Since $\mathbf{E}\{J_{tot}\} = \mathbf{E}\left\{ \mathbf{E}\{J_{tot} | \mathbf{y}_{t_e}^t, i_t, \mathbf{z}_0^{t-1}\}\right\}$, specifying the optimal mapping f_t is equivalent to specifying the set $\mathcal{S}_t(k, \mathbf{z}_0^{t-1})$ such that the inner term $J(\mathbf{y}_{t_e}^t, k, \mathbf{z}_0^{t-1}) = \mathbf{E}\left\{J_{tot} | \mathbf{y}_{t_e}^t, i_t = k, \mathbf{z}_0^{t-1}\right\}$ is minimized over $k \in \mathcal{L}_I$ given $\mathbf{y}_{t_e}^t$ and $\mathbf{z}_0^{t-1}$. That is

$$\mathcal{S}_t(k, \mathbf{z}_0^{t-1}) \triangleq \{\mathbf{y}_{t_e}^t : J(\mathbf{y}_{t_e}^t, k, \mathbf{z}_0^{t-1}) \le J(\mathbf{y}_{t_e}^t, l, \mathbf{z}_0^{t-1}),\ \ 1 \le l \le L_I\}.$$

Since for any $k, l \in \mathcal{L}_I$,

$$\begin{aligned} &\mathbf{E}\left\{ \sum_{s=1}^{t} x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1} \Big| \mathbf{y}_{t_e}^t, i_t = k, \mathbf{z}_0^{t-1} \right\} \\ &= \mathbf{E}\left\{ \sum_{s=1}^{t} x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1} \Big| \mathbf{y}_{t_e}^t, i_t = l, \mathbf{z}_0^{t-1} \right\}, \end{aligned}$$

because the choice of i_t only influences "future" terms, we equivalently get

$$\mathcal{S}_t(k, \mathbf{z}_0^{t-1}) = \{\mathbf{y}_{t_e}^t : J^+(\mathbf{y}_{t_e}^t, k, \mathbf{z}_0^{t-1}) \le J^+(\mathbf{y}_{t_e}^t, l, \mathbf{z}_0^{t-1}),\ \ 1 \le l \le L_I\},$$

where

$$J^+(\mathbf{y}_{t_e}^t, k, \mathbf{z}_0^{t-1}) \triangleq \mathbf{E}\left\{ \sum_{s=t+1}^{T} x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1} \Big| \mathbf{y}_{t_e}^t, i_t = k, \mathbf{z}_0^{t-1} \right\}.$$

This concludes the proof. □

The encoder is specified by the encoder regions $\mathcal{S}_t(k, \mathbf{z}_0^{t-1})$, $k \in \mathcal{L}_I$, $t = 0, \ldots, T-1$. For $M_e > 1$ and $\mathfrak{p} > 1$, it is often difficult to parameterize these regions, as they are subsets of $\mathbb{R}^{\mathfrak{p}\times(M_e+1)}$, for each possible $\mathbf{z}_0^{t-1}$. However, in the scalar case, $M_e = 1$ and $\mathfrak{p} = 1$, assuming "smooth" (continuous or differentiable) initial-state and noise

pdf's, the regions, $\mathcal{S}_t(k, \mathbf{z}_0^{t-1})$, $k \in \mathcal{L}_I$, are non-overlapping subsets of $\mathbb{R}$, for each possible $\mathbf{z}_0^{t-1}$. Therefore, they can be specified by storing the boundaries between them. In the Gaussian case (Gaussian initial-state and noise pdf's) our numerical results indicate that the regions always become intervals when the training is close to convergence. Hence, $\mathcal{S}_t(k, \mathbf{z}_0^{t-1})$ can often be parameterized by storing at most $L_I - 1$ real values for each possible $\mathbf{z}_0^{t-1}$. However, as it is possible to find controllers that result in disconnected optimal sets $\mathcal{S}_t(k, \mathbf{z}_0^{t-1})$, it cannot be said in general that the encoder regions are intervals. This also means that it is not possible to parameterize the encoder regions as intervals without loss. More studies on the properties of the encoder regions are given in Chapter 4.

In Proposition 3.3.2 below, we present a similar result for the open-loop encoder in an open-loop encoder system. This result is useful when looking for locally optimal solutions to Problem 3.2.2 by iteratively alternating between optimizing the open-loop encoder and the CE controller.

Proposition 3.3.2. *Consider a fixed controller* $\mathbf{g}_0^{T-1}$ *and fixed open-loop encoder components* $\bar{\mathbf{f}}_0^{t-1}$ *and* $\bar{\mathbf{f}}_{t+1}^{T-1}$*. Given the linear plant* (2.1) *and the channel* (2.4)*, the open-loop encoder mapping* $\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1})$ *that minimizes the LQ cost* (2.10) *is given by*

$$i_t = \arg\min_{k \in \mathcal{L}_I} \mathbf{E}\left\{ \sum_{s=t+1}^{T} (x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1}) \middle| \bar{\mathbf{y}}_{t_e}^t, i_t = k, \mathbf{j}_0^{t-1} \right\}. \tag{3.9}$$

The proof is similar to the one of Proposition 3.3.1 with the modification that $\{\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1}\}$ is the information available at the open-loop encoder. Since the CE controller in (3.5) is the optimum control strategy in an open-loop encoder system assuming a fixed open-loop encoder $\bar{\mathbf{f}}_0^{T-1}$, the following result is very useful.

Corollary 3.3.3. *Consider an open-loop encoder system. There exists a solution to Problem 3.2.2 that satisfies the following conditions: The controller mapping is given by* $u_t = \ell_t \breve{x}_t$ *with* ℓ_t *as in* (3.6)*; and the open-loop encoder mapping* $\bar{f}_t$ *is given by*

$$i_t = \arg\min_{k \in \mathcal{L}_I} \mathbf{E}\left\{ \sum_{s=t}^{T-1} \tilde{x}_s' \pi_{T-s} \tilde{x}_s \middle| \bar{\mathbf{y}}_{t_e}^t, i_t = k, \mathbf{j}_0^{t-1} \right\}, \tag{3.10}$$

where π_{T-t} *is given in* (3.7)*.*

Proof. By Proposition 3.2.4 we know, given a fixed open-loop encoder (in this case the optimal open-loop encoder), the CE controller (3.5) is the optimal control strategy. Also, it is clear from (3.7) that, the choice of i_t influences only the term $\mathbf{E}\left\{ \sum_{s=t}^{T-1} \tilde{x}_s' \pi_{T-s} \tilde{x}_s \middle| \mathbf{j}_0^t \right\}$ in cost-to-go γ_{t+1}. This concludes the proof. □

In the case of full SI or the open-loop encoder system, the encoding rule (3.9) is replaced with (3.10), in order to reduce computations. Also notice that the encoder

in (3.10) directly minimizes the *weighted estimation error*

$$\mathbf{E}\left\{\tilde{x}_s'\pi_{T-s}\tilde{x}_s\right\}, \quad s=t,\ldots,T-1.$$

Hence there is "separation" between quantization/coding/estimation and control, since the encoder tries to minimize the (weighted) average mean-squared error and since the CE controller splits into computing the estimate $\breve{x}_t$ and the scaling ℓ_t. The encoding is influenced by the control problem only via the matrices $\pi_{T-s}, s=t,\ldots,T-1$, and computing $\breve{x}_t$ at the receiver side can be interpreted as decoding or estimation.

3.3.1 Optimal Encoding for some Special Linear Systems and Channels

This subsection is devoted to a discussion of the encoder design in Proposition 3.3.1 for certain special cases of the general system described in Chapter 2. In particular, we compare the scenarios where process noise and measurement noise are absent and the communication link from the encoder to the controller is error-free. The following results, stated without proof, are a direct consequence of Proposition 3.3.1.

Corollary 3.3.4. *Consider a fixed controller* $\mathbf{g}_0^{T-1}$*, and the fixed encoder components* $\mathbf{f}_0^{t-1}$ *and* $\mathbf{f}_{t+1}^{T-1}$*, for a linear plant* (2.1) *and a memoryless channel* (2.4)*.*

1. *If* $v_t=e_t=0$ *and* $j_t=i_t$*, the encoder mapping* $f_t(\mathbf{y}_{t_e}^t,\mathbf{z}_0^{t-1})$ *that minimizes the LQ cost* (2.10) *is given by*

$$i_t=\arg\min_{k\in\mathcal{L}_I}\mathbf{E}\left\{\sum_{s=t+1}^{T}(x_s'V_sx_s+u_{s-1}'P_{s-1}u_{s-1})\Big|\mathbf{x}_{t_e}^t,i_t=k,\mathbf{z}_0^{t-1}\right\}, \quad (3.11)$$

   where the expectation is over the initial-state distribution.

2. *If* $e_t=0$ *and* $j_t=i_t$*, the encoder mapping* $f_t(\mathbf{y}_{t_e}^t,\mathbf{z}_0^{t-1})$ *that minimizes the LQ cost* (2.10) *is given by* (3.11)*, where the expectation is over the initial-state and the process noise distributions.*

3. *If* $e_t=0$ *and* $j_t\neq i_t$*, the encoder mapping* $f_t(\mathbf{y}_{t_e}^t,\mathbf{z}_0^{t-1})$ *that minimizes the LQ cost* (2.10) *is given by* (3.11)*, where the expectation is over the initial-state distribution, the process noise distribution and the channel distribution.*

In all the three cases above, the optimal encoder is described by (3.11). However, the same expression leads to different levels of computational complexity. In the first case, the initial-state is the only uncertainty involved in the estimation of the current controller state and the prediction of the future evolution. In the second case, the expectation takes not into account only the distribution of the initial-state but also the distribution of the process noise. In the last case, the complexity is further increased, in order to additionally take transmission errors into consideration.

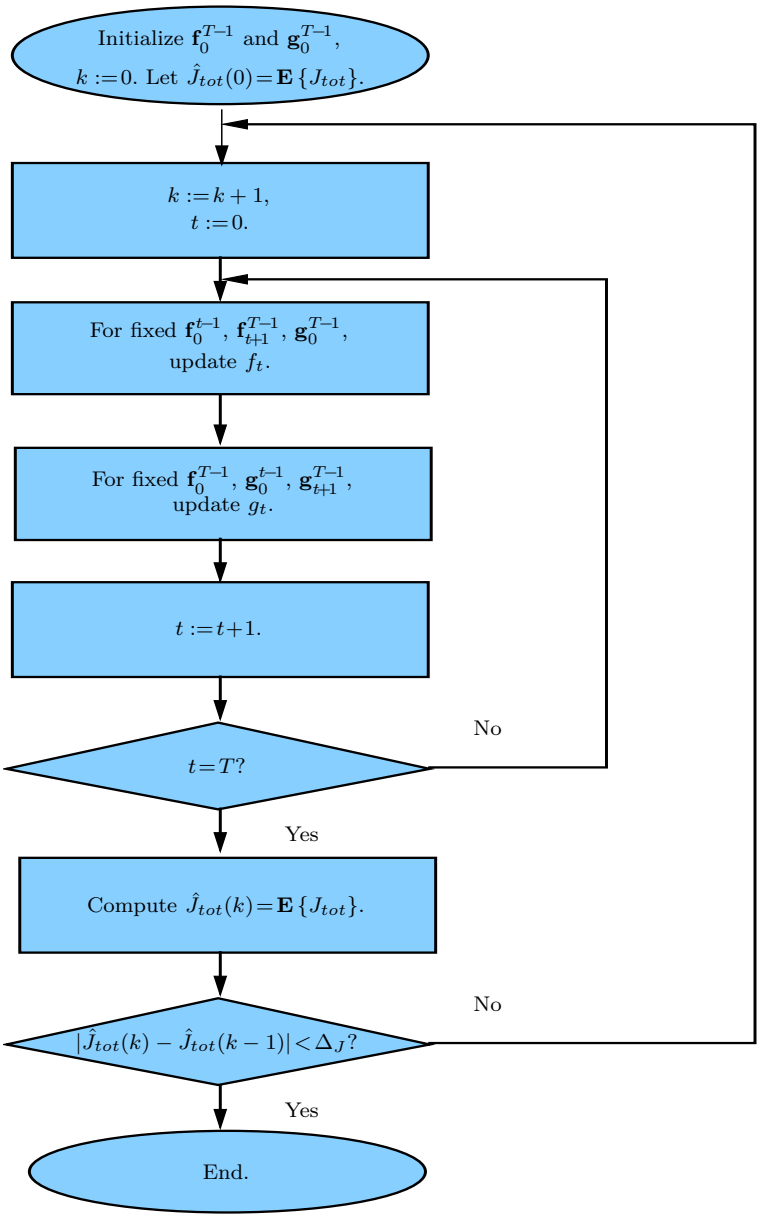


Figure 3.3: The flow-diagram of the iterative encoder–controller optimization procedure. The variable k is a counter for the number of rounds. In each round, all the mappings $\mathbf{f}_0^{T-1}$ and $\mathbf{g}_0^{T-1}$ are updated. The value $\hat{J}_{tot}(k)$ represents the resulting cost $\mathbf{E}\{J_{tot}\}$ after the k^{th} round. The iteration is terminated when the improvement in the system performance is less than a certain threshold Δ_J.

3.4 Iterative Encoder–Controller Design Algorithm

In this section, we propose an encoder–controller design algorithm based on the results from Section 3.2 and Section 3.3. The design is suitable in particular for low transmission rates, and since the algorithm strives to find good encoder–controller pairs for a given noisy channel, it introduces protection against transmission errors. That is, the result is a design for *joint quantization, error protection and control*. Since the controller information $\mathbf{j}_0^t$ has finite resolution, there are only a finite number of possible control commands. Thus, these values can be pre-calculated and stored in a codebook at the controller.

As mentioned, the overall joint encoder–controller optimization problem is typically not tractable, and we therefore propose to optimize the encoder–controller pair iteratively. There are two cases to handle separately:

1. Full encoder SI: In this case, as discussed in Section 3.2.2, we carry out the design for the open-loop encoder system and then use Proposition 3.2.3 to convert the solution to the original problem in Figure 2.1.

2. Partial encoder SI: In this case, we constrain the controller to be a CE controller, as discussed in Section 3.2.3, and carry out the design for the original system in Figure 2.1.

Figure 3.3 depicts a flow-diagram of the design procedure, with notation for the original system (the case of partial SI). An initial encoder–controller pair is specified. Thereafter, each encoder–controller component, f_0, g_0, $\ldots$, f_{T-1}, g_{T-1}, is successively optimized. After one round, if the improvement is not below a predefined threshold Δ_J, a new round is started to update f_0, g_0, $\ldots$, f_{T-1}, g_{T-1}. At time t, first f_t is updated for fixed $\mathbf{f}_0^{t-1}$, $\mathbf{f}_{t+1}^{T-1}$, $\mathbf{g}_0^{T-1}$ and then g_t is updated for fixed $\mathbf{f}_0^{T-1}$, $\mathbf{g}_0^{t-1}$, $\mathbf{g}_{t+1}^{T-1}$. The currently derived f_t will replace the former one to be regarded as a fixed component in optimizing the other components $\mathbf{f}_0^{t-1}$, $\mathbf{f}_{t+1}^{T-1}$ and $\mathbf{g}_0^{T-1}$, until next time when f_t is updated. The rules for updating the encoder–controller components are developed in Section 3.2 and Section 3.3. In particular, we update the encoder component f_t using (3.8) and the controller g_t using (3.5)–(3.6).

In order to jointly optimize the open-loop encoder and controller in an open-loop encoder system, in the case of full encoder SI, we use instead Proposition 3.2.4 to update the CE controller. For the encoding rule, we can use the encoding rule (3.9), or (3.10).

The complexity of the design and the quality of the solution are important practical issues. The complexity issues will be studied in detail in Chapter 4, while a few remarks on the quality are given here.

In principle (and neglecting e.g. problems with numerical accuracy) the iterative design procedure always converges to a local optimum in the case of the open-loop encoder system. This is because the CE controller in Proposition 3.2.4 is optimal for any $\bar{\mathbf{f}}_0^{T-1}$. That is, each time Proposition 3.2.4 is invoked, given an updated encoder, the performance can only remain the same or improve. Similarly, each time the

encoder is updated for a given controller, the resulting performance cannot decline. This is the usual rationale behind proving convergence for iterative training-based designs [SG86, GG92]. Again, in principle the design converges to a stationary point, but nothing can in general be said about the global optimality of this point [SG86].

However, in the general case (with only partial encoder SI) the design does not necessarily converge at all, since in the step where the controller is updated the expressions in (3.5)–(3.6) do not necessarily lead to a new controller with better performance, because the CE controller is not necessarily optimal for the given encoder. Still, in our numerical experiments this has not been a problem, and empirically the design algorithm appears to converge to a solution also in the general case. Loosely speaking, the explanation is that after a few iteration the fact that the new controller is a better CE controller than the previous one makes the performance improve (while during the first few iterations the performance need not improve in each step, depending on the initialization).

3.5 Numerical Examples

In this section we present numerical experiments conduced to demonstrate the performance of the encoder–controller proposed in Section 3.4. For the simplicity of the presentation, we consider a linear scalar plant for which the most recent measurement is encoded and transmitted over a binary symmetric channel. More specifically, the system equations are

$$\begin{aligned} x_{t+1} &= ax_t + u_t + v_t, \quad a > 0, \\ y_t &= x_t + e_t, \end{aligned}$$

and the linear quadratic cost J_{tot} in the performance measure $\mathbf{E}\{J_{tot}\}$ is

$$J_{tot} = \sum_{t=1}^{T} x_t^2 + \rho u_{t-1}^2, \quad \rho \geq 0,$$

where ρ is the weighting parameter specifying the penalty on the control input. The initial-state x_0, process noise v_t and measurement noise e_t are mutually independent and modeled as $\mathcal{N}(0, \sigma_{x_0}^2)$, $\mathcal{N}(0, \sigma_v^2)$, $\mathcal{N}(0, \sigma_e^2)$, respectively.

Because the theoretically optimal performance is not established, in the experiments presented here two special cases are employed as the reference systems. Briefly, we mention them here. First, it is obvious that the best performance is achieved when the channel imperfections are absent. In the case of the ideal channel, Problem 2.4.1 is reduced to the classical LQG problem whose solution can be regarded as the best achievable performance. On the other hand, a control is useless if it results in a system performance worse than without control. Therefore, we adopt the non-control system to verify the inefficiency of an encoder–controller pair.

In what follows, we describe the four types of coding–control schemes which are used in the experiments.

1. U-Kalman I

The first type, referred to as U-Kalman I, uses a time-invariant uniform quantizer to quantize state measurements. At the controller, the received index j_t is mapped into a reconstruction value which is then fed into a Kalman filter for the estimation of the state x_t. The Kalman filter is designed assuming the error cased by measurement noise, quantization distortion, and transmission errors, is white and Gaussian distributed. Thereafter, the control is calculated as a linear function of the Kalman filter output. The linear feedback control law is ℓ_t in (3.6). By this method, the distortion due to the quantization and channel error is treated as parts of measurement noise. Note that, the "extended" measurement noise (including quantization error and channel error) is in fact neither Gaussian nor uncorrelated with the state and the process noise, which makes the Kalman filter a sub-optimal estimator.

2. U-Kalman II

The second type, referred to as U-Kalman II, adopts a time-invariant uniform quantizer to quantize the state measurements. At the controller, the decoded symbols are fed into a Kalman filter to estimate x_t. The Kalman filter is designed assuming that the channel is absent and the system is not exposed to the quantization and transmission errors. Finally, the control is a linear function of the Kalman filter output that the linear control law is ℓ_t (3.6).

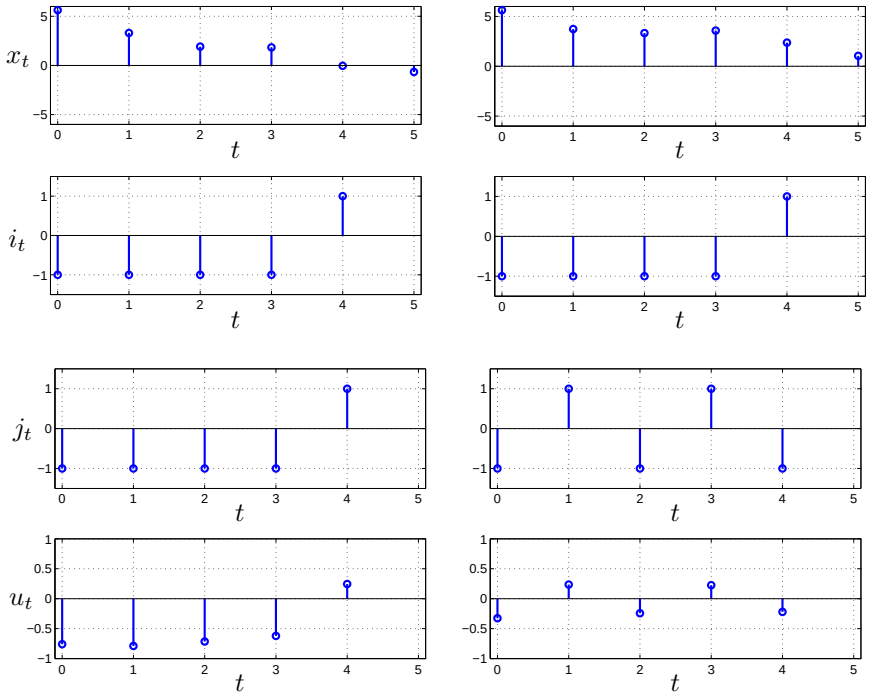
3. U-CE

The third type, referred to as U-CE, utilizes a time-invariant uniform encoder to encode the state measurements, and a CE controller at the receiver side. That is to say $u_t = \ell_t \mathbf{E}\{x_t | \mathbf{j}_0^t\}$, cf., (3.5). The linear control law ℓ_t is recursively computed according to (3.6). This class of coding–control scheme is chosen in the interest of demonstrating the significance of the CE controller. As shown later in the simulations, the CE controller is important to the overall performance.

4. Proposed Encoder–Controller

The last type, referred to as the proposed encoder–controller, is trained iteratively as proposed in Section 3.4, for an encoder of full SI and $M_e = 0$. As explained previously, the initial settings of the encoder–controller will affect the iteration result. In the experiment we use U-CE as the initial setup, which often yields good result.

The overall system performance is determined by a variety of system parameters. The relations among them are complicated, therefore a number of experiments were pursued to investigate the important system parameters, such as data rate, channel errors, SI etc. In what follows, we will discuss each of them in detail.



(a) $\epsilon = 0.04$ (b) $\epsilon = 0.3$

Figure 3.4: System performance is shown with respect to the crossover probability ϵ. The state response x_t, the transmitted symbol i_t, the received symbol j_t and the control u_t are depicted. In this example, $\epsilon = 0.04$ results in no transmission error and $\epsilon = 0.3$ in three errors.

Channel Errors

How does the closed-loop system respond to channel errors is one of the main problems studied in this thesis. First, we assess the impact of the crossover probability ϵ on the state response. In Figure 3.4, the state response of the system, by using the proposed encoder–controller, is depicted together with the transmitted index i_t, the received index j_t, and the control input u_t. The system parameters are chosen in the interest of demonstrating the impact of control. In particular, the system parameters are: $a = 0.9$, $\rho = 0.5$, $T = 5$ and $R = 1$. The initial-state and the noises are modeled as $\mathcal{N}(0, 3)$, $\mathcal{N}(0, 0.5)$ and $\mathcal{N}(0, 0.5)$, respectively. The system has been studied for the crossover probabilities $\epsilon = 0.04$ and $\epsilon = 0.3$. It can be observed from Figure 3.4 that the number of symbol errors increases with ϵ. Then, since a symbol error might result in a control command doing more harm than help, as expected,

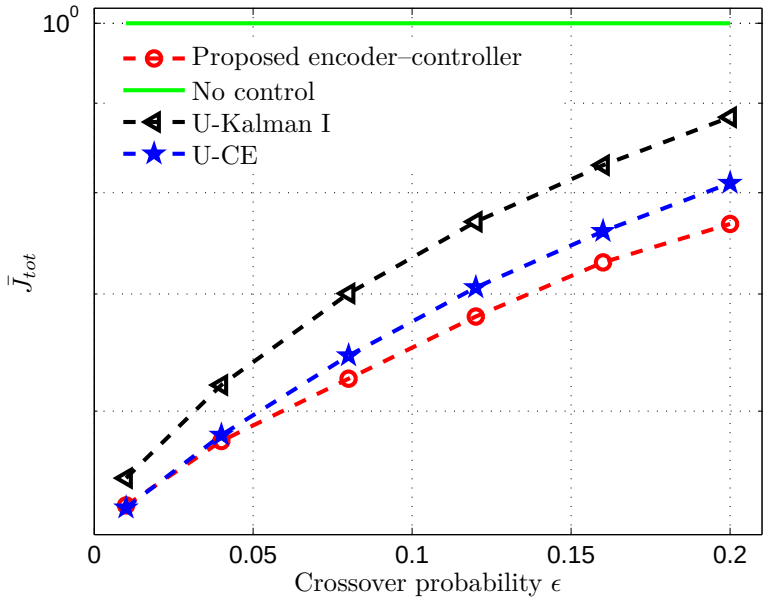


Figure 3.5: A performance comparison among the proposed encoder–controller and two other encoder–controllers, namely U-Kalman I and U-CE.

the magnitudes of the admissible controls become smaller as the number of channel errors increases.

In Figure 3.5, we compare the system performance of the proposed encoder–controller with another two schemes, namely U-Kalman I and U-CE. The system parameters are: $a=1.1$, $\rho=0.5$, $T=2$ and $R=2$. The initial-state and the noises are modeled as $\mathcal{N}(0,5)$, $\mathcal{N}(0,0.5)$ and $\mathcal{N}(0,0.5)$, respectively. The scheme U-Kalman I has the encoder thresholds $\{-2,0,2\}$ and the reconstruction values $\{-3,-1,1,3\}$. The scheme U-CE has also the encoder thresholds $\{-2,0,2\}$. The performance measure $\bar{J}_{tot}$ is given by normalizing $\mathbf{E}\{J_{tot}\}$ with the expected cost obtained when no control action is taken, cf., the horizontal line in Figure 3.5. It can be seen from the figure that the proposed encoder–controller pair outperforms the other two coding–control schemes evidently. Compared with the scenario using U-CE, the improvement in performance of the proposed encoder–controller is moderate. While, compared with the scenario using U-Kalman I, the improvement of the proposed encoder–controller is significant. This observation indicates that the gain obtained by the proposed encoder–controller appears to be mostly attributed to the CE controller.

How do the proposed encoder and controller respond to the increasing channel

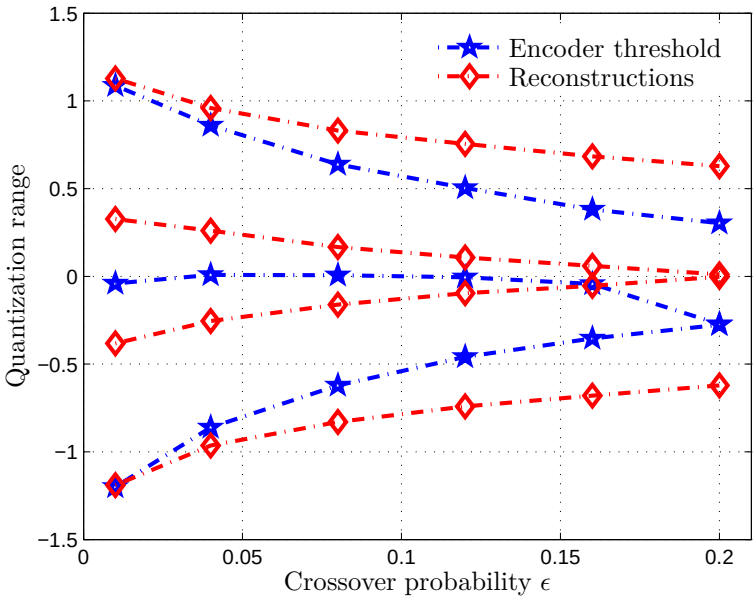


Figure 3.6: The encoding boundaries given by f_0 and the associated reconstructions are shown with respect to the crossover probability ϵ.

error is illustrated in Figure 3.6. The system parameters are the same as in Figure 3.5. In the figure, we demonstrate the partition by the encoder mapping f_0, and the reconstructions $\breve{x}_0$, with respect to the growing ϵ. Recall that the control u_0 is a linear function of $\breve{x}_0$. We note that the number of different control inputs chosen by the encoder decreases with increasing ϵ. This phenomenon, attributed to the varying abilities of binary codewords in combating channel errors, is well-known in quantization for noisy channels. That is, for channels with high error probability, it is beneficial to transmit only the "stronger" codewords [Far90], providing true redundancy for error protection. Note that, the asymmetry at $\epsilon = 0.16$ is also a consequence of the fact that the binary codewords are unequally sensitive to the channel errors. Another impact of increasing ϵ is that the encoder thresholds and the controls are all moved closer to zero, indicating that only small-valued control actions are allowed.

In Figure 3.7, a comparison of all the four types of coding–control schemes, U-Kalman I, U-Kalman II, U-CE, and the proposed encoder–controller, is depicted. Especially, the scheme U-CE is displayed for several step lengths of the uniform encoder. The system parameters are: $a = 0.7$, $\rho = 0.5$, $T = 3$ and $R = 3$. The initial-state and the noises are modeled as $\mathcal{N}(0, 5)$, $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 1)$, respectively.

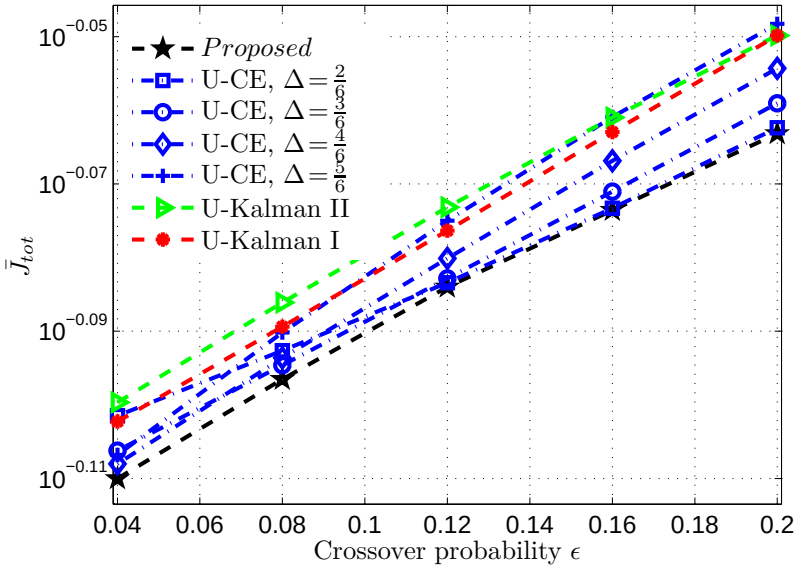


Figure 3.7: A performance comparison among various controller-encoders with respect to the crossover probability ϵ. The scheme U-CE employs a uniform encoder and a CE controller, where Δ is the step length of the uniform encoder. The schemes U-Kalman I and U-Kalman II employ a uniform encoder and a Kalman filter.

From the figure we see that the deterioration in system performance might not be significant if the time-invariant uniform encoder is properly designed. Note that a time-invariant encoder could be viewed as a performance bound for encoders with no SI. All uniform encoders in Figure 3.7 have near-optimal step lengths. However, if the step length is chosen improperly, it can have severe consequences.

Rate

In Figure 3.8 we demonstrate the impact of using different transmission rates together with measurement noise. The cost $\bar{J}_{tot}$ is given by normalizing $\mathbf{E}\{J_{tot}\}$ with the expected cost obtained without any control. The system parameters are: $a=0.7$, $\rho=0.5$, and $T=3$. The initial-state and the process noise are modeled as $\mathcal{N}(0,5)$, $\mathcal{N}(0,1)$ respectively. For all uniform encoders, we let the boundaries be kept equally spaced between -2 and 2. Accordingly, the maximum quantization error in the saturated region decreases with the increasing transmission rate. It can be seen in the figure that the proposed encoder–controller pair outperforms the other three coding–control schemes. Given the same encoder, the system employing a CE controller always performs better than the systems employing the Kalman filters. Still,

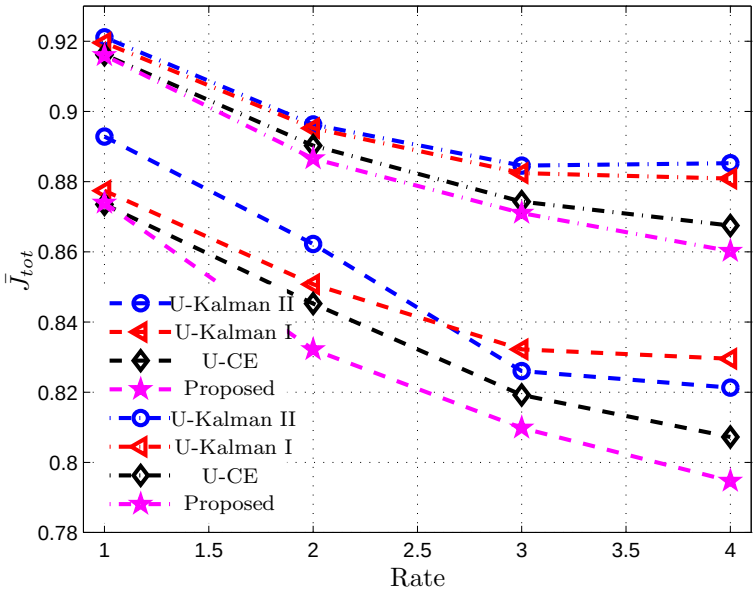


Figure 3.8: A performance comparison among various encoder–controllers with respect to the rate R. The dashed lines are for $\sigma_e^2 = 1$, and the dash–dot lines are for $\sigma_e^2 = 4$. The crossover probability ϵ is 0.1.

the figure shows that the gain obtained by the proposed encoder–controller appears to be mostly attributed to the CE controller. Another interesting observation is that U-Kalman I is not necessarily always superior to U-Kalman II. That means, the way U-Kalman I handles the quantization distortion and transmission errors may do more harm than good.

Side-Information

SI affects the system in several ways. In Figure 3.9, we show a comparison of different degrees of SI when the encoder $\mathbf{f}_0^{T-1}$ is optimized. In particular, we explore no, incomplete and full SI scenarios. In the experiment, the incomplete SI is generated as follows. The least significant bit of the binary received codeword is discarded and the resulting codeword is fed back to the encoder over a noiseless link. The system parameters are the same as in Figure 3.5, i.e., $a = 1.1$, $\rho = 0.5$, $T = 2$ and $R = 2$, and the initial-state and the noises are modeled as $\mathcal{N}(0, 5)$, $\mathcal{N}(0, 0.5)$ and $\mathcal{N}(0, 0.5)$, respectively. We demonstrate the convergence properties of the training algorithm by showing the successive iteration results $\bar{J}_{tot}$ after each round. The figure shows that the improvement given by knowing SI is moderate, since the opti-

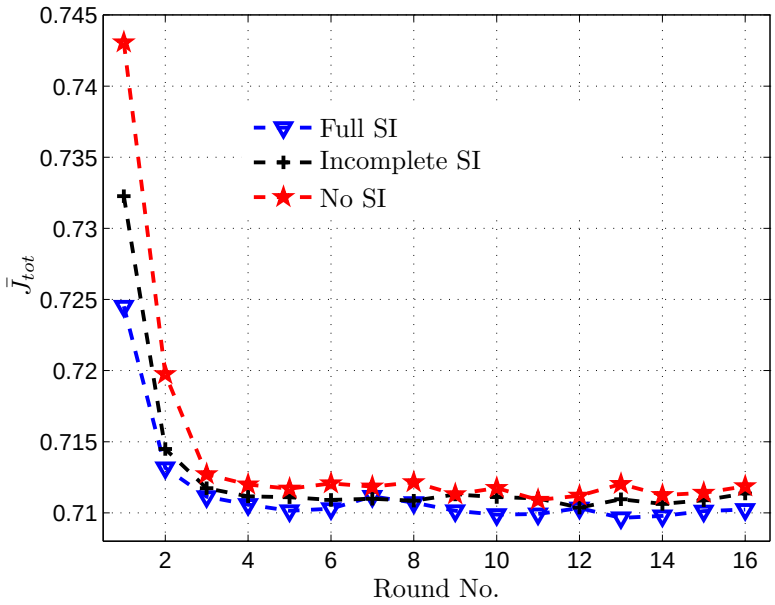


Figure 3.9: A performance comparison among full SI, incomplete SI and no SI scenarios.

mized encoder–controller leads to close densities $p(x_t|\, y_t, \mathbf{z}_0^{t-1})$, irrespective of $\mathbf{z}_0^{t-1}$. However, it is expected that the advantage of having SI becomes more significant when time horizon increases, because SI can reduce the encoder uncertainty in the controller state which accumulates with time.

3.6 Summary

This chapter has investigated the joint optimization of the encoder and the controller in closed-loop control of a linear plant with a low-rate feedback link over a memoryless noisy channel. We introduced an iterative approach to the design of encoder–controller pairs. In deriving design criteria, we arrived at new results regarding the optimality of certainty equivalence controllers in our setting. In the case of full encoder SI, we introduced a "virtual help-system," the open-loop encoder system. We showed that a CE controller is optimal for any given encoder in this system, and we argued that encoder–controller pairs designed for the help-system can be translated to perform well in the original system. In the case of partial SI, we cannot claim that enforcing the CE controller structure is without loss. However, since the general controller problem is challenging in this case,

we used CE controllers as sub-optimal, but practically feasible, approximations. We have performed various numerical investigations. Our results demonstrate the promising performance obtained by employing the proposed design algorithm. We also investigated the impact on system performance of various degree of SI at the encoder.

3.A Proof of Proposition 3.2.1

Proof. According to the principle of optimality, e.g., [Ber76], the truncated control sequence $\mathbf{u}_t^{\star T-1} = \{u_t^\star, \ldots, u_{T-1}^\star\}$ is the optimal policy for the truncated problem where the cost is summarized from time $t{+}1$ to T. Hence, the optimal u_t is the one that minimizes the sum of the future costs, i.e.,

$$u_t^\star = \arg\min_{u_t} \mathbf{E}\left\{ \sum_{s=t+1}^{T} x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1} \Big| \mathbf{j}_0^t \right\}.$$

Let us start the recursive derivation at the last stage $t = T$. Since, the optimal cost-to-go at $t{=}T{+}1$ is zero, i.e., $\gamma_{T+1}^\star{=}0$, the optimal u_{T-1} is the one minimizing λ_T, namely,

$$\begin{aligned}\lambda_T &= \mathbf{E}\left\{ x_T' V_T x_T + u_{T-1}' P_{T-1} u_{T-1} \big| \mathbf{j}_0^{T-1} \right\} \\ &= \mathbf{T_r}\left\{V_T Q_{T-1}\right\} + \int_{\mathbb{R}^n} (x_{T-1}' A' V_T A x_{T-1} + 2u_{T-1}' B' V_T A x_{T-1} \\ &\quad + u_{T-1}'(P_{T-1} + B' V_T B) u_{T-1}) p\big(x_{T-1} \big| \mathbf{j}_0^{T-1}\big)\, \mathrm{d}x_{T-1},\end{aligned}$$

where the term $Q_t = \mathbf{E}\left\{v_t v_t'\right\}$ denotes the covariance matrix of the process noise. Since the noise variance $Q_{T-1} = \mathbf{E}\left\{v_{T-1} v_{T-1}'\right\}$ is independent of the control u_{T-1}, the control u_{T-1} minimizing λ_T is then given by

$$u_{T-1}^\star = -(P_{T-1} + B' V_T B)^\dagger B' V_T A \breve{x}_{T-1}, \tag{3.12}$$

where $\breve{x}_{T-1} = \mathbf{E}\left\{ x_{T-1} \big| \mathbf{j}_0^{T-1} \right\}$ and $(\cdot)^\dagger$ denotes the Moore-Penrose pseudoinverse. Substituting $u_{T-1}^\star$ into λ_T, the optimal cost-to-go at $t{=}T$ is

$$\begin{aligned}\gamma_T^\star = \min_{u_{T-1}} \lambda_T &= \mathbf{E}\left\{ x_{T-1}' I_1 x_{T-1} + \varpi_1 \big| \mathbf{j}_0^{T-1} \right\}, \\ I_1 &\triangleq A' V_T A - \pi_1, \\ \pi_1 &\triangleq A' V_T B (P_{T-1} + B' V_T B)^\dagger B' V_T A, \\ \varpi_1 &\triangleq \mathbf{T_r}\left\{V_T Q_{T-1}\right\} + \mathbf{E}\left\{ \tilde{x}_{T-1}' \pi_1 \tilde{x}_{T-1} \big| \mathbf{j}_0^{T-1} \right\}.\end{aligned} \tag{3.13}$$

At the second last stage $t{=}T-1$, the optimal u_{T-2} is the one minimizing γ_{T-1}, as

$$\begin{aligned}u_{T-2}^\star &= \arg\min_{u_{T-2}} \gamma_{T-1}, \\ \gamma_{T-1} &= \lambda_{T-1} + \mathbf{E}\left\{ \gamma_T^\star \big| \mathbf{j}_0^{T-2} \right\}, \\ \lambda_{T-1} &= \mathbf{E}\left\{ x_{T-1}' V_{T-1} x_{T-1} + u_{T-2}' P_{T-2} u_{T-2} \big| \mathbf{j}_0^{T-2} \right\}.\end{aligned} \tag{3.14}$$

Generalizing to any time t, the optimal control $u_{t-1}^\star$ can be obtained by solving the following equation

$$u_{t-1}^\star = \arg\min_{u_{t-1}} \gamma_t = \arg\min_{u_{t-1}} \left\{ \lambda_t + \mathbf{E}\left\{ \gamma_{t+1}^\star \big| \mathbf{j}_0^{t-1} \right\} \right\},$$

where γ_t is the cost-to-go, given by

$$\begin{aligned}\gamma_t &= \lambda_t + \mathbf{E}\left\{\gamma^\star_{t+1}\middle|\mathbf{j}_0^{t-1}\right\},\\ \lambda_t &= \mathbf{E}\left\{(Ax_{t-1}+Bu_{t-1}+v_{t-1})'V_t(Ax_{t-1}+Bu_{t-1}+v_{t-1})\right.\\ &\qquad \left.+u'_{t-1}P_{t-1}u_{t-1}\middle|\mathbf{j}_0^{t-1}\right\}.\end{aligned}$$

Note that $\gamma_0 = \mathbf{E}\{J_{tot}\}$. This concludes the proof. □

As mentioned previously, it is in general not possible to solve (3.1) efficiently. One main obstruction lies in how the term $\mathbf{E}\left\{\gamma^\star_{t+1}\middle|\mathbf{j}_0^{t-1}\right\}$ is affected by past controls. Consider for example $t=T-1$. The quantity $\mathbf{E}\left\{\tilde{x}'_{T-1}\pi_1\tilde{x}_{T-1}\middle|\mathbf{j}_0^{T-1}\right\}$ in $\gamma^\star_T$, is difficult to analyze, since the received index j_{T-1} is itself a function of u_{T-2} via encoding and transmission. Hence, obtaining an explicit solution to (3.1) is typically not feasible.

3.B Proof of Proposition 3.2.4

Proof. According to Figure 3.2, the open-loop encoder has access to the open-loop measurement $\bar{\mathbf{y}}_{t_e}^t$ and $\mathbf{j}_0^{t-1}$. When the sequence of mappings $\bar{\mathbf{f}}_0^{T-1}$ is fixed, we can verify that the estimation error $\tilde{x}_t$ is not a function of $\mathbf{u}_0^{t-1}$ as follows

$$\tilde{x}_t = x_t - \breve{x}_t = A^t x_0 + \sum_{s=0}^{t-1} A^{t-1-s}v_s - \mathbf{E}\left\{A^t x_0 + \sum_{s=0}^{t-1} A^{t-1-s}v_s\middle|\mathbf{j}_0^t\right\},$$

since the controls $\mathbf{u}_0^{t-1}$ are completely determined by the received symbols $\mathbf{j}_0^{t-1}$. As $\{x_0, \mathbf{v}_0^{T-1}, \mathbf{e}_0^{T-1}\}$ are not affected by $\mathbf{u}_0^{T-1}$, one can show by the following induction that the indices $\mathbf{j}_0^t$ are not functions of $\mathbf{u}_0^{t-1}$. Start the induction at $t=0$. The statement holds true at $t=0$ since

$$i_0 = \bar{f}_0(\bar{y}_0) = \bar{f}_0(Cx_0+e_0), \quad j_0 = E_0(i_0), \quad z_0 = j_0.$$

Assuming for the moment the statement is valid for time t, as already validated for $t=0$, then at time $t+1$,

$$i_{t+1} = \bar{f}_{t+1}(\bar{\mathbf{y}}_{t_e}^{t+1}, \mathbf{j}_0^t), \quad j_{t+1} = E_{t+1}(i_{t+1}), \quad z_{t+1} = j_{t+1},$$

which hence do not involve $\mathbf{u}_0^t$. Therefore, $\mathbf{i}_0^{T-1}$ and $\mathbf{j}_0^{T-1}$ depend only on $\{x_0, \mathbf{v}_0^{T-1}, \mathbf{e}_0^{T-1}\}$ and potential channel errors, but not on $\mathbf{u}_0^{T-1}$.

The fact that the estimation error $\tilde{x}_t = x_t - \breve{x}_t$ is not a function of $\mathbf{u}_0^{t-1}$ for the fixed mappings $\{\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t, \mathbf{j}_0^{t-1})\}_{t=0}^{T-1}$ will significantly simplify the derivation of the optimal control. According to previous calculations, the optimal u_{T-1} is given by (3.12), irrespective of the encoding. To derive the optimal u_{T-2}, let us consider (3.13) and (3.14). Since the covariance of the estimation error $\tilde{x}_{T-1}$ is independent of $\mathbf{u}_0^{T-2}$, we can find the optimal u_{T-2} as

$$\begin{aligned}u^\star_{T-2} &= \ell_{T-2}\breve{x}_{T-2},\\ \ell_{T-2} &= -(P_{T-2} + B'(V_{T-1}+I_1)B)^\dagger B'(V_{T-1}+I_1)A,\end{aligned}$$

and the optimal cost-to-go at $t=T-1$ can be written as

$$\begin{aligned}\gamma^\star_{T-1} &= \mathbf{E}\left\{ x'_{T-2} I_2 x_{T-2} + \varpi_2 \middle| \mathbf{j}_0^{T-2} \right\}, \\ I_2 &= A'(V_{T-1}+I_1)A - \pi_2, \\ \pi_2 &= A'(V_{T-1}+I_1)B(P_{T-2}+B'(V_{T-1}+I_1)B)^\dagger B'(V_{T-1}+I_1)A, \\ \varpi_2 &= \varpi_1 + \mathbf{T_r}\left\{(V_{T-1}+I_1)Q_{T-2}\right\} + \mathbf{E}\left\{ \tilde{x}'_{T-2}\pi_2\tilde{x}_{T-2} \middle| \mathbf{j}_0^{T-2} \right\},\end{aligned} \tag{3.15}$$

with $\{I_1, \pi_1, \varpi_1\}$ as given in (3.13).

Assume $u^\star_{t+1}$ for time $t+1$ is established, as already validated for $t=T-1$, let us continue the derivation of $u^\star_t$. According to Proposition 3.2.1, the optimal control $u^\star_t$ is the one solving the following equation,

$$u^\star_t = \arg\min_{u_t} \left\{ \lambda_{t+1} + \mathbf{E}\left\{ \gamma^\star_{t+2} \middle| \mathbf{j}_0^t \right\} \right\},$$

where $\lambda_{t+1} = \mathbf{E}\left\{ x'_{t+1}V_{t+1}x_{t+1} + u'_t P_t u_t \middle| \mathbf{j}_0^t \right\}$. By (3.15), $\gamma^\star_{t+2}$ is written as

$$\begin{aligned}\gamma^\star_{t+2} &= \mathbf{E}\left\{ x'_{t+1} I_{T-t-1} x_{t+1} + \varpi_{T-t-1} \middle| \mathbf{j}_0^{t+1} \right\}, \\ \varpi_{T-t-1} &= \varpi_{T-t-2} + \mathbf{T_r}\left\{(V_{t+2}+I_{T-t-2})Q_{t+1}\right\} \\ &\qquad + \mathbf{E}\left\{ \tilde{x}'_{t+1}\pi_{T-t-1}\tilde{x}_{t+1} \middle| \mathbf{j}_0^{t+1} \right\}.\end{aligned}$$

Since $\mathbf{E}\left\{ \tilde{x}'_{t+1}\pi_{T-t-1}\tilde{x}_{t+1} \middle| \mathbf{j}_0^t \right\}$ does not depend on $\mathbf{u}_0^{t-1}$ at any time instance t, we can obtain the optimal u_t by solving the following equation,

$$u^\star_t = \arg\min_{u_t} \left\{ \mathbf{E}\left\{ x'_{t+1}V_{t+1}x_{t+1} + u'_t P_t u_t \middle| \mathbf{j}_0^t \right\} + \mathbf{E}\left\{ x'_{t+1} I_{T-t-1} x_{t+1} \middle| \mathbf{j}_0^t \right\} \right\},$$

and the solution is given by

$$u^\star_t = \ell_t \breve{x}_t, \quad \ell_t = -(P_t + B'(V_{t+1}+I_{T-t-1})B)^\dagger B'(V_{t+1}+I_{T-t-1})A.$$

Finally, the resulting optimal cost-to-go $\gamma^\star_{t+1}$ is obtained by replacing u_t in γ_{t+1} with $u^\star_t$,

$$\begin{aligned}\gamma^\star_{t+1} &= \mathbf{E}\left\{ x'_t I_{T-t} x_t + \varpi_{T-t} \middle| \mathbf{j}_0^t \right\}, \\ I_{T-t} &= A'(V_{t+1}+I_{T-t-1})A - \pi_{T-t}, \\ \pi_{T-t} &= A'(V_{t+1}+I_{T-t-1})B(P_t + B'(V_{t+1}+I_{T-t-1})B)^\dagger \\ &\qquad \times B'(V_{t+1}+I_{T-t-1})A, \\ \varpi_{T-t} &= \varpi_{T-t-1} + \mathbf{T_r}\left\{(V_{t+1}+I_{T-t-1})Q_t\right\} + \mathbf{E}\left\{ \tilde{x}'_t \pi_{T-t}\tilde{x}_t \middle| \mathbf{j}_0^t \right\}.\end{aligned}$$

□

Chapter 4

Complexity Considerations

4.1 Introduction

This chapter is devoted to several issues related to the implementation of the training method developed in Chapter 3, as well as the design of complexity reduced systems. In the special case studied in this thesis, where sensor measurements are fed to the controller through a low-rate noisy communication link, typically a few bits per sensor measurement, the validity of the source–channel–control separation principle is questionable. This is a situation where a jointly designed encoder–controller is expected to perform better by carrying out the quantization, channel protection and control simultaneously. Of course, the joint encoder–controller should be designed with much care, since here the overall system performance relies only on a few number of bits. In Chapter 3 we proposed an iterative method to optimize encoder–controller mappings to provide efficient control over noisy channels. The main topic of this chapter is to describe how the training procedure implemented in practice.

This chapter is organized as follows. In Sections 4.2–4.5, the complexity of the training algorithm is discussed. We start in Section 4.2 with a few general remarks on the advantages and disadvantages of the iterative design, and the common assumptions for problems in this chapter are specified. Due to the lack of closed-form solutions, Monte Carlo methods are heavily involved in the optimization of the encoder and controller, as described in Section 4.3. Thereafter, the complexity of the encoder is thoroughly studied since it appears to be a crucial factor in determining whether or not the training is implementable in practice. In particular, two special encoder properties, namely *sufficient statistics* and *regularity*, which are very useful to the practical implementation of the training algorithm, are addressed in Section 4.4 and Section 4.5. Section 4.6 is entirely devoted to the practical encoder–controller designs for long time horizons. Especially, some conventional solutions customized for control over long time horizons are re-examined in the context of control with limited information. Finally, a short summary of the chapter is provided in Section 4.7.

4.2 General Remarks

In this section, we provide a few general remarks on the encoder–controller design described in Chapter 3. Especially, the main advantages and disadvantages of the training-based approach are addressed. Finally, certain common assumptions about the problems studied in this chapter are specified.

4.2.1 Advantages and Disadvantages

As explained previously, finding the globally optimal solution to Problem 2.4.1 is an open problem. We have shown that under certain conditions locally optimal solutions can be obtained by means of iterative training. In general, the training-based algorithm described in Chapter 3 suffers from *the curse of dimensionality* [Bel61]. In our case, the *dimensionality problem* is mainly caused by the exponential increase in complexity as the time horizon increases. Even for systems with low rate, the training process is by no means simple to implement. The difficulties will be described in detail in Section 4.3. Generally speaking, training becomes impractical when confronting any of the following challenges: a long time horizon T, a high system-order, or a high data rate. It is therefore of great importance to find efficient and simple approximations and simplifications to construct practical encoder–controllers which can simultaneously offer high performance and tractable complexity. This issue will be further studied in Section 4.6.

The major drawback of training is the demand of a large number of computations, as described later in Section 4.3. In the meantime, for a large group of sensor nodes subject to the same system parameters, training is an off-line process which only needs to be performed once.

Hence, while carrying out the design can be computationally intensive, the result can often be parameterized as a finite set of parameters, and stored in a look-up table. The table can be implemented in many cheap notes deployed in a large network. In the scalar case, the look-up table typically consists of a finite number of partition boundaries of the encoding regions and control inputs. Thus, the proposed system, when designed, can in general be *used* at low or moderate complexity. Since the training can be carried out off-line, its complexity is not of crucial importance.

Before we start to describe the training algorithm in more detail, let us first in Section 4.2.2 specify the system discussed in this chapter.

4.2.2 Preliminary

For the exposition of the basic principles, our discussion throughout this chapter is restricted to the scalar special case of the general system described in Chapter 2. The system equation is

$$\begin{aligned} x_{t+1} &= ax_t + u_t + v_t, \quad a > 0, \\ y_t &= x_t + e_t, \end{aligned} \tag{4.1}$$

and the quadratic cost J_{tot} is defined by

$$J_{tot} = \sum_{t=1}^{T} x_t^2 + \rho u_{t-1}^2, \quad \rho \geq 0. \tag{4.2}$$

Here, a and ρ are scalar system parameters. The relative weight ρ can be interpreted as a power constraint imposed on control, and typically the average power of u_t reduces as ρ increases. The assumptions about all elements in (4.1) follow Chapter 2. Especially, we assume:

1. The instantaneous rate is time invariant, i.e., $R = \log_2 L$.
2. The entire history of measurements, $\mathbf{y}_0^t$, is available at the encoder.
3. Full SI, $\mathbf{j}_0^{t-1}$, is available at the encoder.
4. The channel output shares the same alphabet with the channel input.

$$\mathcal{L}_I = \mathcal{L}_J = \mathcal{L} = \{0, \ldots, L-1\}, \quad L = 2^R.$$

It should be clear from Chapter 3 that for the general system with partial SI, the optimal encoding and control mappings are interconnected to each other in a highly complex manner. However, in the special case of full SI, there are useful results on the separation of the optimal coding and control. For this reason, throughout this chapter, we will restrict the discussion to the special case of full SI. When full SI is available, the sub-problems involved in the design are often simplified, however they may still be nontrivial. An understanding of full SI systems offers insights that are useful to the practical encoder–controller design for partial SI. Furthermore, the full SI solutions serve as the best achievable performance with respect to all levels of partial SI.

According to Chapter 3, for this setup the optimal encoder and controller mappings at t, for fixed other encoder–controller components, are

$$\begin{aligned} g_t\big(\mathbf{j}_0^t\big) &= \ell_t \mathbf{E}\left\{ x_t \middle| \mathbf{j}_0^t \right\}, \\ f_t\big(\mathbf{y}_0^t, \mathbf{j}_0^{t-1}\big) &= \arg\min_{k \in \mathcal{L}} \mathbf{E}\left\{ \sum_{s=t+1}^{T} x_s^2 + \rho u_{s-1}^2 \middle| \mathbf{y}_0^t, i_t = k, \mathbf{j}_0^{t-1} \right\}, \end{aligned} \tag{4.3}$$

with ℓ_t given by (3.6). How these terms are computed in practice is discussed in the subsequent sections.

4.3 Numerical Solution and Monte Carlo Method

In this section we study how to compute the expressions (4.3) in practice. We start with the controller mapping g_t. First, the linear control law ℓ_t is independent of actual realizations, and hence it is not affected by updating $\mathbf{f}_0^{T-1}$ and $\mathbf{g}_0^{T-1}$. We can thus compute the entire sequence ℓ_0^{T-1} in advance. As a matter of fact, only

the conditional mean estimate, $\mathbf{E}\{x_t|\mathbf{j}_0^t\}$, is required to be re-calculated whenever one of the encoder–controller components f_s or g_s, $s<t$, has been updated. For this reason, we may say that the complexity of the controller is mainly determined by the conditional mean estimate $\mathbf{E}\{x_t|\mathbf{j}_0^t\}$. Conversely, the updating of f_s or g_s, $t<s\leq T-1$, does not affect $\mathbf{E}\{x_t|\mathbf{j}_0^t\}$.

Often, it is not possible to express $\mathbf{E}\{x_t|\mathbf{j}_0^t\}$ in a closed-form. On the other hand, $\mathbf{E}\{x_t|\mathbf{j}_0^t\}$ is completely determined by the system model and system parameters, e.g., the mappings $\mathbf{f}_0^{t-1}$ and $\mathbf{g}_0^{t-1}$, and the pdf's (or pmf's), $p(x_0)$, $p(v_s)$, $p(e_s)$, $\mathcal{P}(j_s|\,i_s)$, $s=0,\ldots,t$. We might think, once all information above is available, we can at least compute $\mathbf{E}\{x_t|\mathbf{j}_0^t\}$ by numerical methods, for example by using Monte Carlo methods to generate a set of samples describing the pdf $p(x_t|\mathbf{j}_0^t)$. Start with the sample set of x_0, and then, successively generate the sample sets for y_0, i_0, j_0, u_0, x_1 etc, until we finally obtain the sample set describing the conditional pdf $p(x_t|\mathbf{j}_0^t)$.

Unfortunately, even using Monte Carlo methods in place of analytical expressions, the design problem may still be infeasible. First, the size of the sample set is important to the accuracy of simulated results. More specifically, in order to have sufficient data for an accurate estimation, the sample set has to increase significantly with T. Therefore, Monte Carlo methods are difficult to use for problems of long time horizons. Consider a simple example where we compute f_0 numerically, according to (4.3), for fixed other encoding–control mappings. In order to derive a decision for a given input y_0, the future cost for each $i_0\in\mathcal{L}$ should be simulated. Start by computing the future cost for a fixed i_0 by encoding $\mathbf{y}_0^1$ using $f_1(\mathbf{y}_0^1,j_0)$. To ensure that there is a sample set of $\mathbf{y}_0^1$ of the size N for each pair $\{i_0,j_0\}$, it requires that the sample set of x_0 should be at least of the size $L^2\times N$, with $L=2^R$. Continuing up to $t=T-1$ and performing $f_{T-1}(\mathbf{y}_0^{T-1},\mathbf{j}_0^{T-2})$, the sample set of x_0 is required to be at least of the size $L^T\times N$.

Second, more seriously, the numerical simulation might fail because of the complexity of the encoder. Even though the above-mentioned problem of the sample size can be handled by a powerful computer, the numerical simulations might still fail because the encoder mappings (4.3) is not practically implementable. Consider the worst case when the encoder mappings are given by the implicit expression (4.3), for all t, without any further simplification. It is not difficult to realize that we will encounter a serious problem because of the dimensionality of the sequence $\mathbf{y}_0^t$, referred hereby to as the *encoder dimensionality problem*. In order to simulate a decision (4.3), all future decisions $f_s(\mathbf{y}_0^s,\mathbf{j}_0^{s-1})$, $s>t$, are required. Each future decision will require in its turn the simulation of their future costs etc. Furthermore, since it is unrealistic to store the resulting i_t for all possible $\{\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\}$, $t=0,\ldots,T-1$, we must perform the simulation whenever an encoding decision is required. Obviously, the overall encoding time will be impractical. We may say generally that the computational complexity of $\mathbf{E}\left\{x_t|\,\mathbf{j}_0^{t-1}\right\}$ is essentially determined by the encoder complexity. Only when the encoding process can be completed within an acceptable time, for all t, then it is possible to estimate $p(x_t|\mathbf{j}_0^t)$ and derive $\mathbf{E}\{x_t|\mathbf{j}_0^t\}$. If that is the case, the controller can be simplified to a look-up table, which consists of a

finite number of possible reconstructions $g_t(\mathbf{j}_0^t)$. In short, a simplified encoder may result in a simplified controller. But if the encoder cannot be simplified, it will be difficult to simplify the controller. For this reason, a further study of the encoder is provided in the next two sections.

4.4 Two Special Encoder Properties

In this section we discuss two special encoder properties that are useful to solve the aforementioned dimensionality problem. As previously pointed out in Section 4.3, the encoder complexity is a crucial factor to whether or not we are able to implement the training algorithm introduced in Chapter 3. Since it is difficult to express f_t of (4.3) in a closed-form, in general, we need to resort to numerical methods. Consider updating f_t according to (4.3) by using Monte Carlo methods, for certain sequences $\mathbf{y}_0^t$ and $\mathbf{j}_0^{t-1}$. First, generate a sample set of x_t, describing $p(x_t|\,\mathbf{y}_0^t, \mathbf{j}_0^{t-1})$. Then, use those samples to simulate an expected future cost. However, the simulations can fail totally because of the encoder dimensionality problem. It is also worth mentioning that since all encoder mappings are interconnected, if any one of the encoder mappings has a serious dimensionality problem, it affects all other mappings.

The encoder dimensionality problem is crucial to the methods based on numerical simulations, and it is mainly caused by two facts: (*i*) the measurement memory increases as time goes, and (*ii*) the measurement alphabet is infinite. However, we note that the dimensionality problem can be solved for certain special encoders. In the subsequent text, we demonstrate two special encoder properties that are useful to deal with the dimensionality problem in training. They are, namely, (*i*) sufficient statistics, by which a high-dimensional mapping $f_t(\mathbf{y}_0^t, \mathbf{j}_0^{t-1})$ can be replaced by a low-dimensional mapping, and (*ii*) regularity, by which the encoder can be represented by a finite number of thresholds (or the endpoints of the quantization cell). These two properties are first demonstrated by examples in the remaining part of this section. Thereafter, the property of regularity will be discussed in more detail in Section 4.5.

4.4.1 Sufficient Statistics

Let us begin with the first-mentioned property, the sufficient statistics, which is an important concept in statistics. This concept is introduced by Ronald Fisher, and can be found in many textbooks, e.g., [Aok67]. Loosely speaking, regarding the encoder mapping in (4.3), a statistic $I_t(\mathbf{y}_0^t, \mathbf{j}_0^{t-1})$ for which it holds that

$$p(x_t|\mathbf{y}_0^t, \mathbf{j}_0^{t-1}) = p(x_t|I_t(\mathbf{y}_0^t, \mathbf{j}_0^{t-1})),$$

is called a sufficient static. We illustrate the utility of the sufficient statistics by using a simple example which is a straightforward application of classical LQG control theory.

Example 4.4.1. ***Sufficient Statistics for a Gaussian System***

Recall the problem setting from Section 4.2 that all past measurements $\mathbf{y}_0^t$ *and full SI are available at the encoder. Pay attention to the fact that knowing* $\mathbf{j}_0^{t-1}$ *is equivalent to knowing* $\mathbf{u}_0^{t-1}$*. Hence, it is straightforward to show that conditioned on knowing* $\mathbf{y}_0^t$ *and* $\mathbf{j}_0^{t-1}$*, the state* x_t *is Gaussian distributed,*

$$p\big(x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\big)\sim\mathcal{N}\Big(\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\},\mathbf{E}\left\{\big(x_t-\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\}\big)^2\right\}\Big).$$

It means that the pdf $p(x_t|\mathbf{y}_0^t,\mathbf{j}_0^{t-1})$ *is completely determined by the conditional expected value* $\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\}$*, and the variance* $\mathbf{E}\left\{(x_t-\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\})^2\right\}$*. They can be calculated by using the classical Kalman filter [Aok67]. For brevity, upon denoting the conditional variance*

$$p_{t|s}\triangleq\mathbf{E}\left\{\big(x_t-\mathbf{E}\left\{x_t|\,\mathbf{y}_0^s,\mathbf{j}_0^{s-1}\right\}\big)^2\right\},$$

the Kalman filter computes $\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\}$ *and* $p_{t|t}$ *recursively as*

$$\begin{aligned}
\mathbf{E}\left\{x_t|\,\mathbf{y}_0^{t-1},\mathbf{j}_0^{t-1}\right\}&=a\mathbf{E}\left\{x_{t-1}|\,\mathbf{y}_0^{t-2},\mathbf{j}_0^{t-2}\right\}+u_{t-1}+aK_{t-1}(y_{t-1}-\mathbf{E}\left\{x_{t-1}|\,\mathbf{y}_0^{t-2},\mathbf{j}_0^{t-2}\right\}),\\
\mathbf{E}\left\{x_t|\,\mathbf{y}_0^{t},\mathbf{j}_0^{t-1}\right\}&=\mathbf{E}\left\{x_t|\,\mathbf{y}_0^{t-1},\mathbf{j}_0^{t-1}\right\}+K_t\left(y_t-\mathbf{E}\left\{x_t|\,\mathbf{y}_0^{t-1},\mathbf{j}_0^{t-1}\right\}\right),\\
K_t&=p_{t|t-1}\left(p_{t|t-1}+\sigma_e^2\right)^{-1},\\
p_{t|t-1}&=a^2p_{t-1|t-2}+\sigma_v^2-aK_{t-1}p_{t-1|t-2},\\
p_{t|t}&=a^2p_{t|t-1}+\sigma_v^2,
\end{aligned}\tag{4.4}$$

where initially $p_{0|-1}=\sigma_{x_0}^2$*. It is worth noting that the variance* $p_{t|t}$ *is independent of actual realizations, and therefore, the conditional mean* $\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\}$ *is a sufficient statistic.*

Remark 4.4.2. *In the classical LQG problem subject to an ideal communication channel, the conditional mean estimate* $\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{u}_0^{t-1}\right\}$ *is computed at the controller. In the case that the control* u_t *is completely determined by* $\mathbf{y}_0^t$*, it holds that* $\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{u}_0^{t-1}\right\}=\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t\right\}$*. In our setting of control over finite-rate noisy channels, two different conditional mean estimates are computed at the encoder and controller. First, the encoder can compute* $\mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\}$*. Here we should observe that unlike in the classical LQG problem, in order to derive* u_t*,* $\mathbf{j}_0^{t-1}$ *is needed, in place of* $\mathbf{y}_0^{t-1}$*, which explains the significance of full SI. Second, because of the finite-rate communication link, the controller at the receiver side can only compute the conditional mean estimate,* $\mathbf{E}\left\{x_t|\,\mathbf{j}_0^t\right\}$*, based on the received symbols.*

For the sake of clarity, in this chapter, the following notations are used for the above-mentioned two conditional mean estimates,

$$\begin{aligned}
\breve{x}_{t|s}&\triangleq\mathbf{E}\left\{x_t|\,\mathbf{j}_0^s\right\},\\
\hat{x}_{t|s}&\triangleq\mathbf{E}\left\{x_t|\,\mathbf{y}_0^s,\mathbf{j}_0^{s-1}\right\},
\end{aligned}$$

as well as the short notations $\breve{x}_t \triangleq \mathbf{E}\{x_t|\,\mathbf{j}_0^t\}$ and $\hat{x}_t \triangleq \mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\}$.

It can be read from (4.4) that the conditional variance $p_{t|t}$ does not depend on the realization of $\mathbf{y}_0^t$. In fact, the conditional mean $\hat{x}_t = \mathbf{E}\left\{x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\right\}$ and $p_{t|t}$ are the sufficient statistics for characterizing the pdf $p(x_t|\,\mathbf{y}_0^t,\mathbf{j}_0^{t-1})$. By means of the sufficient statistics, a high-dimensional encoder, which maps the sequence $\{\mathbf{y}_0^t,\mathbf{j}_0^{t-1}\}$ to i_t, can be replaced by a low-dimensional encoder which maps $\hat{x}_t$ to i_t, for each $\mathbf{j}_0^{t-1}$. In short, for Example 4.4.1, we can simplify the encoder mapping f_t as

$$\begin{aligned} f_t &= \arg\min_{k\in\mathcal{L}} \mathbf{E}\left\{\sum_{s=t+1}^{T}\left(x_s^2+\rho u_{s-1}^2\right)\Big|\mathbf{y}_0^t, i_t=k,\mathbf{j}_0^{t-1}\right\} \\ &= \arg\min_{k\in\mathcal{L}} \mathbf{E}\left\{\sum_{s=t+1}^{T}\left(x_s^2+\rho u_{s-1}^2\right)\Big|\hat{x}_t, i_t=k,\mathbf{j}_0^{t-1}\right\}, \end{aligned}$$

where the encoder dimensionality is significantly reduced. In what follows, we briefly comment on sufficient statistics in the context of open-loop encoders.

Example 4.4.3. ***Sufficient Statistics for Open-Loop Encoders***

Consider the open-loop encoder introduced in Chapter 3 and assume all previous measurements and full SI are available at the encoder. Nevertheless, to optimize the open-loop encoder, we need the conditional pdf $p(x_t|\,\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1})$*, where*

$$\bar{y}_t = \bar{x}_t + e_t, \quad \bar{x}_t = x_t - \sum_{s=0}^{t-1} a^{t-s-1}u_s.$$

First, note that $p(\bar{x}_t|\,\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1})$ *is a Gaussian pdf,*

$$\mathcal{N}\left(\mathbf{E}\left\{\bar{x}_t|\,\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1}\right\},\mathbf{E}\left\{\left(\bar{x}_t-\mathbf{E}\left\{\bar{x}_t\,|\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1}\right\}\right)^2\right\}\right).$$

Therefore $p(x_t|\,\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1})$ *is Gaussian with the following expected value and variance,*

$$\bar{m}_t = \mathbf{E}\left\{\bar{x}_t|\,\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1}\right\} + \sum_{j=0}^{t-1} a^{t-1-j}u_j, \quad \sigma_t^2 = \mathbf{E}\left\{\left(\bar{x}_t-\mathbf{E}\left\{\bar{x}_t|\,\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1}\right\}\right)^2\right\},$$

which indicates that $\bar{m}_t$ *and* σ_t^2 *are sufficient statistics of* $\{\bar{\mathbf{y}}_0^t,\mathbf{j}_0^{t-1}\}$.

4.4.2 Regularity

Regularity is another encoder property which is very useful to solve the encoder dimensionality problem. By *regular* we mean that the encoding regions are non-overlapping and connected, that is, since we discussing the scalar case, each encoding region is an interval containing one of its end points [GG92]. In general, from the implementation point of view, if each encoding region is an interval, then it

can be characterized by the interval end points. Hence, the encoder can be implemented as a look-up table with only the partition thresholds (or the end points of the quantization cells) stored. By regularity, there are at most $L-1$ thresholds for each $\mathbf{j}_0^{t-1}$, $t=0,\ldots,T-1$. As long as the number of the thresholds is small, the difference is not really significant whether the quantization cell is connected or disconnected. On the other hand, from the point of view of training the system, the regularity is a desirable feature because it can simplify the optimization procedure substantially, as will be explained later. For this reason, we discuss the regularity property in a more detailed manner by solving a specific problem, concerning the optimization of the encoder mapping f_0. First, we demonstrate the regularity of the encoder mapping by using numerical examples. Thereafter, the problem of how to assess regularity is addressed in Section 4.5.

Now, let us introduce Problem 4.4.4 which is the problem studied in Section 4.4.2 and Section 4.5.

Problem 4.4.4. *Consider a scalar special case of the general system from Chapter 2 that full state measurement and full SI are available at the encoder. The overall closed-loop system is governed by the following system equation, encoder–controller, and channel,*

$$\begin{array}{lrcl} \text{Plant}: & x_{t+1} & = & ax_t+u_t+v_t, \quad a>0, \\ \text{Encoder}: & i_t & = & f_t\big(x_t,\mathbf{j}_0^{t-1}\big)\,, \\ \text{Channel}: & j_t & = & E_t(i_t)\,, \\ \text{Controller}: & u_t & = & g_t(\mathbf{j}_0^t)\,, \end{array} \tag{4.5}$$

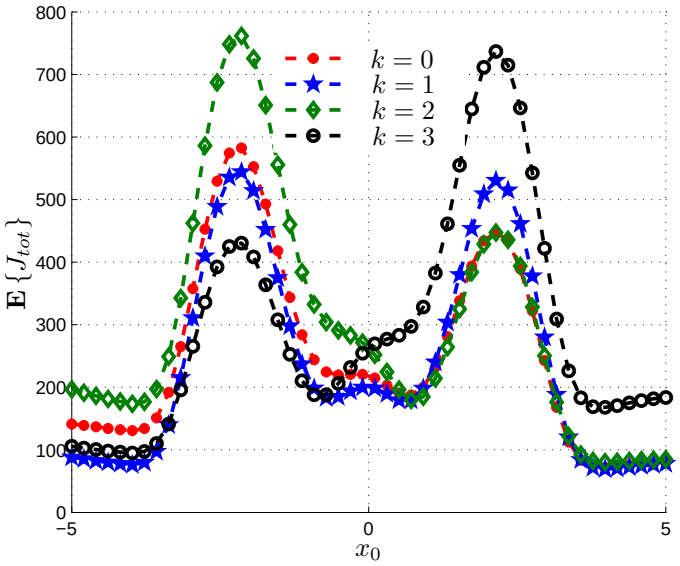
$i_t, j_t \in \mathcal{L}$. All notations in (4.5) *follow Chapter 2. The initial-state x_0 and the process noise v_t are i.i.d. zero-mean Gaussian. The memoryless channel is completely characterized by the transition probability function $\mathcal{P}(j_t|\,i_t)$. Then, for the fixed encoder–controller mappings $\{\mathbf{f}_1^{T-1},\mathbf{g}_0^{T-1}\}$, compute the optimal encoder mapping*

$$f_0(x_0)=\arg\min_{k\in\mathcal{L}}\mathbf{E}\left\{\sum_{t=0}^{T-1}x_{t+1}^2+u_t^2\Big|x_0,i_0=k\right\}.$$

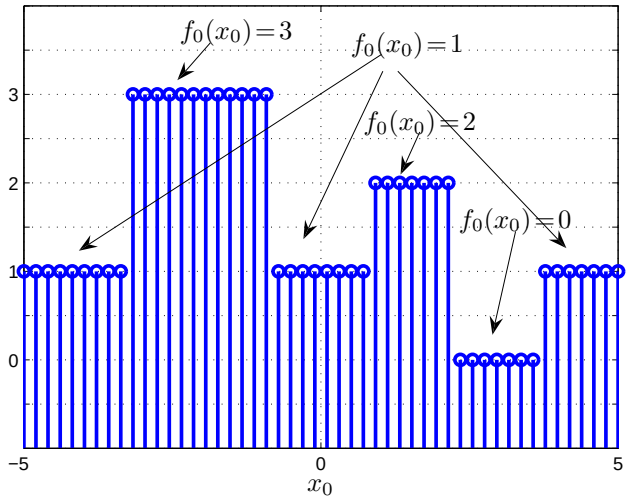
We assume here that all future encoder mappings $\mathbf{f}_1^{T-1}$ are regular since we often initialize with a regular quantizer. It is straightforward to verify that the future terms $\mathbf{E}\left\{x_{t+1}^2\big|\,x_0,i_0\right\}$ and $\mathbf{E}\left\{u_t^2\big|\,x_0,i_0\right\}$, $t=0,\ldots,T-1$, are continuous functions of x_0, as well as the overall cost $\mathbf{E}\left\{\sum_{t=0}^{T-1}x_{t+1}^2+u_t^2\Big|\,x_0,i_0\right\}$. The encoding region $\mathcal{S}_0(k)$, defined as

$$\mathcal{S}_0(k)\triangleq\left\{x_0\in\mathbb{R}:\exists k,\arg\min_{l}\mathbf{E}\left\{\sum_{t=0}^{T-1}x_{t+1}^2+u_t^2\Big|x_0,i_0=l\right\}=k\right\}, \tag{4.6}$$

can be described as a union of disjoint intervals. As a result, instead of a great number of online computations, the encoder can be implemented as a look-up table consisting of thresholds. In Figure 4.1 and Figure 4.2, we use two numerical

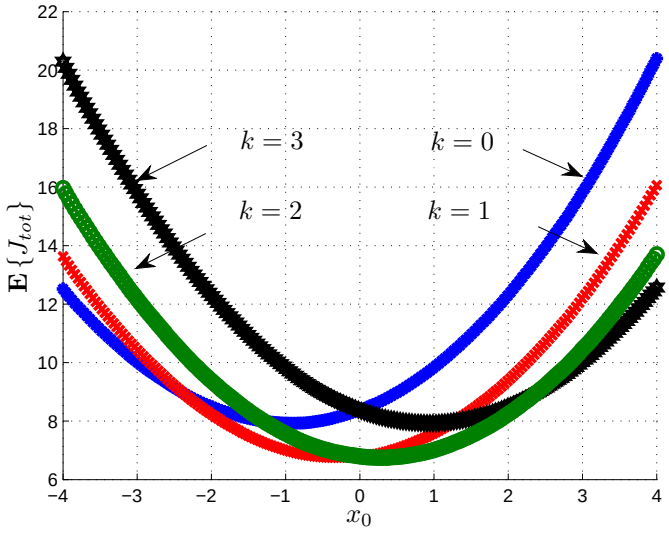


(a) Expected total cost for $i_0 = k$.

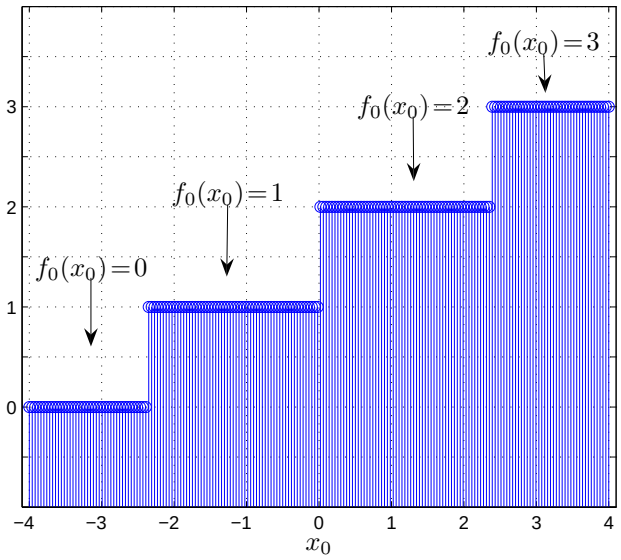


(b) Encoder decision $i_0 = f_0(x_0)$.

Figure 4.1: Encoder mapping $i_0 = f_0(x_0)$ for which the encoding regions can be represented as unions of intervals. (a) The expected total cost for $i_0 = k$, $k \in \mathcal{L}$. (b) The encoder decision $f_0(x_0)$.



(a) Expected total cost if $i_0 = k$.



(b) Encoder decision $i_0 = f_0(x_0)$.

Figure 4.2: A regular encoder mapping $i_0 = f_0(x_0)$, which can be represented by connected intervals. (a) The expected total cost for $i_0 = k$, $k \in \mathcal{L}$. (b) The encoder decision $f_0(x_0)$.

examples to show different classes of $\mathcal{S}_0(k)$. The system parameters for the two examples are almost identical except for the encoder–controller mappings $\mathbf{f}_1^{T-1}$ and $\mathbf{g}_0^{T-1}$. The predicted overall future cost $\mathbf{E}\left\{\sum_{t=0}^{T-1} x_{t+1}^2+u_t^2 \middle| x_0, i_0=k\right\}$ and the associated decision $\arg\min_k \mathbf{E}\left\{\sum_{t=0}^{T-1} x_{t+1}^2+u_t^2 \middle| x_0, i_0=k\right\}$, $k\in\{1,2,3,4\}$, are plotted as functions of x_0, for $R=2$ and $T=4$. We see from the figures that all encoding regions are unions of intervals. However, the encoder mapping in Figure 4.1 is not regular because the encoding region for $f_0(x_0)=1$ is not an interval. On the other hand, the mapping in Figure 4.2 is shown to be regular. This example indicates that the regularity of the optimized encoder mapping is closely related to the future coding–control mappings.

As already mentioned, regarding the trained encoder–controller, the importance of regularity appears to be less obvious. On the other hand, from the training point of view, regular encoding mappings are of great significance. First of all, the regularity sets an upper bound on the number of boundaries of encoding regions. Second, we can use simple and efficient search algorithms to compute the encoding boundaries, for example using binary search. In other words, given fixed computational power, knowledge of the regularity can improve the quality of search results significantly. For these reasons, the concept of the regularity is further studied in the next section, particularly with regard to the impact of x_0 on expected future cost.

4.5 Impact on Expected Future Cost

In Section 4.4, we argued that the regularity plays an important role in the implementation of the iterative training algorithm. In this section, we will study this property in more detail. First in Section 4.5.1 we discuss a special scenario where the future cost is independent of the current decision. Based on the result of Section 4.5.1, a sufficient condition of the regularity is proposed. Thereafter, we apply the proposed sufficient condition to more general cases.

Recall that the optimized encoder assigns the index k to i_0 if x_0 belongs to $\mathcal{S}_0(k)$ of (4.6), which can also be written as

$$\mathcal{S}_0(k)=\bigcap_{\substack{l=0\\ l\neq k}}^{2^R-1}\mathcal{S}_0(k,l),\quad k,l\in\mathcal{L},$$

where the region $\mathcal{S}_0(k,l)$ is defined as

$$\begin{aligned}\mathcal{S}_0(k,l)&\triangleq\left\{x_0:\mathbf{E}\left\{\sum_{s=0}^{T-1}x_{s+1}^2+u_s^2\middle|x_0,i_0=k\right\}\le\mathbf{E}\left\{\sum_{s=0}^{T-1}x_{s+1}^2+u_s^2\middle|x_0,i_0=l\right\}\right\}\\&=\left\{x_0:\mathbf{E}\left\{x_1^2+u_0^2\middle|x_0,i_0=k\right\}+\mathbf{E}\left\{\sum_{s=1}^{T-1}x_{s+1}^2+u_s^2\middle|x_0,i_0=k\right\}\right.\end{aligned}$$

$$\leq \mathbf{E}\left\{ x_1^2 + u_0^2 \middle| x_0, i_0 = l \right\} + \mathbf{E}\left\{ \sum_{s=1}^{T-1} x_{s+1}^2 + u_s^2 \middle| x_0, i_0 = l \right\}, \quad k, l \in \mathcal{L}.$$

That is to say for all elements in $\mathcal{S}_0(k,l)$, choosing k instead of l yields a lower expected future cost. For brevity, let us take $J_0(x_0,k)$ to denote the expected cost, given x_0 and the decision $i_0 = k$, $k \in \mathcal{L}$,

$$J_0(x_0,k) \triangleq \mathbf{E}\left\{ x_1^2 + u_0^2 \middle| x_0, i_0 = k \right\} + \mathbf{E}\left\{ \sum_{s=1}^{T-1} x_{s+1}^2 + u_s^2 \middle| x_0, i_0 = k \right\}, \quad k \in \mathcal{L},$$

where the overall cost is decomposed into two parts. We refer to the first term as the current/direct cost, specifying the dependence of $J_0(x_0,k)$ to u_0 and x_1. The second term is referred to as the future/indirect cost, specifying the dependence of $J_0(x_0,k)$ to all other future controls $\mathbf{u}_1^{T-1}$ and states $\mathbf{x}_2^T$. An intersection point of two encoding regions $\mathcal{S}_0(k)$ and $\mathcal{S}_0(l)$ belongs to the set

$$\mathcal{T}_0(k,l) \triangleq \{ x_0 : J_0(x_0,k) = J_0(x_0,l) \}, \quad k, l \in \mathcal{L}. \tag{4.7}$$

It will be clear from Section 4.5.1 that the number of elements in the set $\mathcal{T}_0(k,l)$ has much to say about the regularity property. In the sequel we will exclude the trivial case that $J_0(x_0,k) = J_0(x_0,l)$, for all x_0. In general, determining the set $\mathcal{T}_0(k,l)$ is a difficult task because the relation between x_0 and $J_0(x_0,k)$ is difficult to formulate in a closed-form, except in certain special cases. Let us start to investigate $\mathcal{T}_0(k,l)$ by expanding the current cost in $J_0(x_0,k)$,

$$\begin{aligned}
\mathbf{E}\left\{ x_1^2 + u_0^2 \middle| x_0, k \right\} =& \mathbf{E}\left\{ (ax_0 + u_0 + v_0)^2 + u_0^2 \middle| x_0, i_0 = k \right\} \\
=& \mathbf{E}\left\{ a^2 x_0^2 + 2x_0 u_0 + 2u_0^2 + v_0^2 \middle| x_0, i_0 = k \right\}, \quad k \in \mathcal{L},
\end{aligned}$$

where $\mathbf{E}\{x_0 v_0\} = 0$ and $\mathbf{E}\{u_0 v_0\} = 0$, due to the independent process noise. Furthermore, the terms $\mathbf{E}\left\{ a^2 x_0^2 \middle| x_0, i_0 = k \right\}$ and $\mathbf{E}\left\{ v_0^2 \middle| x_0, i_0 = k \right\}$ will not affect the decision since x_0 is a constant and v_0 is independent of x_0 and i_0. Define now

$$\begin{aligned}
A_0(x_0,k,l) &\triangleq \mathbf{E}\left\{ \sum_{s=1}^{T-1} x_{s+1}^2 + u_s^2 \middle| x_0, i_0 = l \right\} - \mathbf{E}\left\{ \sum_{s=1}^{T-1} x_{s+1}^2 + u_s^2 \middle| x_0, i_0 = k \right\}, \\
B_0(x_0,k,l) &\triangleq \mathbf{E}\left\{ 2u_0^2 \middle| x_0, i_0 = l \right\} - \mathbf{E}\left\{ 2u_0^2 \middle| x_0, i_0 = k \right\} \\
&= \mathbf{E}\left\{ 2u_0^2 \middle| i_0 = l \right\} - \mathbf{E}\left\{ 2u_0^2 \middle| i_0 = k \right\} \\
&= B_0(k,l), \\
C_0(x_0,k,l) &\triangleq \mathbf{E}\left\{ au_0 \middle| x_0, i_0 = k \right\} - \mathbf{E}\left\{ au_0 \middle| x_0, i_0 = l \right\}, \\
&= \mathbf{E}\left\{ au_0 \middle| i_0 = k \right\} - \mathbf{E}\left\{ au_0 \middle| i_0 = l \right\} \\
&= C_0(k,l).
\end{aligned} \tag{4.8}$$

As shown in (4.8), $B_0(x_0,k,l)=B_0(k,l)$, and $C_0(x_0,k,l)=C_0(k,l)$ are terms independent of x_0. As a matter of fact, the challenge of determining $\mathcal{T}_0(k,l)$ lies in the future term $A_0(x_0,k,l)$, because of its intricate relation to x_0. In the special case that $A_0(x_0,k,l)$ is also independent of x_0, the proof of the regularity is straightforward, as explained next in Section 4.5.1.

4.5.1 Independent Future Cost: $A_0(x_0,k,l)=A_0(k,l)$

In this subsection we discuss the very special case that $A_0(x_0,k,l)=A_0(k,l)$, that is, this term is independent of x_0. Here we should observe that this is a case very similar to optimizing quantizers in classical communications [FV87] with a future cost equaling zero. Following [FV87], we can show that the set $\mathcal{T}_0(k,l)$ has only one solution which is given by,

$$\mathcal{T}_0(k,l)=\left\{\frac{A_0(k,l)+B_0(k,l)}{2C_0(k,l)}\right\}.$$

For ease of the presentation, we introduce $\tau_0(k,l)$ defined by

$$\tau_0(k,l)\triangleq\frac{A_0(k,l)+B_0(k,l)}{2C_0(k,l)}.$$

Then, the region $\mathcal{S}_0(k,l)$ becomes

$$\mathcal{S}_0(k,l)=\begin{cases}(-\infty,\tau_0(k,l)], & C_0(k,l)<0,\\ [\tau_0(k,l),\infty), & C_0(k,l)>0,\\ (-\infty,\infty), & C_0(k,l)=0, A_0(k,l)+B_0(k,l)\geq 0,\\ \emptyset, & C_0(k,l)=0, A_0(k,l)+B_0(k,l)<0,\end{cases} \tag{4.9}$$

which is shown to be an interval. Because of the finite number of interactions of intervals, $\mathcal{S}_0(k)$ is also an interval, which can be represented as

$$\mathcal{S}_0(k)=\begin{cases}\emptyset, & C_0(k,l)=0,\ A_0(k,l)+B_0(k,l)<0 \text{ for some } l,\\ \mathbb{R}, & C_0(k,l)=0,\ A_0(k,l)+B_0(k,l)\geq 0 \text{ for all } l,\\ \left(\tau_0^{low}(k),\tau_0^{up}(k)\right), & otherwise,\end{cases}$$

where the upper and lower boundaries of $\mathcal{S}_0(k)$ are given by

$$\tau_0^{up}(k)=\min_{l:C_0(k,l)>0}\tau_0(k,l),$$
$$\tau_0^{low}(k)=\max_{l:C_0(k,l)<0}\tau_0(k,l).$$

Let the possible 2^R controls be sorted in a decreasing order, i.e.,

$$u_0(k)\geq u_0(l),\quad \text{if } k<l.$$

First we know that $\tau_0^{low}(k) \leq \tau_0^{up}(k)$. Comparing $\tau_0^{up}(k)$ and $\tau_0^{low}(l)$, $k<l$, it follows that

$$\tau_0^{up}(k) = \min_{m>k} \tau_0(k,m) \leq \tau_0(k,l),$$
$$\tau_0^{low}(l) = \max_{m<l} \tau_0(l,m) \geq \tau_0(l,k).$$

Then, because $\tau_0(k,l) = \tau_0(l,k)$, it yields

$$\tau_0^{up}(k) \leq \tau_0^{low}(l). \tag{4.10}$$

Unfortunately, the future term $A_0(x_0,k,l)$ is in most cases a function of x_0, except in certain special cases. However, based on the discussion about the special case where $A_0(x_0,k,l) = A_0(k,l)$, we can show a sufficient condition for the regular encoder.

Lemma 4.5.1. *A sufficient condition for regular encoding mappings is that the set $\mathcal{T}_0(k,l)$ of (4.7) has at most one single element.*

The proof of Lemma 4.5.1 follows simply from (4.9)–(4.10). Lemma 4.5.1 suggests that in order to verify the regularity, we are interested in the number of elements in the set $\mathcal{T}_0(k,l)$, rather than the exact solutions. Here comes another illustrative example which shows that in certain region of x_0, there is at most one element in (4.7), and therefore, the regularity applies.

Example 4.5.2. ***Independent Future Cost***
First, expand the expected cost $J_0(x_0,k)$ for the decision $i_0 = k \in \mathcal{L}$,

$$\begin{aligned} J_0(x_0,k) &= \mathbf{E}\left\{ \sum_{s=0}^{T-1} x_{s+1}^2 + u_s^2 \Big| x_0, i_0 = k \right\} \\ &= \mathbf{E}\left\{ \sum_{s=0}^{T-1} \left(a^s x_0 + \sum_{m=0}^{s-1} a^{s-1-m} u_m + \sum_{n=0}^{s-1} a^{s-1-n} v_n \right)^2 \Big| x_0, i_0 = k \right\} \\ &\quad + \mathbf{E}\left\{ \sum_{s=0}^{T-1} u_s^2 \Big| x_0, i_0 = k \right\}. \end{aligned}$$

Consider the following situation. At a low data rate, there exists certain $x_0^{up}(i_0=k)$, $k \in \mathcal{L}$, for which for all $x_0 > x_0^{up}(i_0=k)$, $\mathbf{E}\{u_s^2 | x_0, i_0=k\}$ and $\mathbf{E}\{u_s | x_0, i_0=k\}$, $s = 0,\ldots,T-1$, are "approximately constant". For example the "approximately constant" can mean

$$\mathcal{P}\big(|\mathbf{E}\{u_s^2 | x_0, i_0 = k\} - \mathbf{E}\{u_s^2 | x_0^{up}(i_0 = k), i_0 = k\}| < \Delta\big) = 1, \quad k \in \mathcal{L},$$

for some small-valued Δ. In other words, the expectations $\mathbf{E}\{u_s^2 | x_0, i_0=k\}$ and $\mathbf{E}\{u_s | x_0, i_0=k\}$, $s=0,\ldots,T-1$, $k \in \mathcal{L}$, are "almost" independent of x_0, for $x_0 > x_0^{up}(i_0=k)$. Similarly, there exists certain $x_0^{low}(i_0)$ for which for all $x_0 < x_0^{low}(i_0)$,

the expectations $\mathbf{E}\left\{ u_s^2 \middle| x_0, i_0 = k \right\}$ *and* $\mathbf{E}\left\{ u_s \middle| x_0, i_0 = k \right\}$, $s = 0, \ldots, T-1$, $k \in \mathcal{L}$ *stay approximately unchanged. Again, we regard those expectations "almost" independent of* x_0. *In the both cases, without much degradation in performance, we can approximate* $J_0(x_0, k)$ *by a second order polynomial,*

$$J_0(x_0, k) \approx A_1(k) x_0^2 + A_2(k) x_0 + A_3(k), \quad x_0 < x_0^{low}(i_0), \;\; x_0 > x_0^{up}(i_0),$$

where the terms $A_1(k)$, $A_2(k)$ *and* $A_3(k)$, $k \in \mathcal{L}$, *are all independent of* x_0. *Especially, it follows that*

$$A_1(k) = \sum_{t=0}^{T-1} a^{2t},$$

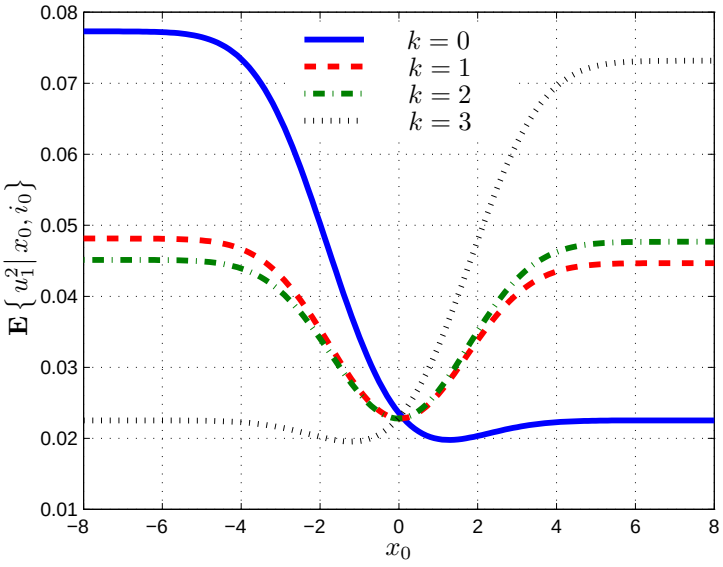
irrespective of k. *As a result, in the regions* $x_0 < x_0^{low}(i_0)$ *and* $x_0 > x_0^{up}(i_0)$, *the difference between any of two functions* $J_0(x_0, k)$ *and* $J_0(x_0, l)$, $k, l \in \mathcal{L}$, $k \neq l$, *is a linear function of* x_0. *Consequently, there are two situations that can happen to* $\mathcal{T}_0(k, l)$: (*i*) *It contains only one element, when* $J_0(x_0, k) \neq J_0(x_0, l)$, *or* (*ii*) *it contains infinitely many elements, when* $J_0(x_0, k) = J_0(x_0, l)$, *for all* x_0.

Conversely, in the interval $x_0^{low}(i_0) < x_0 < x_0^{up}(i_0)$, it will be too coarse to approximate $J_0(x_0, k)$ by a second order polynomial. To get some idea of the computational complexity of deriving $J_0(x_0, k)$, let us compute a few steps of the future costs for Problem 4.4.4. Following the order of time, $\mathbf{E}\left\{ u_t^2 \middle| x_0, i_0 \right\}$ and $\mathbf{E}\left\{ x_{t+1}^2 \middle| x_0, i_0 \right\}$ for different t will be successively investigated. Clearly, the expected costs are determined by the initial-state, the noises, and the future encoder–controller mappings. Loosely speaking, we assume that a proper initial setup is selected. By "proper" we mean that the resulting system performance should outperform the case without any control action. In practice, uniform encoders work often reasonably well as the initial setup.

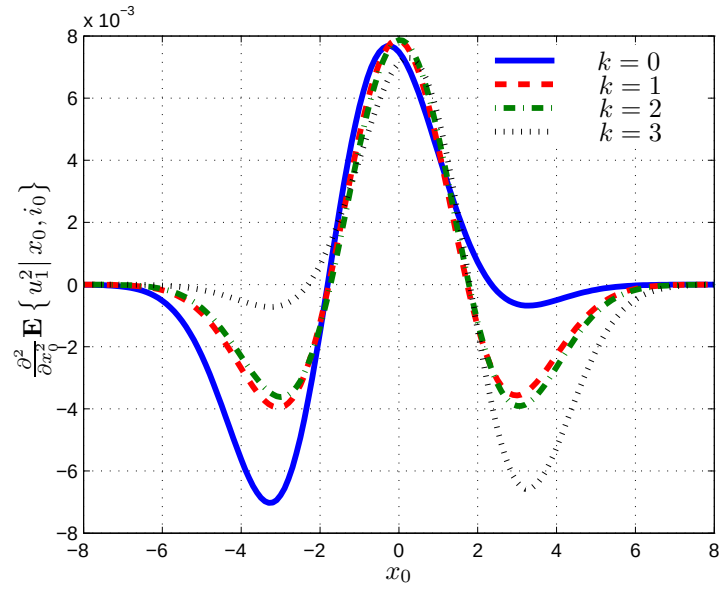
4.5.2 General Case: $A_0(x_0, k, l) \neq A_0(k, l)$

In this subsection, we remove the assumption from Section 4.5.1 that the future cost is independent of x_0, i.e., $A_0(x_0, k, l) = A_0(k, l)$. We discuss how to compute the cost $\mathbf{E}\left\{ x_t^2 + u_{t-1}^2 \middle| x_0, i_0 \right\}$ for the general case where $A_0(x_0, k, l) \neq A_0(k, l)$, $k, l \in \mathcal{L}$. First, expand the term $\mathbf{E}\left\{ u_t^2 \middle| x_0, i_0 \right\}$,

$$\begin{aligned}
\mathbf{E}\left\{ u_t^2 \middle| x_0, i_0 \right\} &= \sum_{j_0=0}^{L-1} \cdots \sum_{j_t=0}^{L-1} u_t^2(\mathbf{j}_0^t) \mathcal{P}\left(\mathbf{j}_0^t \middle| x_0, i_0 \right) \\
&= \sum_{j_0=0}^{L-1} \cdots \sum_{j_t=0}^{L-1} \sum_{i_1=0}^{L-1} \cdots \sum_{i_t=0}^{L-1} u_t^2(\mathbf{j}_0^t) \prod_{s=0}^{t} \mathcal{P}(j_s | i_s) \, \mathcal{P}\left(i_s \middle| x_0, \mathbf{i}_0^{s-1}, \mathbf{j}_0^{s-1} \right),
\end{aligned}$$



(a) $\mathbf{E}\left\{u_1^2 \middle| x_0, i_0 = k\right\}$ as a function of x_0, $k \in \mathcal{L}$.



(b) $\frac{\partial^2}{\partial x_0^2}\mathbf{E}\left\{u_1^2 \middle| x_0, i_0\right\}$ as a function of x_0.

Figure 4.3: The function $\mathbf{E}\left\{u_1^2 \middle| x_0, i_0 = k\right\}$ and its second order derivative with respect to x_0, $k \in \mathcal{L}$.

where we write $u_t(\mathbf{j}_0^t)$ to indicate that u_t is fully determined by $\mathbf{j}_0^t$. The probability $\mathcal{P}(\,i_s|\,x_0,\mathbf{i}_0^{s-1},\mathbf{j}_0^{s-1})$ is given by

$$\mathcal{P}\big(\,i_s|\,x_0,\mathbf{i}_0^{s-1},\mathbf{j}_0^{s-1}\big)=\int_{\mathcal{S}_s(\mathbf{i}_0^s,\mathbf{j}_0^{s-1})} p\big(\,x_s|\,x_0,\mathbf{i}_0^{s-1},\mathbf{j}_0^{s-1}\big)\,\mathrm{d}x_s,$$

and the encoding region $\mathcal{S}_s(\mathbf{i}_0^s,\mathbf{j}_0^{s-1})$ is defined as

$$\mathcal{S}_s(\mathbf{i}_0^s,\mathbf{j}_0^{s-1})\triangleq\big\{x_s:\mathbf{i}_0^s,\mathbf{j}_0^{s-1}\big\}.$$

The conditional pdf $p(x_s\,\big|x_0,\mathbf{i}_0^{s-1},\mathbf{j}_0^{s-1})$ is truly of great importance, since it is also central in computing $\mathbf{E}\big\{\,x_s^2\big|\,x_0,i_0\big\}$,

$$\begin{aligned}\mathbf{E}\big\{\,x_s^2\big|\,x_0,i_0\big\}=&\sum_{j_0=0}^{L-1}\cdots\sum_{j_{s-1}=0}^{L-1}\sum_{i_1=0}^{L-1}\cdots\sum_{i_s=0}^{L-1}\mathcal{P}\big(\,\mathbf{i}_1^{s-1},\mathbf{j}_0^{s-1}\big|\,i_0,x_0\big)\\&\times\int_{\mathcal{S}_s(\mathbf{i}_0^s,\mathbf{j}_0^{s-1})}x_s^2p\big(\,x_s|\,x_0,\mathbf{i}_0^{s-1},\mathbf{j}_0^{s-1}\big)\,\mathrm{d}x_s.\end{aligned}$$

In a short summary, for any t, it follows that

$$\begin{aligned}\mathbf{E}\big\{\,x_{t+1}^2+u_t^2\big|\,x_0,i_0\big\}=&\sum_{j_0=0}^{L-1}\cdots\sum_{j_t=0}^{L-1}\sum_{i_1=0}^{L-1}\cdots\sum_{i_t=0}^{L-1}\mathbf{E}\big\{\,x_{t+1}^2+u_t^2\big|\,x_0,\mathbf{j}_0^t,\mathbf{i}_0^t\big\}\,\mathcal{P}\big(\mathbf{j}_0^t,\mathbf{i}_1^t\big|\,x_0,i_0\big)\\=&\sum_{j_0=0}^{L-1}\cdots\sum_{j_t=0}^{L-1}\sum_{i_1=0}^{L-1}\cdots\sum_{i_t=0}^{L-1}\mathbf{E}\big\{\,x_{t+1}^2+u_t^2\big|\,x_0,\mathbf{j}_0^t,\mathbf{i}_0^t\big\}\,\mathcal{P}(\,j_0|\,i_0)\\&\times\mathcal{P}(\,i_1|\,x_0,i_0,j_0)\,\mathcal{P}(\,j_1|\,i_1)\,\mathcal{P}\big(\,i_2|\,x_0,\mathbf{i}_0^1,\mathbf{j}_0^1\big)\times\ldots\\&\times\mathcal{P}\big(\,i_t|\,x_0,\mathbf{i}_0^{t-1},\mathbf{j}_0^{t-1}\big)\,\mathcal{P}(\,j_t|\,i_t)\\=&\sum_{j_0=0}^{L-1}\cdots\sum_{j_t=0}^{L-1}\sum_{i_1=0}^{L-1}\cdots\sum_{i_t=0}^{L-1}\mathbf{E}\big\{\,x_{t+1}^2+u_t^2\big|\,x_0,\mathbf{j}_0^t,\mathbf{i}_0^t\big\}\\&\times\prod_{s=0}^{t}\mathcal{P}\big(\,i_s|\,x_0,\mathbf{i}_0^{s-1},\mathbf{j}_0^{s-1}\big)\,\mathcal{P}(\,j_s|\,i_s)\,.\end{aligned}$$

Since u_t is completely specified by $\mathbf{j}_0^t$, we can also write

$$\mathbf{E}\big\{\,x_{t+1}^2+u_t^2\big|\,x_0,\mathbf{i}_0^t,\mathbf{j}_0^t\big\}=\sum_{i_{t+1}=0}^{L-1}\int_{\mathcal{S}_{t+1}(\mathbf{i}_0^{t+1},\mathbf{j}_0^t)}x_{t+1}^2p\big(\,x_{t+1}|\,x_0,\mathbf{i}_0^t,\mathbf{j}_0^t\big)\mathrm{d}x_{t+1}+u_t^2.\tag{4.11}$$

More detail about the key factor $p(\,x_{t+1}|\,x_0,\mathbf{i}_0^t,\mathbf{j}_0^t)$ will be given later in this section. Now, let us move on to compute (4.11) for a few steps, following the order of time that signals are generated,

$$x_0\rightarrow u_0\rightarrow x_1\rightarrow\cdots\rightarrow u_{T-1}\rightarrow x_T.$$

It turns out that for $t<2$, we can compute $\mathbf{E}\left\{ u_t^2 \middle| x_0, i_0 \right\}$ and $\mathbf{E}\left\{ x_t^2 \middle| x_0, i_0 \right\}$ explicitly, while for $t \geq 2$, a closed-form solution is extremely difficult. In that case, we will resort to an approximation in terms of exponential functions. In the subsequent text, we show how to compute the future cost for $t<2$ and $t \geq 2$, respectively.

Case 1: $t < 2$

- Computing $\mathbf{E}\left\{ u_0^2 \middle| x_0, i_0 \right\}$

Let us begin by computing $\mathbf{E}\left\{ u_0^2 \middle| x_0, i_0 \right\}$. Conditioned on i_0, u_0 is independent of x_0, i.e., $\mathbf{E}\left\{ u_0^2 \middle| x_0, i_0 \right\} = \mathbf{E}\left\{ u_0^2 \middle| i_0 \right\}$. Particularly, $\mathbf{E}\left\{ u_0^2 \middle| x_0, i_0 \right\}$ can be computed as

$$\mathbf{E}\left\{ u_0^2 \middle| x_0, i_0 \right\} = \sum_{j_0=0}^{L-1} u_0^2(j_0) \mathcal{P}(j_0 \middle| i_0),$$

which is a constant, irrespective of x_0.

- Computing $\mathbf{E}\left\{ x_1^2 \middle| x_0, i_0 \right\}$

Next, move on to x_1 and compute $\mathbf{E}\left\{ x_1^2 \middle| x_0, i_0 \right\}$,

$$\mathbf{E}\left\{ x_1^2 \middle| x_0, i_0 \right\} = \sum_{j_0=0}^{L-1} \sum_{i_1=0}^{L-1} \mathcal{P}(j_0|i_0) \int_{\mathcal{S}_1(\mathbf{i}_0^1, j_0)} x_1^2 p(x_1 \middle| x_0, i_0, j_0) \, \mathrm{d}x_1,$$

which involves the encoder–controller mappings $g_0(j_0)$ and $f_1(x_1, j_0)$. Observe that it holds that $p(x_1 \middle| x_0, i_0, j_0) = p(ax_0 + u_0 + v_0 \middle| x_0, i_0, j_0)$. We know that u_0 is deterministic when j_0 is given, hence, $p(x_1 \middle| x_0, i_0, j_0)$ is a Gaussian pdf, in fact, a shifted function of $p(v_0)$. We can also show that $\mathbf{E}\left\{ x_1^2 \middle| x_0, i_0 \right\}$ is a convex function of x_0, since its second order derivative is always positive.

- Computing $\mathbf{E}\left\{ u_1^2 \middle| x_0, i_0 \right\}$

Following the evolution of the system, the next step is to compute $\mathbf{E}\left\{ u_1^2 \middle| x_0, i_0 \right\}$,

$$\mathbf{E}\left\{ u_1^2 \middle| x_0, i_0 \right\} = \sum_{j_0=0}^{L-1} \sum_{j_1=0}^{L-1} \sum_{i_1=0}^{L-1} u_1^2(\mathbf{j}_0^1) \mathcal{P}(j_1 \middle| i_1) \mathcal{P}(i_1 \middle| x_0, i_0, j_0) \mathcal{P}(j_0 \middle| i_0).$$

The calculation is straightforward. We depict $\mathbf{E}\left\{ u_1^2 \middle| x_0, i_0 \right\}$ and its second order derivatives in Figure 4.3, which shows that as x_0 moves away from the origin, the function $\mathbf{E}\left\{ u_1^2 \middle| x_0, i_0 \right\}$ converges to fixed values. This observation is consistent with the fact that there are a finite number of $u_1(\mathbf{j}_0^1)$.

Case 2: $t \geq 2$

Now we are in the position to compute $\mathbf{E}\left\{ x_2^2 \middle| x_0, i_0 \right\}$,

$$\begin{aligned}\mathbf{E}\left\{ x_2^2 \middle| x_0, i_0 \right\} &= \mathbf{E}\left\{ (ax_1 + u_1 + v_1)^2 \middle| x_0, i_0 \right\} \\ &= \mathbf{E}\left\{ a^2 x_1^2 + u_1^2 + 2ax_1u_1 + 2(ax_1 + u_1)v_1 + v_1^2 \middle| x_0, i_0 \right\}.\end{aligned}$$

Unfortunately the calculation is not straightforward, where the major challenge is caused by the conditional pdf $p(x_2 \,|x_0, \mathbf{i}_0^1, \mathbf{j}_0^1)$.

Generally speaking, by the linear system model, cf., Problem 4.4.4, the conditional pdf $p(x_t \,|x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1})$ can be obtained by the convolution of two pdf's, namely a scaled and shifted version of $p(\,x_{t-1}|\,x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-2})$ and the Gaussian $p(v_{t-1})$. The pdf $p(\,x_t|\,x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1})$ can be deduced recursively in terms of the conditional pdf $p(x_{t-1} \,|x_0, \mathbf{i}_0^{t-2}, \mathbf{j}_0^{t-2})$, the encoder mapping $f_{t-1}(x_{t-1}, \mathbf{j}_0^{t-2})$, the channel $\mathcal{P}(\,j_{t-1}|\,i_{t-1})$ and the process noise pdf $p(v_{t-1})$. Since scaling, shifting and convolution are all operations preserving continuality, by induction, $p(\,x_t|\,x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1})$ is a continuous pdf of x_0.

Back to $p(x_2 \,|x_0, \mathbf{i}_0^1, \mathbf{j}_0^1)$, we already know that $p(\,x_1|\,x_0, i_0, j_0)$ is a Gaussian pdf. Then, because f_1 is regular by assumption, $p(\,x_1|\,x_0, \mathbf{i}_0^1, j_0)$ is a truncated Gaussian. The pdf $p(\,x_2|\,x_0, \mathbf{i}_0^1, \mathbf{j}_0^1)$ is therefore a continuous function of x_0, computed by the convolution of a scaled and shifted version of $p(\,x_1|\,x_0, \mathbf{i}_0^1, j_0)$ and $p(v_1)$,

$$p\big(\,x_2|\,x_0, \mathbf{i}_0^1, \mathbf{j}_0^1\big) = p\big(\,ax_1 + u_1|\,x_0, \mathbf{i}_0^1, \mathbf{j}_0^1\big) * p(v_1),$$

where $*$ denotes the convolution operator. Generalizing to any t, it follows that

$$p\big(\,x_t|\,x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1}\big) = p\big(\,ax_{t-1} + u_{t-1}|\,x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1}\big) * p(v_{t-1}). \tag{4.12}$$

Unfortunately, we do not have a closed-form solution to (4.12) for $t>2$, which will be explained shortly. In order to proceed, we resort to approximations as shown subsequently. First, for the ease of presentation, let us introduce the following polynomial functions,

$$W(x, z, \alpha) \triangleq \alpha(1)x^2 + \alpha(2)z^2 + \alpha(3)x + \alpha(4)z + \alpha(5)xz + \alpha(6), \tag{4.13}$$

$$W_1(x, \alpha) \triangleq \alpha(1)x^2 + \alpha(2)x + \alpha(3), \tag{4.14}$$

$$W_2(x, z, \alpha) \triangleq \alpha(1)x + \alpha(2)z + \alpha(3), \tag{4.15}$$

where $\alpha = \{\alpha(k)\}$, $k \in \mathbb{N}$, is a vector independent of x and z. Note that, $W_1(x, \alpha)$ and $W_2(x, z, \alpha)$ are simple special cases of $W(x, z, \alpha)$. Let us also introduce the $\mathrm{erf}(x)$ function [CD02], defined as,

$$\mathrm{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} \mathrm{d}y.$$

According to [CD02], we can approximate the $\operatorname{erf}(x)$ function as

$$\begin{aligned}\operatorname{erf}(x) &= \begin{cases} 1-\operatorname{erfc}(x), & x>0,\\ -(1-\operatorname{erfc}(-x)), & x\le 0,\end{cases}\\ &\approx \begin{cases} 1-(\frac{1}{6}e^{-x^2}+\frac{1}{2}e^{-\frac{4}{3}x^2}), & x>0,\\ -1+(\frac{1}{6}e^{-x^2}+\frac{1}{2}e^{-\frac{4}{3}x^2}), & x\le 0.\end{cases}\end{aligned} \tag{4.16}$$

Based on the approximation (4.16), in the following lemma we show a useful closed-form approximation of the conditional pdf $p\big(x_t\big|\, x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1}\big)$ and the expectation $\mathbf{E}\left\{x_t^2\big|\, x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1}\right\}$.

Lemma 4.5.3. *Consider Problem 4.4.4. Let $t\ge 2$, it holds that*

1. *A closed-form approximation of the conditional pdf $p(x_t|\, x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1})$ is given by a sum of functions of the form*

$$B_1 e^{W(x_0, x_t, B_2)} \left(\operatorname{erf}(W_2(x_0, x_t, B_3)) - \operatorname{erf}(W_2(x_0, x_t, B_4))\right), \tag{4.17}$$

   where W and W_2 are as defined in (4.13) *and* (4.15)*, and the vector terms B_1–B_4 are independent of x_0 and x_t. As a matter of fact,* (4.17) *can also be expressed as a sum of functions of the form $A_1 e^{W(x_0, x_t, A_2)}$, where A_1 and A_2 are vectors independent of x_0 and x_t.*

2. *A closed-form approximation of the expectation $\mathbf{E}\left\{x_t^2\big|\, x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1}\right\}$ can be written as a sum of functions of the form,*

$$W_1(x_0, C_1) e^{W_1(x_0, C_2)}, \tag{4.18}$$

   where W_1 is defined in (4.14)*, and the vector terms C_1 and C_2 are independent of x_0.*

The proof of Lemma 4.5.3 is given in Appendix 4.A.

4.5.3 Numerical Example

In this subsection, we demonstrate a numerical example which solves Problem 4.4.4 based on the results of Section 4.5.2. First in Figure 4.5, the expected cost

$$J_0(x_0, k) = \mathbf{E}\left\{\sum_{t=0}^{1} x_{t+1}^2 + u_t^2 \Big| x_0, i_0 = k\right\}, \quad k \in \mathcal{L},$$

is depicted for $k \in \{0, 1, 2, 3\}$. We can read from the figure that all the encoding regions of this example are regular. Then, the difference between two predictions is computed,

$$\mathbf{E}\left\{\sum_{t=0}^{1} x_{t+1}^2 + u_t^2 \Big| x_0, i_0 = k\right\} - \mathbf{E}\left\{\sum_{t=0}^{1} x_{t+1}^2 + u_t^2 \Big| x_0, i_0 = l\right\}, \quad k, l \in \mathcal{L},\ k \ne l.$$

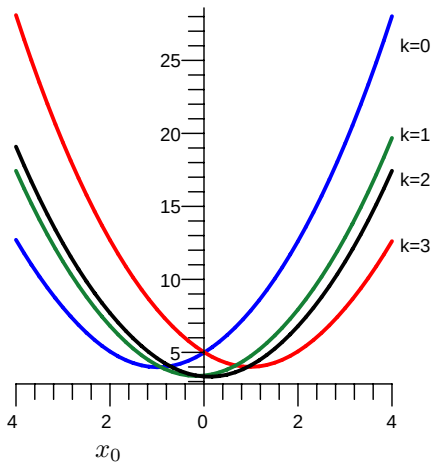


Figure 4.4: The expected future cost for $i_0 = k$, $k \in \mathcal{L}$, calculated by using the approximations of $\mathrm{erf}(x)$.

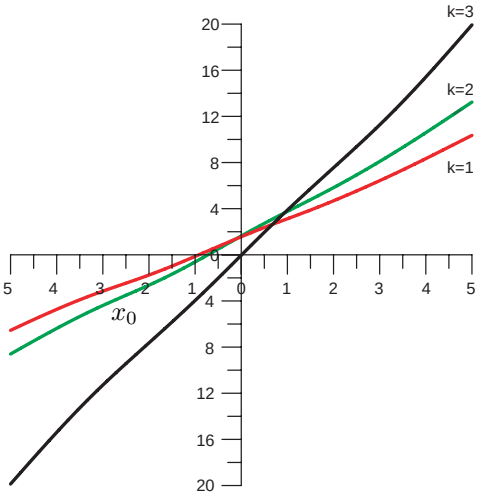


Figure 4.5: Differences among $J_0(x_0, 0)$ and $J_0(x_0, k)$, $k \in \mathcal{L}$, $k \neq 0$.

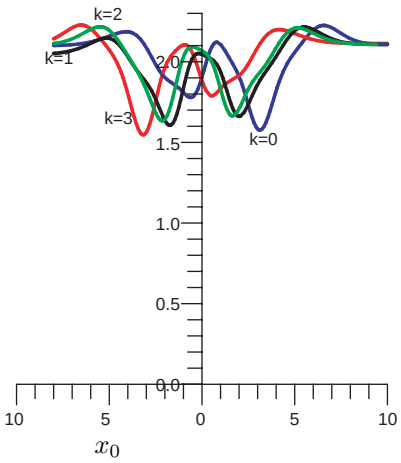


Figure 4.6: The second order derivatives of $\sum_{t=0}^{1} \mathbf{E}\left\{ x_{t+1}^2 + u_t^2 \middle| x_0, i_0 = k \right\}$, $k \in \mathcal{L}$, as a function of x_0.

Figure 4.5 shows the differences between $J_0(x_0, 1)$ and $J_0(x_0, k)$, $k \neq 1$. All three curves cross zero only once. Furthermore, in Figure 4.6, the second order derivatives of $J_0(x_0, k)$, $k \in \{0, 1, 2, 3\}$, are depicted. In the region far from the origin, the second order derivative is approximately constant, which is in accordance with the previous discussions. On the other hand, in the neighborhood of the origin, the expected cost is not a second order polynomial. Yet it is a convex function shown by the positive second order derivative.

Finally, the performance–complexity trade-off is shown in Figure 4.7. In particular, the figure depicts the impact on the optimized encoder by taking more future costs into account. We sort the thresholds in the increasing order, and label the threshold by $\tau_0^t(x_0, k)$. Regarding the notation $\tau_0^t(x_0, k)$, first, $t \in \{0, \ldots, T\}$ specifies the length of the time horizon involved in the overall cost. In other words, the threshold $\tau_0^t(x_0, k)$ is computed based on the future terms $\mathbf{x}_1^t$ and $\mathbf{u}_0^{t-1}$. Second, the index $k \in \{1, \ldots, 2^R - 1\}$ specifies the k^{th} threshold, and

$$\tau_0^t(x_0, k) \leq \tau_0^t(x_0, l), \quad \text{if } k < l.$$

In the example, by taking more future costs into consideration, the thresholds are shifted towards the origin. It is also worth mentioning that in the example in Figure 4.7, the impact on the optimized encoder by the future costs decreases rapidly with time. The figure shows that even though the future costs are significant to the overall cost $\mathbf{E}\{J_{tot}\} = \sum_{t=1}^{T} \mathbf{E}\left\{ x_t^2 + u_{t-1}^2 \right\}$, their impact on the optimized threshold is limited. We also know that the encoder complexity increases significantly with

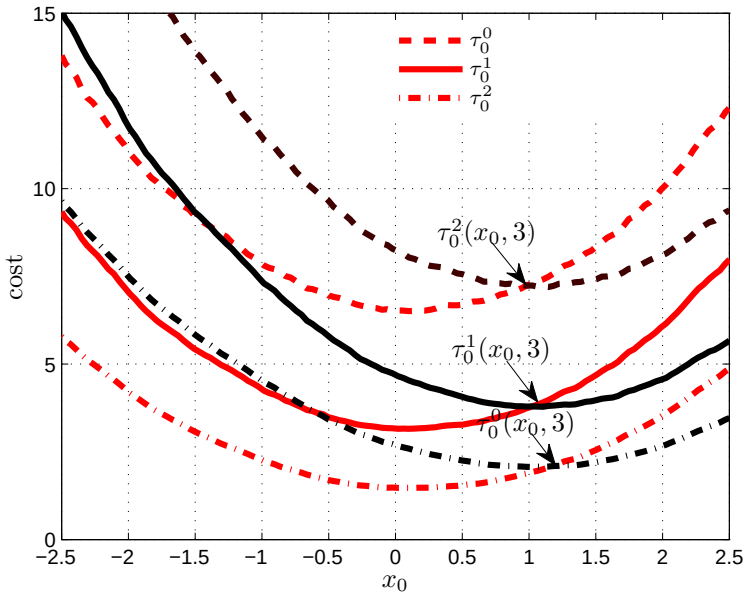


Figure 4.7: The impact on f_0 by the future states. The notation $\tau_0^2(x_0, 3)$ means the third threshold obtained by minimizing $\sum_{t=1}^{2} \mathbf{E}\left\{x_t^2 + u_{t-1}^2\right\}$.

the future terms. Hence, from a practical point of view, we can significantly reduce the computational burden by only taking into account the near-future cost $\arg\min_{k\in\mathcal{L}} \sum_{s=t}^{t+\Delta_t} \mathbf{E}\left\{u_s^2 + x_{s+1}^2 | x_t, i_t = k, \mathbf{j}_0^{t-1}\right\}$, for properly selected step length Δ_t, without causing serious performance degradation. How to select a suitable Δ_t is worth further investigation.

In this section, we proposed an approximation to compute the expected future cost. The approximation can be used to assess the regularity property of the encoder. In our simulations of Example 4.4.1, we always observe regular encoder mappings if we initialize with regular encoder mappings. For the general system in which the initial-state, process noise or measurement noise have arbitrary pdf's, and initialized with arbitrary encoder–controller mappings, the situation is much more complex. It is not difficult to find such examples that the encoding mappings are not regular.

4.6 Control over Long Time Horizon

In this section the problem of designing encoder–controllers for long time horizon is studied. As explained previously, in the case of a long time horizon, even

for a scalar system of a low data rate, a solution by iterative training requires extensive computation. The high computational complexity of the training procedure is a driving factor for the development of lower-complexity high-performance coding–control schemes. In striving toward practical encoder–controller designs of high performance and low complexity, it is important to find efficient and useful simplifications and approximations. This section is mainly devoted to an empirical study of several practical encoder–controllers constructed based on the results from Chapter 3. The objective is to seek a good compromise between computational effort and efficient controls.

As discussed in the first part of this chapter, the iterative encoder–controller design from Chapter 3 suffers from the dimensionality problem. Generally speaking, the design becomes impractical when confronting any of the following challenges: a long time horizon T, a high system-order, or a high data rate. However, our experiments have shown that satisfactory performance can be achieved in practice even though the encoder–controller pair is trained for a low communication rate and a short time horizon. Here we mention two observations.

1. Typically in control systems, the state response can be divided into two phases, the transient-state phase and the steady state phase. During the transient-state phase, in order to catch up the rapid time variation of the state, the encoder–controller mappings f_t and g_t change correspondingly with time. On the other hand, in the steady state phase, the mappings f_t and g_t are relatively constant. Therefore, in practice, satisfactory performance can be expected by only optimizing the encoder–controller components for steps until the steady state is reached, and thereafter applying a time-invariant encoder–controller mapping.

2. The experiments have shown that above a certain value, the performance gain attained by increasing data rate drops significantly. Unlike the situation when the channel is error-free, in order to maintain robustness against channel errors, training typically results in cautious control inputs, meaning: (i) Magnitudes of controls are small, and (ii) differences among controls are small. In fact, the higher the crossover probability is, the smaller and closer are the values of the feasible controls. Consequently, the advantage of having a high rate, is substantially reduced as the error-level increases.

4.6.1 Problem Formulation

Since the key point is to deal with long time horizon, we consider a scalar system, which is governed by the linear equation system (4.1), i.e.,

$$\begin{aligned} x_{t+1} &= ax_t + u_t + v_t, \quad a > 0, \\ y_t &= x_t + e_t. \end{aligned}$$

Process noise v_t and measurement noise e_t are modeled as i.i.d. Gaussian zero-mean processes with variances σ_v^2 and σ_e^2. They are mutually independent of the

initial-state x_0, which is also i.i.d. zero-mean Gaussian, i.e., $x_0 \sim \mathcal{N}(0, \sigma_{x_0}^2)$. State measurements will be conveyed to the controller via a low-rate noisy channel. The coding–control scheme will be described in the subsequent text. Here, we first specify the design goal which is to minimize the expected cost $\mathbf{E}\{J_{tot}\}$ where J_{tot} is given by (4.2), i.e.,

$$J_{tot} = \sum_{t=1}^{T} x_t^2 + \rho u_{t-1}^2.$$

Focusing on explaining basic concepts, we consider the special scenario that full SI and all previous measurements are available at the encoder, where the Kalman filter can be adopted at the encoder to compute the conditional mean squared estimate. As discussed previously, unlike in the classical LQG problem, the control u_t is now completely determined by the channel outputs $\mathbf{j}_0^t$, and the Kalman filter employed at the encoder will compute the conditional mean estimate $\hat{x}_t = \mathbf{E}\{x_t | \mathbf{y}_0^t, \mathbf{j}_0^t\}$, cf., (4.4).

According to the results from Chapter 3, the optimal controller for a fixed encoder is given by

$$u_t = \ell_t \mathbf{E}\left\{x_t | \mathbf{j}_0^t\right\}, \quad \ell_t = \frac{a\,\phi_{t+1}}{\phi_{t+1}+\rho}, \quad \phi_t = 1 + \frac{a^2\phi_{t+1}\rho}{\phi_{t+1}+\rho}, \text{ with } \phi_T = 1, \tag{4.19}$$

where the separation principle applies. As discussed in the first part of this chapter, deriving the estimate $\mathbf{E}\{x_t | \mathbf{j}_0^t\}$ is a challenging task, mostly because the computational complexity is significantly increased as the time horizon increases. In what follows, we are going to present several alternative designs which use different approximations of $\mathbf{E}\{x_t | \mathbf{j}_0^t\}$, grouped with respect to the memory access pattern. In all designs, we employ a separate decoder at the receiver side, and restrict the control to take the following form,

$$u_t = l_t d_t,$$

where l_t denotes a linear control law and d_t denotes a state estimator. Next we introduce a class of decoders with limited memory, the deterministic mapping

$$d_t = D_t(\mathbf{j}_{t_c}^t), \quad t_c = t - M_c, \;\; M_c \in \{0, \ldots, t\},$$

where M_c specifies the memory of the past received symbols at the controller. The decoded symbol d_t is viewed as an estimate of the state x_t. As before, $\breve{x}_t$ denotes the state estimator $\mathbf{E}\{x_t | \mathbf{j}_0^t\}$, and $\hat{x}_t$ the state estimator $\mathbf{E}\{x_t | \mathbf{y}_0^t, \mathbf{j}_0^t\}$.

4.6.2 Type I: Coding–Control Schemes with limited memory

As discussed previously, the main challenge of implementing (4.19) lies in the significantly increasing complexity of f_t and g_t as the time horizon T grows, since both f_t and g_t utilize information from the past, and as well, they influence the future

evolution. According to the system model, the state x_t and past received symbols $\mathbf{j}_0^t$ are the most important terms to the future evolution at each t. Based on experimental study we observe that a substantial part of information about x_t and $\mathbf{j}_0^{t-1}$ is often carried in the latest state measurements and channel outputs. Similarly, the major impact of the current state is on near-future states. Therefore, a reasonable simplification is to ignore part of information from far past, and also, part of impact on far future. In addition, when the system approaches the steady state the time variation of encoder–controller mappings typically becomes insignificant. For these reasons, we can simplify the decoder–controller by introducing a restriction on the decoder–controller memory. In other words, the decoder–controller is only given access to the latest M_c+1 channel outputs. Aiming at providing good performance at both the transient and steady state, we present the following coding–control (CCS) policies.

CCS 4.6.1. *We divide the entire time horizon into two phases. The first phase, including the beginning M_c+1 pairs of encoder–controller mappings, $\mathbf{f}_0^{M_c}$ and $\mathbf{g}_0^{M_c}$, answers for the transient behavior of the system. They are designed using the training method proposed in Chapter 3. More precisely, they are designed for the case that the horizon is T, and the costs contributed by the states after t_{M_c} are zero. In the second phase, $t_{M_c}+1,\ldots,T$, a pair of time-invariant encoder–controller mappings are used. A reasonable candidate is the encoder–controller pair designed for $t=M_c$, i.e., f_{M_c} and $g_{M_c}(\mathbf{j}_{t-M_c}^t)$.*

CCS 4.6.1 is somehow an ad hoc approach, especially the encoder–control mappings for $t > t_{M_c}$. However, if the system quickly approaches the steady state in the first M_c+1 steps, the performance degradation is expected to be insignificant. Below, we introduce two other memory-restricted encoder–controllers, based on the idea of moving horizon. They will be compared with CCS 4.6.1.

CCS 4.6.2. *Divide the entire time horizon into short time slots of equal-length, and let M_c denote the length of one time slot. At each time slot a M_c-step encoder–controller optimized using the training method from Chapter 3 is employed.*

Note that, the last state in the current time slot is also the initial-state of the next time slot. To avoid ambiguity, we distinguish ℓ_t computed for the longer horizon T and the shorter horizon M_c by the notations $\ell_{T,s}$, $s\in\{0,\ldots,T\}$, and $\ell_{M_c,s}$, $s\in\{0,\ldots,M_c\}$. It follows that $\ell_t=\ell_{T,t}$, if nothing else is stated. Regarding CCS 4.6.2, at each $t=kM_c$, $k\in\mathbb{Z}^+$, the linear control law l_{T,kM_c} is reset to $\ell_{M_c,0}$, and similarly, $l_{T,kM_c+1}=\ell_{M_c,1}$, and so on.

The linear control law ℓ_t (4.19) plays an important role in the minimization over the entire time horizon in the classical LQG problem. Therefore, we present CCS 4.6.3 by modifying CCS 4.6.2 slightly.

CCS 4.6.3. *Divide the entire time horizon into short time slots of equal-length, and let M_c denote the length of one time slot. In the k^{th} time slot, i.e., $t=kM_c,\ldots,(k+$*

$1)M_c-1$, *we apply an* M_c*-step encoder–controller, trained according to Chapter 3, but with the linear law* $\ell_{T,kM_c},\ldots,\ell_{T,(k+1)M_c-1}$.

One obvious drawback of CCS 4.6.2 and CCS 4.6.3 is the inefficient use of memory. Loosely speaking, at a low rate, the memory plays a significant role in increasing the resolution of the measurement. CCS 4.6.1 is superior to CCS 4.6.2 and CCS 4.6.3 since it is better at trading the memory for data rate. This will be illustrated by the numerical examples in Section 4.6.5. Finally, the following two systems are also used as reference systems.

4.6.3 Type II: Coding–Control Schemes with No Memory

Consider the special case that the encoder–controller has only access to the current measurement y_t and received symbol j_t. We design encoder–controllers according to the following criterion.

CCS 4.6.4. *The encoder–controller mappings take only the nearest future into account, based on the latest state measurement and channel output,*

$$\begin{aligned} f_t(y_t) &= \arg\min_{k} \mathbf{E}\left\{ x_{t+1}^2 + \rho u_t^2 \middle| y_t, i_t = k \right\}, \\ g_t(j_t) &= \arg\min_{u_t} \mathbf{E}\left\{ x_{t+1}^2 + \rho u_t^2 \middle| j_t \right\}. \end{aligned} \tag{4.20}$$

It is straightforward to show that we can also write the design criterion (4.20) as the following one,

$$\begin{aligned} f_t(y_t) &= \arg\min_{k} \mathbf{E}\left\{ (x_t - d_t)^2 \middle| y_t, i_t = k \right\}, \\ g_t(j_t) &= -\frac{a}{\rho+1}\mathbf{E}\left\{ x_t \middle| j_t \right\}, \end{aligned} \tag{4.21}$$

which means $l_t = -a/(\rho+1)$ and $d_t = \mathbf{E}\{x_t|\, j_t\}$. Here, the linear control law l_t is time-invariant. In fact, it is exactly the same one as ℓ_T in (4.19). Further, due to the mutual dependence of the encoder and controller, the mappings (4.21) are obtained iteratively in practice. The training process is performed in a similar manner to that used in Chapter 3, i.e., alternating between the optimization of encoder and controller mappings. Here, the complexity of the training is substantially reduced because of the ignorance of the past and future terms. Note also, the function $\mathbf{E}\left\{ (x_t - d_t)^2 \middle| y_t, i_t \right\}$ can be rewritten as,

$$\begin{aligned} &\mathbf{E}\left\{ (x_t - d_t)^2 \middle| y_t, i_t \right\} \\ &= \sum_{l=0}^{L-1} \mathcal{P}(j_t = l|\, i_t) \left((\mathbf{E}\{x_t|\, y_t\} - d_t)^2 + \mathbf{E}\left\{ x_t^2 \middle| y_t \right\} - (\mathbf{E}\{x_t|\, y_t\})^2 \right), \end{aligned}$$

which makes use of the facts that (i) conditioned on i_t, j_t is independent of y_t, and (ii), x_t conditioned on y_t is independent of i_t and j_t. The last two terms

$\mathbf{E}\left\{ x_t^2 \middle| y_t \right\}$ and $(\mathbf{E}\{ x_t | y_t \})^2$ have no impact on the decision of i_t, therefore the encoder mapping (4.21) can be reduced to

$$f_t(y_t) = \arg\min_k \sum_{l=0}^{L-1} \mathcal{P}(j_t = l | i_t = k) \left(\mathbf{E}\{ x_t | y_t \} - d_t\right)^2 .$$

Finally, to demonstrate the significance of the linear control law ℓ_t of (4.19), the following encoder–controller pair is included.

CCS 4.6.5. *The encoder–controller is given by the equations*

$$f_t(y_t) = \arg\min_k \mathbf{E}\left\{ (x_t - d_t)^2 \middle| y_t, i_t = k \right\},$$
$$g_t(j_t) = \ell_t \mathbf{E}\{ x_t | j_t \} .$$

4.6.4 Type III: Coding–Control Schemes with Infinite Memory

Type II encoder–controllers, CCS 4.6.4 and CCS 4.6.5, utilize only information carried in the current measurement y_t and channel output j_t, to minimize the average cost one step ahead. Consequently, these approaches may degrade the overall system performance seriously, since the useful information in the history is not fully exploited. To study the significance of memory, we consider two full-memory encoder–controllers, and particularly assume $\mathbf{j}_0^t$ and $\mathbf{y}_0^t$ are available to the controller and the encoder, respectively.

However since the optimal full-memory encoder–controller is prohibited by complexity limitations, only practical schemes are considered here. More specifically, we adopt a Kalman filter to compute an estimate of the expected value $\mathbf{E}\{ x_t | \mathbf{j}_0^t \}$. Recall that the Kalman filter can perfectly compute the conditional mean estimate $\mathbf{E}\{ x_t | \mathbf{y}_0^t, \mathbf{u}_0^t \}$, as well as $\mathbf{E}\{ x_t | \mathbf{y}_0^t, \mathbf{j}_0^t \}$ in the special case that u_t is completely determined by $\mathbf{j}_0^t$. Unfortunately, the Kalman filter is not able to produce $\mathbf{E}\{ x_t | \mathbf{j}_0^t \}$. In other words, we can not use the Kalman filter to implement an optimal controller when there is a channel between the sensor and the controller. However, we can implement the Kalman filter to obtain a computationally feasible solution. In short, the coding–controller scheme works as follows. At the encoder, a time-invariant uniform encoder is employed, and then at the receiver side, the reconstructed values are fed into a Kalman filter. Thereafter, the control law is computed as a linear function of the Kalman filter output where we let the linear control law l_t be equal to ℓ_t of (4.19). In CCS 4.6.6 and CCS 4.6.7, we present two system designs in which the Kalman filter is designed differently. As a matter of fact, they are U-Kalman I and U-Kalman II from Section 3.5.

CCS 4.6.6. *Let l_t be ℓ_t of* (4.19)*. Design the Kalman filter assuming the channel is absent. In other words, the Kalman filter is designed entirely based on the statistical knowledge of the initial-state, the process noise and the measurement noise, as in the classical LQG problem.*

CCS 4.6.7. *Let l_t be ℓ_t of* (4.19)*. Design the Kalman filter assuming the total distortion due to measurement noise, quantization and transmission error is white and Gaussian distributed.*

By CCS 4.6.7, the distortion caused by quantization and transmission error is treated as a part of measurement noise. Note that, the "extended" measurement noise (including quantization error and channel error) is neither Gaussian nor uncorrelated with the state and the process noise, making the Kalman filter a sub-optimal estimator.

4.6.5 Numerical Examples

In this sub-section, numerical experiments are carried out to demonstrate the performance of the various encoder–controllers presented in this section. For the simplicity of presentation, in the experiments we consider a linear scalar plant for which the encoded measurement is transmitted over a binary symmetric channel. The system equations and the LQ cost follow (4.1) and (4.2). The initial-state x_0, process noise v_t and measurement noise e_t are all i.i.d. zero-mean Gaussian. The system parameters are chosen in the interest of demonstrating both the transient phase and the steady-state phase. In particular, the parameters are: $a=0.9$, $\sigma_{x_0}^2=5$, $\sigma_v^2=2$, $\sigma_e^2=2$, $\rho=5$, $T=30$, $\epsilon=0.08$, and $R=2$.

In Figure 4.8, we compare the Type I schemes for which the controllers are restricted to have a memory-length of $M_c = 2$. On the y-axis, the instantaneous expected cost is depicted, as a function of t. As expected, the best performance is achieved by CCS 4.6.1, because of the efficient use of the memory. Regarding the other two schemes, CCS 4.6.2 and CCS 4.6.3, the latter one performs better. In fact, it can be read from Figure 4.8 that CCS 4.6.2 is superior to CCS 4.6.3 only in the initial steps, and thereafter, CCS 4.6.2 converges to a level higher than CCS 4.6.3 does. For both CCS 4.6.2 and CCS 4.6.3, a sawtooth behavior is observed at the steady state. This is attributed to the periodic memory building process. Both CCS 4.6.2 and CCS 4.6.3 employ a new M_c-step encoder–controller in each time slot, so the memory is emptied at each $t = kM_c + 1$, $k \in \mathbb{Z}^+$, and then built up as the time goes by. Consequently, the number of admissible controls is periodically reduced to L, and then gradually increases as memory accumulates. However, the memory re-building problem is eased up by using CCS 4.6.1, for which the instantaneous cost is decreasing smoothly.

As stated previously, the computational complexity increases substantially with the channel rate R and the controller memory M_c. The degradation in performance is expected to be small if M_c is large enough that the steady state is quickly reached. In the study of the impact of M_c, our experiments often show that even for small M_c, the performance improvement along with increasing M_c is practically negligible. In a similar manner we also investigate the significance of transmission rate R. Empirically, it shows that the performance gain along with the increasing transmission rate also drops rapidly, since in order to protect against channel errors,

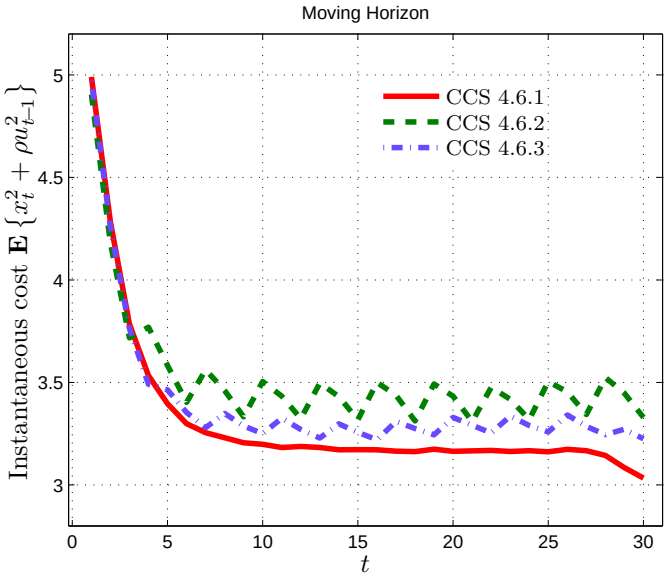


Figure 4.8: The performance of the Type I coding–control schemes, CCS 4.6.1-CCS 4.6.3.

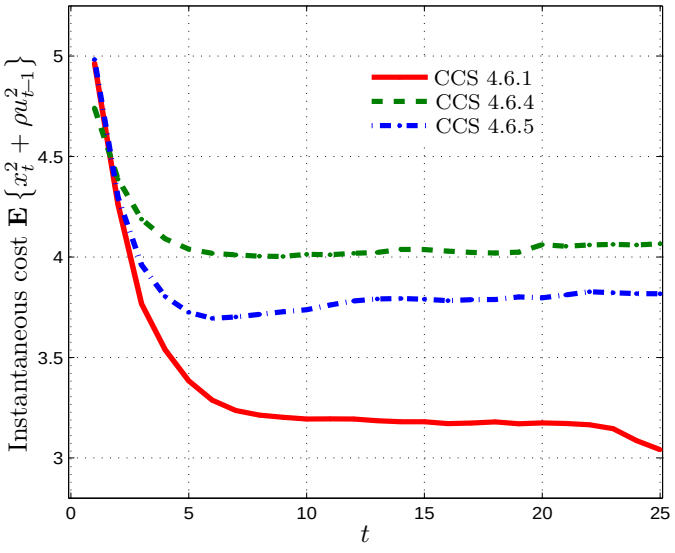


Figure 4.9: The performance of the Type II coding–control schemes, compared with CCS 4.6.1.

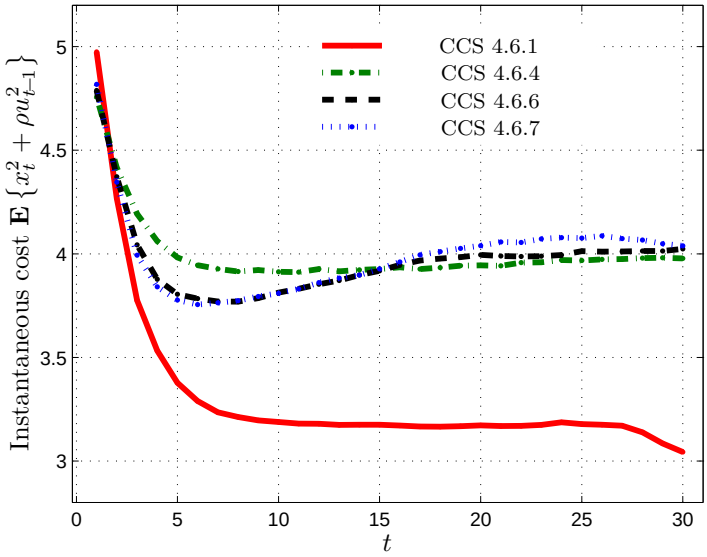


Figure 4.10: The performance of the Type III coding–control schemes, compared with CCS 4.6.1 and CCS 4.6.4.

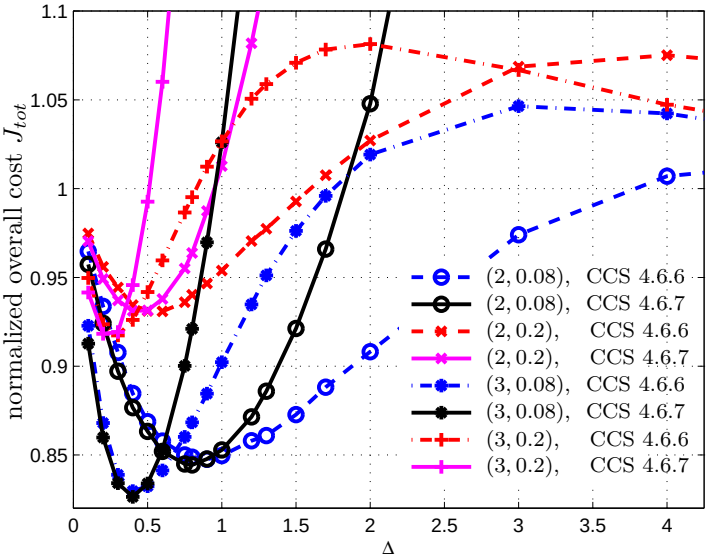


Figure 4.11: The performance of CCS 4.6.6 and CCS 4.6.7 in terms of the step length of the uniform encoder. The notation $(2, 0.08)$ means $R=2$ and $\epsilon=0.08$.

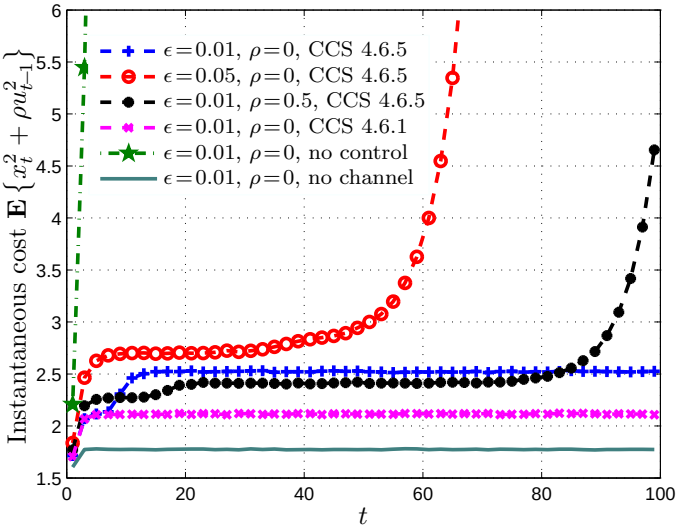
redundancy is introduced by the jointly designed encoder–controller to function as a sort of channel code. In the case of high crossover probabilities, this fact explains why the number of codewords the encoder chooses is less than the available codewords.

In Figure 4.9, Type II coding–control schemes, CCS 4.6.4 and CCS 4.6.5, are evaluated. Compared with CCS 4.6.1, there is an evident gap, attributed to the ignorance of the information in memory. Figure 4.9 also shows that by replacing ℓ_T of CCS 4.6.4 with ℓ_t the performance can be considerably improved without increasing the memory. This is because ℓ_t is derived with respect to the total cost over the entire time horizon, while ℓ_T is obtained by only taking the cost at next step into consideration. As expected, CCS 4.6.4 outperforms the modified control only at the first step. But, the modified system converges to an average cost at a lower level.

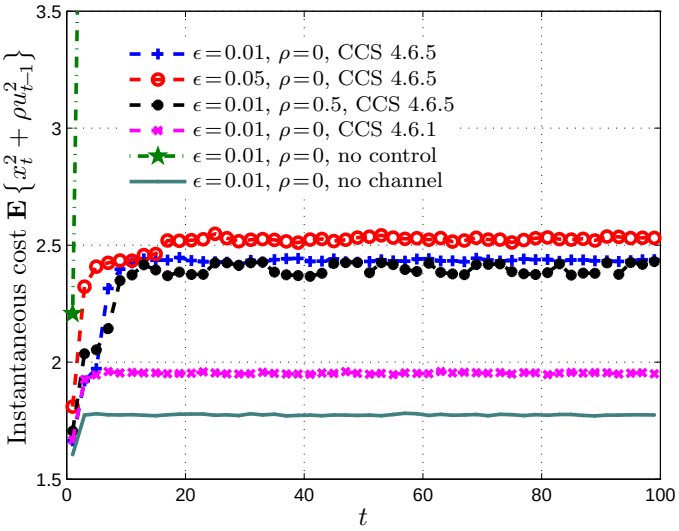
In Figure 4.10, Type III coding–control schemes, CCS 4.6.6 and CCS 4.6.7 are compared with CCS 4.6.1 and CCS 4.6.4. First of all, since both CCS 4.6.6 and CCS 4.6.7 employ optimal time-invariant uniform encoders, their overall performance appear to be equally good. Second, there is also an evident gap between the Type III coding–control schemes and CCS 4.6.1, which illustrates the situation that information in the memory is wasted when it can not be properly exploited. CCS 4.6.1 outperforms the Type III coding–control schemes, although its accessible memory is substantially limited.

Further comparison of the Kalman-filter-based approaches are shown in Figure 4.11. We let the step length of the uniform encoder vary along the x-axis, and the normalized overall cost, with respect to the non-control case, is depicted along the y-axis. Other parameters are the same as in Figure 4.10. The figure reveals that severe consequences can be expected if the information in the memory is not properly exploited. When the step length is small, CCS 4.6.6 performs slightly better than CCS 4.6.7. On the other hand, the expected cost of CCS 4.6.6 grows unbounded as the step length increases. Relatively, the performance of CCS 4.6.7 deteriorates slowly as the step length increases, and the corresponding instantaneous cost converges to a certain value. In general, in the presence of channel errors the impact of data rate becomes more complicated that increasing the rate R does not necessarily always lead to a better performance. Figure 4.11 shows also that low-rate mappings could be more robust to quantization and transmission errors than certain high-rate mappings. For CCS 4.6.6, increasing or reducing the step length, both will worsen the mismatch between the true noise variance and the one the Kalman filter is designed for. If the mismatch is serious, it can have dire consequences. On the other hand, in CCS 4.6.7, all distortions contributing to the difference between the true state and the message decoded at the controller are treated as measurement noise, i.i.d. zero-man Gaussian. Although this assumption differs from reality, CCS 4.6.7 is aware of the true noise variance. In the worst case, the system turns to an open-loop system, without any control action.

Finally in Figure 4.12a, we compare our methods with the coding–control scheme proposed in [GN08]. We adopt the same system parameters used in the numerical



(a) R=2



(b) R=3

Figure 4.12: The performance of CCS 4.6.1 and CCS 4.6.5, applied to the system from [GN08], for different rate R. The common system parameters are: $a = 1.1$, T=100, R=2, σ_w^2=1, $\sigma_{x_0}^2$=1, σ_e^2=1. (a) The rate R is 2. (b The rate R is 3.

example in [GN08] where the system parameters are: $a=1.1$, $T=100$, $R=2$, $\rho=0$, $\epsilon=0.01$, $x_0\sim\mathcal{N}(0,1)$, $v_t\sim\mathcal{N}(0,1)$, $e_t\sim\mathcal{N}(0,1)$. The two reference systems, without controls and without channels, are included in the same figure. Note that, the open-loop system, without any control, is unstable. In [GN08], the authors showed numerically that their scheme was able to stabilize the unstable plant. Here we show in Figure 4.12 that our methods CCS 4.6.1 and CCS 4.6.5 are able to stabilize the unstable plant as well. Compared with [GN08], our schemes achieve a more smooth and lower steady-state level. In the same figure, CCS 4.6.5 is also simulated for higher ρ and ϵ values to demonstrate the negative impact on the overall system performance caused by increasing ρ or ϵ. In the both cases, the stability is violated because of the reduced control power. Figure 4.12b is used to demonstrate that a higher rate may improve the stability of the closed-loop system. At $\rho=0.5$ or $\epsilon=0.05$, CCS 4.6.5 is not able to stabilize the closed-loop system for $R=2$. By increasing the transmission rate R from 2 to 3, the closed-loop system is stabilized.

4.7 Summary

In this chapter, we discussed the implementation of the training algorithm proposed in Chapter 3. A major part was devoted to a discussion of the encoder complexity which was revealed to play an important role in the optimization process. In general, the proposed training algorithm is complicated and the implementation may encounter the dimensionally problem. Overcoming the complexity barrier is really one of the major challenges from the practical point of view. In the first part of this chapter, we discussed two special encoder properties, the sufficient statistics and the regularity, which appeared to be very useful in solving the dimensionality problem in training. We derived a closed-form approximation to compute the expected future costs, which can be used to determine the regularity of the encoder. Thereafter, in the second part the complexity problem encountered when the horizon is large was studied. In particular, a number of low-complexity coding–control strategies subject to various memory restrictions, were empirically compared.

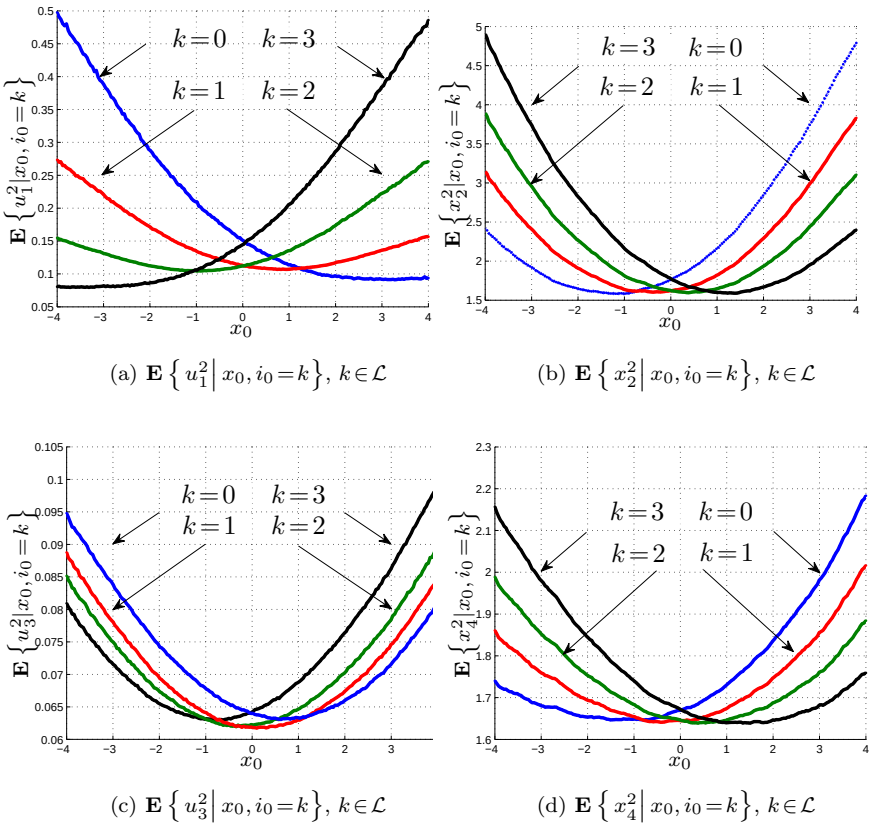
4.A Proof of Lemma 4.5.3

Proof. We prove Lemma 4.5.3 by using induction. Start at $t=2$, the derivation of $p(x_2|\,x_0,\mathbf{i}_0^1,\mathbf{j}_0^1)$ involves a convolution of two pdf's, which can be written as

$$\int_{x_1} A_1 e^{-\frac{W_2^2(x_1,x_0,A_2)}{A_3}} e^{-\frac{W_2^2(x_2,x_1,A_4)}{A_5}} \mathrm{d}x_1, \tag{4.22}$$

where W_2 is as defined in (4.15) and the terms A_1–A_5 are independent of $\mathbf{x}_0^2$. The resulting $p(x_2|\,x_0,\mathbf{i}_0^1,\mathbf{j}_0^1)$ can be written as a sum of terms of the form (4.17), i.e.,

$$B_1 e^{W(x_0,x_2,B_2)} \left(\operatorname{erf}\left(W_2\left(x_0,x_2,B_3\right)\right)-\operatorname{erf}\left(W_2\left(x_0,x_2,B_4\right)\right)\right),$$



(a) $\mathbf{E}\left\{u_1^2\middle|x_0,i_0=k\right\}$, $k\in\mathcal{L}$ (b) $\mathbf{E}\left\{x_2^2\middle|x_0,i_0=k\right\}$, $k\in\mathcal{L}$

(c) $\mathbf{E}\left\{u_3^2\middle|x_0,i_0=k\right\}$, $k\in\mathcal{L}$ (d) $\mathbf{E}\left\{x_4^2\middle|x_0,i_0=k\right\}$, $k\in\mathcal{L}$

Figure 4.13: Predicted costs obtained by numerical simulation.

where B_1–B_4 are vector terms independent of x_0 and x_2, and W and W_2 are as defined in (4.13) and (4.15). As a matter of fact, (4.17) can be expressed as a sum of functions of the form $A_1 e^{W(x_0,x_2,A_2)}$, where A_1 and A_2 are independent of x_0 and x_2.

In order to obtain $\mathbf{E}\left\{x_2^2\middle|x_0,\mathbf{i}_0^1,\mathbf{j}_0^1\right\}$ and $p(x_3|x_0,\mathbf{i}_0^2,\mathbf{j}_0^2)$, it involves the integration of the $\mathrm{erf}(\cdot)$ function. Unfortunately, there is no explicit expression to this integration. Resorting to numerical methods, we depict some $\mathbf{E}\left\{x_t^2\middle|x_0,i_0\right\}$ and $\mathbf{E}\left\{u_t^2\middle|x_0,i_0\right\}$ in Figure 4.13. However, the numerical computation is not only time consuming, but also unable to reveal certain useful properties of the functions. Given the above-mentioned challenges, we solve the problem by approximating the $\mathrm{erf}(x)$ function with certain exponential functions, as shown in (4.16), and thereby deriving a closed-form approximation of the overall cost.

First, consider the derivation of $\mathbf{E}\left\{x_2^2\middle|x_0,\mathbf{i}_0^1,\mathbf{j}_0^1\right\}$. According to (4.22), it is

straightforward to show that the derivation involves the following type of integration,

$$\int_{x_2} x_2^2 A_1 e^{W(x_0, x_2, A_2)} \mathrm{d}x_2,$$

where A_1 and A_2 are independent of x_0 and x_2. The result of the above integration is a sum of functions which can all be written into the form (4.18), i.e.,

$$W_1(x_0, C_1) e^{W_1(x_0, C_2)},$$

where C_1 and C_2 are vectors independent of x_0. Similarly, the derivation of the conditional pdf $p(x_3 | x_0, \mathbf{i}_0^2, \mathbf{j}_0^2)$ leads to integrations of the form

$$\int_{x_2} A_1 e^{W(x_0, x_2, A_2)} e^{W(x_2, x_3, A_3)} \mathrm{d}x_2,$$

and finally a sum of functions of (4.17), i.e.,

$$B_1 e^{W(x_0, x_3, B_2)} \left(\mathrm{erf}(W_2(x_0, x_3, B_3)) - \mathrm{erf}(W_2(x_0, x_3, B_4)) \right),$$

where A_1–A_3 and B_1–B_4 are vector terms independent of x_0 and x_3. This is exactly the same type of function as the approximation of $p(x_2 | x_0, \mathbf{i}_0^1, \mathbf{j}_0^1)$.

As a matter of fact, the above result can be generalized to all t, i.e., any $p(x_t | x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1})$ can be written as a sum of functions of the form (4.17). Similarly, the expectation $\mathbf{E}\left\{ x_t^2 \middle| x_0, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1} \right\}$ can be written as a sum of functions of the form (4.18). □

Chapter 5

Optimized Rate Allocation

5.1 Introduction

In Chapter 3, the problem of how to optimize encoder–controller mappings to improve the efficiency of control over a finite-rate noisy channel was addressed. In this chapter we present an alternative method to advance the overall control performance, given limited communication resources. That is to optimize the allocation of the instantaneous rates over time. How to allocate communication resources over space and time is important. For feedback control system this is a largely open problem. In the literature of control with quantized feedback, it has often been assumed that bits (rates) are evenly distributed to sensor measurements, e.g., [TSM04, NFZE07, BSJ08], mainly for the reason of simplicity. Some work about how to assign bits among the elements of a state vector of the plant, while imposing a constraint on the number of bits over time, can be found in e.g., [LL05a, XJH+05]. However, owing to the non-stationarity of the state observations, it is natural to expect considerable gains by employing time-varying communication resources, i.e., a non-uniform allocation of transmission rates over time. Hence, an even distribution of bits to all instantaneous rates is often not the most efficient solution.

How to achieve the optimal rate allocation in control systems is a challenging task. One obstacle is to find a tractable distortion function, which we need to use as objective functions for the rate optimization problem. Furthermore, such an optimization problem is often non-convex and non-linear, which implies that it is difficult to compute the optimal solution in practice.

The main contribution of this chapter is a novel method of rate allocation for state feedback control of a linear system over a noisy channel. Specifically, we are interested to the rate allocation problem in two important cases in closed-loop control: linear feedback control and state estimation. By resorting to an approximation based on high-rate quantization theory, we are able to derive a computationally feasible scheme that seeks to minimize the overall distortion over a finite time horizon. The resulting rate allocation is not necessarily evenly distributed. Practical con-

siderations concerning integer-rate constraints and the accuracy of high-rate approximations are discussed and illustrated through numerical examples. It is worth remarking that although high-rate theory requires high rates to be valid, this theory can often be used also at lower rates. Experience has shown that high-rate theory can make useful predictions at low rates. Overall good performance of our method is shown by numerical simulations, even the rate is as low as 3, 4 bits per sample.

The rate allocation problem studied in this chapter is related to classical rate allocation problems in communications, e.g., [GG92, FZ06]. Inspired by [GG92, FZ06], we resort to high-rate quantization theory [Ger79, GG92, MN93, GR95, RRM03] to quantify the relation between rate and performance for a general class of quantizers, while previous work has often focused on the special case of optimized quantizers. For example in [Lim05], the problem is studied in the context of transform codes, where the objective function is convex, and a closed-form optimal solution can be derived. In our setting we will show that the overall distortion is a non-convex function of the instantaneous rates, which makes more difficult the computation of the optimal solution.

The rest of this chapter is organized as follows. First, in Section 5.2, the closed-loop control system studied in this chapter is described and the rate allocation problem is presented. Since many results in this chapter are based on high-rate theory, Section 5.3 is devoted to a brief review of some useful results on high-rate quantization. In Section 5.4, a state estimation problem under rate allocation constraints is posed and solved for cases with and without channel errors. Thereafter, Section 5.5 deals with the state feedback control problem by following the same procedure as in Section 5.4. However, the solution derived in Section 5.5 is limited by the fact that the number of equations increases as the horizon T grows, and it is therefore not practically useful when T is large. For this reason, in Section 5.6, a fast bit-rate allocation algorithm is suggested based on certain additional approximations. In many cases the low-complexity solution from Section 5.6 works sufficiently well. In addition, search methods based on testing are discussed in Section 5.7. In the same section, we propose an efficient search algorithm, specially customized for the rate allocation in control systems. Practical issues such as how to deal with non-negativity and integer constraints can be found in Section 5.8, as well as some remarks on the performance degradation caused by various approximations and simplifications. In Section 5.9, numerical simulations are carried out to demonstrate the performance of the proposed bit-rate allocations. Finally, a summary of the chapter is given in Section 5.10.

5.2 System Description and Problem Statement

The goal of this chapter is to arrive at a practical rate allocation scheme for state feedback control over a noisy channel. Figure 5.1 shows a block-diagram of the control systems studied in this chapter. Following Figure 5.1, we first briefly introduce each building block of the system, and then describe the rate constrained optimiza-

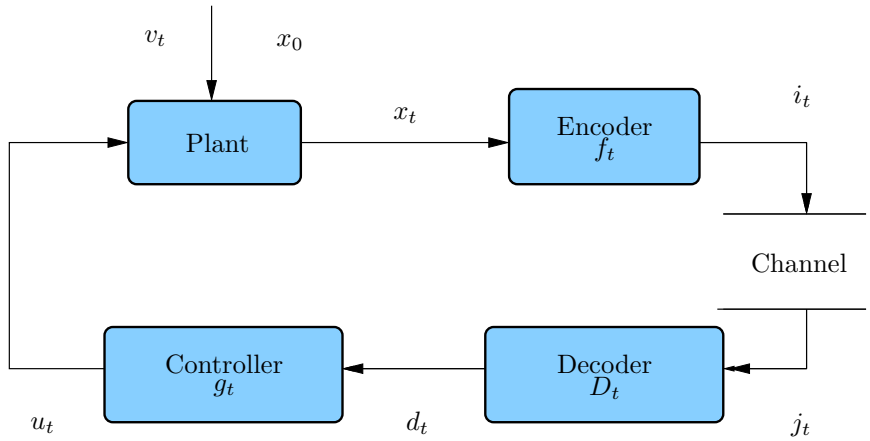


Figure 5.1: Block-diagram for the closed-loop system studied in this chapter. The system has a separate decoder unit and a controller.

tion problem. Finally, the coding unit and the channel will be further explained.

5.2.1 System Description

We throughout this chapter consider a special case of the general model of Chapter 2, namely a scalar system with full state observation (a system without measurement noise), to focus mainly on the influence of a current decision on future events, which is one of the most fundamental properties of a control system. More specifically, the linear plant is governed by the equation

$$x_{t+1} = ax_t + u_t + v_t, \quad a > 0, \tag{5.1}$$

where x_t, u_t, $v_t \in \mathbb{R}$. Process noise v_t is modeled as an i.i.d. Gaussian zero-mean process with a time-invariant variance σ_v^2. The process noise is mutually independent of the initial-state x_0, which is also i.i.d. zero-mean Gaussian, i.e., $x_0 \sim \mathcal{N}(0, \sigma_{x_0}^2)$.

At the encoder, the full state measurement is coded by a memoryless time-varying encoder, which takes only the current state x_t as input, and produces an index i_t,

$$i_t = f_t(x_t) \in \mathcal{L}_t = \left\{0, \ldots, 2^{R_t} - 1\right\}, \quad R_t \in \mathbb{Z}^+, \ t = 0, \ldots, T-1, \tag{5.2}$$

where we allow the rate R_t to be a time-varying non-negative integer. The index i_t will be mapped into a binary codeword before being fed into a binary channel. Details concerning the channel model will be given later. Recall, the mapping from an index to a codeword is commonly referred to as the index assignment (IA). Unlike in the error-free scenario where all IA's perform equally well, in the presence of channel errors different IA's have different impact on the system performance.

Finding the optimal IA is a combinatorial problem which is known to be NP-hard [Far90]. In this chapter, we therefore average out the dependence on a specific IA by randomization. At each transmission, a random assignment is generated and revealed to the encoder and decoder. Previous work that assumed a random IA to facilitate further analysis includes [ZM94, MR06]. Of course, to assign IA randomly for each transmission is impractical in real communication systems. However, using the random IA in the analysis can characterize the average performance for a given rate allocation, and one can always find at least one IA which performs as good as the random assignment. Often, it is possible to find IA's which outperform the random assignment. Therefore, in practice, we can first use random IA in the analysis to optimize the rate allocation. Then, for the optimized rate allocation we can use an IA which performs better than the random assignement.

At the receiver side, there is a separate decoder unit and a controller. The decoder takes the instantaneous channel output $j_t \in \mathcal{L}_t$ as the input, and produces an estimate of x_t, denoted by d_t,

$$d_t = D_t(j_t) \in \mathbb{R}, \tag{5.3}$$

where $D_t(\cdot)$ is a deterministic function. The estimate d_t can take on one of 2^{R_t} values, referred to as the reconstructions. For brevity, we also use d_t to represent the decoding codebook, in particular, the notation $d_t(k)$ specifies the decoded value associated with the integer value k. Accordingly, $d_t(i_t)$ is the reconstruction chosen by the encoder, and $d_t(j_t)$ by the decoder. Finally, the control u_t is computed based on the decoded symbol, i.e.,

$$u_t = g_t(d_t) \in \mathbb{R}.$$

We will be more specific about the control law g_t after the rate allocation problem is presented.

5.2.2 Problem Statement

In short, the goal is to minimize the expected overall cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$, subject to a total rate constraint. More specifically, the overall performance measure $J_{tot}(\mathbf{R}_0^{T-1})$ is given by

$$J_{tot}(\mathbf{R}_0^{T-1}) = \sum_{t=1}^{T} J_t(\mathbf{R}_0^{t-1}) = \sum_{t=1}^{T} x_t^2 + \rho u_{t-1}^2, \quad \rho \geq 0, \tag{5.4}$$

where J_t denotes the instantaneous cost, and ρ is the importance factor of the control input with respect to the state. The rate constraint is

$$\sum_{t=0}^{T-1} R_t \leq R_{tot}, \quad R_t \in \mathbb{Z}^+, \ t = 0, \ldots, T-1. \tag{5.5}$$

That is to say, the sum of the instantaneous rates cannot exceed R_{tot}, the *total rate*. The sequence of rates is denoted by $\mathbf{R}_0^{T-1} = \{R_0, \ldots, R_{T-1}\}$. We refer to $\mathbf{R}_0^{T-1}$ as the *bit-rate allocation*. Throughout this chapter, we say "for all t" when we mean "for $t = 0, \ldots, T-1$". Notice that, the implicit relation of the rate allocation $\mathbf{R}_0^{T-1}$ and the cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ is closely related to the channel and coding–control scheme, which will be specified next.

Generally speaking, it is a hard problem to optimize the average cost $\mathbf{E}\{J_{tot}\}$ with respect to both the coding–control scheme and the rate allocation. Therefore, we take a more problematic approach and assume that the controller is given by the classical LQG state feedback control law. That is, the control is taken to be a linear function of the decoded symbol d_t,

$$u_t = \ell_t d_t, \tag{5.6}$$

where the linear control law ℓ_t is calculated as

$$\ell_t \triangleq -\frac{a\phi_{t+1}}{\phi_{t+1} + \rho}, \quad \phi_t = 1 + \frac{a^2\phi_{t+1}\rho}{\phi_{t+1} + \rho}, \text{ with } \phi_T = 1. \tag{5.7}$$

That is to say, if the estimate d_t is close to the true state x_t then classical linear quadratic Gaussian (LQG) theory [Aok67, Ber76] is expected to give good results, even though this theory does not account for channel errors and quantization distortion.

Next, we specify the communication units in the closed-loop system, i.e., the channel and encoder–decoder pair.

Binary Symmetric Channel and Random Index Assignment

Let the discrete memoryless channel have the input $i_t \in \mathcal{L}_t$ and the output $j_t \in \mathcal{L}_t$. A discrete memoryless channel is described by the transition probability function $\mathcal{P}(j_t|i_t)$. In this chapter, two types of discrete channels are considered: (i) the finite-rate error-free channel, and (ii) the noisy channel consisting of a BSC and a random IA.

The combination of the random IA and a BSC forms the end-to-end channel seen by the encoder and decoder. The end-to-end channel is completely specified by the symbol transition probability function $\mathcal{P}(j_t|\, i_t)$. At the bit level, the channel is characterized by the crossover probability $\epsilon = \mathcal{P}(0|\,1) = \mathcal{P}(1|\,0)$ of the BSC. Because of the symmetry, it is reasonable to consider only $0 \le \epsilon \le 0.5$. The overall symbol error probability $\mathcal{P}(j_t|i_t)$ of the end-to-end channel is determined by both ϵ and the randomized IA, according to

$$\mathcal{P}(j_t|\, i_t) = \begin{cases} \alpha(R_t), & j_t \neq i_t, \\ 1 - (2^{R_t} - 1)\alpha(R_t), & j_t = i_t. \end{cases} \tag{5.8}$$

Here, $\alpha(R_t)$ is obtained by averaging over all possible IA's [ZM94],

$$\begin{aligned}\alpha\left(R_t\right) &\triangleq \frac{1}{\prod_{k=1}^{2^{R_t}} k} \sum_{b_t \in \mathcal{B}_t} \mathcal{P}(b_t(j_t)|b_t(i_t)) \\ &= \frac{1}{\prod_{k=1}^{2^{R_t}} k}(2^{R_t}-2)! \left(\sum_{n=0}^{2^{R_t}-1} \sum_{m=0, m\neq n}^{2^{R_t}-1} \mathcal{P}(j_t = m|i_t = n) \right) \\ &= \frac{1-(1-\epsilon)^{R_t}}{2^{R_t}-1},\end{aligned}$$

with $(\cdot)!$ denoting the factorial. The set $\mathcal{B}_t$ contains all the functions $b_t : \mathcal{L}_t \mapsto \{0,1\}^{R_t}$, also referred to as IA's. For the channel (5.8), all symbol errors are equally probable.

Clearly, the error-free channel is the special case with $\epsilon = 0$ and $j_t = i_t$, i.e., the channel output is always identical to the channel input. The only restriction imposed by this channel is the finite-rate R_t. The reason for considering error-free channels is mainly to study the impact of quantization.

Two Examples of Encoder–Decoders

Throughout the chapter, encoder–decoder pairs are memoryless mappings equivalent to a quantizer. We will use two conventional quantizers for demonstrating concepts and techniques. They are: (i) the uniform quantizer, and (ii) the memoryless source-optimized quantizer. We choose to study these two quantizers mostly because they are simple to implement and feasible to analyze. Also, we use them to represent different degrees of source information the controller can take advantage of. Usually, more statistical information is exploited by the source-optimized quantizer than by the uniform quantizer.

Example 5.2.1. ***Uniform Quantizer***

Owing to its simplicity, the uniform quantizer is thoroughly studied in the literature and commonly used in practice. We describe the step length Δ_t of a uniform quantizer as a function of the rate R_t and the quantizer range $[-\nu_t, \nu_t]$,

$$\Delta_t = \frac{2\nu_t}{2^{R_t}}.$$

The quantizer works as follows,

$$i_t = \begin{cases} 0, & x_t < -\nu_t + \Delta_t, \\ k, & -\nu_t + k\Delta_t \leq x_t \leq -\nu_t + (k+1)\Delta_t, \quad k \in \mathbb{N}, \\ 2^{R_t}-1, & x_t > \nu_t - \Delta_t. \end{cases}$$

The range ν_t can be selected with respect to the source, see Example 5.2.2 next.

Example 5.2.2. ***Uniform Quantizer and Gaussian source***

Consider a source x_t, and let p_{ν_t} denote the probability that x_t is within the range of the quantizer, i.e., $p_{\nu_t} \triangleq \mathcal{P}(x_t \in [-\nu_t, \nu_t])$. If x_t is zero-mean Gaussian with variance $\sigma^2_{x_t}$, ν_t is related to p_{ν_t} and σ^2_x as follows,

$$\nu_t = \sigma_{x_t} Q^{-1}\left(\frac{1-p_{\nu_t}}{2}\right),$$

where $Q^{-1}(\cdot)$ is the inverse function of the Q-function [Pro95], with the Q-function defined as

$$Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} \mathrm{d}y.$$

Example 5.2.3. ***Source-Optimized Quantizer***

Besides the uniform quantizer, we will also study a so-called source-optimized quantizer. The index i_t is chosen according to the following rule

$$\begin{aligned} i_t &= \arg\min_k \left\{(x_t - d_t(k))^2\right\}, \\ d_t(k) &= \arg\min_{y\in\mathbb{R}} \left\{\mathbf{E}\left\{(x_t - y)^2 \middle| x_t \in \mathcal{S}_t(k)\right\}\right\}, \end{aligned}$$

where $\mathcal{S}_t(k) \triangleq \{x_t : i_t = k\}$ denotes a quantization cell. The decoding rule is

$$d_t = d_t(k), \quad \text{if } j_t = k,$$

where j_t is the index received at the decoder. Here, the quantization cell $\mathcal{S}_t(k)$ and the reconstruction $d_t(k)$ are optimized only with respect to the pdf $p(x_t)$ and the quadratic cost $\mathbf{E}\left\{(x_t - d_t(i_t))^2\right\}$, without taking any potential channel error into account. However, if transmission error occurs seldom, the source-optimized quantizer is expected to still achieve a good performance.

Summarizing the above discussions, Problem 5.2.4 below specifies the rate allocation problems studied in this chapter.

Problem 5.2.4. *Given the linear plant* (5.1)*, the memoryless channel* (5.8)*, the memoryless encoder–decoder pair* (5.2)–(5.3)*, and the control law* (5.6)–(5.7)*, find the optimal bit-rate allocation $\mathbf{R}_0^{T-1}$, $R_t \in \mathbb{Z}^+$, $\forall t$, which minimizes the expected cost of* (5.4)*, subject to the total bit-rate constraint* (5.5)*, i.e.,*

$$\begin{aligned} \min_{\mathbf{R}_0^{T-1}} \quad & \mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\} \\ \text{s.t.} \quad & \sum_{t=0}^{T-1} R_t \le R_{tot}, \quad R_t \in \mathbb{Z}^+, \ t = 0, \ldots, T-1. \end{aligned}$$

As stated in Problem 5.2.4, the rate R_t is a non-negative integer. In this work, we will first solve a relaxed problem by optimizing the rate allocation for $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$,

Figure 5.2: Block-diagram for the communication over a BSC. The symbol Π is a notation for the IA unit.

and thereafter, treat the non-negativity and integer constraints separately. One of the main challenges of Problem 5.2.4 is that the cost function does not have a closed-form expression in terms of $\mathbf{R}_0^{T-1}$. In the next section, we propose an approximation, which will then be used to derive the solution of the bit-rate allocation problem.

5.3 High-Rate Approximation of MSE

The rate allocation proposed in this chapter can be generalized, by which we mean that it can be readily adopted to a variety of quantizers under certain assumptions. To achieve this goal, we need to formulate a general objective function. The first major challenge lies in deriving a useful expression for the mean squared error, which appears to be a central figure-of-merit not only in the state estimation problem but also in the state feedback control problem. In general, it is difficult to formulate closed-form expressions, even in the case of simple uniform quantizers. Inspired by the classical works, e.g., [GG92], we resort to high-rate quantization to compute MSE. For this reason, some results on high-rate quantization theory are briefly reviewed in this section. For further detail about high-rate quantization theory, we refer the reader to e.g., [ZM94, MR06]. It is worth remarking that although high-rate quantization requires high rates to be valid. However, in practice this theory is also useful at low rates, such as $R_t = 3$, 4 bits.

Figure 5.2 illustrates the signal path from the source to the destination, through a noisy channel. Consider transmitting one sensor measurement over the channel. The source signal x_t is first fed into an encoder, producing the coded index $i_t \in \{0, \ldots, 2^{R_t} - 1\}$, with R_t denoting the instantaneous rate. Each index value will be mapped to a unique binary codeword of length R_t, by means of the IA unit. As stated previously, IA is important to the overall system performance. A good IA exploits knowledge about the source, the channel, and the encoder–decoder. However, finding the optimal IA is a combinatorial problem which is known to be NP-hard. As discussed above, we therefore average out the dependence on a specific IA by randomization. At each transmission, a random assignment is generated and revealed to the encoder and decoder. At the receiver side, the inverse procedures of IA and quantization are applied to channel outputs to produce an estimate d_t. Given a source with the pdf $p(x_t)$, a channel $\mathcal{P}(j_t | i_t)$, and an encoder–decoder pair

of the rate R_t, the MSE can be calculated as

$$\mathbf{E}\left\{(x_t - d_t)^2\right\} = \sum_{k=0}^{2^{R_t}-1} \sum_{l=0}^{2^{R_t}-1} \mathcal{P}(j_t = l \,|\, i_t = k) \int_{x_t \in \mathcal{S}_t(k)} |x_t - d_t(l)|^2 p(x_t) \mathrm{d}x_t, \tag{5.9}$$

where $\mathcal{S}_t(k) \triangleq \{x_t : i_t = k\}$ is the quantization cell that all $x_t \in \mathcal{S}_t(k)$ will be assigned the index value k.

To quantify the impact of the rate R_t on the MSE (5.9), we need to be more specific about the source, the channel, and the quantizer. In general, it is difficult to formulate a closed-form expression for (5.9). Given this difficulty, we show below a useful approximation of (5.9), derived under the high-rate assumption. Roughly speaking, the high-rate assumption requires that the pdf of the source is approximately constant within the same quantization cell. Following [MR06], for a symmetric channel, e.g., (5.8), a high-rate approximation of the MSE is

$$\begin{aligned}\mathbf{E}\left\{(x_t - d_t)^2\right\} \approx 2^{R_t}\alpha(R_t)\sigma_{x_t}^2 + \varphi_t \alpha(R_t) \int_{y\in\mathbb{R}} y^2 \lambda_t(y) \mathrm{d}y \\ + \frac{G^{-2}}{3} \varphi_t^{-2} \int_{x\in\mathbb{R}} \lambda_t^{-2}(x) p(x_t = x) \mathrm{d}x,\end{aligned} \tag{5.10}$$

where the source x_t is zero-mean with variance $\sigma_{x_t}^2$. A brief explanation of (5.10) is given here, and more detail of this expression can be found in Appendix 5.A. First, the constant G represents the volume of a unit sphere, and for a scalar quantizer $G = 2$. Second, the function $\lambda_t(x)$ is referred to as the *point density function*, specifying the density of reconstruction values of the quantizer. Resembling a probability density function, it follows that $\lambda_t(x) \geq 0$, for all x, and $\int_{\mathbb{R}} \lambda_t(x) \mathrm{d}x = 1$. Finally, φ_t, $1 \leq \varphi_t \leq 2^{R_t}$, specifies the number of codewords the encoder will chose. If the crossover probability ϵ is large-valued, in order to protect against the channel error, a good encoder may only use a part of the available codewords. Both the uniform quantizer and the source-optimized quantizer have $\varphi_t = 2^{R_t}$, for all t.

Essentially, we are interested in an efficient approximation to describe the relation between the MSE and the rate R_t. By a further approximation, $2^{R_t}\alpha(R_t) \approx 1 - (1-\epsilon)^{R_t}$, we rewrite (5.10) and introduce the high-rate approximation $\hat{J}_t$,

$$\mathbf{E}\left\{(x_t - d_t)^2\right\} \approx \hat{J}_t(\beta_t, \kappa_t, R_t) \triangleq \beta_t(1 - (1-\epsilon)^{R_t}) + \kappa_t 2^{-2R_t}, \tag{5.11}$$

where β_t and κ_t are,

$$\beta_t \triangleq \sigma_{x_t}^2 + \int_{y\in\mathbb{R}} y^2 \lambda_t(y) \mathrm{d}y, \tag{5.12}$$

$$\kappa_t \triangleq \bar{G} \int_{x\in\mathbb{R}} \lambda_t^{-2}(x) p(x) \mathrm{d}x, \tag{5.13}$$

with $\bar{G} \triangleq G^{-2}/3$. According to (5.11)–(5.13), β_t and κ_t can take on any non-negative real value, including 0 and ∞. However, for practical sources and encoder–decoder

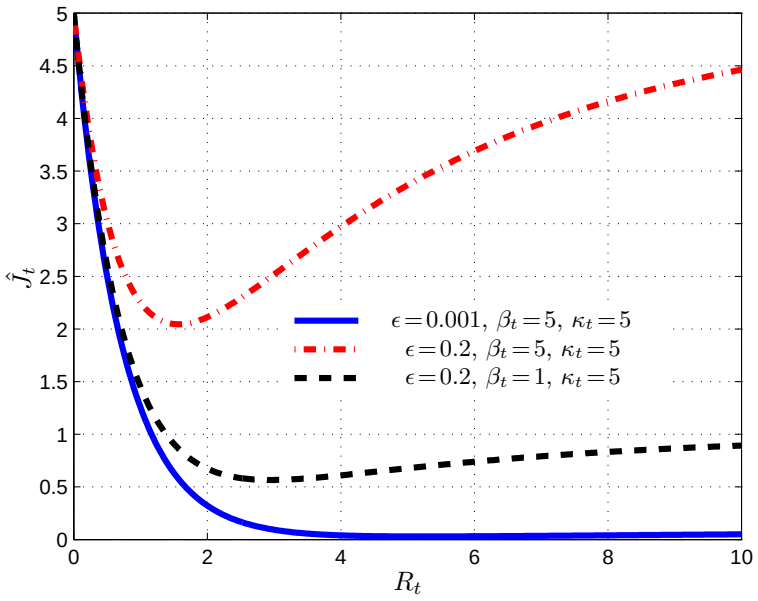


Figure 5.3: The impact of R_t, κ_t, β_t and ϵ on the distortion $\hat{J}_t$ in (5.11).

pairs, it follows that $0<\beta_t<\infty$, $0<\kappa_t<\infty$, which is considered throughout this chapter. The expression of the distortion $\hat{J}_t$ in (5.11) is rather general for a large variety of quantizers, described by means of the point density function, and derived under the high-rate assumption.

The expression in (5.11) has certain useful properties that will allow us to solve the rate allocation problem. In Figure 5.3 we illustrate the impact of the parameters R_t, κ_t, β_t, and ϵ on the distortion $\hat{J}_t$ for three cases. First, it should be observed that the crossover probability ϵ is instrumental to the convexity of the function. When $\epsilon=0$, $\hat{J}_t$ is always monotonically decreasing. In fact, $\hat{J}_t$ is a convex function with respect to R_t. On the other hand, for noisy channels, convexity only holds for certain $\{\beta_t,\kappa_t\}$ pairs. For the general case of an arbitrary $\{\beta_t,\kappa_t\}$ pair, (5.11) is a so-called quasi-convex function, as explained below in Lemma 5.3.2.

Definition 5.3.1. ***Quasi-convex function [BV04]***

A function $f:\mathbb{R}^n\mapsto\mathbb{R}$ is quasi-convex if its domain, $domf$, and all its sub-level sets $\mathcal{S}_\alpha=\{x\in domf : f(x)\leq\alpha\}$, for $\alpha\in\mathbb{R}$, are convex.

Lemma 5.3.2. *The distortion function*

$$\hat{J}_t(\beta_t,\kappa_t,R_t)=\beta_t(1-(1-\epsilon)^{R_t})+\kappa_t 2^{-2R_t},\quad 0<\beta_t<\infty,\ \ 0<\kappa_t<\infty,$$

is a quasi-convex function and has a unique global minimum.

Proof. Compute the first order derivative of $\hat{J}_t$, with respect to R_t,

$$\frac{\partial \hat{J}_t}{\partial R_t}(\beta_t, \kappa_t, R_t) = -\beta_t \ln(1-\epsilon)(1-\epsilon)^{R_t} - 2\kappa_t \ln(2) 2^{-2R_t}.$$

The first part, $-\beta_t \ln(1-\epsilon)(1-\epsilon)^{R_t}$, is strictly decreasing towards 0 as R_t goes to infinity, i.e., $\lim_{R_t\to\infty} -\beta_t \ln(1-\epsilon)(1-\epsilon)^{R_t} =\downarrow 0$. The second part $-2\kappa_t \ln(2) 2^{-2R_t}$ is strictly increasing towards 0 as R_t grows, i.e., $\lim_{R_t\to\infty} -2\kappa_t \ln(2) 2^{-2R_t} =\uparrow 0$. Note also, since $0 \leq \epsilon \leq 0.5$, $(1-\epsilon)^{R_t}$ decreases more slowly than 2^{-2R_t}. Accordingly, we can conclude that $\partial \hat{J}_t / \partial R_t$ has at most one critical point $R_t^\star$, which solves the following equation

$$\frac{\partial \hat{J}_t}{\partial R_t}(\beta_t, \kappa_t, R_t^\star) = -\beta_t \ln(1-\epsilon)(1-\epsilon)^{R_t^\star} - 2\ln(2)\kappa_t 2^{-2R_t^\star} = 0.$$

In case that $\epsilon = 0$, the critical point is always at infinity, i.e., $R_t^\star = \infty$, for all $0 < \kappa_t < \infty$, since $\lim_{R_t\to\infty} \partial \hat{J}_t / \partial R_t = 0$. Similarly, compute the second order derivative of $\hat{J}_t$, with respect to R_t,

$$\frac{\partial^2 \hat{J}_t}{\partial R_t^2}(\beta_t, \kappa_t, R_t) = -\beta_t (\ln(1-\epsilon))^2 (1-\epsilon)^{R_t}) + 4(\ln 2)^2 \kappa_t 2^{-2R_t}. \tag{5.14}$$

We can show that the critical point is a minimum, since $\lim_{R_t\to 0} \partial^2 \hat{J}_t / \partial R_t^2 > 0$. Then, for all $R_t < R_t^\star$, $\partial \hat{J}_t / \partial R_t$ is negative, and $\hat{J}_t$ is monotonically decreasing. Conversely, for all $R_t > R_t^\star$, $\partial \hat{J}_t / \partial R_t$ is positive, and $\hat{J}_t$ is monotonically increasing, cf., Figure 5.3. □

As will be shown later, Lemma 5.3.2 is instrumental to solve the rate allocation problems studied in this chapter.

5.3.1 Examples

In the remaining part of this section, we compute $\hat{J}_t$ for two examples: the uniform quantizer and the source-optimized quantizer, introduced in Section 5.2.

Example 5.3.3. ***Uniform Quantizer***

For a uniform quantizer with a quantization range $[-\nu_t, \nu_t]$, the point density function is

$$\lambda_t(x_t) = \frac{1}{2\nu_t}.$$

If the source signal and the uniform quantizer share the same range $[-\nu_t, \nu_t]$, the high-rate approximation $\hat{J}_t$, according to (5.11)*, is*

$$\hat{J}_t = \left(\sigma_{x_t}^2 + \frac{\nu_t^2}{3}\right)\left(1 - (1-\epsilon)^{R_t}\right) + 4\nu_t^2 \bar{G} 2^{-2R_t}, \tag{5.15}$$

which means that the parameters β_t and κ_t of (5.12)–(5.13) *are*

$$\beta_t = \sigma_{x_t}^2 + \frac{\nu_t^2}{3}, \quad \kappa_t = 4\nu_t^2 \bar{G}. \tag{5.16}$$

The first order derivative of $\hat{J}_t$, with respect to R_t, becomes

$$\frac{\partial \hat{J}_t}{\partial R_t} = -\left(\sigma_{x_t}^2 + \frac{\nu_t^2}{3}\right)(1-\epsilon)^{R_t} \ln(1-\epsilon) - 8\nu_t^2 \bar{G} 2^{-2R_t}.$$

In the absence of channel errors, the high-rate distortion $\hat{J}_t$ and its first order derivative with respect to R_t, are

$$\hat{J}_t = 4\nu_t^2 \bar{G} 2^{-2R_t}, \quad \frac{\partial \hat{J}_t}{\partial R_t} = -8\nu_t^2 \bar{G} 2^{-2R_t}.$$

Example 5.3.4. ***Source-Optimized Quantizer***

A source-optimized quantizer minimizes the MSE distortion (5.9) *for the special case where channel errors are absent. As shown in literature, e.g., [GG92, NN95], the point density function for this class of quantizers is given by*

$$\lambda_t(x_t) = \frac{(p(x_t))^{1/3}}{\int_{\mathbb{R}} (p(x_t))^{1/3} \mathrm{d}x_t}.$$

For Gaussian distributed sources (shown in Appendix 5.B), the high-rate approximation of the MSE distortion is

$$\hat{J}_t = 4\sigma_{x_t}^2 (1 - (1-\epsilon)^{R_t}) + \mu \sigma_{x_t}^2 2^{-2R_t}, \tag{5.17}$$

where μ, determined by the pdf of the normalized source signal $y = x_t/\sigma_{x_t}^2$, is

$$\mu \triangleq \frac{1}{12}\left(\int_{y\in\mathbb{R}} (p(y))^{\frac{1}{3}}\, \mathrm{d}y\right)^3 = \frac{1}{12}\left(\int_{y\in\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}}\right)^{\frac{1}{3}} \mathrm{d}y\right)^3 = \frac{\sqrt{3}\pi}{2}. \tag{5.18}$$

Accordingly, the parameters β_t and κ_t, are

$$\beta_t = 4\sigma_{x_t}^2, \quad \kappa_t = \mu \sigma_{x_t}^2, \tag{5.19}$$

with μ as defined in (5.18). *The first order derivative is simply*

$$\frac{\partial \hat{J}_t}{\partial R_t} = -4\sigma_{x_t}^2 \ln(1-\epsilon)(1-\epsilon)^{R_t} - 2\mu\sigma_{x_t}^2 2^{-2R_t}.$$

For an error-free channel ($\epsilon = 0$), $\hat{J}_t$ and $\partial \hat{J}_t/\partial R_t$ are

$$\hat{J}_t = \mu \sigma_{x_t}^2 2^{-2R_t}, \quad \frac{\partial \hat{J}_t}{\partial R_t} = -2\mu\sigma_{x_t}^2 2^{-2R_t}.$$

Here we mention a special class of $\hat{J}_t$, which can be written as

$$\hat{J}_t = \sigma_{x_t}^2 \left(\tilde{\beta}_t(1-(1-\epsilon)^{R_t}) + \tilde{\kappa}_t 2^{-2R_t}\right) = \sigma_{x_t}^2 \tilde{J}_t(\tilde{\beta}_t, \tilde{\kappa}_t, R_t), \tag{5.20}$$

where $\tilde{J}_t(\tilde{\beta}_t, \tilde{\kappa}_t, R_t) \triangleq \tilde{\beta}_t(1-(1-\epsilon)^{R_t}) + \tilde{\kappa}_t 2^{-2R_t}$, and $0 < \tilde{\beta}_t < \infty$ and $0 < \tilde{\kappa}_t < 0$ are independent of R_t and $\sigma_{x_t}^2$. As will be shown later in Section 5.5, this class of $\hat{J}_t$ is central to our solutions to the state feedback control problems. The first and the second order derivatives of $\tilde{J}_t(\tilde{\beta}_t, \tilde{\kappa}_t, R_t)$ with respect to R_t are

$$\frac{\partial \tilde{J}_t}{\partial R_t}(\tilde{\beta}_t, \tilde{\kappa}_t, R_t) = -\tilde{\beta}_t \ln(1-\epsilon)(1-\epsilon)^{R_t} - 2\ln(2)\tilde{\kappa}_t 2^{-2R_t}, \tag{5.21}$$

$$\frac{\partial^2 \tilde{J}_t}{\partial R_t^2}(\tilde{\beta}, \tilde{\kappa}, R_t) = -\tilde{\beta}_t (\ln(1-\epsilon))^2 (1-\epsilon)^{R_t} + 4(\ln 2)^2 \tilde{\kappa}_t 2^{-2R_t}. \tag{5.22}$$

Owing to the fact that $\tilde{J}_t$ is a special case of $\hat{J}_t$, Lemma 5.3.2 applies directly to $\tilde{J}_t$. The next two examples are used to demonstrate the utility of (5.20).

Example 5.3.5. ***Gaussian Source and Source-Optimized Quantizer***

Consider a zero-mean Gaussian source and a source-optimized quantizer, as described in Example 5.3.4. The MSE under the high-rate assumption can be approximated by (5.20)*, where*

$$\tilde{\beta}_t = 4, \quad \tilde{\kappa}_t = \mu,$$

with μ *as defined in* (5.18).

Example 5.3.6. ***Gaussian Source and Uniform Quantizer***

Consider a zero-mean Gaussian source and a uniform quantizer as described in Example 5.3.3. If the distortion caused by signals out of the quantizer support $[-\nu_t, \nu_t]$ *is negligible, the MSE under the high-rate assumption can be approximated by* (5.20) *with the following* $\tilde{\beta}_t$ *and* $\tilde{\kappa}_t$

$$\tilde{\beta}_t = 1 + \frac{\left(Q^{-1}\left(\frac{1-p_{\nu_t}}{2}\right)\right)^2}{3}, \quad \tilde{\kappa}_t = 4\bar{G}\left(Q^{-1}\left(\frac{1-p_{\nu_t}}{2}\right)\right)^2.$$

The high-rate approximation of MSE described in this section is instrumental to formulate useful objective functions for the rate allocation problems, as shown in the next two sections. Again, we stress that in practice high-rate theory has been shown to be also very useful at low rates.

5.4 Rate Allocation for State Estimation

This section discusses the special case of the general system (5.1) when $u_t = 0$. For this dynamic system, we can formulate a rate allocation problem for state

estimation, where the criterion is motivated by the closed-loop control, as clarified later in Section 5.5. The estimation problem is easier to solve, compared with the analogues control problem. For this reason, we will first in this section study the rate allocation problem for state estimation before tackling Problem 5.2.4.

Consider now a linear plant governed by the equation

$$x_{t+1} = ax_t + v_t, \quad a > 0. \tag{5.23}$$

Following Section 5.2, the mutually independent initial-state and process noise are i.i.d. zero-mean Gaussian with variances $\sigma_{x_0}^2$ and σ_v^2, respectively. Now, the goal is to minimize the expected overall estimation error $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ with $J_{tot}(\mathbf{R}_0^{T-1})$ given by

$$J_{tot}(\mathbf{R}_0^{T-1}) = \sum_{t=0}^{T-1} J_t(R_t) = \sum_{t=0}^{T-1} (x_t - d_t)^2. \tag{5.24}$$

Problem 5.4.1 below specifies the rate allocation problem studied in this section.

Problem 5.4.1. *Given the linear plant* (5.23)*, the channel* (5.8)*, and the encoder–decoder mapping* (5.2)–(5.3)*, find the optimal bit-rate allocation* $\mathbf{R}_0^{T-1}$ *which minimizes the expected value of the LQ cost* (5.24)*, subject to the total bit-rate constraint* (5.5)*, i.e.,*

$$\begin{aligned} \min_{\mathbf{R}_0^{T-1}} \quad & \mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}, \\ \text{s.t.} \quad & \sum_{t=0}^{T-1} R_t \leq R_{tot}, \quad R_t \in \mathbb{Z}^+, t = 0, \ldots, T-1, \end{aligned}$$

with $J_{tot}(\mathbf{R}_0^{T-1})$ *given by* (5.24)*.*

Note that the average instantaneous distortion

$$\mathbf{E}\left\{J_t(R_t)\right\} = \mathbf{E}\left\{(x_t - d_t)^2\right\}, \tag{5.25}$$

is a function of R_t because of the reconstruction d_t. According to (5.23), we can write the state x_t as a function of the initial-state x_0 and the process noises $\mathbf{v}_0^{t-1}$

$$x_t = a^t x_0 + \sum_{s=0}^{t-1} a^{t-1-s} v_s.$$

Since x_0 and $\mathbf{v}_0^{t-1}$ are i.i.d. zero-mean Gaussian, consequently, x_t is also zero-mean Gaussian with the variance

$$\sigma_{x_t}^2 = a^{2t}\sigma_{x_0}^2 + \sum_{s=0}^{t-1} \left(a^{t-1-s}\right)^2 \sigma_v^2.$$

We recall that the state x_t does not depend on the communication over the noisy link. Therefore, x_t is not affected by the rate allocation, and consequently, the instantaneous distortion functions are separable, i.e., the instantaneous distortion $J_t(R_t)$ depends only on the current rate R_t. As a result, the major challenge lies in deriving a useful expression of the MSE, cf., the instantaneous distortion (5.25). In general, it is difficult to formulate closed-form expressions, even in the case of simple uniform quantizers. In order to proceed, we resort to approximations based on high-rate theory. More specifically, we approximate the distortion $\mathbf{E}\{J_t(R_t)\}$ of (5.25) by the high-rate expression $\hat{J}_t(\beta_t, \kappa_t, R_t)$ of (5.11), and solve the rate allocation problem with respect to the optimization instantaneous distortion,

$$\mathbf{E}\{J_t(R_t)\} \approx \hat{J}_t(\beta_t, \kappa_t, R_t) = \beta_t(1-(1-\epsilon)^{R_t}) + \kappa_t 2^{-2R_t}. \tag{5.26}$$

Next, we introduce the rate unconstrained and constrained optimization problems which are approximate versions of Problem 5.4.1.

Problem 5.4.2. *Find the rate allocation* $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$ *which solves the problem,*

$$\min_{\mathbf{R}_0^{T-1}} \quad \sum_{t=0}^{T-1} \hat{J}_t(\beta_t, \kappa_t, R_t),$$

where $\hat{J}_t(\beta_t, \kappa_t, R_t)$ *is given by* (5.26).

Problem 5.4.3. *Find the rate allocation* $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$ *which solves the problem,*

$$\begin{aligned} \min_{\mathbf{R}_0^{T-1}} \quad & \sum_{t=0}^{T-1} \hat{J}_t(\beta_t, \kappa_t, R_t), \\ \text{s. t.} \quad & \sum_{t=0}^{T-1} R_t \leq R_{tot}, \end{aligned}$$

where $\hat{J}_t(\beta_t, \kappa_t, R_t)$ *is given by* (5.26).

The solution to Problem 5.4.3, summarized in Theorem 5.4.4 below, states the main result of this section.

Theorem 5.4.4. *Suppose* $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$.

- *For noisy channels* ($\epsilon > 0$)*, it holds that*

  1. *If* $R_{tot} \geq \sum_{t=0}^{T-1} R_t^\star$*, where* $\mathbf{R}_0^{\star T-1}$ *is a solution to the system of equations*

$$\begin{aligned} \frac{\partial \hat{J}_0}{\partial R_0}(\beta_0, \kappa_0, R_0^\star) &= 0, \\ &\vdots \\ \frac{\partial \hat{J}_{T-1}}{\partial R_{T-1}}(\beta_{T-1}, \kappa_{T-1}, R_{T-1}^\star) &= 0, \end{aligned} \tag{5.27}$$

then $\mathbf{R}_0^{\star T-1}$ *solves Problem 5.4.3.*

2. *If* $R_{tot} < \sum_{t=0}^{T-1} R_t^\star$, *where* $\mathbf{R}_0^{\star T-1}$ *solves* (5.27), *then a solution* $\{\mathbf{R}_0^{T-1}, \theta\}$ *to the system of equations*

$$\begin{aligned} -\frac{\partial \hat{J}_0}{\partial R_0}(\beta_0, \kappa_0, R_0) &= \theta, \\ &\vdots \\ -\frac{\partial \hat{J}_{T-1}}{\partial R_{T-1}}(\beta_{T-1}, \kappa_{T-1}, R_{T-1}) &= \theta, \\ \sum_{t=0}^{T-1} R_t &= R_{tot}, \end{aligned} \tag{5.28}$$

solves Problem 5.4.3, with θ *denoting the associated Lagrange multiplier.*

- *For error-free channels* ($\epsilon = 0$), *it holds that*

$$R_t = \frac{R_{tot}}{T} + \frac{1}{2}\log_2\left(\frac{\kappa_t}{\left(\prod_{t=0}^{T-1}\kappa_t\right)^{\frac{1}{T}}}\right), \quad t = 0, \ldots, T-1, \tag{5.29}$$

solves Problem 5.4.3.

To prove Theorem 5.4.4, we need to use Lemma 5.4.5–Lemma 5.4.8, as shown subsequently. First, we deal with the general case that $\epsilon > 0$, following the standard approach for constrained optimization problems. Thereafter, the special case that $\epsilon = 0$ is discussed. We should mention that the rates given by Theorem 5.4.4 are real values. How to cope with the non-negativity and integer constraints will be discussed later in Section 5.8.

Noisy Channels

We start by discussing the general case that $\epsilon > 0$. First, we note that the unconstrained problem for the noisy scenario has a unique global minimum that is not necessarily achieved at infinity, i.e., $R_t = \infty$, as stated in the following lemma.

Lemma 5.4.5. *Let* $\epsilon > 0$. *Problem 5.4.2 has a unique global minimum,* $\mathbf{R}_0^{\star T-1}$, *which solves* (5.27), *i.e.,*

$$\begin{aligned} -\beta_0 \ln(1-\epsilon)(1-\epsilon)^{R_0^\star} - 2\ln(2)\kappa_0 2^{-2R_0^\star} &= 0, \\ &\vdots \\ -\beta_{T-1} \ln(1-\epsilon)(1-\epsilon)^{R_{T-1}^\star} - 2\ln(2)\kappa_{T-1} 2^{-2R_{T-1}^\star} &= 0. \end{aligned}$$

Proof. Compute the critical point, at which the gradient $\mathrm{G}(\mathbf{R}_0^{\star T-1})$ is a zero vector,

$$\mathrm{G}(\mathbf{R}_0^{\star T-1}) \triangleq \begin{bmatrix} \frac{\partial}{\partial R_0} \sum_{t=0}^{T-1} \hat{J}_t(\beta_t, \kappa_t, R_t^\star) \\ \vdots \\ \frac{\partial}{\partial R_{T-1}} \sum_{t=0}^{T-1} \hat{J}_t(\beta_t, \kappa_t, R_t^\star) \end{bmatrix} = 0.$$

Straightforward calculation yields (5.27). It is interesting to note that the system of equations is decoupled and the variables $\mathbf{R}_0^{T-1}$ are separable. We can argue that, since each decoupled function $\hat{J}_t(\beta_t, \kappa_t, R_t)$ is quasi-convex and has one unique global minimum, as shown in Lemma 5.3.2, the overall distortion $\sum_{t=0}^{T-1} \hat{J}_t(\beta_t, \kappa_t, R_t)$ has a unique global minimum. □

From Lemma 5.4.5, we know that when $R_{tot} \geq \sum_{t=0}^{T-1} R_t^\star$, where $\mathbf{R}_0^{\star T-1}$ is a solution to (5.27), the same allocation $\mathbf{R}_0^{\star T-1}$ simultaneously solves Problem 5.4.3. On the other hand if $R_{tot} < \sum_{t=0}^{T-1} R_t^\star$, where $\mathbf{R}_0^{\star T-1}$ solves (5.27), the solution to the system of equations (5.28) solves Problem 5.4.3, as stated in the lemma below.

Lemma 5.4.6. *Let $\epsilon > 0$. A solution to the system of equations* (5.28) *solves Problem 5.4.3.*

Proof. The proof is based on Lagrange duality theory. First, we note that strong duality applies, because the constraint is a positive linearly independent combination of R_t, the Mangasarian-Fromowitz constraint qualification applies [Hor95]. More discussion on the strong duality will be given later. Second, we minimize the Lagrangian

$$\begin{aligned} \eta(\mathbf{R}_0^{T-1}, \theta) &= \sum_{t=0}^{T-1} \mathbf{E}\left\{J_t(R_t)\right\} + \theta \left(\sum_{t=0}^{T-1} R_t - R_{tot} \right) \\ &= \sum_{t=0}^{T-1} \left(\beta_t (1 - (1-\epsilon)^{R_t}) + \kappa_t 2^{-2R_t} \right) + \theta \left(\sum_{t=0}^{T-1} R_t - R_{tot} \right). \end{aligned}$$

The first order derivatives of $\eta(\mathbf{R}_0^{T-1}, \theta)$ with respect to the variables R_t and θ, are

$$\begin{aligned} \frac{\partial}{\partial R_t} \eta(\mathbf{R}_0^{T-1}, \theta) &= -\beta_t \ln(1-\epsilon)(1-\epsilon)^{R_t} - 2\ln(2)\kappa_t 2^{-2R_t} + \theta, \\ &= \frac{\partial \hat{J}_t}{\partial R_t}(\beta_t, \kappa_t, R_t) + \theta, \quad t = 0, \ldots, T-1, \\ \frac{\partial}{\partial \theta} \eta(\mathbf{R}_0^{T-1}, \theta) &= \sum_{t=0}^{T-1} R_t - R_{tot}. \end{aligned}$$

The minimum is achieved when all above equations are equal to 0. Observe that, we arrive at a system of equations with $T+1$ equations and $T+1$ unknowns, i.e., $\mathbf{R}_0^{T-1}$ and θ,

$$
\begin{aligned}
-\frac{\partial \hat{J}_0}{\partial R_0}(\beta_0, \kappa_0, R_0) &= \theta, \\
&\vdots \\
-\frac{\partial \hat{J}_{T-1}}{\partial R_{T-1}}(\beta_{T-1}, \kappa_{T-1}, R_{T-1}) &= \theta, \\
\sum_{t=0}^{T-1} R_t &= R_{tot},
\end{aligned}
$$

as given in (5.28). □

It is worth noting that the solution to (5.28) belongs to the set $\{\mathbf{R}_0^{T-1}: 0 \leq R_t \leq R_t^\star, \forall t\}$, where $\mathbf{R}_0^{\star T-1}$ is the unconstrained global minimum that solves (5.27). This can be realized by the following facts. First, observe that $\partial \hat{J}_t/\partial R_t < 0$ when $R_t < R_t^\star$, and $\partial \hat{J}_t/\partial R_t > 0$ when $R_t > R_t^\star$, $\forall t$. The solution to (5.27) requires that $\partial \hat{J}_t/\partial R_t$, $\forall t$, have the same sign. However, if $\partial \hat{J}_t/\partial R_t > 0$, $\forall t$, it will violate the total rate constraint $R_{tot} \leq \sum_{t=0}^{T-1} R_t$. Therefore, only $\mathbf{R}_0^{T-1} \in \{\mathbf{R}_0^{T-1} : 0 \leq R_t \leq R_t^\star, \forall t\}$ can be a solution to (5.28). Moreover, the rates are always non-negative numbers which in practice can be ensured by excluding the instantaneous cost $\mathbf{E}\{J_t(R_t)\}$ associated with the negative rate and resolving the rate optimization problems with respect to the new overall cost. Consequently, a solution to the constrained problem always belongs to the convex set $\{\mathbf{R}_0^{T-1} : 0 \leq R_t \leq R_t^\star, \forall t\}$. As a result, we arrive at a convex optimization problem since both the objective function and the variable set are convex, then the strong duality applies.

In general, we do not have a closed-form solution to (5.28). However, this nonlinear system of equations can be solved by using numerical methods, e.g., [BT97]. Below, we briefly discuss one algorithm based on Newton's method. We start by defining the vector Φ, constructed by all unknown variables,

$$
\Phi \triangleq \begin{bmatrix} R_0 \\ \vdots \\ R_{T-1} \\ \theta \end{bmatrix},
$$

and the system of equations

$$
Z(\Phi) \triangleq \begin{cases}
Z_0 &= -\beta_0 \ln(1-\epsilon)(1-\epsilon)^{R_0} - 2\ln(2)\kappa_0 2^{-2R_0} + \theta, \\
&\vdots \\
Z_{T-1} &= -\beta_{T-1} \ln(1-\epsilon)(1-\epsilon)^{R_{T-1}} - 2\ln(2)\kappa_{T-1} 2^{-2R_{T-1}} + \theta, \\
Z_T &= \sum_{t=0}^{T-1} R_t - R_{tot}.
\end{cases}
$$

Newton's method performs an iterative search for the optimal vector Φ, which solves the equation

$$Z(\Phi) = 0.$$

The results of the k^{th} and $(k-1)^{th}$ iterations, $\Phi_{[k]}$ and $\Phi_{[k-1]}$, are related by

$$\Phi_{[k]} = \Phi_{[k-1]} - J_F^{-1} Z(\Phi_{[k-1]}),$$

where J_F denotes the Jacobian matrix,

$$J_F\left(\mathbf{R}_0^{T-1}, \theta\right) = \begin{bmatrix} \frac{\partial Z_0}{\partial R_0} & \cdots & \frac{\partial Z_0}{\partial R_{T-1}} & \frac{\partial Z_0}{\partial \theta} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial Z_{T-1}}{\partial R_0} & \cdots & \frac{\partial Z_{T-1}}{\partial R_{T-1}} & \frac{\partial Z_{T-1}}{\partial \theta} \\ \frac{\partial Z_T}{\partial R_0} & \cdots & \frac{\partial Z_T}{\partial R_{T-1}} & \frac{\partial Z_T}{\partial \theta} \end{bmatrix}.$$

The elements in J_F are calculated according to

$$\frac{\partial Z_t}{\partial R_s} = \begin{cases} \frac{\partial^2 \hat{J}_s}{\partial R_s^2}(\beta_s, \kappa_s, R_s), & t = s,\ t \neq T, \\ 0, & t \neq s,\ t \neq T, \\ 1, & t = T, \end{cases}$$

$$\frac{\partial Z_t}{\partial \theta} = \begin{cases} 1, & t \neq T, \\ 0, & t = T, \end{cases}$$

where $\partial^2 \hat{J}_s / \partial R_s^2$ is given by (5.14). We can see below that many elements in the Jacobian matrix are zero,

$$J_F = \begin{bmatrix} \frac{\partial^2 \hat{J}_0}{\partial R_0^2}(\beta_0, \kappa_0, R_0) & \cdots & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \frac{\partial^2 \hat{J}_{T-1}}{\partial R_{T-1}^2}(\beta_{T-1}, \kappa_{T-1}, R_{T-1}) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}.$$

Error-Free Channels

For an error-free channel ($\epsilon = 0$), we can show that the system of equations (5.28) has a closed-form solution. This is because when $\epsilon = 0$, $\beta_t \ln(1-\epsilon)(1-\epsilon)^{R_t} = 0$, $\forall t$. Let us first take a look at the unconstrained problem, and the solution is formulated in Lemma 5.4.7.

Lemma 5.4.7. *Let $\epsilon = 0$. Problem 5.4.2 is convex and the global minimum is achieved at infinity, i.e., $R_t^\star = \infty$, $\forall t$.*

Proof. When $\epsilon = 0$, the instantaneous distortion of Problem 5.4.2 becomes

$$\mathbf{E}\left\{J_t(R_t)\right\} = \kappa_t 2^{-2R_t}, \quad t = 0, \ldots, T-1.$$

Taking the derivative of the overall cost with respect to R_t, gives

$$\frac{\partial}{\partial R_s} \sum_{t=0}^{T-1} \mathbf{E}\left\{J_t(R_t)\right\} = -2\ln(2)\kappa_s 2^{-2R_s}, \quad s = 0, \ldots, T-1.$$

The function $-2\ln(2)\kappa_t 2^{-2R_t}$ is monotonically increasing with the rate R_t, and $\lim_{R_t \to \infty} -2\ln(2)\kappa_t 2^{-2R_t} = 0$. Computing the second order derivatives,

$$\frac{\partial^2}{\partial R_k^2} \sum_{t=0}^{T-1} \mathbf{E}\left\{J_t(R_t)\right\} = 4(\ln 2)^2 \kappa_k 2^{-2R_k},$$
$$\frac{\partial^2}{\partial R_k \partial R_l} \sum_{t=0}^{T-1} \mathbf{E}\left\{J_t(R_t)\right\} = 0, \quad k \neq l,$$

the Hessian of the overall cost $\sum_{t=0}^{T-1} \mathbf{E}\left\{J_t(R_t)\right\}$ is

$$\mathrm{H}(\mathbf{R}_0^{T-1}) = \begin{bmatrix} 4(\ln 2)^2\kappa_0 2^{-2R_0} & 0 & \cdots & 0 \\ 0 & 4(\ln 2)^2\kappa_1 2^{-2R_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 4(\ln 2)^2\kappa_{T-1} 2^{-2R_{T-1}} \end{bmatrix}.$$

We note that the above Hessian matrix is positive definite, for all $R_t < \infty$ and $0 < \kappa_t < \infty$, because all the elements on the diagonal are positive. As a result, the optimization problem is convex, and the minimum is achieved at $R_t = \infty$, $\forall t$. □

According to Lemma 5.4.7, there is no finite-valued R_{tot} that achieves the global minimum when $\epsilon = 0$. Next, we move on to the constrained optimization problem. The solution to (5.28) is summarized in Lemma 5.4.8.

Lemma 5.4.8. *Let $\epsilon = 0$. A solution $\mathbf{R}_0^{T-1}$ to the system of equations* (5.28) *is* (5.29)*, i.e.,*

$$R_t = \frac{R_{tot}}{T} + \frac{1}{2}\log_2\left(\frac{\kappa_t}{\left(\prod_{t=0}^{T-1}\kappa_t\right)^{\frac{1}{T}}}\right), \quad t = 0, \ldots, T-1.$$

Proof. According to (5.28), it is straightforward to write R_t as a function of θ,

$$R_t = -\frac{1}{2}\log_2\frac{\theta}{2\ln(2)\kappa_t} = \frac{1}{2}\log_2\left(2\ln(2)\kappa_t\right) - \frac{1}{2}\log_2\theta. \tag{5.30}$$

We can solve θ by means of the total bit-rate constraint (5.5), and the answer is

$$\theta = 2^{\frac{1}{T}\left(\sum_{t=0}^{T-1} \log_2(2\ln(2)\kappa_t) - R_{tot}\right)}. \tag{5.31}$$

Substituting (5.31) into (5.30), (5.29) follows immediately. Note that $\left(\prod_{t=0}^{T-1} \kappa_t\right)^{\frac{1}{T}}$ is the geometric mean of the sequence κ_t. □

Example 5.4.9. ***Gaussian Source and Source-Optimized Quantizer***

Applying (5.19) *and* (5.29) *to a zero-mean Gaussian source and its associated source-optimized quantizer, the solution is*

$$R_t = \frac{R_{tot}}{T} + \frac{1}{2}\log_2\left(\frac{\sigma_{x_t}^2}{\left(\prod_{t=0}^{T-1}\sigma_{x_t}^2\right)^{\frac{1}{T}}}\right), \quad t = 0, \ldots, T-1,$$

where $\left(\prod_{t=0}^{T-1}\sigma_{x_t}^2\right)^{\frac{1}{T}}$ *is the geometric mean of the sequence* $\sigma_{x_t}^2$.

Now we are in the position to prove Theorem 5.4.4.

Proof. **(Theorem 5.4.4)**

In short, we prove the general case for noisy channels in two steps. First, we can show that Problem 5.4.2 has a global minimum at $\mathbf{R}_0^{\star T-1}$, which solves the system of equations (5.27), as shown by Lemma 5.4.5. Second, we can show that the solution to (5.28) solves Problem 5.4.3, by using Lagrange duality theory, cf., Lemma 5.4.6. The error-free special case is also proved in two steps where the unconstrained global minimum is discussed in Lemma 5.4.7, while the constrained solution is derived in Lemma 5.4.8. Based on Lemma 5.4.5–Lemma 5.4.8, we conclude that Theorem 5.4.4 is proved. □

Before we move on to the feedback control problem, let us consider a special case of Problem 5.4.2 and Problem 5.4.3 where the instantaneous cost can be written to the form

$$\hat{J}_t(\tilde{\beta}, \tilde{\kappa}, R_t) = \sigma_{x_t}^2(\tilde{\beta}(1-(1-\epsilon)^{R_t}) + \tilde{\kappa}2^{-2R_t}) = \sigma_{x_t}^2 \tilde{J}_t(\tilde{\beta}, \tilde{\kappa}, R_t), \tag{5.32}$$

where $\tilde{\beta}$ and $\tilde{\kappa}$ are time-invariant and $\tilde{J}_t(\tilde{\beta}, \tilde{\kappa}, R_t)$ is as defined in (5.20). Applying Lemma 5.4.5 and Theorem 5.4.4 to this special case, i.e., with $\hat{J}_t(\tilde{\beta}, \tilde{\kappa}, R_t)$ as given by (5.32), the results are summarized in Corollary 5.4.10 and Corollary 5.4.11.

Corollary 5.4.10. *Consider the special case with* $\hat{J}_t(\tilde{\beta}, \tilde{\kappa}, R_t)$ *as given by* (5.32). *Problem 5.4.2 has a unique global minimum* $R_t = R^\star$, $\forall t$, *with* $R^\star$ *solving the equation*

$$0 = \tilde{\beta}\ln(1-\epsilon)(1-\epsilon)^{R^\star} + 2\ln(2)\tilde{\kappa}2^{-2R^\star}. \tag{5.33}$$

The proof of Corollary 5.4.10 follows the proof of Lemma 5.4.5. In this special case the optimal rates are identical for all t, i.e.,

$$R_0^\star = R_1^\star = \cdots = R_{T-1}^\star = R^\star.$$

The optimal value $R^\star$ can be obtained numerically by solving the equation (5.33). This result suggests that when $R_{tot} \geq TR^\star$, where $R^\star$ is the solution to (5.33), the optimal bit allocation is $R_t = R^\star$, $\forall t$. On the other hand, when $R_{tot} < TR^\star$, we should solve the unconstrained optimization problem as described previously, cf., Corollary 5.4.11 below.

Corollary 5.4.11. *Consider the special case with $\hat{J}_t(\tilde{\beta}, \tilde{\kappa}, R_t)$ as given by (5.32). Suppose $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$.*

- *For noisy channels ($\epsilon > 0$), it holds that*

  1. *If $R_{tot} \geq TR^\star$, where $R^\star$ is the solution to (5.33), then $R_t = \mathbb{R}^\star$, $\forall t$, solves Problem 5.4.3.*
  2. *If $R_{tot} < TR^\star$ where $R^\star$ solves (5.33), then the solution $\{\mathbf{R}_0^{T-1}, \theta\}$ to (5.28) solves Problem 5.4.3, with θ denoting the associated Lagrange multiplier.*

- *For error-free channels ($\epsilon = 0$), it holds that*

$$R_t = \frac{R_{tot}}{T} + \frac{1}{2}\log_2\left(\frac{\tilde{\kappa}_t}{\left(\prod_{t=0}^{T-1}\tilde{\kappa}_t\right)^{\frac{1}{T}}}\right), \quad t = 0, \ldots, T-1,$$

  solves Problem 5.4.3.

The proof of Corollary 5.4.11 follows straightforwardly the proofs of Theorem 5.4.4 and Corollary 5.4.10.

Back to the system of equations (5.28), under certain conditions, further approximations can be introduced to solve systems of equations, somewhat simpler than (5.28). An example is given here. First, it could be observed that ϵ is typically a small number. Thus, by setting $(1-\epsilon)^{R_t} \approx 1$, we obtain the following system of equations,

$$\begin{aligned} \beta_t \ln(1-\epsilon) + 2\ln(2)\kappa_t 2^{-2R_t} &= \theta, \qquad t = 0, \ldots, T-1, \\ \sum_{t=0}^{T-1} R_t &= R_{tot}. \end{aligned} \tag{5.34}$$

Rewrite the system of equations (5.34) and let R_t be a function of θ,

$$\begin{aligned} -\frac{1}{2}\log_2\left(\frac{\theta-\beta_t\ln(1-\epsilon)}{2\ln(2)\kappa_t}\right) &= R_t, \quad t=0,\ldots,T-1, \\ \sum_{t=0}^{T-1} R_t &= R_{tot}. \end{aligned} \tag{5.35}$$

The Lagrange multiplier θ in (5.35) can be solved numerically, for example by Newton's method, as explained below. According to (5.35), R_{tot} can be written as

$$R_{tot}=\sum_{t=0}^{T-1} -\frac{1}{2}\log_2\left(\frac{\theta-\beta_t\ln(1-\epsilon)}{2\ln(2)\kappa_t}\right)=\log_{\frac{1}{4}}\left(\prod_{t=0}^{T-1}\left(\frac{\theta}{2\ln(2)\kappa_t}-\frac{\beta_t\ln(1-\epsilon)}{2\ln(2)\kappa_t}\right)\right). \tag{5.36}$$

Let us introduce Z, a function of θ

$$Z(\theta) \triangleq \prod_{t=0}^{T-1}\left(\frac{\theta}{2\ln(2)\kappa_t}-\frac{\beta_t\ln(1-\epsilon)}{2\ln(2)\kappa_t}\right)-\left(\frac{1}{4}\right)^{R_{tot}}.$$

It is straightforward to verify that (5.36) is equivalent to require

$$Z(\theta)=0.$$

Taking the first order derivative of $Z(\theta)$ with respect to θ, we obtain

$$\frac{\partial Z}{\partial\theta}=\sum_{t=0}^{T-1}\frac{1}{2\ln(2)\kappa_t}\prod_{\substack{s=0\\ s\neq t}}^{T-1}\left(\frac{\theta}{2\ln(2)\kappa_t}-\frac{\beta_s\ln(1-\epsilon)}{2\ln(2)\kappa_s}\right).$$

Let $\theta_{[k]}$, $k\in\mathbb{N}$, be taken to denote the result given by the k^{th} iteration, and it is related to $\theta_{[k-1]}$ as

$$\theta_{[k]}=\theta_{[k-1]}-\frac{Z(\theta_{[k-1]})}{\frac{\partial Z}{\partial\theta}(\theta_{[k-1]})}.$$

Finally, substituting the iteration result into (5.35), the optimized rates are obtained.

In this section, we studied the special case of optimizing the rate allocation for *state estimation*, as a first fundamental step in solving the rate allocation problem for state feedback control. First, we approximated the overall distortion function by means of high-rate approximation theory. Second, we showed that the unconstrained optimization problem has a global minimum, which solves the rate allocation problem if such a global minimum does not violate the rate constraint. On the other hand, if the global minimum violates the rate constraint, we solved the rate constrained optimization problem by means of Lagrangian duality for non-linear non-convex problems. Based on the result in this section, we will in the next section solve the analogous problem of bit allocation for controlling a dynamic system.

5.5 Rate Allocation for State Feedback Control

Now we are in the position to study the rate allocation problem for state feedback control. We follow exactly the same optimization procedure as the one used for the state estimation problem in Section 5.4: (*i*) Express the overall cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ explicitly as a function of the rates $\mathbf{R}_0^{T-1}$, and (*ii*) introduce Lagrange multipliers to solve the constrained optimization problem. Compared with the state estimation problem in Section 5.4, optimizing rate allocation for state feedback control becomes much more complicated. The *essential challenge* is that the communication between the sensor and the controller will affect all future states. This fact has some negative effects on the optimization of the rate allocation. Most importantly, the nice property that x_t is independent of all past $\mathbf{R}_0^{t-1}$, valid for systems in Section 5.4, is violated. Estimating $\mathbf{E}\left\{x_t^2\right\}$ becomes a formidable task also because the estimation error propagates with time. Errors from the past unfortunately influence all future states. However, in practice, systems of interest are mostly stable in the closed-loop, where error propagation is expected to be less problematic.

Recall the quadratic cost of Problem 5.2.4, i.e., $J_{tot}(\mathbf{R}_0^{T-1})$ given by (5.4),

$$J_{tot}(\mathbf{R}_0^{T-1}) = \sum_{t=1}^{T} \left(x_t^2 + \rho u_{t-1}^2\right).$$

Because the process noise v_t is white, uncorrelated with x_t and u_t, we can write $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ as

$$\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\} = \mathbf{E}\left\{(\phi_0 - 1)x_0^2 + \sum_{t=0}^{T-1} \phi_{t+1} v_t^2 + \sum_{t=0}^{T-1} (\phi_{t+1} + \rho)(-x_t\ell_t + u_t)^2\right\}, \tag{5.37}$$

where ϕ_t and ℓ_t are as given in (5.6)–(5.7), i.e.,

$$\phi_t = 1 + \frac{a^2\phi_{t+1}\rho}{\phi_{t+1} + \rho}, \text{ with } \phi_T = 1; \quad \ell_t = -\frac{a\phi_{t+1}}{\phi_{t+1} + \rho}.$$

Clearly, only the last sum of (5.37) is affected by control. As explained previously in Section 5.2, throughout this chapter we restrict the control u_t to take the form

$$u_t = \ell_t d_t,$$

where ℓ_t is calculated according to (5.6)–(5.7). This control is motivated by the observation that satisfactory performance is expected if d_t is close to x_t. By using $u_t = \ell_t d_t$, minimizing $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ of (5.4) is equivalent to minimizing the expected value of the following $J_{tot}(\mathbf{R}_0^{T-1})$,

$$J_{tot}(\mathbf{R}_0^{T-1}) = \sum_{t=0}^{T-1} \pi_t (x_t - d_t)^2, \quad \pi_t \triangleq (\phi_{t+1} + \rho)\ell_t^2, \tag{5.38}$$

subject to the total bit-rate constraint (5.5). Hereby, we replace the cost function (5.4) with (5.38), and the instantaneous objective function is now

$$\mathbf{E}\left\{J_t(\mathbf{R}_0^t)\right\} = \mathbf{E}\left\{\pi_t(x_t - d_t)^2\right\}. \tag{5.39}$$

Compared with the distortion function for state estimation (5.25), there are two main differences: (*i*) the objective function (5.39) has a time-varying weighting factor; and (*ii*), the state x_t depends on all past rate allocations $\mathbf{R}_0^{t-1}$.

Before dealing with Problem 5.2.4, we will first discuss a special but important case of Problem 5.2.4, with $\rho=0$, known as the *minimum variance control* [Åst70] in the literature. This special case has a considerably simplified solution, compared with the general case. However, this solution reveals a couple of important features of the rate allocation in control systems. Thereafter, the general system of Problem 5.2.4 is treated using basically similar techniques, both with and without channel errors.

5.5.1 Minimum Variance Control: $\rho=0$

In this section, we reveal some key features of the rate allocation in control systems, by solving a simple special case of Problem 5.2.4. That is the *minimum variance control* problem, with $\rho=0$. In particular, we seek the optimal rate allocation that minimizes the average cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$, where $J_{tot}(\mathbf{R}_0^{T-1})$ is given by

$$J_{tot}(\mathbf{R}_0^{T-1}) = \sum_{t=1}^{T} J_t(\mathbf{R}_0^{t-1}) = \sum_{t=1}^{T} x_t^2, \tag{5.40}$$

and the associated average instantaneous cost is

$$\mathbf{E}\left\{J_t(\mathbf{R}_0^{t-1})\right\} = \mathbf{E}\left\{x_t^2\right\}. \tag{5.41}$$

The state equation is still (5.1), and the initial-state and process noise are i.i.d. zero-mean Gaussian, mutually independent. According to (5.6), if $\rho=0$, then $\ell_t=-a$, i.e., the control law becomes

$$u_t = -ad_t. \tag{5.42}$$

The cost $J_{tot}(\mathbf{R}_0^{T-1})$ in (5.40) is not explicitly dependent of $\mathbf{u}_0^{T-1}$, but implicitly through $\mathbf{x}_0^{T-1}$. Furthermore, to focus on explaining the challenges in a state feedback control system, we consider here a finite-rate error-free channel ($\epsilon=0$). The encoder takes the fully observed state measurement x_t as the input, and the decoder takes the received symbol, $j_t=i_t$.

In contrast to the estimation problem in Section 5.4, here, the true pdf of x_t is in advance no longer available. We have to estimate it in certain ways. In order to proceed, we will approximate the state x_t by a zero-mean Gaussian source, because the initial-state and the process noise are zero-mean Gaussian. By imposing such a Gaussian approximation, we only need to estimate the state variance, which is denoted by $\hat{\sigma}_{x_t}^2$.

Based on the model (5.40)–(5.42), $\mathbf{E}\left\{x_t^2\right\}$ can be deduced from $\mathbf{E}\left\{x_{t-1}^2\right\}$ as,

$$\mathbf{E}\left\{x_t^2\right\} = \mathbf{E}\left\{(ax_{t-1} - ad_{t-1} + v_{t-1})^2\right\} = a^2\mathbf{E}\left\{(x_{t-1} - d_{t-1})^2\right\} + \sigma_v^2,$$

because v_{t-1} is uncorrelated with $\mathbf{x}_0^{t-1}$ and $\mathbf{d}_0^{t-1}$. For the simplicity of the presentation, we also assume that $\mathbf{E}\left\{(x_{t-1}-d_{t-1})^2\right\}$ can be approximated by the high-rate form (5.20) with $\tilde{\kappa}_t = \tilde{\kappa}$, $\forall t$, i.e.,

$$\mathbf{E}\left\{(x_{t-1} - d_{t-1})^2\right\} \approx \sigma_{x_{t-1}}^2 \tilde{J}_{t-1}(\tilde{\kappa}, R_{t-1}) = \sigma_{x_{t-1}}^2 \tilde{\kappa} 2^{-2R_{t-1}}.$$

Thus, $\sigma_{x_t}^2$ can be recursively expressed in terms of $\sigma_{x_{t-1}}^2$ and R_{t-1}. Likewise, $\sigma_{x_{t-1}}^2$ can be recursively expressed in terms of $\sigma_{x_{t-2}}^2$ and $\mathbf{R}_{t-2}^{t-1}$, etc. Continuing up, $\sigma_{x_t}^2$ can finally be expressed in terms of $\sigma_{x_0}^2$ and $\mathbf{R}_0^{t-1}$. Based on this observation, a high-rate approximation of $\mathbf{E}\left\{x_t^2\right\}$ can be recursively computed as

$$\hat{\sigma}_{x_t}^2 = \hat{\sigma}_{x_{t-1}}^2 \tilde{\kappa} 2^{-2R_{t-1}}. \tag{5.43}$$

For the ease of presentation, let us introduce the parameters $A_t > 0$ and $B_t > 0$,

$$A_t \triangleq a^{2(t+1)} \sigma_{x_0}^2 \tilde{\kappa}^{(t+1)}, \quad B_t \triangleq a^{2t} \sigma_v^2 \tilde{\kappa}^t. \tag{5.44}$$

When $\tilde{\kappa}$ is known, A_t and B_t can be computed in advance. By substituting (5.44) to (5.43), we can write $\hat{\sigma}_{x_t}^2$ as a function of $\mathbf{R}_0^{t-1}$, $\sigma_{x_0}^2$ and σ_v^2,

$$\hat{\sigma}_{x_t}^2 = A_{t-1} 2^{-2\sum_{s=0}^{t-1} R_s} + \sum_{m=1}^{t-1} B_m 2^{-2\sum_{n=t-m}^{t-1} R_n} + \sigma_v^2. \tag{5.45}$$

The high-rate approximation leads to the following approximate version of the instantaneous cost (5.41)

$$\mathbf{E}\left\{J_t(\mathbf{R}_0^{t-1})\right\} = \hat{\sigma}_{x_t}^2, \tag{5.46}$$

where $\hat{\sigma}_{x_t}^2$ is given by (5.45), and with $\hat{\sigma}_{x_0}^2 = \sigma_{x_0}^2$. Hereby, we will optimize the rate allocation with respect to (5.46), instead of (5.41). The constrained rate allocation problems based on (5.45) is formulated as follows.

Problem 5.5.1. *Find the rate allocation $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$ which solves the problem,*

$$\begin{aligned} \min_{\mathbf{R}_0^{T-1}} \quad & \sum_{t=1}^{T} \hat{\sigma}_{x_t}^2, \\ \text{s. t.} \quad & \sum_{t=0}^{T-1} R_t \leq R_{tot}, \end{aligned}$$

where $\hat{\sigma}_{x_t}^2$ is given by (5.45).

By arguing similarly as for Lemma 5.4.7, the rate unconstrained version of Problem 5.5.1 has a unique global minimum that cannot be achieved by finite-valued $\mathbf{R}_0^{T-1}$. On the other hand, the solution to the constrained optimization problem can be computed as stated below in Lemma 5.5.2,

Lemma 5.5.2. *Let $\epsilon=0$. A solution to the system of equations,*

$$\begin{aligned}\sum_{s=t+1}^{T}\left(2A_{s-1}2^{-2\sum_{k=0}^{s-1}R_k}\right)+\sum_{n=t+1}^{T}\sum_{m=1}^{n-1}2B_m2^{-2\sum_{l=n-m}^{n-1}R_l}&=\ \theta,\quad\forall t,\\ \sum_{t=0}^{T-1}R_t&=\ R_{tot},\end{aligned}\tag{5.47}$$

solves Problem 5.5.1, where A_s and B_n are defined in (5.44).

Proof. The proof is based on Lagrange duality theory. The strong duality holds, because the constraint is a positive linearly independent combination of the rates R_t, $t=0,\ldots,T-1$, then the Mangasarian-Fromowitz constraint qualification applies. The next step is to minimize the Lagrangian,

$$\eta(\mathbf{R}_0^{T-1},\theta)=\sum_{t=0}^{T-1}\hat{\sigma}_{x_t}^2+\theta\left(\sum_{t=0}^{T-1}R_t-R_{tot}\right),$$

which is done by setting the first order derivatives of $\eta(\mathbf{R}_0^{T-1},\theta)$ to 0. Note that, $\hat{\sigma}_{x_t}^2$ in (5.45) depends upon all past $\mathbf{R}_0^{t-1}$. Similarly, all future $\hat{\sigma}_{x_s}^2$, for which $s>t$, depends on the current R_t. Straightforward calculation of $\partial\eta(\mathbf{R}_0^{T-1},\theta)/\partial R_t=0$ leads to the equation,

$$\sum_{s=t+1}^{T}\left(2A_{s-1}2^{-2\sum_{k=0}^{s-1}R_k}\right)+\sum_{n=t+1}^{T}\sum_{m=1}^{n-1}2B_m2^{-2\sum_{l=n-m}^{n-1}R_l}=\theta.$$

Hence, we obtain a system of $T+1$ equations and $T+1$ unknowns, i.e., $\mathbf{R}_0^{T-1}$ and θ. □

Sort the first T equations, $\partial\eta(\mathbf{R}_0^{T-1},\theta)/\partial R_t=0$, $\forall t$, after t. Observe that, each product term that contains both 2^{-2R_t} and $2^{-2R_{t+1}}$ is a common product term for the $(t+1)^{th}$ and $(t+2)^{th}$ equations. We can simplify the system of equations by setting the left-hand side of the $(t+1)^{th}$ equation equal to the left-hand side of the $(t+2)^{th}$ equation. By doing this, the Lagrange multiplier θ is eliminated, and as well as the common terms for the $(t+1)^{th}$ and $(t+2)^{th}$ equations. A new system of equations, consisting of T equations and T unknowns, $\mathbf{R}_0^{T-1}$, is obtained, where the t^{th} equation is

$$A_{t-1}2^{-2\sum_{n=0}^{t-1}R_n}+\sum_{s=1}^{t-1}B_{t-1}2^{-2\sum_{n=s}^{t-1}R_n}-\sum_{s=t}^{T-1}B_s2^{-2\sum_{n=t}^{s}R_n}=0,\tag{5.48}$$

$t = 1, \ldots, T-1$, and the T^{th} equation is the total bit-rate constraint (5.5). An example will make the system of equations clear. We demonstrate (5.47)–(5.48) for $T = 5$, see Figure 5.4 and (5.49). The system of equations with the Lagrange multiplier θ can be found in Figure 5.4, while the system of equations without the Lagrange multiplier θ is shown in (5.49).

$$\begin{aligned}
0 &= A_0 2^{-2R_0} - B_1 2^{-2R_1} - B_2 2^{-2(\sum_{t=1}^{2} R_t)} - B_3 2^{-2(\sum_{t=1}^{3} R_t)} - B_4 2^{-2(\sum_{t=1}^{4} R_t)}, \\
0 &= A_1 2^{-2(\sum_{t=0}^{1} R_t)} + B_1 2^{-2R_1} - B_2 2^{-2R_2} - B_3 2^{-2(\sum_{t=2}^{3} R_t)} - B_4 2^{-2(\sum_{t=2}^{4} R_t)}, \\
0 &= A_2 2^{-2(\sum_{t=0}^{2} R_t)} + B_2 2^{-2(\sum_{t=1}^{2} R_t)} + B_2 2^{-2R_2} - B_3 2^{-2R_3} - B_4 2^{-2(\sum_{t=3}^{4} R_t)}, \\
0 &= A_3 2^{-2(\sum_{t=0}^{3} R_t)} + B_3 2^{-2(\sum_{t=1}^{3} R_t)} + B_3 2^{-2(\sum_{t=2}^{3} R_t)} + B_3 2^{-2R_3} - B_4 2^{-2R_4}, \\
0 &= \sum_{t=0}^{4} R_t - R_{tot}.
\end{aligned} \tag{5.49}$$

In the system of equations in Figure 5.4, there are six unknowns $\{\mathbf{R}_0^4, \theta\}$. The common terms of two adjacent equations are marked with different frames. After removing the common terms, we arrive at (5.49), a new system of equations with five unknowns, $\mathbf{R}_0^4$. When T is a very small number we could solve the system of equations explicitly. However, as T increases, it becomes impractical to derive a closed-form solution. Alternatively, the system of equations (5.49), can be solved by numerical methods, for example Newton's method, cf., Section 5.4.

5.5.2 General Case

Now, we move on to solve Problem 5.2.4 for the general case of arbitrary ϵ and ρ. Similar to the special case where $\rho = 0$, the terms $\mathbf{E}\{x_t^2\}$ and $\mathbf{E}\{(x_t - d_t)^2\}$ are essential to the solution. In order to proceed, we still approximate the state x_t by a zero-mean Gaussian variable, where $\hat{\sigma}_{x_t}^2$ denotes the estimated variance. The next challenge lies in the derivation of $\hat{\sigma}_{x_t}^2$. To facilitate the derivation of a tractable overall cost for optimization, we consider an upper bound for $\hat{\sigma}_{x_t}^2$ by simplifying the correlation between x_t and d_t, so that it holds

$$\hat{\sigma}_{x_t}^2 = (A_t + B_t \tilde{J}_{t-1}(\tilde{\beta}_{t-1}, \tilde{\kappa}_{t-1}, R_{t-1}))\hat{\sigma}_{x_{t-1}}^2 + \sigma_v^2, \tag{5.50}$$

where $\tilde{J}_{t-1}(\tilde{\beta}_{t-1}, \tilde{\kappa}_{t-1}, R_{t-1})$ is as defined in (5.20), and $A_t > 0$ and $B_t > 0$ are independent of R_{t-1}, $\hat{\sigma}_{x_{t-1}}^2$ and σ_v^2. The above assumption is reasonable, and the following two examples are used to illustrate the utility and explain the motivation of (5.50).

Example 5.3. ***Decoder*** $d_t = \mathbf{E}\{x_t | j_t\}$

Consider the decoder $d_t = \mathbf{E}\{x_t | j_t\}$. *In this case, the estimation error* $x_t - d_t$ *is uncorrelated with the estimate* d_t, *which gives*

$$\mathbf{E}\{(x_t - d_t)^2\} = \mathbf{E}\{x_t^2\} - \mathbf{E}\{d_t^2\}.$$

$$
\begin{aligned}
&2A_0 2^{-2R_0} \\
&\qquad +2A_1 2^{-2\left(\sum_{t=0}^{1} R_t\right)} \\
&2A_1 2^{-2\left(\sum_{t=0}^{1} R_t\right)} \\
&\qquad +2B_1 2^{-2R_1} \\
&2A_2 2^{-2\left(\sum_{t=0}^{2} R_t\right)} \\
&\qquad +2B_2 2^{-2\left(\sum_{t=1}^{2} R_t\right)} \\
&\qquad +2B_2 2^{-2R_2} \\
&2A_3 2^{-2\left(\sum_{t=0}^{3} R_t\right)} \\
&\qquad +2B_3 2^{-2\left(\sum_{t=1}^{3} R_t\right)} \\
&\qquad +2B_3 2^{-2\left(\sum_{t=2}^{3} R_t\right)} \\
&\qquad +2B_3 2^{-2R_3} \\
&2A_4 2^{-2\left(\sum_{t=0}^{4} R_t\right)} \\
&\qquad +2B_4 2^{-2\left(\sum_{t=1}^{4} R_t\right)} \\
&\qquad +2B_4 2^{-2\left(\sum_{t=2}^{4} R_t\right)} \\
&\qquad +2B_4 2^{-2\left(\sum_{t=3}^{4} R_t\right)} \\
&\qquad +2B_4 2^{-2R_4}
\end{aligned}
\begin{aligned}
&+2A_1 2^{-2\left(\sum_{t=0}^{1} R_t\right)} +2A_2 2^{-2\left(\sum_{t=0}^{2} R_t\right)} +2A_3 2^{-2\left(\sum_{t=0}^{3} R_t\right)} +2A_4 2^{-2\left(\sum_{t=0}^{4} R_t\right)} = \theta \\
&+2A_2 2^{-2\left(\sum_{t=0}^{2} R_t\right)} +2A_3 2^{-2\left(\sum_{t=0}^{3} R_t\right)} +2A_4 2^{-2\left(\sum_{t=0}^{4} R_t\right)} \\
&+2B_2 2^{-2\left(\sum_{t=1}^{2} R_t\right)} +2B_3 2^{-2\left(\sum_{t=1}^{3} R_t\right)} +2B_4 2^{-2\left(\sum_{t=1}^{4} R_t\right)} = \theta \\
&+2A_3 2^{-2\left(\sum_{t=0}^{3} R_t\right)} +2A_4 2^{-2\left(\sum_{t=0}^{4} R_t\right)} \\
&+2B_3 2^{-2\left(\sum_{t=1}^{3} R_t\right)} +2B_4 2^{-2\left(\sum_{t=1}^{4} R_t\right)} \\
&+2B_3 2^{-2\left(\sum_{t=2}^{3} R_t\right)} +2B_4 2^{-2\left(\sum_{t=2}^{4} R_t\right)} = \theta \\
&+2A_4 2^{-2\left(\sum_{t=0}^{4} R_t\right)} \\
&+2B_4 2^{-2\left(\sum_{t=1}^{4} R_t\right)} \\
&+2B_4 2^{-2\left(\sum_{t=2}^{4} R_t\right)} \\
&+2B_4 2^{-2\left(\sum_{t=3}^{4} R_t\right)} = \theta \\
&= \theta
\end{aligned}
$$

Figure 5.4: Equation systems for $T = 4$. Between two adjacent equations, the common terms are marked with the same type of frame.

Accordingly, we can approximate $\mathbf{E}\left\{x_t^2\right\}$ *as*

$$\begin{aligned}\mathbf{E}\left\{x_t^2\right\} &= a^2\mathbf{E}\left\{x_{t-1}^2\right\} + \ell_{t-1}^2\mathbf{E}\left\{d_{t-1}^2\right\} + 2a\ell_{t-1}\mathbf{E}\left\{x_{t-1}d_{t-1}\right\} + \sigma_v^2 \\ &\approx \left(a^2+\ell_{t-1}^2+2a\ell_{t-1}-(\ell_{t-1}^2+2a\ell_{t-1})\tilde{J}_{t-1}(\tilde{\beta}_{t-1},\tilde{\kappa}_{t-1},R_{t-1})\right)\mathbf{E}\left\{x_{t-1}^2\right\}+\sigma_v^2,\end{aligned} \tag{5.51}$$

by using $\mathbf{E}\left\{(x_{t-1}-d_{t-1})^2\right\}\approx\mathbf{E}\left\{x_{t-1}^2\right\}\tilde{J}_{t-1}(\tilde{\beta}_{t-1},\tilde{\kappa}_{t-1},R_{t-1})$. *Based on* (5.51), *we relate the approximations* $\hat{\sigma}_{x_t}^2$ *and* $\hat{\sigma}_{x_{t-1}}^2$ *as*

$$\hat{\sigma}_{x_t}^2 = \left(a^2+\ell_{t-1}^2+2a\ell_{t-1}-(\ell_{t-1}^2+2a\ell_{t-1})\tilde{J}_{t-1}(\tilde{\beta}_{t-1},\tilde{\kappa}_{t-1},R_{t-1})\right)\hat{\sigma}_{x_{t-1}}^2+\sigma_v^2,$$

which implies

$$A_t = a^2+\ell_{t-1}^2+2a\ell_{t-1}, \quad B_t = -(\ell_{t-1}^2+2a\ell_{t-1}).$$

Clearly, both A_t *and* B_t *are independent of* R_{t-1}, $\hat{\sigma}_{x_{t-1}}^2$ *and* σ_v^2.

In general, we can write $\mathbf{E}\left\{x_t^2\right\}$ as,

$$\begin{aligned}\mathbf{E}\left\{x_t^2\right\} &= \mathbf{E}\left\{(ax_{t-1}+\ell_{t-1}d_{t-1}+v_{t-1})^2\right\} \\ &= \mathbf{E}\left\{(-\ell_{t-1}x_{t-1}+\ell_{t-1}d_{t-1}+(a+\ell_{t-1})x_{t-1})^2\right\}+\sigma_v^2 \\ &= \ell_{t-1}^2\mathbf{E}\left\{(x_{t-1}-d_{t-1})^2\right\}+(a+\ell_{t-1})^2\mathbf{E}\left\{x_{t-1}^2\right\} \\ &\quad -2(a+\ell_{t-1})\ell_{t-1}\mathbf{E}\left\{x_{t-1}(x_{t-1}-d_{t-1})\right\}+\sigma_v^2.\end{aligned}$$

The term $\mathbf{E}\left\{x_{t-1}(x_{t-1}-d_{t-1})\right\}$ depends on the source, the quantizer and the channel. It is often difficult to formulate a closed-form expression. Example 5.5.4 below shows another case that (5.50) applies.

Example 5.5.4. $\mathbf{E}\left\{x_{t-1}(x_{t-1}-d_{t-1})\right\}=\Gamma(\epsilon)\mathbf{E}\left\{x_{t-1}^2\right\}$.

Consider the scenario that $\mathbf{E}\left\{x_{t-1}(x_{t-1}-d_{t-1})\right\}$ *can be written as*

$$\mathbf{E}\left\{x_{t-1}(x_{t-1}-d_{t-1})\right\} = \Gamma(\epsilon)\mathbf{E}\left\{x_{t-1}^2\right\}, \tag{5.52}$$

where Γ *depends only on* ϵ. *Then* $\hat{\sigma}_{x_t}^2$ *can be expressed in the form of* (5.50), *with*

$$A_t = (a+\ell_{t-1})^2-2(a+\ell_{t-1})\ell_{t-1}\Gamma(\epsilon), \quad B_t = \ell_{t-1}^2. \tag{5.53}$$

In practice, the linear approximation (5.52) can be applied generally to all systems in Section 5.2.2 by finding suitable A_t and B_t to approximate the true instantaneous costs. Therefore, (5.50) is always applicable. In what follows, we rewrite the overall cost by taking (5.50) into consideration. According to the high-rate approximation (5.20) and (5.50), we define a new instantaneous cost, an approximate version of the instantaneous cost (5.39),

$$\mathbf{E}\left\{J_t(\mathbf{R}_0^t)\right\} \approx \mathfrak{J}_t(\mathbf{R}_0^t) \triangleq \pi_t\hat{\sigma}_{x_t}^2\tilde{J}_t(\tilde{\beta}_t,\tilde{\kappa}_t,R_t), \tag{5.54}$$

where $\hat{\sigma}^2_{x_t}$ can be written as (5.50). That is

$$\hat{\sigma}^2_{x_t} = (A_t + B_t \tilde{J}_{t-1}(\tilde{\beta}_{t-1}, \tilde{\kappa}_{t-1}, R_{t-1}))\hat{\sigma}^2_{x_{t-1}} + \sigma_v^2, \tag{5.55}$$

where A_t and B_t are independent of R_{t-1}, $\hat{\sigma}^2_{x_{t-1}}$, and σ_v^2. The unconstrained and constrained rate allocation problems based on (5.54) are formulated as follows.

Problem 5.5.5. *Find the rate allocation* $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$ *which solves the problem,*

$$\min_{\mathbf{R}_0^{T-1}} \quad \sum_{t=0}^{T-1} \mathfrak{J}_t(\mathbf{R}_0^t),$$

where $\mathfrak{J}_t(\mathbf{R}_0^t)$ *is given by* (5.54)–(5.55).

Problem 5.5.6. *Find the rate allocation* $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$ *which solves the problem,*

$$\begin{aligned} \min_{\mathbf{R}_0^{T-1}} \quad & \textstyle\sum_{t=0}^{T-1} \mathfrak{J}_t(\mathbf{R}_0^t), \\ \text{s. t.} \quad & \sum_{t=0}^{T-1} R_t \leq R_{tot}, \end{aligned}$$

where $\mathfrak{J}_t(\mathbf{R}_0^t)$ *is given by* (5.54)–(5.55).

Similarly to Section 5.4, we first present Theorem 5.5.7, the main result of this section. Then the proof of Theorem 5.5.7 is detailed subsequently.

Theorem 5.5.7. *Suppose* $\mathbf{R}_0^{T-1} \in \mathbb{R}^T$.

- *For error-free channels* ($\epsilon = 0$)*, it holds that a solution* $\mathbf{R}_0^{T-1}$ *to the system of equations*

$$\begin{aligned} \sum_{s=t}^{T-1} \left(2 \sum_{b_0=0}^{1} \cdots \sum_{b_t=1}^{1} \cdots \sum_{b_{s-1}=0}^{1} \Psi_s(b_0, \ldots, b_{s-1}) \right) &= \theta, \quad t = 0, \ldots, T-1, \\ \sum_{t=0}^{T-1} R_t &= R_{tot}, \end{aligned} \tag{5.56}$$

solves Problem 5.5.6, with θ *denoting the associated Lagrange multiplier, and* $b_k \in \{0, 1\}$, $k \in \{0, \ldots, T-2\}$*, denoting a binary variable. The term* $\Psi_t(b_0, \ldots, b_{t-1})$ *is defined as*

$$\Psi_t(b_0, \ldots, b_{t-1}) \triangleq \pi_t \bar{B} \left(\prod_{s=\bar{s}+1}^{t-1} \bar{B}_s \right) \left(\prod_{m=0}^{t} \tilde{\kappa}_m^{b_m} \right) 2^{-2\left(\sum_{n=0}^{t-1} b_n R_n + R_t\right)}. \tag{5.57}$$

Here,

1. *The terms* π_t *and* $\tilde{\kappa}_t$ *are specified in* (5.7) *and* (5.20).
2. *The term* $\bar{s}$ *is the smallest integer* s *that* $b_s=1$, *(i.e.,* $b_s=0$, $s<\bar{s}$*).*
3. *The term* $\bar{B}$ *is*

$$\bar{B} \triangleq \begin{cases} \tau_{\bar{s}-1}, & \bar{s} > 0, \\ B_0\sigma_{x_0}^2, & \bar{s} = 0, \end{cases} \tag{5.58}$$

where B_t *is defined as in* (5.50), *and* τ_s *is calculated recursively as,*

$$\tau_s \triangleq A_s\tau_{s-1} + \sigma_v^2, \quad \tau_0 \triangleq A_0\sigma_{x_0}^2 + \sigma_v^2.$$

4. *The parameter* $\bar{B}_s$ *is*

$$\bar{B}_s \triangleq \begin{cases} A_s, & b_s = 0, \\ B_s, & b_s = 1. \end{cases} \tag{5.59}$$

- *For noisy channels* ($\epsilon>0$), *it holds that*

1. *If* $R_{tot} \geq \sum_{t=0}^{T-1} R_t^\star$, *where* $\mathbf{R}_0^{\star T-1}$ *is a solution to the system of equations*

$$\begin{aligned} \tilde{J}_0(\tilde{\beta}_0, \tilde{\kappa}_0, R_0^\star) &= 0, \\ &\vdots \\ \tilde{J}_{T-1}(\tilde{\beta}_{T-1}, \tilde{\kappa}_{T-1}, R_{T-1}^\star) &= 0, \end{aligned} \tag{5.60}$$

with $\tilde{J}_t(\tilde{\beta}_t, \tilde{\kappa}_t, R_t)$ *given by* (5.20), *then the same* $\mathbf{R}_0^{\star T-1}$ *solves Problem 5.5.6.*

2. *If* $R_{tot} < \sum_{t=0}^{T-1} R_t^\star$, *where* $\mathbf{R}_0^{\star T-1}$ *is a solution to* (5.60), *then the solution to the system of equations*

$$\begin{aligned} -\sum_{s=t}^{T-1} \Psi_{t,s} &= \theta, \quad t = 0, \ldots, T-1, \\ \sum_{t=0}^{T-1} R_t &= R_{tot}, \end{aligned} \tag{5.61}$$

solves Problem 5.5.6. Here, the term $\Psi_{t,s}$ *is defined as*

$$\Psi_{t,s} \triangleq \sum_{b_0=0}^{1} \cdots \sum_{b_t=1}^{1} \cdots \sum_{b_{s-1}=0}^{1} \pi_s \bar{\Psi}(b_0, \ldots, b_s), \tag{5.62}$$

where $b_k \in \{0,1\}$, $k \in \{0, \ldots, T-2\}$, *is a binary variable, and* $\bar{\Psi}(b_0, \ldots, b_s)$ *is given by*

$$\bar{\Psi}(b_0, \ldots, b_s) \triangleq \bar{B} \left(\prod_{m=\bar{s}+1}^{s-1} \bar{B}_m \right) \left(\prod_{n=\bar{s}+1}^{s-1} (C_n)^{b_n} \right).$$

The terms $\bar{B}$ *and* $\bar{B}_n$ *are given by* (5.58)–(5.59)*, and* C_n *is defined as*

$$C_n \triangleq \begin{cases} \frac{\partial \tilde{J}_n}{\partial R_n}(\tilde{\beta}_n, \tilde{\kappa}_n, R_n), & n = t, \\ \tilde{J}_n(\tilde{\beta}_n, \tilde{\kappa}_n, R_n), & n \neq t. \end{cases} \tag{5.63}$$

To prove Theorem 5.5.7, we need to use Lemma 5.5.8–Lemma 5.5.14, as shown subsequently. We start by describing the error-free scenario.

Error-Free Channels

In this section we derive a solution in the case of error-free channels. We recall that the key to the problem is to find a useful objective function. Therefore, in Lemma 5.5.8 we rewrite (5.54) and express it in a more handy form.

Lemma 5.5.8. *Let* $\epsilon = 0$*. The instantaneous cost* (5.54) *can be written as a sum of* 2^t *product terms, denoted by* $\Psi_t(b_0, \ldots, b_{t-1})$*,*

$$\mathfrak{J}_t(\mathbf{R}_0^t) = \sum_{b_0=0}^{1} \cdots \sum_{b_{t-1}=0}^{1} \Psi_t(b_0, \ldots, b_{t-1}), \tag{5.64}$$

where $b_s \in \{0, 1\}$*,* $s = 0, \ldots, t-1$*, is a binary variable, and the term* $\Psi_t(b_0, \ldots, b_{t-1})$ *is as defined in* (5.57)*.*

Proof. Let $\epsilon = 0$. The instantaneous cost (5.54) becomes

$$\mathfrak{J}_t(\mathbf{R}_0^t) = \pi_t \hat{\sigma}^2_{x_t} \tilde{\kappa}_t 2^{-2R_t}, \tag{5.65}$$

where $\hat{\sigma}^2_{x_t}$ can be written in the form

$$\hat{\sigma}^2_{x_t} = (A_t + B_t \tilde{\kappa}_t 2^{-2R_{t-1}}) \hat{\sigma}^2_{x_{t-1}} + \sigma_v^2,$$

with A_t and B_t as defined in (5.50). By recursively replacing $\hat{\sigma}^2_{x_s}$ with $\hat{\sigma}^2_{x_{s-1}}$ and R_{s-1}, for $s = t, \ldots, 1$, up to $\hat{\sigma}^2_{x_0} = \sigma^2_{x_0}$, we can write $\hat{\sigma}^2_{x_t}$ as a function of $\mathbf{R}_0^{t-1}$. In particular, $\mathfrak{J}_t(\mathbf{R}_0^t)$ can be written as a sum of 2^t product terms, $\Psi_t(b_0, \ldots, b_{t-1})$ from (5.57), i.e.,

$$\Psi_t(b_0, \ldots, b_{t-1}) = \pi_t \bar{B} \left(\prod_{s=\bar{s}+1}^{t-1} \bar{B}_s \right) \left(\prod_{m=0}^{t} \tilde{\kappa}_m^{b_m} \right) 2^{-2\left(\sum_{n=0}^{t-1} b_n R_n + R_t\right)},$$

where all notations are referred to Theorem 5.5.7. □

Next, we show the solution to Problem 5.5.5, with $\mathfrak{J}_t(\mathbf{R}_0^t)$ as given by (5.65).

Lemma 5.5.9. *Let* $\epsilon = 0$*. Problem 5.5.5 has a unique global minimum, which is achieved at* $R_t^\star = \infty$*,* $\forall t$*.*

Proof. Based on (5.64), it is straightforward to show that

$$(1)\ \frac{\partial}{\partial R_s}\mathfrak{J}_t(\mathbf{R}_0^t) < 0,$$
$$(2)\ \lim_{R_s\to\infty}\frac{\partial}{\partial R_s}\mathfrak{J}_t(\mathbf{R}_0^t) = 0,$$

$s=0,\ldots,t$, for all t, which concludes the proof. Note that $\partial\mathfrak{J}_t(\mathbf{R}_0^t)/\partial R_s = 0$, $s>t$, for all t. □

Next, we show that given a finite-valued R_{tot}, the constrained optimization problem can be solved according to Lemma 5.5.10 below.

Lemma 5.5.10. *Let $\epsilon=0$. A solution to the system of equations* (5.56)*, i.e.,*

$$\sum_{s=t}^{T-1}\left(2\sum_{b_0=0}^{1}\cdots\sum_{b_t=1}^{1}\cdots\sum_{b_{s-1}=0}^{1}\Psi_s(b_0,\ldots,b_{s-1})\right)=\theta,\quad t=0,\ldots,T-1,$$
$$\sum_{t=0}^{T-1}R_t = R_{tot},$$

solves Problem 5.5.6, where $\Psi_t(b_0,\ldots,b_{t-1})$ is given by (5.57)*.*

Proof. First, we note that strong duality holds, because the constraint is positive and linearly independent of the rate R_t, the Mangasarian-Fromowitz constraint qualification applies. When $\epsilon=0$, the instantaneous cost (5.54) becomes

$$\mathfrak{J}_\mathfrak{t} = \pi_t\hat{\sigma}_{x_t}^2\tilde{\kappa}_t 2^{-2R_t} = \sum_{b_0=0}^{1}\cdots\sum_{b_{t-1}=0}^{1}\Psi_t(b_0,\ldots,b_{t-1}),$$

with $\Psi_t(b_0,\ldots,b_{t-1})$ given by (5.57). Then, we minimize the Lagrangian

$$\eta(\mathbf{R}_0^{T-1},\theta) = \sum_{t=0}^{T-1}\mathfrak{J}_t(\mathbf{R}_0^t) + \theta\left(\sum_{t=0}^{T-1}R_t - R_{tot}\right).$$

Taking the derivative of $\eta(\mathbf{R}_0^{T-1},\theta)$ with respect to each R_t, it leads to

$$\frac{\partial}{\partial R_t}\eta(\mathbf{R}_0^{T-1},\theta) = \sum_{s=t}^{T-1}\left(-2\sum_{b_0=0}^{1}\cdots\sum_{b_t=1}^{1}\cdots\sum_{b_{s-1}=0}^{1}\Psi_s(b_0,\ldots,b_{s-1})\right)+\theta.$$

Setting $\partial\eta(\mathbf{R}_0^{T-1},\theta)/\partial R_t = 0$ and $\partial\eta(\mathbf{R}_0^{T-1},\theta)/\partial\theta = 0$, yields immediately the system of equations (5.56). □

Sort the equations of (5.56) by time. We note that a $\Psi_s(b_0,\ldots,b_{s-1})$ with $b_t = b_{t+1} = 1$ is a common term for the adjacent $(t+1)^{th}$ and $(t+2)^{th}$ equations. A substraction between the two equations can eliminate the Lagrange multiplier θ and the other common terms. As a matter of fact, a $\Psi_s(b_0,\ldots,b_{s-1})$ with $b_m = b_n = 1$ is a common term for the $(m+1)^{th}$ and $(n+1)^{th}$ equations. Likewise, we can identify a common term shared by more than two equations using a similar rule.

When T is very small, we can solve the system of equations (5.56) by following the standard approach, i.e., successively eliminating unknown parameters. However, as the horizon grows, it becomes hard to eliminate all unknown parameters. Instead, we solve the overall system of equations numerically, as discussed in the preceding section.

Noisy Channels

In this section, we deal with Problem 5.5.5 and Problem 5.5.6 for cases where $\epsilon > 0$. At this point it should be clear that the approximated variance $\hat{\sigma}^2_{x_t}$ is essential to the derivation of the overall cost. As similar to the error-free scenario, starting at $s = t$, by successively replacing $\hat{\sigma}^2_{x_s}$ with $\hat{\sigma}^2_{x_{s-1}}$, right up to $\hat{\sigma}^2_{x_0} = \sigma^2_{x_0}$, we can formulate $\hat{\sigma}^2_{x_t}$ as a function of $\mathbf{R}_0^{t-1}$, $\sigma^2_{x_0}$ and σ^2_v. In particular, $\mathfrak{J}_t(\mathbf{R}_0^t)$ is a sum of 2^t terms, as shown in Lemma 5.5.11,

Lemma 5.5.11. *Let $\epsilon > 0$. The instantaneous cost* (5.54) *can be rewritten as*

$$\mathfrak{J}_t(\mathbf{R}_0^t) = \sum_{b_0=0}^{1} \cdots \sum_{b_{t-1}=0}^{1} \pi_t \Psi_t(b_0,\ldots,b_{t-1}) \tilde{J}_t(\tilde{\beta}_t, \tilde{\kappa}_t, R_t), \tag{5.66}$$

where $\tilde{J}_t(\tilde{\beta}_t, \tilde{\kappa}_t, R_t)$ is as defined in (5.20)*, and $b_s \in \{0,1\}$, $s \in \{0,\ldots,t-1\}$, is a binary variable. The term $\Psi_t(b_0,\ldots,b_{t-1})$ is defined as*

$$\Psi_t(b_0,\ldots,b_{t-1}) \triangleq \bar{B}\left(\prod_{s=\bar{s}+1}^{t-1} \bar{B}_s \left(\tilde{J}_s(\tilde{\beta}_s, \tilde{\kappa}_s, R_s)\right)^{b_s}\right),$$

with π_t, $\bar{B}$, $\bar{B}_s$ given by (5.38) *and* (5.58)–(5.59)*, and $\bar{s}$ is the smallest integer s for which $b_s = 1$.*

Proof. Lemma 5.5.11 is proved by straightforward calculation.

$$\mathfrak{J}_t(\mathbf{R}_0^t) = \sum_{b_0=0}^{1} \cdots \sum_{b_{t-1}=0}^{1} \pi_t \Psi_t(b_0,\ldots,b_{t-1}) \left(\tilde{\beta}_t \left(1-(1-\epsilon)^{R_t}\right) + \tilde{\kappa}_t 2^{-2R_t}\right),$$

where $\Psi_t(b_0,\ldots,b_{t-1})$ is defined as

$$\Psi_t(b_0,\ldots,b_{t-1}) = \bar{B}\left(\prod_{s=\bar{s}+1}^{t-1} \bar{B}_s \left(\tilde{\beta}_t \left(1-(1-\epsilon)^{R_s}\right) + \tilde{\kappa}_s 2^{-2R_s}\right)^{b_s}\right).$$

□

According to Lemma 5.5.11, each instantaneous cost consists of 2^t terms, and all the 2^t terms have the common factors π_t and $\tilde{J}_t(\tilde{\beta}_t, \tilde{\kappa}_t, R_t)$. Figure 5.5 demonstrates an efficient method to compute the middle term $\Psi_t(b_0, \ldots, b_{t-1})$, where the binary sequence $\mathbf{b}_0^{t-1}$ plays an important role. Note that, there are totally 2^t different binary sequences of length t. Each sequence $\mathbf{b}_0^{t-1}$ has a corresponding $\Psi_t(b_0, \ldots, b_{t-1})$, which can be computed by following the binary tree in Figure 5.5 from the root to the branch nodes. As a matter of fact, the 2^t branch nodes at the t^{th} level represent the 2^t $\Psi_t(b_0, \ldots, b_{t-1})$ terms. We illustrate the computation of one $\Psi_t(b_0, \ldots, b_{t-1})$ term in Example 5.5.12.

Example 5.5.12. ***Computing*** $\Psi_t(0, 0, 1, 1)$

In this example, we demonstrate how to compute $\Psi_t(0, 0, 1, 1)$ *by following the binary tree in Figure 5.5 from the root to the branch note. Start by setting* $s = \sigma_{x_0}^2$*, and then successively obtain*

$$
\begin{aligned}
b_0 = 0 \Rightarrow \;& s := A_0\sigma_0^2 + \sigma_v^2, \\
b_1 = 0 \Rightarrow \;& s := A_1(A_0\sigma_0^2 + \sigma_v^2) + \sigma_v^2, \\
b_2 = 1 \Rightarrow \;& s := B_2(A_1(A_0\sigma_0^2 + \sigma_v^2) + \sigma_v^2)\tilde{J}_2(\tilde{\beta}_2, \tilde{\kappa}_2, R_2), \\
b_3 = 1 \Rightarrow \;& s := B_3(B_2(A_1(A_0\sigma_0^2 + \sigma_v^2) + \sigma_v^2)\tilde{J}_2(\tilde{\beta}_2, \tilde{\kappa}_2, R_2))\tilde{J}_3(\tilde{\beta}_3, \tilde{\kappa}_3, R_3).
\end{aligned}
$$

Finally, $\Psi_t(0, 0, 1, 1) = B_3(B_2(A_1(A_0\sigma_0^2 + \sigma_v^2) + \sigma_v^2)\tilde{J}_2(\tilde{\beta}_2, \tilde{\kappa}_2, R_2))\tilde{J}_3(\tilde{\beta}_3, \tilde{\kappa}_3, R_3)$.

Next, in Lemma 5.5.13 we show the solution to the rate unconstrained problem.

Lemma 5.5.13. *Let* $\epsilon > 0$*. Problem 5.5.5 has a unique global minimum* $\mathbf{R}_0^{\star T-1}$*, which solves the following system of equations,*

$$
\begin{aligned}
\frac{\partial \tilde{J}_0}{\partial R_0}(\tilde{\beta}_0, \tilde{\kappa}_0, R_0^\star) &= 0, \\
&\vdots \\
\frac{\partial \tilde{J}_{T-1}}{\partial R_{T-1}}(\tilde{\beta}_{T-1}, \tilde{\kappa}_{T-1}, R_{T-1}^\star) &= 0,
\end{aligned}
$$

with $\partial \tilde{J}_t / \partial R_t$ *as defined in* (5.21).

The proof of Lemma 5.5.13 can be found in Appendix 5.C. Following the standard procedure, we show how to solve the rate constrained optimization problem in Lemma 5.5.14.

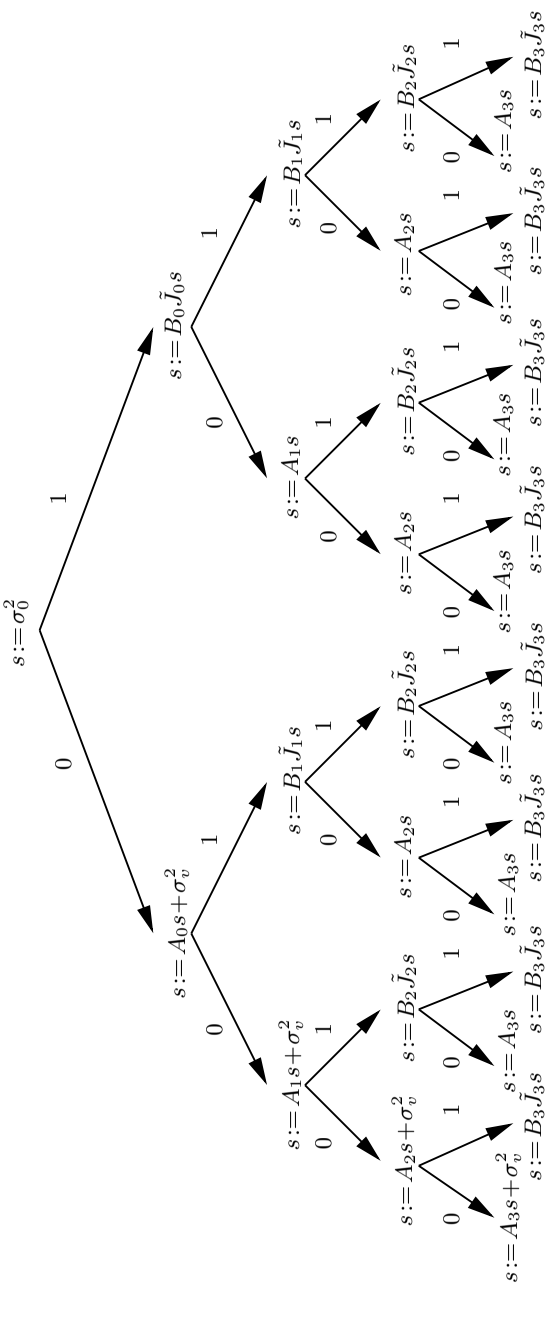


Figure 5.5: The break-down terms of $\Psi_t(b_0, \ldots, b_{t-1})$, where $\tilde{J}_1$, $\tilde{J}_2$ and $\tilde{J}_3$ are short notations for $\tilde{J}_t(\tilde{\beta}_1, \tilde{\kappa}_1, R_1)$, $\tilde{J}_t(\tilde{\beta}_2, \tilde{\kappa}_2, R_2)$ and $\tilde{J}_t(\tilde{\beta}_3, \tilde{\kappa}_3, R_3)$.

Lemma 5.5.14. *Let $\epsilon > 0$. A solution to the system of equations* (5.61)*, i.e.,*

$$-\sum_{s=t}^{T-1} \Psi_{t,s} = \theta, \quad t = 0, \ldots, T-1,$$
$$\sum_{t=0}^{T-1} R_t = R_{tot},$$

solves Problem 5.5.6, with θ denoting the associated Lagrangian multiplier, and with $\Psi_{t,s}$ given by (5.62)*.*

Proof. The proof is still based on Lagrange duality theory. First, strong duality applies, because the constraint is positive and linearly independent of the rate R_t, the Mangasarian-Fromowitz constraint qualification applies. Second, we minimize the Lagrangian,

$$\eta(\mathbf{R}_0^{T-1}, \theta) = \sum_{t=0}^{T-1} \mathfrak{J}_t(\mathbf{R}_0^t) + \theta \left(\sum_{t=0}^{T-1} R_t - R_{tot} \right),$$

where $\hat{\sigma}_{x_t}^2$ is given by (5.55). The derivative $\partial\eta(\mathbf{R}_0^{T-1}, \theta)/\partial R_t$ is

$$\frac{\partial}{\partial R_t} \eta(\mathbf{R}_0^{T-1}, \theta) = \sum_{s=t}^{T-1} \Psi_{t,s} + \theta,$$

where $\Psi_{t,s}$ is given by (5.62), i.e.,

$$\Psi_{t,s} = \sum_{b_0=0}^{1} \cdots \sum_{b_t=1}^{1} \cdots \sum_{b_{s-1}=0}^{1} \pi_s \bar{\Psi}(b_0, \ldots, b_s).$$

The binary variable b_m, $m = 0, \ldots, s$, and $m \neq t$, takes the value in $\{0, 1\}$, except b_t, which takes only the value 1. The term $\bar{\Psi}(b_0, \ldots, b_s)$ is defined as

$$\bar{\Psi}(b_0, \ldots, b_s) = \bar{B} \left(\prod_{m=\bar{s}+1}^{s-1} \bar{B}_m \right) \left(\prod_{n=\bar{s}+1}^{s-1} (C_n)^{b_n} \right),$$

where $\bar{B}$ and $\bar{B}_n$ are given by (5.58)–(5.59), and C_n is given by (5.63). Then $\partial\eta(\mathbf{R}_0^{T-1}, \theta)/\partial R_t = 0$ and $\partial\eta(\mathbf{R}_0^{T-1}, \theta)/\partial\theta = 0$ together lead to (5.61), as stated in Lemma 5.5.14. □

An example will make the system of equations (5.61) more clear. Let $T = 3$ which yields 4 equations and 4 unknown parameters, $\{\mathbf{R}_0^2, \theta\}$. Given $\{\pi_0^2, \mathbf{A}_0^1, \mathbf{B}_0^1\}$

as defined previously in Theorem 5.5.7, (5.61) leads to the following system of equations,

$$\begin{aligned}
\theta =& -\pi_0\sigma_{x_0}^2\frac{\partial\tilde{J}_0}{\partial R_0}(\beta_0,\tilde{\kappa}_0,R_0)-\pi_1 B_0\sigma_{x_0}^2\frac{\partial\tilde{J}_0}{\partial R_0}(\tilde{\beta}_0,\tilde{\kappa}_0,R_0)\tilde{J}_1(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)\\
&-\pi_2 A_1 B_0\sigma_{x_0}^2\frac{\partial\tilde{J}_0}{\partial R_0}(\tilde{\beta}_0,\tilde{\kappa}_0,R_0)\tilde{J}_2(\tilde{\beta}_2,\tilde{\kappa}_2,R_2)\\
&-\pi_2 B_1 B_0\sigma_{x_0}^2\frac{\partial\tilde{J}_0}{\partial R_0}(\tilde{\beta}_0,\tilde{\kappa}_2,R_0)\tilde{J}_1(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)\tilde{J}_2(\tilde{\beta}_2,\tilde{\kappa}_2,R_2),\\
\theta =& -\pi_1(A_0\sigma_{x_0}^2+\sigma_v^2)\frac{\partial\tilde{J}_1}{\partial R_1}(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)-\pi_1 B_0\sigma_{x_0}^2\tilde{J}_0(\tilde{\beta}_0,\tilde{\kappa}_0,R_0)\frac{\partial\tilde{J}_1}{\partial R_1}(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)\\
&-\pi_2 B_1(A_0\sigma_{x_0}^2+\sigma_v^2)\frac{\partial\tilde{J}_1}{\partial R_1}(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)\tilde{J}_2(\tilde{\beta}_2,\tilde{\kappa}_2,R_2)\\
&-\pi_2 B_1 B_0\sigma_{x_0}^2(\tilde{J}_0(\tilde{\beta}_0,\tilde{\kappa}_0,R_0)\frac{\partial\tilde{J}_1}{\partial R_1}(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)\tilde{J}_2(\tilde{\beta}_2,\tilde{\kappa}_2,R_2),\\
\theta =& -\pi_2(A_1(A_0\sigma_{x_0}^2+\sigma_v^2)+\sigma_v^2)\frac{\partial\tilde{J}_2}{\partial R_2}(\tilde{\beta}_2,\tilde{\kappa}_2,R_2)\\
&-\pi_2 A_1 B_0\sigma_{x_0}^2\tilde{J}_0(\tilde{\beta}_0,\tilde{\kappa}_0,R_0)\frac{\partial\tilde{J}_2}{\partial R_2}(\tilde{\beta}_2,\tilde{\kappa}_2,R_2)\\
&-\pi_2 B_1(A_0\sigma_{x_0}^2+\sigma_v^2)\tilde{J}_1(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)\frac{\partial\tilde{J}_2}{\partial R_2}(\tilde{\beta}_2,\tilde{\kappa}_2,R_2)\\
&-\pi_2 B_1 B_0\sigma_{x_0}^2\tilde{J}_0(\tilde{\beta}_0,\tilde{\kappa}_0,R_0)\tilde{J}_1(\tilde{\beta}_1,\tilde{\kappa}_1,R_1)\frac{\partial\tilde{J}_2}{\partial R_2}(\tilde{\beta}_2,\tilde{\kappa}_2,R_2),\\
R_{tot} =& \sum_{t=0}^{2} R_t.
\end{aligned}$$

As discussed previously, the resulting system of equations can be solved by numerical methods. Finally, the proof of Theorem 5.5.7 is given below.

Proof. **(Theorem 5.5.7)**

Based on Lemma 5.5.8–Lemma 5.5.10, we note that when $\epsilon = 0$, given any $R_{tot} < \infty$, a solution to Problem 5.5.6 is given by (5.56), since the global optimum to the unconstrained problem is $R_t^\star = \infty$, $\forall t$. For $\epsilon > 0$, a solution to Problem 5.5.6 is (5.60), if the global optimum to the unconstrained problem does not violate the rate constraint, cf., Lemma 5.5.13. Otherwise the solution is given by (5.61), as shown by Lemma 5.5.14. Hence, we conclude that the proof of Theorem 5.5.7 follows Lemma 5.5.8–Lemma 5.5.14. □

It is worth noting that when $\tilde{\beta}_t = \tilde{\beta}$ and $\tilde{\kappa}_t = \tilde{\kappa}$, the unconstrained estimation problem and control problem, Problem 5.4.2 and Problem 5.5.5, have the same global minimum.

In this section, we studied the problem to assign optimally R_{tot} bits to T time units for state feedback control over noisy channels. The solutions are derived first for the minimum variance control scenario ($\rho = 0$) and then for the general case ($\rho > 0$). In the both cases, the rate allocation problems are solved in two steps. First, we approximated the objective functions by means of high-rate quantization theory. Second, we showed that the unconstrained optimization problem has a global minimum, which solves the rate allocation problem if such a global minimum does not violate the rate constraint. On the other hand, if the global minimum violates the rate constraint, then we solved the rate constrained optimization problem by means of Lagrangian duality for non-convex non-linear problems.

5.6 Low Complexity Solutions

The complexity of the rate allocation optimization algorithm presented in Section 5.5 for feedback state control is significantly increased, compared with the state estimation problem in Section 5.4. This is due to the fact that $\mathbf{E}\left\{x_t^2\right\}$ and its approximation $\hat{\sigma}_{x_t}^2$ depend on all previous $\mathbf{R}_0^{t-1}$, in a complicated manner. As a result, deriving $\hat{\sigma}_{x_t}^2$ as a function of $\mathbf{R}_0^{t-1}$ requires a large number of computations, as T increases. In this section a fast rate allocation algorithm is presented by which we aim for a method of reasonably good performance, without heavy computations. The idea is to take advantage of some realistic approximations of $\sigma_{x_t}^2$ that are not dependent on the past rates $\mathbf{R}_0^{t-1}$. Recall the cost (5.37), introduced in Section 5.5,

$$J_{tot}(\mathbf{R}_0^{T-1}) = \sum_{t=0}^{T-1} \mathbf{E}\left\{\pi_t (x_t - d_t)^2\right\}, \quad \pi_t = (\phi_{t+1} + \rho)\ell_t^2.$$

Based on the preceding discussions in Section 5.5 we know that the true statistics of x_t is obtainable first after the control sequence $\mathbf{u}_0^{t-1}$ is known. However, it is in general impossible to express the pdf $p(x_t)$ in a closed-form, even if $\mathbf{u}_0^{t-1}$ is available. In order to proceed, certain approximations and simplifications will be made. First of all, we simplify the problem by approximating x_t using a zero-mean Gaussian pdf, denoted by $\mathcal{N}(0, \hat{\sigma}_{x_t}^2)$, as done previously in Section 5.5. Certainly, we are interested in controls that have positive impact on the system performance. Loosely speaking, by positive impact we mean that the control should ensure the true state to stay within a bounded region. First, the upper bound is given by the case that no control action is taken, for all t. It is obvious that a useful coding–control scheme should outperform the uncontrolled system. Second, the lower bound is given by the case that there is no channel in the system. In the absence of the channel imperfections (quantizer distortion and transmission errors), Problem 5.2.4 is reduced to the classical LQG problem, and the optimal solution is well-known. Naturally, this lower bound serves as the best achievable performance. A solution of a satisfactory controller should stay close to the optimal solution of LQG problem. For ease of presentation, we refer to the upper bound as the non-controlled scenario, and the lower bound to as the optimally-controlled scenario.

Based on the above discussion, we first introduce the upper bound J_{up} for the non-controlled scenario,

$$J_{up}(\mathbf{R}_0^{T-1}) = \sum_{t=0}^{T-1} \mathbf{E}\left\{\pi_t \left(x_t^{up} - d_t\right)^2\right\},$$

where x_t^{up} denotes the state obtained when no control action is applied, in particular

$$x_t^{up} = a^t x_0 + \sum_{s=0}^{t-1} a^{t-1-s} v_s.$$

Note that, since x_0 and $\mathbf{v}_0^{t-1}$ are i.i.d. zero-mean Gaussian variables, x_t^{up} is consequently zero-mean Gaussian with the variance

$$\sigma^2_{x_t^{up}} = a^2 \sigma^2_{x_{t-1}^{up}} + \sigma_v^2 = a^{2t}\sigma^2_{x_0} + \sum_{s=0}^{t-1} \left(a^{t-1-s}\right)^2 \sigma_v^2.$$

In the second place, the lower bound J_{low} is given by,

$$J_{low}(\mathbf{R}_0^{T-1}) = \sum_{t=0}^{T-1} \mathbf{E}\left\{\pi_t \left(x_t^{low} - d_t\right)^2\right\},$$

where x_t^{low} is obtained by the LQG control assuming x_t is directly available at the controller. Also, x_t^{low} is zero-mean Gaussian with the variance

$$\begin{aligned}\sigma^2_{x_t^{low}} &= (a+\ell_{t-1})^2 \sigma^2_{x_{t-1}^{low}} + \sigma_v^2 \\ &= \left(\prod_{s=0}^{t-1} (a+\ell_s)^2\right) \sigma^2_{x_0} + \sum_{n=0}^{t-2} \left(\prod_{m=0}^{n} (a+\ell_{t-1-m})^2\right) \sigma_v^2.\end{aligned}$$

If we compute the instantaneous cost using (5.54) and approximating $\hat{\sigma}^2_{x_t}$ with $\sigma^2_{x_t^{low}}$ or $\sigma^2_{x_t^{up}}$, the rate allocation problem is substantially simplified. In that case, we find ourselves confronted by a problem resembling Problem 5.4.1 in Section 5.4. The main advantage of the system of Problem 5.4.1 is that the instantaneous cost is independent of $\mathbf{R}_0^{t-1}$, and consequently separable.

Once the cost function is established, we can straightforwardly apply the method of Lagrange multipliers. Below, we summarize the results for the error-free scenario in Corollary 5.6.1, and the noisy scenario in Corollary 5.6.2.

Corollary 5.6.1. *For error-free channels ($\epsilon=0$), it holds that*

$$R_t = \frac{R_{tot}}{T} + \frac{1}{2}\log_2\left(\frac{\hat{\sigma}^2_{x_t}\tilde{\kappa}_t\pi_t}{\left(\prod_{t=0}^{T-1}\hat{\sigma}^2_{x_t}\tilde{\kappa}_t\pi_t\right)^{\frac{1}{T}}}\right), \quad t=0,\ldots,T-1,$$

is a solution which minimizes the Lagrangian

$$\eta(\mathbf{R}_0^{T-1},\theta)=\sum_{t=0}^{T-1}\pi_t\hat{\sigma}_{x_t}^2\tilde{\kappa}_t 2^{-2R_t}+\theta\left(\sum_{t=0}^{T-1}R_t-R_{tot}\right).$$

Following the standard procedure, Corollary 5.6.1 is proved by finding $\mathbf{R}_0^{T-1}$ which simultaneously satisfies $\partial\eta(\mathbf{R}_0^{T-1},\theta)/\partial\theta=0$ and $\partial\eta(\mathbf{R}_0^{T-1},\theta)/\partial R_t=0$, $\forall t$.

Corollary 5.6.2. *For noisy channels* ($\epsilon>0$)*, it holds that a solution* $\{\mathbf{R}_0^{T-1},\theta\}$ *to the system of equations*

$$\begin{aligned}\pi_t\hat{\sigma}_{x_t}^2\tilde{\beta}_t(1-\epsilon)^{R_t}+\pi_t\hat{\sigma}_{x_t}^2\tilde{\kappa}_t 2^{-2R_t}=\theta,\quad t=0,\ldots,T-1,\\ \sum_{t=0}^{T-1}R_t=R_{tot},\end{aligned}\tag{5.67}$$

minimizes the Lagrangian

$$\eta(\mathbf{R}_0^{T-1},\theta)=\sum_{t=0}^{T-1}\pi_t\hat{\sigma}_{x_t}^2\tilde{J}_t(\tilde{\beta}_t,\tilde{\kappa}_t,R_t)+\theta\left(\sum_{t=0}^{T-1}R_t-R_{tot}\right).$$

The proof of Corollary 5.6.2 follows the proof of Lemma 5.4.6, since the system of equations (5.67) is almost identical to (5.28), besides the weighting factor π_t. The imposed approximations and the simplifications are expected to cause performance degradation. In many cases, we can choose one of the above approximations to compute a simple solution. However, which one of the two approximations is closer to the true solution is not easy to answer. In the best situation the both approximations perform so close that they result in exactly the same rate allocation. Otherwise, listed below are some observations that can us to choose between $\sigma^2_{x_t^{low}}$ and $\sigma^2_{x_t^{up}}$:

1. In the absence of channel errors, the influence of control is increased with the magnitude of the system parameter a. If $|a|\gg 1$, then the performance is close to J_{low}.

2. For an error-free channel, if $|a|\ll 1$, then the performance is dominated by the plant's dynamics. Consequently, J_{low} and J_{up} are rather close.

3. A noisy communication channel has a variety of negative impacts on the system. If the rate is low or the crossover probability of the channel is high, only cautious control actions are suitable, thus the resulting performance might be far from the optimal-controlled scenario. In those cases the true system will perform more closely to J_{up}.

5.7 Exhaustive Search

As explained in Section 5.5, the exact solution to the optimal bit-rate allocation problem over a noisy channel is difficult to derive. Given this difficulty, we proposed a rate allocation method developed based on a number of approximations and simplifications, and consequently, performance degradation is expected.

However, we know that for discrete problems, the globally optimal solution can always be found by examining all possible cases, which is commonly referred to as the *exhaustive search*. For the reason of completeness, this section is devoted to a brief discussion of designing rate allocations by testing all possible cases. Especially, we will present a search order which is reasonably efficient in practice. This search algorithm is developed for situations when the optimal rate allocation is monotonously decreasing, which is rather common for stable closed-loop systems. First, for ease of presentation, we define the search set,

$$\mathcal{R}_{T,R_{tot}} = \left\{ \mathbf{R}_0^{T-1} : \sum_{t=0}^{T-1} R_t \leq R_{tot}, R_t \in \mathbb{Z}^+ \right\}.$$

A constrained monotonously decreasing rate allocation has $R_t \geq R_s$, for all $t < s$, $\mathbf{R}_0^{T-1} \in \mathcal{R}_{T,R_{tot}}$. Next, let us establish certain rules that are used in the forthcoming comparisons of the rate allocations. Consider two rate allocations, the integer sequences $\mathbf{R}_0^{T-1}(1) \in \mathcal{R}_{T,R_{tot}}$ and $\mathbf{R}_0^{T-1}(2) \in \mathcal{R}_{T,R_{tot}}$, for a fixed R_{tot}.

1. First, the sequence $\mathbf{R}_0^{T-1}(1)$ is said to be *larger* than the sequence $\mathbf{R}_0^{T-1}(2)$, if

$$\sum_{t=0}^{s_{min}} R_t(1) > \sum_{t=0}^{s_{min}} R_t(2), \tag{5.68}$$

where,

$$s_{min} \triangleq \min_s \left\{ \sum_{t=0}^{s} R_t(1) \neq \sum_{t=0}^{s} R_t(2) \right\}. \tag{5.69}$$

In other words, s_{min} is the smallest integer s at which the two sequences differ (or the first time the two sequences differ). The smallest sequence, by our definition, has the mostly uniform rate allocation. In view of some additional properties, the smallest sequence has:

a) The smallest maximum instantaneous rate, i.e.,

$$\min_{\mathbf{R}_0^{T-1} \in \mathcal{R}_{T,R_{tot}}} \{\max\{R_0, R_1, \ldots, R_{T-1}\}\}.$$

b) The smallest difference between the maximum and minimum rates, i.e.,

$$\min_{\mathbf{R}_0^{T-1} \in \mathcal{R}_{T,R_{tot}}} \{\max\{R_0, R_1, \ldots, R_{T-1}\} - \min\{R_0, R_1, \ldots, R_{T-1}\}\}.$$

Algorithm 5.7.1 Bit allocation by Exhaustive Search

1. Set the minimum cost $J_{min} = \mathbf{E}\{J_{tot}(\mathbf{0})\}$, where $\mathbf{0}$ is a vector of zeros. Set the existing best rate allocation R_{min} to $\mathbf{0}$.
2. For $R_0, \ldots, R_{T-1} \in \{0, \ldots, R_{tot}\}$,
   2.a. If the rate constraint $\sum_{t=0}^{T-1} R_t \leq R_{tot}$ is fulfilled, compute the expected cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$.
   2.b. If $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\} < J_{min}$, replace the exiting minimum cost J_{min} and the existing best rate allocation R_{min} by $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ and $\mathbf{R}_0^{T-1}$.

Algorithm 5.7.2 Bit allocation by Binary Search

1. Let R_{max} be the largest value R_t can take.
2. Start with setting $t = 0$. For $k = 0, \ldots, R_{max}$,
   2a. Construct two rate allocations as follows:
   (i) Let the both allocations have $R_0 = k$;
   (ii) Among all the feasible sequences $\mathbf{R}_1^{T-1}$, which fulfill the rate constraint, select the largest and the smallest sequences according to (5.68)–(5.69).
   2b. Compute $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ for the selected two rate allocations. Denote $J_0^{min}(k)$ the cost associated to the rate allocation with the smallest $\mathbf{R}_1^{T-1}$, and $J_0^{max}(k)$ the largest.
3. Assign $R_0 = \arg\min_k \left\{\min\left(J_0^{max}(k), J_0^{min}(k)\right)\right\}$.
4. Let $t = t+1$. For $k = 1, \ldots, R_{t-1}$,
   4a. Construct two rate allocations which both have $\mathbf{R}_0^{t-1}$ and $R_t = k$. They are assigned the largest and the smallest sequences among the feasible sequences $\mathbf{R}_{t+1}^{T-1}$.
   4b. Compute $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ for the two rate allocations, and denote them $J_t^{min}(\mathbf{R}_0^{t-1}, k)$ and $J_t^{max}(\mathbf{R}_0^{t-1}, k)$.
   4c. Assign $R_t = \arg\min_k \{\min\left(J_t{}^{max}(\mathbf{R}_0^{t-1}, k), J_t^{min}(\mathbf{R}_0^{t-1}, k)\right)\}$.
5. If $t = T-1$, then stop, otherwise go to Step 5.

2. On the other hand, we say that the allocation $\mathbf{R}_0^{T-1}(1)$ performs *better* than the allocation $\mathbf{R}_0^{T-1}(2)$, meaning that

$$\mathbf{E}\left\{J_{tot}\left(\mathbf{R}_0^{T-1}(1)\right)\right\} < \mathbf{E}\left\{J_{tot}\left(\mathbf{R}_0^{T-1}(2)\right)\right\},$$

where $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ is the overall cost that we want to minimize. The cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ can either be calculated according to certain analytical expression, or simulated numerically.

A regular exhaustive search can be simply implemented as T loops, as described

in Algorithm 5.7.1. Let all T parameters, $R_0, \ldots, R_{T-1}$, take on values from 0 to R_{tot}. If the sum of the instantaneous rates does not exceed R_{tot}, then the expected cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ will be calculated, either analytically or numerically. If the expected cost $\mathbf{E}\left\{J_{tot}(\mathbf{R}_0^{T-1})\right\}$ is smaller than the existing minimum J_{min}, then replace the existing minimum and store the corresponding rate allocation as the best allocation up to now.

Owing to the fact that the horizon T and the total rate R_{tot} are both finite numbers, the globally optimal rate allocation can always be found, by an exhaustive search over all R_{tot}^T sequences. Also, it is worth mentioning that when the channel is error-free, the total bit-rate constraint in Algorithm 5.7.1 can be replaced by $\sum_{t=0}^{T-1} R_t = R_{tot}$. Clearly, the search set given by the rate constraint $\sum_{t=0}^{T-1} R_t = R_{tot}$ is considerably smaller than the original search set in $\sum_{t=0}^{T-1} R_t \leq R_{tot}$. The reduction of the search space is because in the error-free case increasing data rate always yields less distortion. On the other hand, in the presence of channel errors, the situation becomes more complicated. It may happen that increasing data rate does more harm than good. Therefore, we should indeed search among all sequences which fulfill $\sum_{t=0}^{T-1} R_t \leq R_{tot}$. Obviously, as T and R_{tot} increase, a full search according to Algorithm 5.7.1 becomes impractical, since there are R_{tot}^T combinations to test.

For this reason, we present an alternative search method in Algorithm 5.7.2 which is fairly efficient in practice. As a matter of fact, what we suggest is a binary search with a special search order. Moreover, we introduce R_{max} to provide a flexility of varying the size of the search set. We can increase or decrease the search set by adjusting R_{max}.

5.8 Practical Considerations

This section is devoted to several considerations regarding the implementation of the rate allocation algorithms developed in Section 5.4 and Section 5.5. In Section 5.8.1, we deal with the non-negativity and integer constraints imposed on $\mathbf{R}_0^{T-1}$. In Section 5.8.2, performance degradation cased by various approximations and simplifications are discussed.

5.8.1 Non-negativity and Integer Constraint

In this subsection we deal with the assumption of Theorem 5.4.4 and Theorem 5.5.7 that the rate R_t is allowed to be negative and real numbers. In practice, R_t is of course a non-negative integer. As regards the classical rate allocation problems, there are many works dealing with the non-negativity and integer constraints, e.g., [FZ05b]. It is beyond our intention to engage in this issue here. In this section, we briefly present the solutions that are used in our numerical experiments.

First, we should mention that the solutions to the system of equations in Theorem 5.4.4 and Theorem 5.5.7 might contain negative rates. We deal with the problem of negative rates as follows. Set the negative rates to 0. Then, exclude the

instantaneous costs associated with the negative rates from the overall cost function and resolve the optimization problem following Theorem 5.4.4 or Theorem 5.5.7. Repeat the procedure until all negative rates are settled.

The solution given by Theorem 5.4.4 and Theorem 5.5.7 are real-valued rates. In the simplest case, we can handle the integer constraint by applying the rule,

$$\tilde{R}_t = \lceil R_t \rfloor, \quad t = 0, \ldots, T-1,$$

where $\lceil \cdot \rfloor$ is taken to mean *rounding to* the nearest integer, and R_t is a solution given by Theorem 5.4.4 or Theorem 5.5.7. This operation might violate the total bit-rate constraint (5.5). In many cases the deviation is however only a few bits, mostly only one single bit. Hence, it brings up an interesting question, namely, for a given rate allocation $\tilde{\mathbf{R}}_0^{T-1}$, which bit is the least significant one we can exclude? The answer to the state estimation problem (cf., Problem 5.4.1) is rather straightforward, attributed to the fact that $\mathbf{E}\left\{x_t^2\right\}$ is not affected by the communication system. In that case, we are looking for the time $\hat{t}$ satisfying

$$\hat{t} = \arg\min_t \left\{ \hat{J}_t\big(\beta_t, \kappa_t, \tilde{R}_t - 1\big) - \hat{J}_t\left(\beta_t, \kappa_t, \tilde{R}_t\right) \right\},$$

with the instantaneous cost $\hat{J}_t(\beta_t, \kappa_t, \tilde{R}_t) = \beta_t(1-(1-\epsilon)^{\tilde{R}_t}) + \kappa_t 2^{-2\tilde{R}_t}$. Since the cost $\hat{J}_t(\beta_t, \kappa_t, \tilde{R}_t)$ does not depend on past $\tilde{\mathbf{R}}_0^{t-1}$, the additional bit at t only has impact on the current cost $\mathbf{E}\left\{J_t(\tilde{R}_t)\right\}$. The new rate allocation $\bar{\mathbf{R}}_0^{T-1}$ follows simply the rule

$$\begin{cases} \bar{R}_t &= \tilde{R}_t, & t \neq \hat{t}, \\ \bar{R}_t &= \tilde{R}_t - 1, & t = \hat{t}. \end{cases} \tag{5.70}$$

Finding the least significant bit in a state feedback control system is a more challenging task. To compare the influence of each R_t on the total cost, it is not enough to consider only their impact on the current time instant. We should also take into account their impact on all future states. In that case we are looking for $\hat{t}$ which is

$$\hat{t} = \arg\min_t \left\{ \sum_{s=t+1}^{T-1} \mathfrak{J}_s\big(\tilde{\mathbf{R}}_0^{t-1}, \tilde{R}_t - 1, \tilde{\mathbf{R}}_{t+1}^{s}\big) + \mathfrak{J}_t\big(\tilde{\mathbf{R}}_0^{t-1}, \tilde{R}_t - 1\big) - \sum_{s=t}^{T-1} \mathfrak{J}_s\big(\tilde{\mathbf{R}}_0^{s}\big) \right\}, \tag{5.71}$$

where $\mathfrak{J}_t$, $t = 0, \ldots, T-1$, is the instantaneous cost given by (5.54). In other words, we seek the least significant bit which minimally increases the overall expected cost. The new rate allocation is then finalized by (5.70).

Based on the preceding discussion, we can remove the least significant bit by means of (5.71). If there are several bits to discard, we can successively adopt (5.71) to exclude one bit at a time. However, we should keep in mind that the successive deduction does not ensure the optimal solution, resembling the well known fact that an optimal control over a finite horizon is not necessarily also optimal in terms of the instantaneous cost.

Conversely, the analogues problem of adding the most significant bit can be solved by finding the following $\hat{t}$,

$$\hat{t} = \arg\max_{t} \left\{ \sum_{s=t}^{T-1} \mathfrak{J}_s\big(\tilde{\mathbf{R}}_0^s\big) - \sum_{s=t+1}^{T-1} \mathfrak{J}_s\big(\tilde{\mathbf{R}}_0^{t-1}, \tilde{R}_t - 1, \tilde{\mathbf{R}}_{t+1}^s\big) - \mathfrak{J}_t\big(\tilde{\mathbf{R}}_0^{t-1}, \tilde{R}_t - 1\big) \right\},$$

which maximally reduces the overall expected cost.

Next, we present a more sophisticate method by formulating the rounding as a binary optimization problem. The rounded rate $\tilde{R}_t$ is related to the real-valued rate R_t as

$$\tilde{R}_t = b_t \lceil R_t \rceil + (1 - b_t) \lfloor R_t \rfloor, \quad b_t \in \{0, 1\}, \;\; t = 0, \ldots, T-1,$$

where $\lceil \cdot \rceil$ denotes the rounding upwards to the nearest integer, and $\lfloor \cdot \rfloor$ denotes the rounding downwards to the nearest integer. We optimize the rounding by searching for the binary sequence $\mathbf{b}_0^{T-1}$ which minimizes the overall cost $J_{tot}(\tilde{\mathbf{R}}_0^{T-1})$,

$$\begin{aligned} \min_{\mathbf{b}_0^{T-1}} \quad & \mathbf{E}\left\{ J_{tot}\left(\tilde{\mathbf{R}}_0^{T-1}\right)\right\}, \\ \text{s. t.} \quad & \sum_{t=0}^{T-1} \tilde{R}_t \leq R_{tot}. \end{aligned} \tag{5.72}$$

A solution to the problem (5.72) can always be obtained by exhaustive search or combinatorial algorithms [PS98].

5.8.2 Performance Degradation

In Sections 5.4–5.7 we have suggested several methods to optimize rate allocation for state estimation and state feedback control over finite-rate noisy channels. In order to formulate useful overall objective functions, different levels of approximations and simplifications are introduced. In this subsection, a few remarks are given on the impacts of the diverse approximations and simplifications on the quality of the solutions. Before starting the discussion, it is worth mentioning that even though the absolute accuracy of the estimate of the overall cost is flawed, the proposed algorithms may still be able to provide a satisfactory solution, because what really matters is often the ratios among the costs at all time instants.

We start with the remark on the Lagrangian-based method. First, since the primal function is formulated based on several approximations and simplifications, even though we can find a solution to the relaxed optimization problem, the solution is not necessarily optimal for the original problem. Second, despite the performance degradation caused by all approximations and simplifications, what we can say about the solution to the Lagrangian dual problem is closely related to the objective function. If the objective function is convex, then given the linear constraint (5.5), the strong duality applies if the solution $\{\mathbf{R}_0^{T-1}, \theta\}$ fulfills the Kaursh–Kuhn–Tucker

(KKT) condition [BV04]. On the other hand, if the objective function is non-convex, the KKT condition is in many cases only a necessary condition for strong duality. In general, the convexity does not hold for the rate allocation problem discussed in this chapter. For example, consider the objective function of the state estimation problem from Section 5.4. It is a sum of quasi-convex functions of the form

$$\beta_t(1-(1-\epsilon)^{R_t})+\kappa_t 2^{-2R_t}, \quad 0<\beta_t<\infty,\ \ 0<\kappa_t<\infty.$$

The quasi-convexity is commonly not preserved by summation. In Figure 5.6, the typical behavior of the overall cost function is demonstrated by a simple example with the overall cost

$$\beta_1(1-(1-\epsilon)^{R_1})+\kappa_1 2^{-2R_1}+\beta_2(1-(1-\epsilon)^{R_2})+\kappa_2 2^{-2R_2}. \tag{5.73}$$

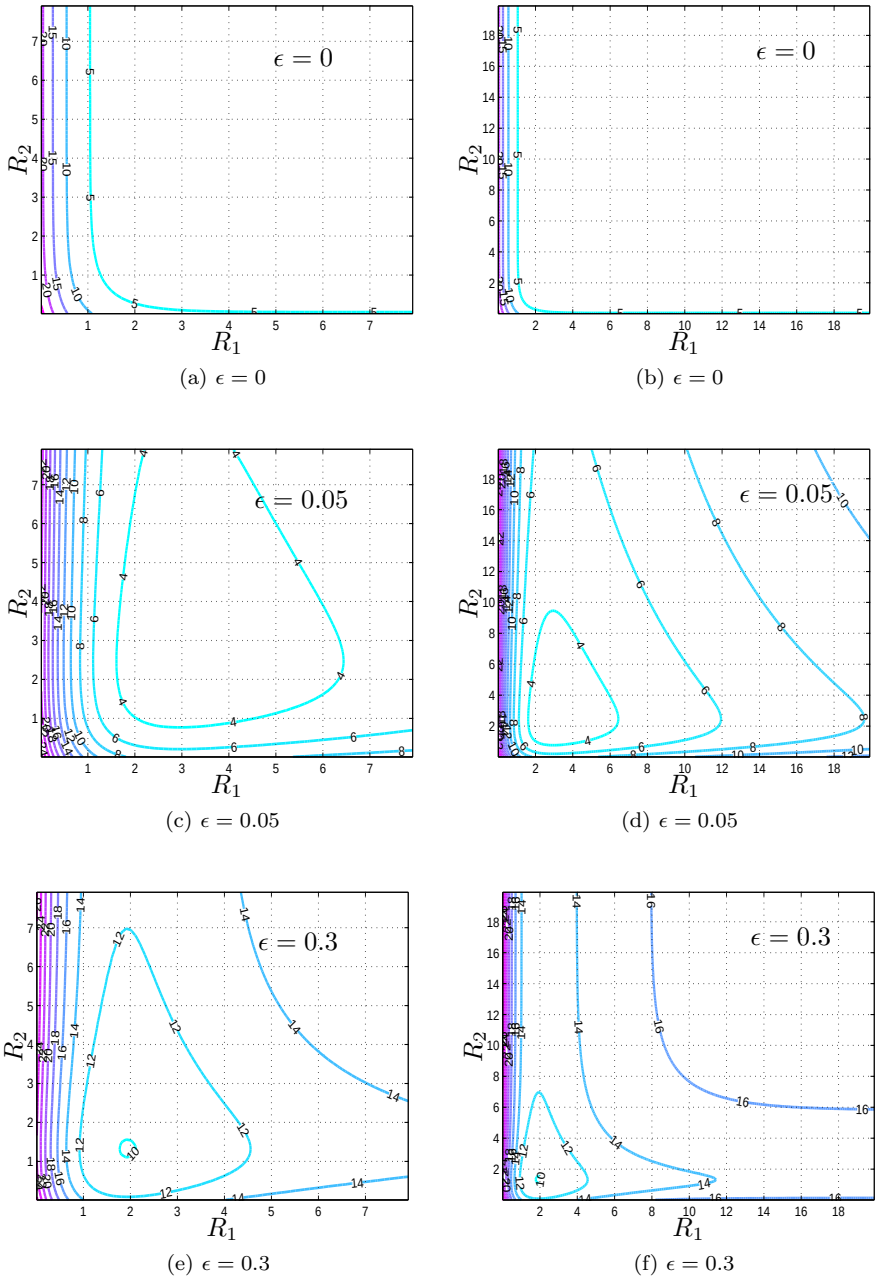
The quasi-convexity is verified by the sub-level sets. In particular, Figure 5.6 shows that the quasi-convexity can be violated by increasing the crossover probability ϵ or the rate R_t. It is worth mentioning that the rate constraint is an inequality in our minimization problem. Therefore, an optimal solution can be ensured by the Lagrangian-based method if we are able to find the subset of $\{\mathbf{R}_0^{T-1}: 0\leq R_t<\infty, \forall t\}$ that not only embodies the optimal solution but also formulates a convex optimization problem. Because of the property of $\hat{J}_t$, we could show that our optimization problems have a global minimum, and if the global minimum does not violate the rate constraint, then it is the solution to the rate constrained optimization problem. On the other hand, if the global minimum violates the rate constraint, a convex optimization problem is formulated and solved by strong duality.

Next, we discuss the high-rate approximation of the MSE. First, even though the high-rate assumption requires that the pdf of the source is approximately constant over one quantization cell, however, the quantization works fairly well in practice for low rates as much as 3, 4 bits. On the other hand, the accuracy of the quantization decreases when the rate approaches 0. That is to say, at low rates the proposed rate allocation algorithm does not work as well as in the high-rate region, attributed to all approximations made in the derivation of the high-rate expression (5.11). The worst case occurs at $R_t=0$, where the estimation errors given by (5.15) and (5.17) are even worse than $\mathbf{E}\{x_t^2\}=\sigma_x^2$, obtained by setting $d_t=0$.

Finally, the Gaussian approximation becomes flawed as the rate decreases. The problem becomes more serious for unstable systems because errors accumulate as time goes on. However, the Gaussian process noise in the system plays a role in alleviating the conflict between the model and the true system. As the rate increases, the problem of accuracy is quickly solved.

5.9 Numerical Experiments

In this section, we present the numerical experiments conducted to evaluate the performance of the proposed bit-rate allocation algorithms.



(a) $\epsilon = 0$ (b) $\epsilon = 0$

(c) $\epsilon = 0.05$ (d) $\epsilon = 0.05$

(e) $\epsilon = 0.3$ (f) $\epsilon = 0.3$

Figure 5.6: The impact of R_t and ϵ on the quasi-convexity of the cost function (5.73). The quasi-convexity can be violated by increasing the crossover probability ϵ or the rate R_t.

Let us first address a few issues common for all experiments in this section. In general, we optimize the rate allocation according to the algorithms proposed in Section 5.4 or Section 5.5, and then assess their performance by means of numerical simulations. For state feedback control problems, the optimized rate allocation is obtained by applying Theorem 5.5.7 and the binary rounding algorithm described in Section 5.8. More specifically, we optimize the rate allocation for state feedback control with respect to the objective function (5.54) of Problem 5.5.6, the overall performance is on the other hand evaluated in terms of the objective function (5.4) of Problem 5.2.4. Consequently, performance degradation caused by various approximations and simplifications is expected. Likewise, for state estimation problems, the optimized rate allocation is referred to as the sequence $\mathbf{R}_0^{T-1}$ obtained by applying Theorem 5.4.4 and the binary rounding algorithm described in Section 5.8.

Throughout this section, the initial-state and the process noise are zero-mean Gaussian with variances $\sigma_{x_0}^2$ and σ_v^2, respectively. For the exposition of the basic design concept and procedure, we choose to use a time-varying uniform quantizer for which the quantizer range is related to the estimated signal variance as $\nu_t = 4\hat{\sigma}_{x_t}$. At the same time, we also use this quantizer to expose the importance of a carefully designed coding–controller. As revealed later, this simple quantizer is far from optimal in view of the efficiency of communications. Besides the high-rate approximation, we derive the instantaneous cost (5.66) also based on the following simplifications. First, x_t is assumed to be zero-mean Gaussian. Second, the high-rate approximation (5.16) is adopted, particularly assuming that the distortion caused by the signals outside the support of the quantizer is negligible. Finally, (5.50) and (5.53) are utilized, where we let $\Gamma(\epsilon)$ be a linear function heuristically obtained by numerical experiments. Actually, the exact value of $\Gamma(\epsilon)$ is not influent since a multiplicative constant of the cost function does not change the optimal solution.

Roughly speaking, since x_t is assumed to be zero-mean Gaussian, the rate allocation becomes meaningful if the state variance varies significantly over time. For example we can use the difference of $\sigma_{x_0}^2$ and $\sigma_{x_{T-1}}^2$ to measure the variation. If the variation is considerably large, a non-uniform allocation can gain by exploiting the flexibility of adjusting quantizers according to the statistical knowledge of the state. Otherwise, an even distribution of the bit-rates is expected to work satisfactorily. Unfortunately, the system is affected by a number of parameters: the system dynamic a, the relative weight ρ, the variances $\sigma_{x_0}^2$, σ_v^2, the total bit-rate budget R_{tot}, and the crossover probability ϵ, etc. Even if we know a great deal about the impact of each parameter separately, when all of them influence simultaneously, it becomes difficult to predict the system behavior. Listed below are some situations where $\sigma_{x_0}^2$ and $\sigma_{x_{T-1}}^2$ could differ considerably in magnitude.

1. When the system parameter a is apart from 1. Note that the system is highly stable when $|a| \ll 1$, or highly unstable when $|a| \gg 1$.
2. When the process noise is on average small compared with the initial-state. For large-valued process noise, the system converges to a high steady-state level.

	Rate allocation	R_{tot}
RA_1	8 8 8 8 8 8 8 8 8 8	80
RA_2	7 7 7 7 7 7 7 7 7 7	70
RA_3	6 6 6 6 6 6 6 6 6 6	60
RA_4	5 5 5 5 5 5 5 5 5 5	50
RA_5	4 4 4 4 4 4 4 4 4 4	40
RA_6	3 3 3 3 3 3 3 3 3 3	30
RA_7	2 2 2 2 2 2 2 2 2 2	20
RA_8	1 1 1 1 1 1 1 1 1 1	10
RA_9	8 8 8 6 0 0 0 0 0 0	30
RA_{10}	7 7 7 7 2 0 0 0 0 0	30
RA_{11}	6 6 6 6 6 0 0 0 0 0	30
RA_{12}	5 3 3 3 3 3 3 3 2 2	30
RA_{13}	5 4 3 3 3 3 3 2 2 2	30
RA_{14}	5 4 4 3 3 3 2 2 2 2	30

Table 5.1: Various rate allocations used in Figure 5.7 and Figure 5.8. RA_{12} is the optimized rate allocation for the state feedback control problem and RA_{14} is the optimized rate allocation for the state estimation problem. The others are arbitrarily selected rate allocations.

3. When the crossover probability of the channel is small. A high crossover probability favors conservative control actions and consequently slows down the convergence rate.
4. When the penalty on the control input is mild. A large-valued ρ corresponds to a hard power constraint on the control input, which leads to a slow convergence rate.

First we demonstrate the performance of the proposed rate allocation scheme by comparing it with some arbitrarily selected allocations. The system parameters are chosen in the interest of demonstrating non-uniform rate allocation. In particular, the system parameters are: $a = 0.5$, $\rho = 0.1$, $T = 10$, $R_{tot} = 30$, $\epsilon = 0.001$, $\sigma_{x_0}^2 = 10$, and $\sigma_v^2 = 0.1$. As explained in the beginning of this section, a time-varying uniform quantizer is adopted and the quantizer range is specified by the estimated variances of the states.

Figure 5.7 shows the performance of state feedback control systems. The performance measure used to derive the rate allocation is (5.66), while the performance measure used to verify the rate allocation is (5.54). The simulated costs are obtained by averaging over 100 IA's and each IA 150 000 samples. In Table 5.1, all the 14 rate allocations are listed. The allocation RA_{12} is obtained by applying the

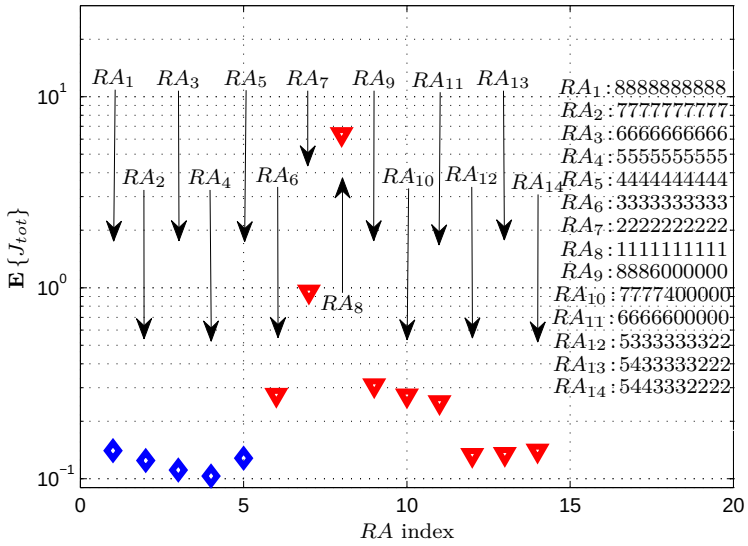


Figure 5.7: A performance comparison among various rate allocations for state feedback control. Allocations marked with a diamond do not satisfy the total rate constraint.

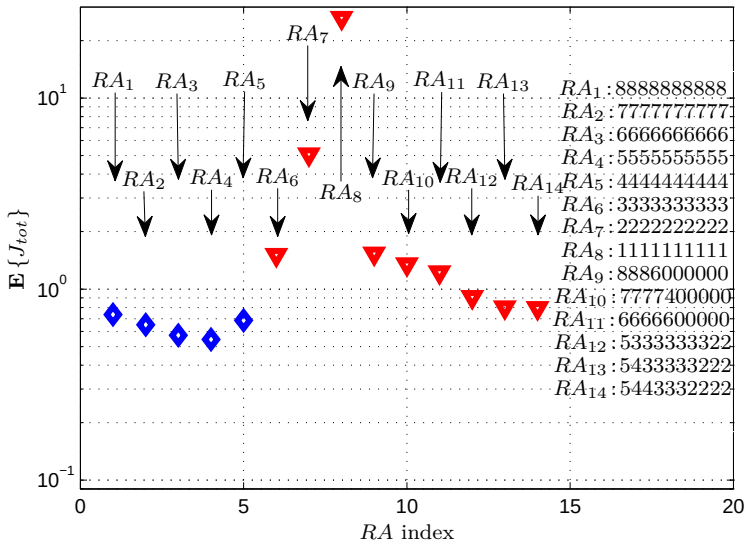


Figure 5.8: A performance comparison among various rate allocations for state estimation. Allocations marked with a diamond do not satisfy the total rate constraint.

proposed optimization algorithm from Theorem 5.5.7 and the binary rounding algorithm from Section 5.8. Concerning the optimized allocation RA_{12}, R_t is fairly evenly distributed over t. Compared with the uniform allocation RA_6, we see that our method gives an evident gain.

The uniform allocations RA_1–RA_8 have a time-invariant instantaneous rate, varying from 8 bits to 1 bit. In fact, based on our analysis, $\tilde{\beta}_t = \tilde{\beta}$, $\tilde{\kappa}_t = \tilde{\kappa}$, and the solution to the unconstraint problem, Problem 5.5.5, is $R_t^\star = 5$, $\forall t$. This is consistent with the simulation result that RA_4 is even superior to allocations with a higher total rate. In the presence of the channel errors, more bits can sometimes do more harm than good. However, RA_4 does not satisfy the total rate constraint, and therefore, the system of equations (5.61) is solved, which yields RA_{12}.

The allocations RA_9–RA_{13} are used to represent the strategy that more bits are assigned to the initial-states. This strategy is not efficient in the current example, because, as discussed previously, the additional bits exceeding the critical point, $R^\star = 5$, do more harm than good. Furthermore, the degradation caused by reducing one bit at a lower rate is more significant than the improvement along with increasing one bit at a higher rate. As a simple example, the degradation caused by using 1 bit instead of 2 is more serious than the improvement caused by using 5 bits instead of 4.

In Figure 5.8, we demonstrate the simulation results corresponding to state feedback control for the state estimation problem. Here, the optimized rate allocation is obtained by applying Theorem 5.4.4 and the binary rounding algorithm. The involved system parameters are the same as in Figure 5.7. The performance for the same 14 allocations, RA_1–RA_{14} as in Figure 5.7, are depicted. In particular, RA_4 is still the global optimum which solves the unconstrained optimization problem; while RA_{14} is the optimized rate allocation for state estimation. Performance in Figure 5.8 is measured by the objective function of Problem 5.4.1, and it is obtained by averaging over 100 IA's and each IA 150 000 samples. Compared with the uniform allocation RA_6, we see that our method gives an evident gain. Note that, here RA_{14} outperforms RA_{12}, which is the optimized allocation for state feedback control. An explanation for the gain of RA_{14} is that without control action the trajectory approaches 0 slowly, therefore it requires more bits in the initial-states.

The next example is pursued to demonstrate the impact of ρ. As mentioned previously, ρ regulates the power of control inputs. More precisely, on average the magnitude of the control input decreases as ρ increases. In other words, a large ρ yields on average small-valued control inputs, consequently, a slow state response and a high steady-state level. This experiment uses mostly the same system parameters as previous experiments, i.e., $a = 0.5$, $T = 10$, $R_{tot} = 30$, $\epsilon = 0.001$, $\sigma_{x_0}^2 = 10$, $\sigma_v^2 = 0.1$, and the time-varying uniform quantizer. In Figure 5.9, the rate allocations, optimized according to Theorem 5.5.7, are demonstrated for several ρ-values. Based on our analysis, the global minimum to the rate unconstrained problem, according to Lemma 5.5.13, is $R_t^\star = R^\star = 5$, $\forall t$, irrespective of ρ. This is consistent with Figure 5.9(a)–(d) that there is no R_t larger than 5. When ρ is small, for example $\rho = 0.1$, large-valued control inputs are allowed and the steady state is

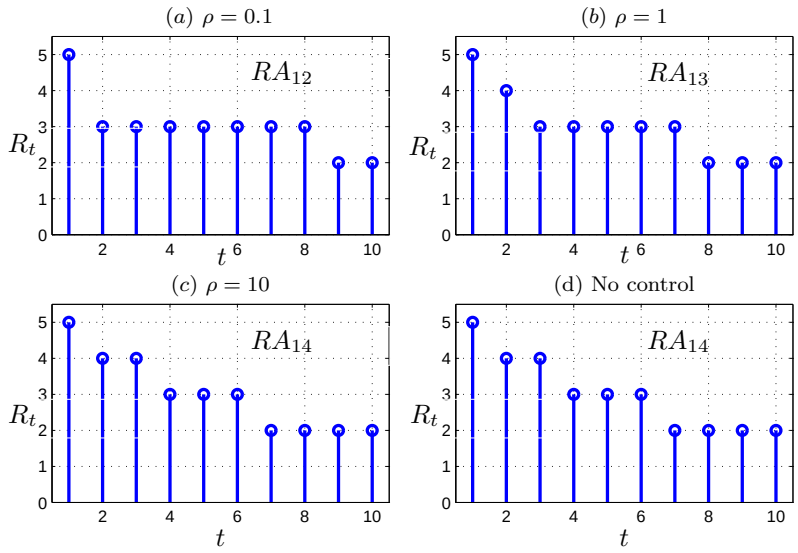


Figure 5.9: Optimized rate allocations with respect to ρ.

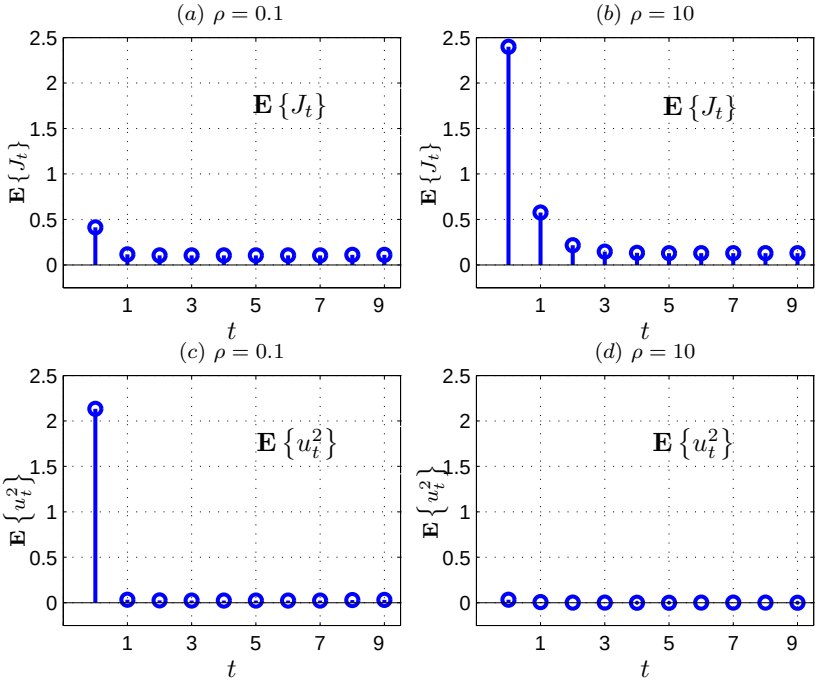


Figure 5.10: A performance comparison between ρ=0.1 and ρ=10.

quickly reached. As ρ increases, only small-valued control inputs are allowed and it takes longer time for the system to reach the steady state. This explains the results in Figure 5.9(b)–(c) that more bits are needed in the initial-states when ρ is large. The rate allocation in Figure 5.9(d) is obtained by solving Problem 5.4.3 for state estimation, where Theorem 5.4.4 is used together with the binary rounding algorithm. Interestingly, the optimized rate allocation in this case is the same as when $\rho = 10$, cf., Figure 5.9(c). It is a reasonable observation, since, when $\rho = 10$, first, the control inputs are extremely small and have hardly impact on the state evolution; second, π_t is nearly constant over time. As a result, Problem 5.5.6 for state control becomes almost identical to Problem 5.4.3 for state estimation. The simulated instantaneous costs and the control inputs for $\rho = 0.1$ and $\rho = 10$ are depicted in Figure 5.10. As expected, the instantaneous cost for $\rho = 10$ is remarkably higher than the instantaneous cost for $\rho = 0.1$. Moreover, when $\rho = 10$, the system performs similarly to the case without any control.

In Figure 5.11, the impact of the parameter ϵ is studied, by fixing the other parameters and only varying ϵ. In particular, the system parameters are: $T = 10$, $R_{tot} = 30$, $\rho = 1$, $\sigma_{x_0}^2 = 10$ and $\sigma_v^2 = 0.1$. Applying Theorem 5.5.7, the optimized rate allocation for $\epsilon = 0.001$ and $\epsilon = 0.1$ are RA_{13} and RA_7, respectively. At $\epsilon = 0.001$, the global minimum to the unconstrained problem, according to Lemma 5.5.13, is $R_t^\star = 5$, $\forall t$, which means the rate constraint is violated at the global minimum. On the other hand, at $\epsilon = 0.1$, the global minimum is $R_t = 2$, $\forall t$, so that the rate constraint is fulfilled. Recall that the quantizer range is fixed irrespective of the rate, a reduction in the rate leads to larger quantization errors, but it might result in more robust codewords against transmission errors. As a matter of fact, beyond a certain rate, the harm caused by transmission errors is much more serious than the reduction of quantization error along with an additional bit. In other words, the "additional bits" will do harm than good, as demonstrated in Figure 5.11. In the figure, the simulated instantaneous costs and the control inputs obtained by using RA_{13} and RA_7 at $\epsilon = 0.001$ and $\epsilon = 0.1$, are depicted. At $\epsilon = 0.001$, for $R_t < R^\star = 5$, we can always improve the performance by increasing R_t, and therefore, RA_{13} outperforms RA_7. At $\epsilon = 0.1$, the situation is different. When $R_t > 2$, the performance is degraded by increasing the rate, which is consistent to the simulation result in Figure 5.11.

In Section 5.2, we have posed Problem 5.2.4 to assign totally R_{tot} bits optimally to T time units. In fact, the solutions to the optimal rate allocation problem have partly answered the question the other way round. That is to say how much data it is truly needed to achieve a certain system performance, and the limitation of the system's performance is exposed. In the absence of channel errors, increasing the data rate typically means more accurate information, consequently, a better control performance. This is true even though the encoder–decoder is not optimal to the system. Unfortunately, in the presence of channel errors, the situation is complicated. The channel error has several negative impacts on the system performance. First of all, if the encoder–controller is not optimized, increasing the data rate does not necessarily improve the system performance. We can easily find examples in

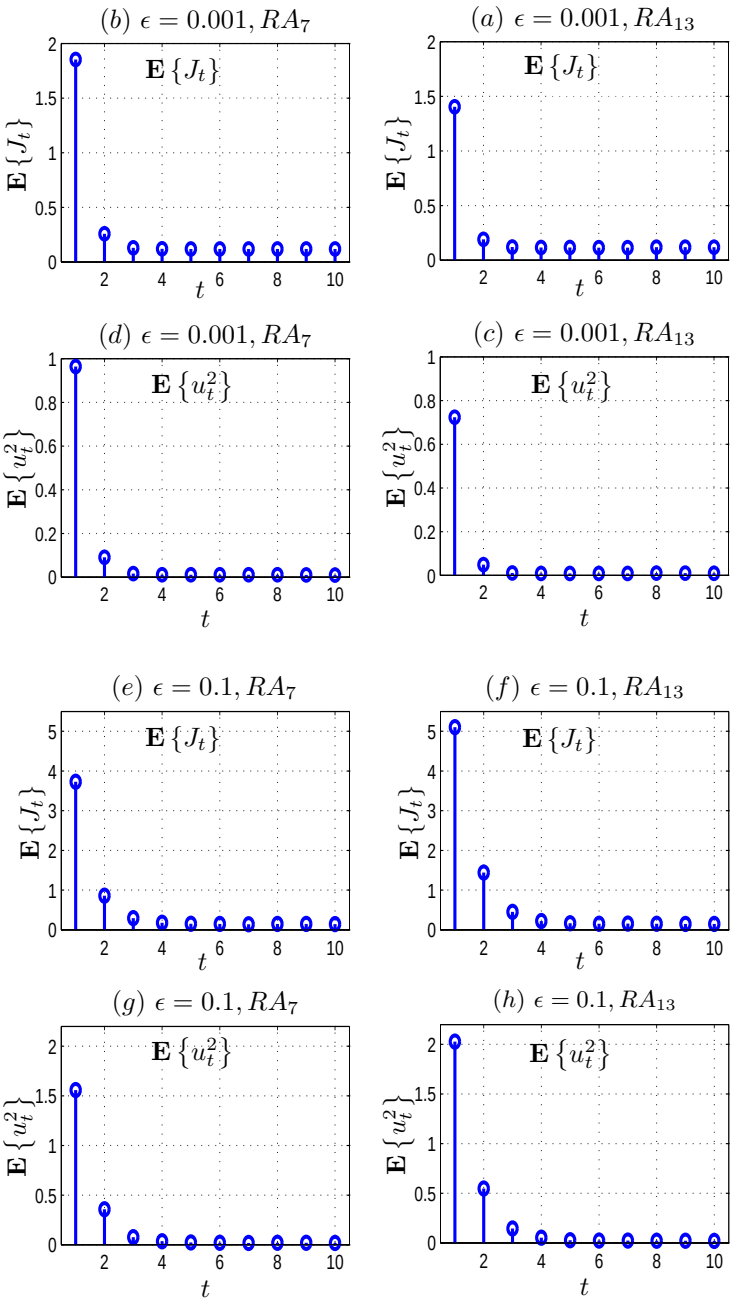


Figure 5.11: A performance comparison with respect to ϵ.

which enhancing the data rate does more harm than good. Moreover, the improvement given by the rate is significantly reduced if the crossover probability of the channel is high. It is worth noticing that the solution to the unconstraint problem can be considered as a measure of the quality of the quantizer. In particular, for optimal quantizers, the solution should be at $R=\infty$. In other words, by optimizing the encoder–decoder pair, the solution of Lemma 5.5.13 is moved towards ∞ to enhance the efficiency of the available communication resources.

As discussed in Section 5.8, performance degradation is expected due to the various simplifications and approximations. Experiments were pursued to compare the proposed rate allocation scheme with more than 100 other allocations which fulfill the rate constraint, using the same system parameters as for Figure 5.9. We observed that the optimized allocation is among the best allocations and the performance is satisfactory. In Figure 5.12, a comparison of the pdf's of the estimated x_t and the true x_t is depicted, for x_1 and x_2. This experiment used as well the same system parameters as Figure 5.9, i.e., $a=0.5$, $T=10$, $R_{tot}=30$, $\epsilon=0.001$, $\sigma_{x_0}^2=10$, $\sigma_v^2=0.1$, and the time-varying uniform quantizer. The comparison is carried out for three ρ values: $\rho=0.1$, $\rho=1$ and $\rho=10$. As explained previously, for large-valued ρ, the influence of control is moderate. Therefore the Gaussian assumption is more correct. On the other hand, for small ρ, the influence of control is significant, which has reduced the Gaussian assumption of the state x_t.

Finally, we mention here that the results in this chapter are also useful when formulating new interesting rate allocation problems with diverse communication constraints. Problem 5.2.4 is formulated that the cost of communication is implicitly considered by the number of total bits R_{tot}. An interesting variation could be to take into account the number of communications into the objective function. Below we show one simple example that the problem is formulated for periodic control.

Example 5.9.1. ***Periodic Control***

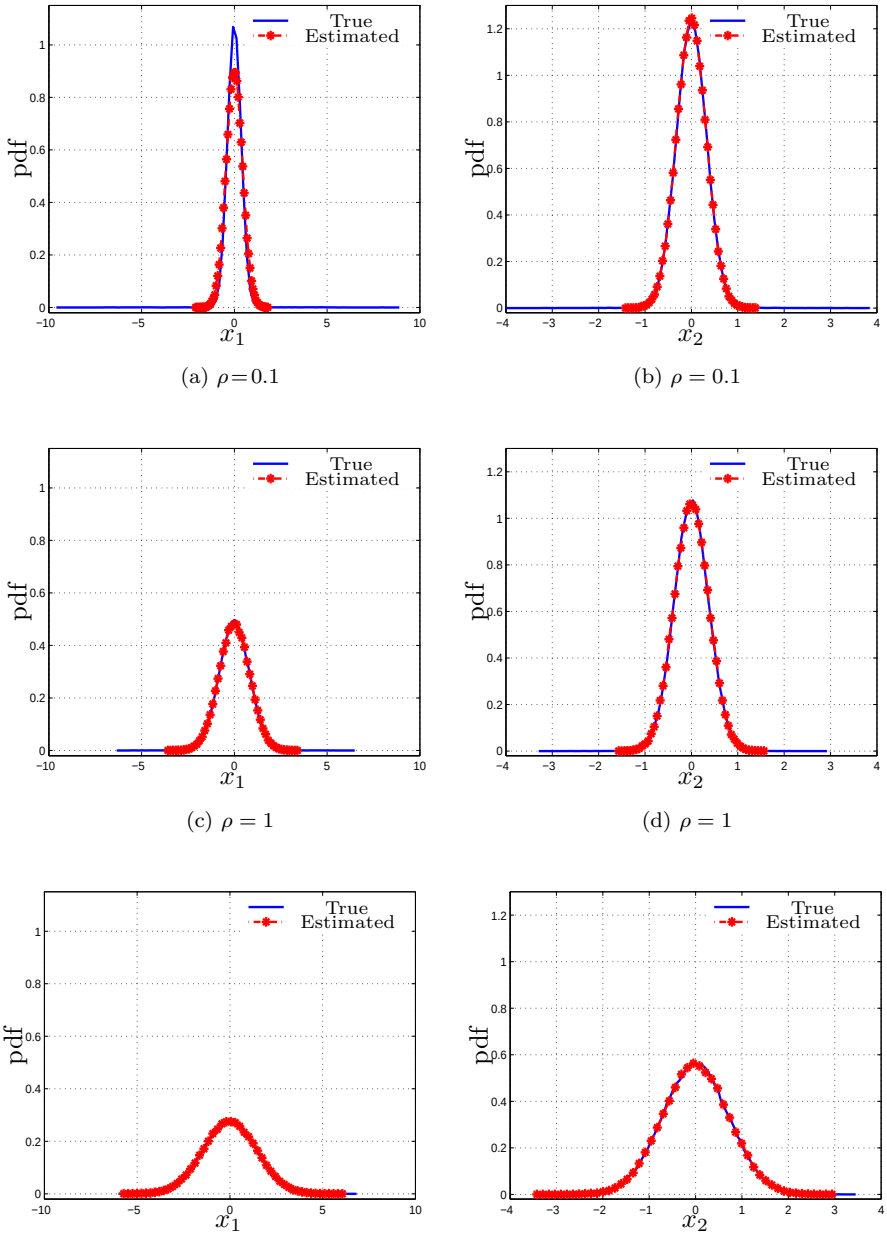
Consider a linear plant as described in Section 5.2.1 and the following periodic control law. We reduce the communications between the sensor and controller by not transmitting the state measurement as frequently as every time unit. Let t_s specify the time duration (or the number of time units) between two transmissions that the sensor at each t either transmits the measurement using R bits, or stays silent,

$$R_t = \begin{cases} R, & t = k(t_s+1), \\ 0, & t \neq k(t_s+1), \quad k \in \mathbb{Z}^+. \end{cases}$$

The number of transmissions N_{tr} is related to the horizon T and the samplings interval t_s as

$$N_{tr} = \begin{cases} \left\lfloor \frac{T}{t_s+1} \right\rfloor + 1, & T \geq t_s+1, \\ 1, & T < t_s+1. \end{cases}$$

The time-invariant instantaneous rate R is related to the total bit-rate R_{tot} and the



(a) $\rho = 0.1$ (b) $\rho = 0.1$

(c) $\rho = 1$ (d) $\rho = 1$

(e) $\rho = 10$ (f) $\rho = 10$

Figure 5.12: The pdf's of the estimated x_t and the true x_t.

number of transmissions as

$$R = \left\lfloor \frac{R_{tot}}{N_{tr}} \right\rfloor . \tag{5.74}$$

The total rate constraint $\sum_{t=0}^{T-1} R_t \leq R_{tot}$ *is always fulfilled given* (5.74)*. At the controller, a hybrid control law is adopted*

$$u_t = \begin{cases} \ell_t d_t, & t = k(t_s+1), \\ 0, & t \neq k(t_s+1), \quad k \in \mathbb{Z}^+, \end{cases} \tag{5.75}$$

i.e., only when receiving a measurement, $u_t = \ell_t d_t$*, with* ℓ_t *as given in* (5.7)*; while* $u_t = 0$ *otherwise. In short, the optimization problem is that given a linear plant* (5.1)*, a discrete memoryless channel, a memoryless encoder–decoder* (5.2)–(5.3)*, a control law* (5.75)*, and a total bit-rate budget* R_{tot}*, find the sampling rate* t_s *and the channel rate* R *that minimize an objective function that involves* x_t*,* u_t *and* N_{tr}*. Even though a solution based on testing different* t_s *is straightforward, while how to formulate a useful cost function that takes into account the communication cost is nontrivial.*

5.10 Summary

In this chapter, we formulated a rate allocation problem to assign totally R_{tot} bits optimally to T time units for control over noisy channels. First, we approximated the overall distortion function by means of high-rate quantization theory. Second, we showed that the unconstrained optimization problem has a global minimum, which solves the rate allocation problem if such a global minimum does not violate the rate constraint. On the other hand, if the global minimum violates the rate constraint, we solved the rate constrained optimization problem by means of Lagrangian duality for non-convex non-linear problems. Finally, numerical simulations showed good performance of the proposed rate allocation scheme. In the presence of the channel errors, more bits can sometimes do more harm than good, the encoder–controller mapping is therefore instrumental to achieve satisfactory overall performance when the communication resources are limited. How to optimize the rate allocation and the encoder–controller mappings jointly is a challenging problem for future research.

5.A High-Rate Approximation of MSE

Here we briefly review some useful results on high-rate approximation of the mean square error distortion. More detail about this topic is referred to e.g., [ZM94, MR06]. For brevity, we drop the time index throughout this section. As depicted in Figure 5.2, let the source signal x, with a pdf $p(x)$, be transmitted to the destination via a BSC with a crossover probability ϵ. Each transmission randomly selects an IA and revealed to both the encoder and decoder. At the encoder, the coded index i is produced based on x, and at the decoder, the channel output j will be mapped to one of 2^R values in the codebook $\{d(0), \ldots, d(2^R-1)\}$, with R referred to as the

rate. Let $d(i)$ be taken to denote the reconstruction value chosen by the encoder, and $d(j)$ is the true decoded value. Using randomized IA, the symbol transition probability function is given by (5.8), i.e.,

$$\mathcal{P}(j|\,i) = \begin{cases} \alpha(R), & j \neq i, \\ 1-(2^R-1)\alpha(R), & j = i, \end{cases} \qquad \alpha(R) \triangleq \frac{1-(1-\epsilon)^R}{2^R-1}.$$

As shown in (5.9), the mean squared estimation error of x with respect to d is

$$\mathbf{E}\left\{(x-d)^2\right\} = \sum_{l=0}^{2^R-1} \int_x \sum_{k=0}^{2^R-1} \mathcal{P}(\,j=l|\,i=k)(x-d(l))^2 p(x)\mathrm{d}x.$$

First, the source pdf at high-rate is approximately constant over one quantization cell, i.e., $p(x) \approx p(x=d(k))$, for all $x \in \mathcal{S}(k)$, where $\mathcal{S}(k) \triangleq \{x : i=k\}$ is the k^{th} quantization cell. Hence, we can approximate the MSE as

$$\mathbf{E}\left\{(x-d)^2\right\} \approx \sum_{k=0}^{2^R-1} p(x=d(k)) \sum_{l=0}^{2^R-1} \mathcal{P}(\,j=l|\,i=k) \int_x (x-d(l))^2 \mathrm{d}x. \tag{5.76}$$

Second, since the error $e=x-d(i)$ is typically a small number at high-rate, with $d(i)$ denoting the reconstruction chosen by the encoder, a Taylor expansion of $(x-d(j))^2$ gives

$$\begin{aligned}(x-d(j))^2 &= (d(i)+e-d(j))^2 \\ &= (d(i)-d(j))^2 + \mathrm{d}^1(d(i),d(j))e + \frac{1}{2}\mathrm{d}^2(d(i),d(j))e^2 + \mathcal{O}(|e|^3),\end{aligned} \tag{5.77}$$

where $\mathrm{d}^1(d(i),d(j))$ and $\mathrm{d}^2(d(i),d(j))$ are defined as

$$\begin{aligned}\mathrm{d}^1(d(i),d(j)) &\triangleq \left.\frac{\partial(x-d(j))^2}{\partial x}\right|_{x=d(i)}, \\ \mathrm{d}^2(d(i),d(j)) &\triangleq \left.\frac{\partial^2(x-d(j))^2}{\partial x^2}\right|_{x=d(i)}.\end{aligned} \tag{5.78}$$

Substituting (5.77)–(5.78) into (5.76) and neglecting $\mathcal{O}(|e|^3)$ terms, it yields

$$\begin{aligned}\mathbf{E}\left\{(x-d)^2\right\} \approx \sum_{k=0}^{2^R-1} p(d(k)) \sum_{l=0}^{2^R-1} \mathcal{P}(\,j=l|\,i=k) \\ \times \int_e \left((d(k)-d(l))^2 + \mathrm{d}^1(d(k),d(l))e + \frac{1}{2}\mathrm{d}^2(d(k),d(l))e^2\right)\mathrm{d}e.\end{aligned} \tag{5.79}$$

Third, the point density at high-rate is $\lambda(x)=1/(2^R V(\mathcal{S}(k)))$, where $V(\mathcal{S}(k))$ denotes the volume of $\mathcal{S}(k)$. Based on the above discussions, [MR06] showed that (5.79) can be approximated by the following expression,

$$\mathbf{E}\left\{(x-d)^2\right\} \approx \int_x J_x p(x)\mathrm{d}x, \quad J_x \triangleq 2^R\alpha(R)\int_y (x-y)^2\lambda(y)\mathrm{d}y + 2^{-2R}\bar{G}\lambda^{-2}(x).$$

The authors showed also that for a zero-mean source signal and a quantizer of 2^R cells, we can approximate the MSE as

$$\mathbf{E}\left\{(x-d)^2\right\} \approx 2^R\alpha(R)\sigma_x^2 + 2^R\alpha(R)\int_y y^2\lambda(y)\mathrm{d}y + \bar{G}2^{-2R}\int_x \lambda^{-2}(x)p(x)\mathrm{d}x, \tag{5.80}$$

cf., (5.10). For the detail about the derivations is referred to [MR06]. In this chapter, we simplify (5.80) further by using $2^R\alpha(R) \approx 1-(1-\epsilon)^R$. Following (5.11), define $\hat{J}(\beta,\kappa,\epsilon)$ as

$$\hat{J}(\beta,\kappa,\epsilon) \triangleq \beta(1-(1-\epsilon)^R) + \kappa 2^{-2R}, \tag{5.81}$$

where, β and κ are given by

$$\begin{aligned}
\beta &\triangleq \sigma_x^2 + \int_y y^2\lambda(y)\mathrm{d}y,\\
\kappa &\triangleq \bar{G}\int_x \lambda^{-2}(x)p(x)\mathrm{d}x, \quad \bar{G} \triangleq \frac{G^{-2}}{3}.
\end{aligned}$$

5.B High-rate Approximation for the Gaussian Case

Consider a zero-mean Gaussian source and a source-optimized encoder, we approximate $\hat{J}(\beta,\kappa,\epsilon)$ according to (5.81),

$$\hat{J}_t(\beta,\kappa,R) = (1-(1-\epsilon)^R)\sigma_x^2 + \underbrace{(1-(1-\epsilon)^R)\int_y y^2\lambda(y)\mathrm{d}y}_{\mathrm{I}} + \underbrace{\bar{G}2^{-2R}\int_x \lambda^{-2}(x)p(x)\mathrm{d}x}_{\mathrm{II}}. \tag{5.82}$$

Recall, the point density function for a scala source-optimized quantizer,

$$\lambda(x) = \frac{(p(x))^{\frac{1}{3}}}{\int_{-\infty}^{\infty}(p(x))^{\frac{1}{3}}\mathrm{d}x}.$$

Let us first calculate $\int_{-\infty}^{\infty} y^2\lambda(y)\mathrm{d}y$ in part I of (5.82),

$$\begin{aligned}\int_{-\infty}^{\infty} y^2\lambda(y)\mathrm{d}y &= \frac{\displaystyle\int_{-\infty}^{\infty} y^2\left(\frac{1}{\sqrt{2\pi\sigma_x^2}}e^{-\frac{y^2}{2\sigma_x^2}}\right)^{\frac{1}{3}}\mathrm{d}y}{\displaystyle\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi\sigma_x^2}}e^{-\frac{y^2}{2\sigma_x^2}}\right)^{\frac{1}{3}}\mathrm{d}y} \\ &= \frac{(6\pi\sigma_x^2)^{\frac{1}{2}}}{(2\pi\sigma_x^2)^{\frac{1}{6}}\cdot 2\cdot\frac{1}{6\sigma_x^2}}\cdot\frac{(2\pi\sigma_x^2)^{\frac{1}{6}}}{(6\pi\sigma_x^2)^{\frac{1}{2}}} = 3\sigma_x^2. \qquad (5.83)\end{aligned}$$

We have used the relation

$$\int_{-\infty}^{\infty} x^2 e^{-ax^2}\mathrm{d}x = 2\int_0^{\infty} x^2 e^{-ax^2}\mathrm{d}x = \frac{1}{2a}\sqrt{\frac{\pi}{a}}.$$

Let us move on to part II of (5.82),

$$\begin{aligned}\bar{G}\int_x \lambda^{-2}(x)p(x)\mathrm{d}x &= \frac{1}{12}\left(\int_{-\infty}^{\infty}(p(x))^{\frac{1}{3}}\mathrm{d}x\right)^3 \\ &= \frac{1}{12}\left(\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\pi\sigma_x^2}}e^{-\frac{x^2}{2\sigma_x^2}}\right)^{\frac{1}{3}}\mathrm{d}x\right)^3 \qquad (5.84) \\ &= \frac{3^{\frac{1}{2}}\cdot 3\cdot 2\pi\sigma_x^2}{12} = \underbrace{\frac{\sqrt{3}\pi}{2}}_{\mu}\sigma_x^2.\end{aligned}$$

Combining (5.83) and (5.84), the distortion $\hat{J}_t$ is given by

$$\hat{J}_t = 4\sigma_x^2(1-(1-\epsilon)^R) + \mu\sigma_x^2 2^{-2R}.$$

5.C Proof of Lemma 5.5.13

Proof. According to Lemma 5.5.11, $\mathfrak{J}_t(\mathbf{R}_0^t)$ can be written as

$$\mathfrak{J}_t(\mathbf{R}_0^t) = \sum_{b_0=0}^{1}\cdots\sum_{b_{t-1}=0}^{1} W(b_0,\ldots,b_{t-1})\left(\prod_{s=0}^{t-1}\left(\tilde{J}_s(\tilde{\beta}_s,\tilde{\kappa}_s,R_s)\right)^{b_s}\right)\tilde{J}_t(\tilde{\beta}_t,\tilde{\kappa}_t,R_t).$$

The coefficient $W(b_0,\ldots,b_{t-1})$, independent of $\mathbf{R}_0^{t-1}$, is defined as

$$W(b_0,\ldots,b_{t-1}) \triangleq \pi_t \bar{B}\left(\prod_{s=\bar{s}+1}^{t-1}\bar{B}_s\right),$$

where $\mathbf{b}_0^{t-1}$, π_t, $\bar{B}$, $\bar{B}_s$ and $\bar{s}$ are as specified in Theorem 5.5.7. Taking the first order derivative of $\mathfrak{J}_t(\mathbf{R}_0^t)$ with respect to R_k, it gives

$$\begin{aligned}
\frac{\partial}{\partial R_k}\mathfrak{J}_t(\mathbf{R}_0^t) &= \frac{\partial \tilde{J}_k}{\partial R_k}(\tilde{\beta}_k,\tilde{\kappa}_k,R_k)\left(\sum_{b_0=0}^{1}\cdots\sum_{b_k=1}^{1}\cdots\sum_{b_{t-1}=0}^{1}W(b_0,\ldots,b_{t-1})\right.\\
&\qquad\left.\times\prod_{s=0,s\neq k}^{t-1}\left(\tilde{J}_s(\tilde{\beta}_s,\tilde{\kappa}_s,R_s)\right)^{b_s}\tilde{J}_t(\tilde{\beta}_t,\tilde{\kappa}_t,R_t)\right),\quad k<t,\\
\frac{\partial}{\partial R_t}\mathfrak{J}_t(\mathbf{R}_0^t) &= \frac{\partial \tilde{J}_t}{\partial R_t}(\tilde{\beta}_t,\tilde{\kappa}_t,R_t)\left(\sum_{b_0=0}^{1}\cdots\sum_{b_k=1}^{1}\cdots\sum_{b_{t-1}=0}^{1}W(b_0,\ldots,b_{t-1})\right.\\
&\qquad\left.\times\prod_{s=0}^{t-1}\left(\tilde{J}_s(\tilde{\beta}_s,\tilde{\kappa}_s,R_s)\right)^{b_s}\right),
\end{aligned}\tag{5.85}$$

where the sum-terms are positive. Note that $\partial\mathfrak{J}_t(\mathbf{R}_0^t)/\partial R_k=0$, for $k>t$. Since (5.85) applies for all t, it follows that at the critical point

$$\frac{\partial}{\partial R_k}\sum_{t=0}^{T-1}\mathfrak{J}_t(\mathbf{R}_0^t)=0 \;\Leftrightarrow\; \frac{\partial \tilde{J}_k}{\partial R_k}(\tilde{\beta}_k,\tilde{\kappa}_k,R_k^\star)=0.$$

Computing the second order derivatives, implies

$$\begin{aligned}
\frac{\partial^2\mathfrak{J}_t(\mathbf{R}_0^t)}{\partial R_k^2} &= \frac{\partial^2 \tilde{J}_k}{\partial R_k^2}(\tilde{\beta}_k,\tilde{\kappa}_k,R_k)\left(\sum_{b_0=0}^{1}\cdots\sum_{b_k=1}^{1}\cdots\sum_{b_{t-1}=0}^{1}W(b_0,\ldots,b_{t-1})\right.\\
&\qquad\left.\times\left(\prod_{s=0,s\neq k}^{t-1}\left(\tilde{J}_s(\tilde{\beta}_s,\tilde{\kappa}_s,R_s)\right)^{b_s}\right)\tilde{J}_t(\tilde{\beta}_t,\tilde{\kappa}_t,R_t)\right),\quad k<t,\\
\frac{\partial^2\mathfrak{J}_t(\mathbf{R}_0^t)}{\partial R_k\partial R_l} &= \frac{\partial \tilde{J}_k}{\partial R_k}(\tilde{\beta}_k,\tilde{\kappa}_k,R_k)\frac{\partial \tilde{J}_l}{\partial R_l}(\tilde{\beta}_l,\tilde{\kappa}_l,R_l)\left(\sum_{b_0=0}^{1}\cdots\sum_{b_k=1}^{1}\cdots\sum_{b_{t-1}=0}^{1}W(b_0,\ldots,b_{t-1})\right.\\
&\qquad\left.\times\left(\prod_{s=0,s\neq k,l}^{t-1}\left(\tilde{J}_s(\tilde{\beta}_s,\tilde{\kappa}_s,R_s)\right)^{b_s}\right)\tilde{J}_t(\tilde{\beta}_t,\tilde{\kappa}_t,R_t)\right),\quad k,l<t,\;\; k\neq l,\\
\frac{\partial^2\mathfrak{J}_t(\mathbf{R}_0^t)}{\partial R_t^2} &= \frac{\partial^2 \tilde{J}_t}{\partial R_t^2}(\tilde{\beta}_t,\tilde{\kappa}_t,R_t)\left(\sum_{b_0=0}^{1}\cdots\sum_{b_k=1}^{1}\cdots\sum_{b_{t-1}=0}^{1}W(b_0,\ldots,b_{t-1})\right.
\end{aligned}$$

$$\times \left(\prod_{s=0}^{t-1} \left(\tilde{J}_s(\tilde{\beta}_s, \tilde{\kappa}_s, R_s) \right)^{b_s} \right) \Bigg),$$

$$\frac{\partial^2 \mathfrak{J}_t(\mathbf{R}_0^t)}{\partial R_t \partial R_l} = \frac{\partial \tilde{J}_t}{\partial R_t}(\tilde{\beta}_t, \tilde{\kappa}_t, R_t) \frac{\partial \tilde{J}_l}{\partial R_l}(\tilde{\beta}_l, \tilde{\kappa}_l, R_l) \left(\sum_{b_0=0}^{1} \cdots \sum_{b_k=1}^{1} \cdots \sum_{b_{t-1}=0}^{1} W(b_0, \ldots, b_{t-1}) \right.$$
$$\left. \times \left(\prod_{s=0, s\neq l}^{t-1} \left(\tilde{J}_s(\tilde{\beta}_s, \tilde{\kappa}_s, R_s) \right)^{b_s} \right) \right), \quad l < t.$$

Note that when k or $l > t$, the second order derivative is 0. At the critical point, $\partial \tilde{J}_k / \partial R_k = 0$ and $\partial^2 \tilde{J}_k / \partial R_k^2 > 0$, hence, all elements at the diagonal of the Hessian matrix are positive. We can therefore conclude that the Hessian matrix,

$$\mathrm{H}(\mathbf{R}_0^{\star T-1}) = \begin{bmatrix} \frac{\partial^2}{\partial R_0^2} \sum_{t=0}^{T-1} \mathfrak{J}_t & 0 & \cdots & 0 \\ 0 & \frac{\partial^2}{\partial R_1^2} \sum_{t=0}^{T-1} \mathfrak{J}_t & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2}{\partial R_{T-1}^2} \sum_{t=0}^{T-1} \mathfrak{J}_t \end{bmatrix},$$

is positive definite and the critical point is a global minimum. □

Chapter 6

Coding–Control for the Gaussian Channel

6.1 Introduction

In Chapter 3, an iterative method is developed to optimize encoder–controller mappings for feedback control over finite-input finite-output channels. Here, by a finite-input channel we mean that the channel input has a finite alphabet. This terminology will also be used for the channel output. In the present chapter, we extend the system model in Chapter 2 to include infinite-output channels. More precisely, we investigate how the generalization to infinite-output channels affects the optimization of the encoder–controller mappings, theoretically and practically, by studying one special type of infinite-output channels, namely, the binary Gaussian channel (BGC). In communication engineering, the BGC is a common channel model for situations where digital signals are corrupted by additive Gaussian noise. Beside the applicability, we choose to study the BGC also because it can provide structural and functional insights of the solution, by means of instructive and relatively simple calculations.

The rest of this chapter is organized as follows. In Section 6.2 the model of the extended system is described and modifications due to the generalization to infinite-output channels are specified. Section 6.3 is devoted mainly to controller design. We start by discussing the optimal controller which exploits all past channel outputs. However, due to the computational complexity and the memory demand, the above-mentioned controller is in general not practically implementable. In practice, certain approximations and simplifications have to be imposed. After exposing the recursive structure of the optimal controller, we propose a low-complexity controller which can take advantage of both the soft and hard information of channel outputs. Here in this chapter, we refer to a soft channel output as a real-valued channel output, while a hard channel output to an integer-valued channel output. Thereafter, a brief discussion of the encoder design is included. Finally, numerical simulations are carried out to demonstrate the performance of the various coding–control schemes presented in Section 6.3. It should be observed that the main difference among the various controllers studied in this chapter is the different levels of the past

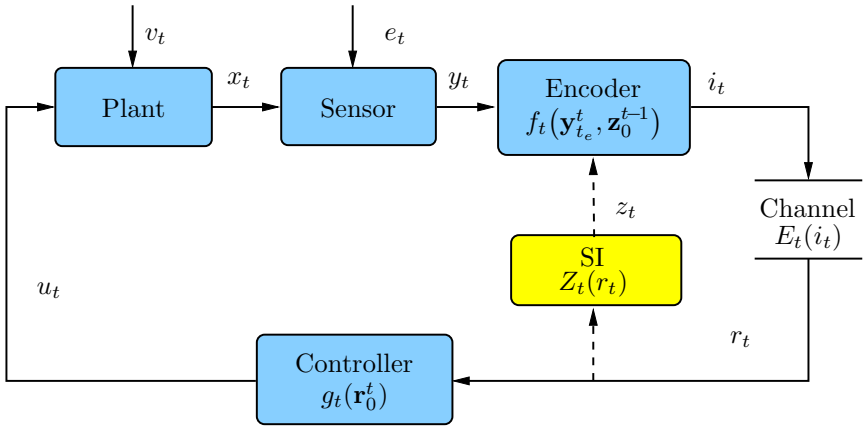


Figure 6.1: A general system for feedback control over an infinite-output channel. The dashed line indicates potential SI from the controller to the encoder.

information they can exploit.

6.2 Problem Formulation

In this section, we introduce an extended version of the control system described in Chapter 2 by considering infinite-output channels. Here, infinite-output channels are a special case of finite-output channels where we allow the output alphabet to be of infinite, or even uncountable, size. Therefore, we will mostly discuss the modifications brought by the extension, taking the system from Chapter 2 as a reference. In the meantime, system components which are not affected by the extension retain the same form, and the reader is referred to Chapter 2 for their detailed descriptions.

In the most general case, we consider a control system with a communication channel as depicted in Figure 6.1. The multi-variable linear plant is governed by the same equation system as (2.1), i.e.,

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + v_t, \\ y_t &= Cx_t + e_t, \end{aligned}$$

where $x_t \in \mathbb{R}^{\mathtt{n}}, u_t \in \mathbb{R}^{\mathtt{m}}, y_t \in \mathbb{R}^{\mathtt{p}}$, are the state, the control, and the measurement, respectively. The matrices $A \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$, $B \in \mathbb{R}^{\mathtt{n}\times\mathtt{m}}$, $C \in \mathbb{R}^{\mathtt{p}\times\mathtt{n}}$, are known, where (A, C) is state observable and (A, B) is state controllable. The process noise $v_t \in \mathbb{R}^{\mathtt{n}}$ and the measurement noise $e_t \in \mathbb{R}^{\mathtt{p}}$ are modeled as i.i.d. zero-mean Gaussian processes. They are mutually independent of the initial-state $x_0 \in \mathbb{R}^{\mathtt{n}}$, which is an i.i.d. zero-mean Gaussian vector.

The memory-based time-varying encoder takes the previous measurements and potential SI as input. In particular, the mapping is described by the function,

$$i_t = f_t\big(\mathbf{y}_{t_e}^{t}, \mathbf{z}_0^{t-1}\big)\,, \quad t_e = t - M_e,\ \ M_e \in \{0,\ldots,t\}, \tag{6.1}$$

where $i_t \in \mathcal{L}_I = \{0,\ldots,L_I-1\}$, $L_I \in \mathbb{N}$, is the integer index, and M_e specifies the encoder memory of past measurements, cf., (2.2). In this chapter, we define the side-information z_t at the encoder to be

$$z_t = Z_t(r_t)\,,$$

where Z_t is a deterministic and memoryless function, and r_t is the soft channel output which will be specified later. Accordingly, $z_t = r_t$, if full SI is available; while, $z_t = 0$, if there is no SI at the encoder. Between the extremes, there are a variety of cases with incomplete SI. Let the memoryless channel have the input variable i_t and output r_t, defined by

$$r_t = E_t(i_t)\,, \tag{6.2}$$

where E_t is a random memoryless mapping. Note that, conditioned on the index i_t, the mapping to r_t is independent of all past events. The channel output (6.2), $r_t \in \mathbb{R}^{R_I}$, is a real-valued vector which differs from the integer output (2.3) from Chapter 2. As commonly used in literature, we refer to a *soft channel output* as a real-valued channel output, and a *hard channel output* to an integer-valued channel output. Consequently, we refer to a *soft controller* (or a soft-information-based controller) as a controller which exploits only soft channel outputs. Likewise, we refer to a *hard controller* (or a hard-information-based controller) as a controller which exploits only hard channel outputs. As an example of (6.2), throughout this chapter we consider the special case of the binary Gaussian channel. As mentioned in the beginning of this chapter, the BGC is also chosen in the interest of illustrating the basic principles using insightful and simple calculations.

Next, we describe the operation of the BGC in this chapter by first specifying the channel input. At time t, based on $\mathbf{z}_0^{t-1}$, the history of the past SI, the coded index $i_t \in \mathcal{L}_I$ is mapped into b_t, a binary codeword of R_I bits. More specifically,

$$b_t(i_t, \mathbf{z}_0^{t-1}) = \Big[\ b_t^{[1]}\big(i_t, \mathbf{z}_0^{t-1}\big) \quad b_t^{[2]}\big(i_t, \mathbf{z}_0^{t-1}\big) \quad \ldots \quad b_t^{[R_I]}\big(i_t, \mathbf{z}_0^{t-1}\big)\ \Big]\,, \tag{6.3}$$

where, $b_t^{[k]}\big(i_t, \mathbf{z}_0^{t-1}\big)$, $k \in \{1,\ldots,R_I\}$, represents a binary bit. Here, the notation $b_t^{[k]}\big(i_t, \mathbf{z}_0^{t-1}\big)$, or the like, works as follows. The subscript t indicates that the vector b is a *time-varying* entity. The superscript k in the square bracket indicates that this binary bit is the k^{th} element of the vector b. Finally, in the round bracket, we describe the dependence of $b_t^{[k]}$ on the current channel input i_t and SI, $\mathbf{z}_0^{t-1}$, i.e., $b_t^{[k]}(l, \mathbf{z}_0^{t-1}) = b_t^{[k]}(i_t = l, \mathbf{z}_0^{t-1})$. Terms in the round bracket will be left out if they are not relevant to the context.

The BGC produces a real-valued output vector $r_t \in \mathbb{R}^{R_I}$ such that the k^{th} element, $r_t^{[k]}$, $k \in \{1, \ldots, R_I\}$, is governed by

$$r_t^{[k]} = b_t^{[k]} + w_t^{[k]}. \tag{6.4}$$

The additive noise $w_t^{[k]}$ is i.i.d. zero-mean Gaussian, with a finite time-invariant variance σ_w^2. Clearly, the channel has a finite input alphabet, but an infinite output alphabet, $\mathcal{L}_I \neq \mathcal{L}_J$.

At the receiver side, we consider a controller that causally utilizes the entire history of past channel outputs $\mathbf{r}_0^t$ to produce the control command,

$$u_t = g_t\left(\mathbf{r}_0^t\right). \tag{6.5}$$

Note that the control u_t is completely determined by the past channel outputs $\mathbf{r}_0^t$. Next, we introduce Problem 6.2.1 which specifies the problem studied in this chapter.

Problem 6.2.1. *Consider the system* (2.1) *and the channel* (6.3)–(6.4)*, find the encoder–controller mappings* (6.1) *and* (6.5) *which minimize the expected value* $\mathbf{E}\{J_{tot}\}$*, where*

$$J_{tot} = \sum_{t=1}^{T} \left(x_t' V_t x_t + u_{t-1}' P_{t-1} u_{t-1}\right).$$

The matrices V_t and P_t are symmetric and positive definite.

Throughout this chapter, we denote the conditional mean estimate of the state x_s, based on the history of the received vectors $\mathbf{r}_0^t$,

$$\breve{x}_{s|t} \triangleq \mathbf{E}\left\{x_s | \mathbf{r}_0^t\right\}, \quad s \leq t.$$

In the following, we use $\breve{x}_t$ as a short notation for $\breve{x}_{t|t} = \mathbf{E}\{x_t | \mathbf{r}_0^t\}$.

6.3 Controller Design

In this section, we are concerned with the controller design. We first show the optimal controller derived based on the result from Chapter 3. Because of the complexity, the implementation of the optimal controller is impossible. Given this difficulty, several low-complexity controllers are presented, which are useful in practice.

Theoretically, the optimal encoder–controller from Chapter 3 has no special restriction on the size of the alphabets of the channel input and output. Therefore, the results from Chapter 3 apply straightforwardly to finite-input infinite-output channels. As a direct result of Proposition 3.2.4 from Chapter 3, we present the optimal controller in Proposition 6.3.1 below, for a fixed full SI open-loop encoder.

Proposition 6.3.1. *Consider an open-loop encoder system with a fixed open-loop encoder* $\overline{\mathbf{f}}_0^{T-1} = \{\bar{f}_t(\bar{\mathbf{y}}_{t_e}^t, \mathbf{r}_0^{t-1})\}_{t=0}^{T-1}$. *Given the plant* (2.1) *and the memoryless channel* (6.2)*, the controller* $u_t = g_t(\mathbf{r}_0^t)$ *that minimizes the LQ cost* (2.10) *is given by*

$$u_t = \ell_t \mathbf{E}\left\{ x_t | \, \mathbf{r}_0^t \right\}. \tag{6.6}$$

The linear control law ℓ_t *can be recursively computed according to* (3.6).

The proof of Proposition 6.3.1 follows from Proposition 3.2.4. Unfortunately, this optimal controller (6.6) is difficult to compute in practice. In the first place, there is no closed-form expression for (6.6). In the second place, we will find ourselves confronted by a crucial dimensionality problem, if we attempt to compute (6.6) numerically. As a result, to perform an iterative training as proposed in Chapter 3, is even less possible. Moreover, owing to the requirement on SI feedback channels, full SI is obviously an unrealistic assumption in the context of BGC. Despite this fact, we will still study the full SI scenario for the following reasons. First, by having access to full SI, the encoder can perfectly deduce the past controls and the encoder complexity is significantly reduced. Second, based on the results from Chapter 3 we know that locally optimal solutions can be obtained for certain full SI scenarios. Finally, full SI achieves the best performance compared with all levels of partial SI. Given the above considerations, we in the next step study the optimal control (6.6) for full SI cases. In particular, (6.6) is discussed in terms of a recursive structure, which is very useful to the design of practical controllers. As a matter of fact, although the recursive calculation is not practically implementable, it will guide us to design practical controllers with realistic information patterns. Thereafter, we will in Section 6.3 propose several practical encoder–controller designs which exploit different levels of channel output information.

6.3.1 Soft-Information-Based Controller

In contrast to a finite-output channel, to implement the optimal control (6.6) as a look-up table is certainly no longer practically possible, even for a small T. By observing that (6.6) is an estimator-based controller, the Hadamard-based decoding techniques can be used. The Hadamard transform has been shown to be very useful in representing functions that map integers to real values. Concerning decoding real-valued channel outputs, this method elegantly exposes the connection between the estimation of a source symbol and the estimation of individual bits. For a detailed description of the Hadamard-based soft decoding, we refer the reader to [Sko99a] and references therein.

Following [Sko99a], we briefly describe the implementation of the controller (6.6) in terms of Hadamard matrices. We start by elaborating $\mathbf{E}\{x_t | \, \mathbf{r}_0^t\}$, where $\mathbf{r}_0^t$ is a fixed sequence of channel outputs. Especially, we focus on the recursive structure of $\mathbf{E}\{x_t | \, \mathbf{r}_0^t\}$ which appears to be very useful to the implementation of efficient controllers. Based on the system model described in Section 6.2 and Bayes' rule,

we can write $\mathbf{E}\{x_t|\,\mathbf{r}_0^t\}$ as follows,

$$\begin{aligned}\mathbf{E}\{x_t|\,\mathbf{r}_0^t\} &= \sum_{l=0}^{L_I-1}\mathcal{P}(i_t=l|\,\mathbf{r}_0^t)\,\mathbf{E}\{x_t|\,i_t=l,\mathbf{r}_0^{t-1}\} \qquad (6.7)\\ &= \frac{\sum_{l=0}^{L_I-1}\mathcal{P}(i_t=l|\,\mathbf{r}_0^{t-1})\,p(r_t|\,i_t=l,\mathbf{r}_0^{t-1})\,\mathbf{E}\{x_t|\,i_t=l,\mathbf{r}_0^{t-1}\}}{\sum_{k=0}^{L_I-1}\mathcal{P}(i_t=k|\,\mathbf{r}_0^{t-1})\,p(r_t|\,i_t=k,\mathbf{r}_0^{t-1})}\\ &= \frac{\sum_{l=0}^{L_I-1}p(r_t|\,i_t=l)\,\mathcal{P}(i_t=l|\,\mathbf{r}_0^{t-1})\,\mathbf{E}\{x_t|\,i_t=l,\mathbf{r}_0^{t-1}\}}{\sum_{k=0}^{L_I-1}p(r_t|\,i_t=k)\,\mathcal{P}(i_t=k|\,\mathbf{r}_0^{t-1})}.\end{aligned}$$

Note that, because of the assumption of a memoryless channel (6.2), conditioned on the coded index i_t, the current channel output r_t is independent of the past channel outputs $\mathbf{r}_0^{t-1}$. As a matter of fact, the major challenge of (6.7) lies in the terms $\mathbf{E}\{x_t|\,i_t,\mathbf{r}_0^{t-1}\}$ and $\mathcal{P}(i_t|\,\mathbf{r}_0^{t-1})$, due to the lack of efficient methods to calculate or store them. On the other hand, the term $p(r_t|\,i_t)$ can be computed efficiently, because it is a Gaussian pdf. In what follows we briefly describe the recursive calculation of $p(x_t|\mathbf{r}_0^t)$ given $p(x_{t-1}|\mathbf{r}_0^{t-1})$.

Again consider (6.7). First, the pdf $p(r_t|\,i_t)$ is specified by the channel. Second, the pmf $\mathcal{P}(i_t|\,\mathbf{r}_0^{t-1})$ is specified by the encoder mapping f_t and $p(x_t|\,\mathbf{r}_0^{t-1})$, which is related to $p(x_{t-1}|\,\mathbf{r}_0^{t-1})$ as

$$p(x_t|\,\mathbf{r}_0^{t-1}) = p(ax_{t-1}+u_{t-1}+v_{t-1}|\,\mathbf{r}_0^{t-1}). \qquad (6.8)$$

The computation of (6.8) is straightforward, since u_{t-1} is deterministic given $\mathbf{r}_0^{t-1}$, and v_{t-1} is independent of x_{t-1} and u_{t-1}. Finally, $\mathbf{E}\{x_t|\,i_t,\mathbf{r}_0^{t-1}\}$ is computed by using $p\{x_t|\,i_t,\mathbf{r}_0^{t-1}\}$, which can be expressed as

$$p(x_t|\,i_t=l,\mathbf{r}_0^{t-1}) = \frac{p(x_t|\,\mathbf{r}_0^{t-1})\,\mathcal{P}(i_t=l|\,x_t,\mathbf{r}_0^{t-1})}{\int_{x_t}p(x_t|\,\mathbf{r}_0^{t-1})\,\mathcal{P}(i_t=l|\,x_t,\mathbf{r}_0^{t-1})\,\mathrm{d}x_t},$$

where the pdf $p(x_t|\,\mathbf{r}_0^{t-1})$ is related to $p(x_{t-1}|\,\mathbf{r}_0^{t-1})$ as shown in (6.8), and the pmf $\mathcal{P}(i_t=l|\,x_t,\mathbf{r}_0^{t-1})$ is specified by the encoder mapping f_t. Above, we have shown the recursive derivation of $p(x_t|\,\mathbf{r}_0^t)$ based on $p(x_{t-1}|\,\mathbf{r}_0^{t-1})$.

However, a closed-form expression of $p(x_{t-1}|\,\mathbf{r}_0^{t-1})$ is extremely hard to derive; and moreover, to store $p(x_{t-1}|\,\mathbf{r}_0^{t-1})$ is practically unrealistic because of the dimensionality problem.

At this moment let us ignore the dimensionality problem of $\mathbf{E}\{x_t|\,i_t,\mathbf{r}_0^{t-1}\}$ and $\mathcal{P}(i_t|\,\mathbf{r}_0^{t-1})$, and assume that the terms are available when they are needed. Following [Sko99a], we show another way to view the impact of the current channel output r_t by rewriting $\mathbf{E}\{x_t|\,\mathbf{r}_0^t\}$ in terms of Hadamard matrices. In the first place, we construct a matrix $\bar{C}_t(\mathbf{r}_0^{t-1})$, which carries a priori information about r_t, since the l^{th} column of $\bar{C}_t(\mathbf{r}_0^{t-1})$, denoted by $c_t(l,\mathbf{r}_0^{t-1})$, is the conditional centroid $\mathbf{E}\{x_t|\,i_t=l,\mathbf{r}_0^{t-1}\}$, i.e.,

$$c_t\left(l,\mathbf{r}_0^{t-1}\right) = \mathbf{E}\{x_t|\,i_t=l,\mathbf{r}_0^{t-1}\}, \quad l\in\mathcal{L}_I. \qquad (6.9)$$

The matrix $\bar{C}_t(\mathbf{r}_0^{t-1})$ can be written as a product of two matrices (cf., Section 1.4),

$$\bar{C}_t\big(\mathbf{r}_0^{t-1}\big) = \bar{T}_t\big(\mathbf{r}_0^{t-1}\big)\, H_t\big(\mathbf{r}_0^{t-1}\big)\,, \tag{6.10}$$

where the Hadamard matrix $H_t(\mathbf{r}_0^{t-1})$ has the l^{th} column, $h_t(l, \mathbf{r}_0^{t-1})$, $l \in \mathcal{L}_I$, formed by the binary codeword

$$\left[\; b_t^{[R_I]}\big(i_t = l, \mathbf{r}_0^{t-1}\big) \quad b_t^{[R_I-1]}\big(i_t = l, \mathbf{r}_0^{t-1}\big) \quad \ldots \quad b_t^{[1]}\big(i_t = l, \mathbf{r}_0^{t-1}\big) \;\right],$$

to which the coded index $i_t(\mathbf{r}_0^{t-1})$ is mapped, cf., (6.3). More precisely, the column $h_t(l, \mathbf{r}_0^{t-1})$ is computed as follows,

$$h_t\left(i_t = l, \mathbf{r}_0^{t-1}\right) = \begin{bmatrix} 1 \\ b_t^{[R_I]}\big(i_t = l, \mathbf{r}_0^{t-1}\big) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ b_t^{[1]}\big(i_t = l, \mathbf{r}_0^{t-1}\big) \end{bmatrix}, \tag{6.11}$$

where the symbol $\otimes$ denotes the Kronecker product. Back to (6.10), given $\bar{C}_t(\mathbf{r}_0^{t-1})$ and $H_t(\mathbf{r}_0^{t-1})$, we can compute the matrix $\bar{T}_t(\mathbf{r}_0^{t-1})$, which we refer it to as the *encoding matrix*, since it specifies the relation between the encoded index i_t and the centroid c_t. Note that each sequence $\mathbf{r}_0^{t-1}$ has an associated encoding matrix $\bar{T}_t(\mathbf{r}_0^{t-1})$. The encoding matrix $\bar{T}_t(\mathbf{r}_0^{t-1})$ is also instrumental to the decoding process, which can be realized by rewriting (6.7) as

$$\begin{aligned} \mathbf{E}\left\{ x_t \middle| \mathbf{r}_0^t \right\} &= \bar{T}_t\big(\mathbf{r}_0^{t-1}\big) \underbrace{\frac{\sum_{l=0}^{L_I-1} p(\, r_t |\, i_t = l)\, \mathcal{P}\big(i_t = l \big| \mathbf{r}_0^{t-1}\big)\, h_t\big(l, \mathbf{r}_0^{t-1}\big)}{\sum_{k=0}^{L_I-1} p(\, r_t |\, i_t = k)\, \mathcal{P}\big(i_t = k \big| \mathbf{r}_0^{t-1}\big)}}_{\hat{h}_t\left(\mathbf{r}_0^t\right)} \\ &= \bar{T}_t\left(\mathbf{r}_0^{t-1}\right) \hat{h}_t\big(\mathbf{r}_0^t\big)\,. \end{aligned} \tag{6.12}$$

Notice that the current channel output r_t only affects the estimate $\hat{h}_t(\mathbf{r}_0^t)$, but not the encoding matrix $\bar{T}_t(\mathbf{r}_0^{t-1})$. Following Theorem 1 from [Sko99a], we introduce the following lemma.

Lemma 6.3.2. *The term $\hat{h}_t(\mathbf{r}_0^t)$ can be computed according to*

$$\hat{h}_t\left(\mathbf{r}_0^t\right) = \left[m'_{h_t}\left(\mathbf{r}_0^{t-1}\right) \cdot \hat{p}_t\left(\mathbf{r}_0^t\right)\right]^{-1} R_{h_t h_t}\left(\mathbf{r}_0^{t-1}\right) \hat{p}_t\left(\mathbf{r}_0^t\right), \tag{6.13}$$

where the terms $R_{h_t h_t}(\mathbf{r}_0^{t-1})$, $m_{h_t}(\mathbf{r}_0^{t-1})$, and $\hat{p}_t(\mathbf{r}_0^t)$ are defined as,

$$R_{h_t h_t}(\mathbf{r}_0^{t-1}) \triangleq \sum_{l=0}^{L_I-1} \mathcal{P}\big(i_t = l \big| \mathbf{r}_0^{t-1}\big)\, h_t\big(l, \mathbf{r}_0^{t-1}\big)\, h_t\big(l, \mathbf{r}_0^{t-1}\big)' \tag{6.14}$$

$$m_{h_t}(\mathbf{r}_0^{t-1}) \triangleq \sum_{l=0}^{L_I-1} \mathcal{P}\big(i_t = l \big| \mathbf{r}_0^{t-1}\big)\, h_t\big(l, \mathbf{r}_0^{t-1}\big) \tag{6.15}$$

$$\hat{p}_t\left(\mathbf{r}_0^t\right) \triangleq \mathbf{E}\left\{ h_t\big(\mathbf{r}_0^{t-1}\big) \,\middle|\, r_t, \mathcal{P}\big(i_t = l \big| \mathbf{r}_0^{t-1}\big) = \frac{1}{L_I}, \forall l \right\}. \tag{6.16}$$

Here, $\hat{p}_t(\mathbf{r}_0^t)$ in (6.16) is the a posteriori expectation of $h_t(\mathbf{r}_0^{t-1})$, conditioned on the occurrence that all indices in $\mathcal{L}_I$ are equally likely. The proof of Lemma 6.3.2 can be found in Appendix 6.A. Note that, we can write the term $\hat{p}_t(\mathbf{r}_0^t)$ as

$$\begin{aligned}\hat{p}_t\big(\mathbf{r}_0^t\big) &\triangleq \mathbf{E}\left\{ h_t\big(\mathbf{r}_0^{t-1}\big)\big|\, r_t, \mathcal{P}\big(\, i_t = k|\, \mathbf{r}_0^{t-1}\big) = \frac{1}{L_I}, \forall k \right\} \\ &= \frac{\dfrac{1}{L_I}\displaystyle\sum_{k=0}^{L_I-1} h_t\big(k, \mathbf{r}_0^{t-1}\big)\, p(\, r_t|\, i_t = k)}{\dfrac{1}{L_I}\displaystyle\sum_{m=0}^{L_I-1} p(\, r_t|\, i_t = m)}.\end{aligned}$$

It should be observed that $\hat{p}_t(\mathbf{r}_0^t)$ is the only term in (6.14)–(6.16) which is affected by the current channel output r_t. According to (6.16), the binary bits $b_t^{[k]}(\mathbf{r}_0^{t-1})$, $k \in \{1, \ldots, R_I\}$ are statically independent, because of the fact that the conditional pdf's $\mathcal{P}(\, i_t = l|\, \mathbf{r}_0^{t-1})$ are equal, irrespective of $l \in \mathcal{L}_I$. Together with the assumption of memoryless channels, we are able to compute $\hat{p}_t(\mathbf{r}_0^t)$ by individually estimating each bit $\hat{b}_t^{[k]}(\mathbf{r}_0^{t-1}, r_t)$, $k \in \{1, \ldots, R_I\}$. In fact, $\hat{p}_t(\mathbf{r}_0^t)$ can be calculated according to

$$\hat{p}_t\big(\mathbf{r}_0^t\big) = \begin{bmatrix} 1 \\ \hat{b}_t^{[R_I]}\Big(\mathbf{r}_0^{t-1}, r_t^{[R_I]}\Big) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ \hat{b}_t^{[1]}\Big(\mathbf{r}_0^{t-1}, r_t^{[1]}\Big) \end{bmatrix},$$

where $\hat{b}_t^{[k]}(\mathbf{r}_0^{t-1}, r_t^{[k]})$ is defined as

$$\mathbf{E}\left\{ b_t^{[k]} \Big|\, \mathbf{r}_0^{t-1}, r_t^{[k]}, \mathcal{P}\Big(b_t^{[k]} = 1\Big) = \mathcal{P}\Big(b_t^{[k]} = 0\Big) = \frac{1}{2} \right\}. \tag{6.17}$$

Notice that, conditioned on the transmitted codeword b_t, the received codeword r_t is independent of all past outputs $\mathbf{r}_0^{t-1}$. Therefore, $\hat{b}_t^{[k]}(\mathbf{r}_0^{t-1}, r_t^{[k]}) = \hat{b}_t^{[k]}(r_t^{[k]})$, where $\hat{b}_t^{[k]}(r_t^{[k]})$ is simply

$$\begin{aligned}\hat{b}_t^{[k]}\left(r_t^{[k]}\right) =& \mathbf{E}\left\{ b_t^{[k]} \Big|\, r_t^{[k]}, \mathcal{P}\Big(b_t^{[k]} = 1\Big) = \mathcal{P}\Big(b_t^{[k]} = 0\Big) = \frac{1}{2} \right\} \\ =& \frac{e^{\frac{r_t^{[k]}}{\sigma_w^2}} - e^{-\frac{r_t^{[k]}}{\sigma_w^2}}}{e^{\frac{r_t^{[k]}}{\sigma_w^2}} + e^{-\frac{r_t^{[k]}}{\sigma_w^2}}} = \tanh\left(\frac{r_t^{[k]}}{\sigma_w^2}\right).\end{aligned} \tag{6.18}$$

Above, we have shown that the Hadamard framework can be used to implement an efficient soft controller. Especially, it is worth noticing that $\hat{h}_t(\mathbf{r}_0^t)$ in (6.12) can be viewed as a channel decoder, while $\bar{T}_t(\mathbf{r}_0^{t-1})$ functions as a source decoder. However, the terms $\mathbf{E}\left\{ x_t|\, i_t, \mathbf{r}_0^{t-1} \right\}$ in (6.9) and $\mathcal{P}(\, i_t|\, \mathbf{r}_0^{t-1})$ in (6.12) are obstacles to the implementation of the Hadamard-based soft controller. First, it is exceedingly

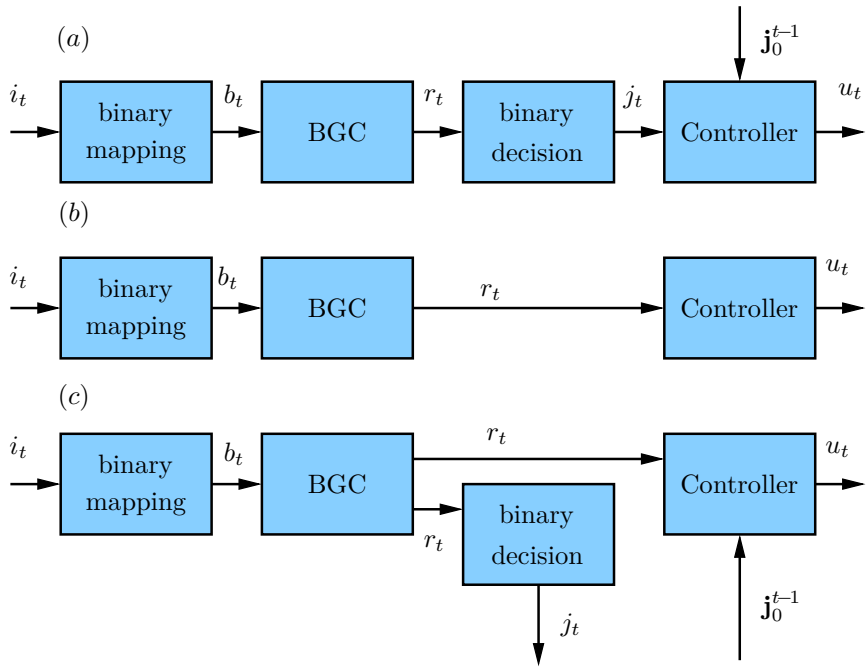


Figure 6.2: Three types of controllers: (a) the controller $g_t(\mathbf{j}_0^t)$ utilizes only hard information; (b) the controller $g_t(r_t)$ utilizes only the latest soft channel output; (c) the controller $g_t(r_t, \mathbf{j}_0^{t-1})$ utilizes both soft and hard information.

difficult to derive closed-form expressions for them. Second, it is neither possible to compute them numerically nor to implement them as look-up tables, as done for the finite-output channels in Chapter 3. This is attributed to the dimensionality of r_t. Third, it is also unrealistic to feedback to the encoder full SI since this would require an infinite-resolution feedback channel. Given the above challenges, in order to proceed, we will in the next step replace $\mathbf{E}\left\{ x_t | \, i_t, \mathbf{r}_0^{t-1} \right\}$ and $\mathcal{P}(i_t | \, \mathbf{r}_0^{t-1})$ by approximations of lower complexity. The remaining part of this section is therefore devoted to some practical encoder–controllers for BGC constructed based on the results of the soft-information-based controller (6.12).

6.3.2 Hard-Information-Based Controller

Recall the signal flow through the channel: the encoded index i_t is first mapped into a binary codeword b_t. Then, antipodal signaling is utilized to transmit the binary codeword bit-by-bit. The information-carrying signal is deteriorated by additive Gaussian noise when passing through the BGC channel, where the channel output is a vector of real values. Unlike the estimator-based controller (6.12), we will in

this subsection describe a detector-based controller which operates on the hard information of channel outputs. In short, the controller works as follows. First, let a binary decision be made on the real-valued channel output according to

$$\hat{r}_t^{[k]} = \begin{cases} 1, & r_t^{[k]} > 0, \\ -1, & r_t^{[k]} \leq 0, \end{cases}$$

where, $r_t^{[k]}$ is the k^{th} element of the channel output vector r_t. As commonly addressed in the literature, we refer to this binary decision as a *hard decision*. Correspondingly, a decision based on the real-valued channel output r_t, is referred to as a *soft decision*. The binary codeword $\hat{r}_t$ is then mapped to a so-called received index $j_t \in \mathcal{L}_I$, of the same finite alphabet as the channel input. Finally the control command is derived based on all received indices, i.e., $u_t = g_t(\mathbf{j}_0^t)$.

In fact, the BGC and the hard decision can be regarded as a combined channel, operating approximately as a binary symmetric channel. The main advantages in employing a hard-information-based controller are: (i) it is practically implementable because of the low complexity, and (ii), the results of the iterative training method from Chapter 3 can be applied directly. Of course, since the hard-information-based controller has not taken into consideration all information carried by channel outputs, such a solution is expected to cause a degradation in system performance.

6.3.3 Combined Soft-Hard Controller

As stated previously, a straightforward implementation of the controller (6.6) is practically impossible. On the other hand, completely relying on hard decisions can result in serious performance degradation, especially when the channel is highly noisy. Therefore, we are motivated to improve the hard-information-based controller by additionally exploiting certain soft information of the channel outputs. More precisely, let the controller to be on the form $g_t(r_t, \mathbf{z}_0^{t-1})$, which takes as input the current channel output r_t and the past SI, $\mathbf{z}_0^{t-1}$. The entire history of SI $\mathbf{z}_0^{t-1}$ can be viewed as a low-rate approximation of the channel outputs $\mathbf{r}_0^{t-1}$. Only exploiting r_t and $\mathbf{z}_0^{t-1}$, we follow the structure of the soft controller (6.12) and adopt the following controller

$$g_t\big(r_t, \mathbf{z}_0^{t-1}\big) = \ell_t \mathbf{E}\left\{ x_t \middle| r_t, \mathbf{z}_0^{t-1} \right\} = \ell_t \bar{T}_t\big(\mathbf{z}_0^{t-1}\big)\, \mathbf{E}\left\{ h_t\big(\mathbf{z}_0^{t-1}\big) \middle| r_t \right\}, \tag{6.19}$$

where the encoding matrix $\bar{T}_t(\mathbf{z}_0^{t-1})$ solves the equation,

$$\bar{C}_t\big(\mathbf{z}_0^{t-1}\big) = \bar{T}_t\big(\mathbf{z}_0^{t-1}\big)\, H_t\big(\mathbf{z}_0^{t-1}\big)\,.$$

Here, the matrix $\bar{C}_t(\mathbf{z}_0^{t-1})$ consists of all conditional centroids $\mathbf{E}\left\{ x_t \middle| i_t, \mathbf{z}_0^{t-1} \right\}$. The l^{th} column of $\bar{C}_t(\mathbf{z}_0^{t-1})$, denoted by $c_t(l, \mathbf{z}_0^{t-1})$, is

$$c_t\big(l, \mathbf{z}_0^{t-1}\big) = \mathbf{E}\left\{ x_t \middle| i_t = l, \mathbf{z}_0^{t-1} \right\}.$$

Denoted by $h_t(l, \mathbf{z}_0^{t-1})$, the l^{th} column of the Hadamard matrix $H_t(\mathbf{z}_0^{t-1})$ is formed in a similar manner as (6.11), with $b_t(i_t = l, \mathbf{r}_0^{t-1})$ replaced by $b_t(i_t = l, \mathbf{z}_0^{t-1})$, $l \in \mathcal{L}_I$. In order to complete (6.19), we also need $\mathbf{E}\left\{ h_t(\mathbf{z}_0^{t-1}) \middle| r_t \right\}$, which can be obtained as follows

$$\begin{aligned}\mathbf{E}\left\{ h_t\big(\mathbf{z}_0^{t-1}\big) \middle| r_t \right\} =& \hat{h}_t(\mathbf{z}_0^{t-1}, r_t) \\ =& \left[m'_{h_t}(\mathbf{z}_0^{t-1}) \cdot \hat{p}_t\left(r_t, \mathbf{z}_0^{t-1}\right) \right]^{-1} R_{h_t h_t}(\mathbf{z}_0^{t-1}) \hat{p}_t\left(r_t, \mathbf{z}_0^{t-1}\right),\end{aligned}$$

with $R_{h_t h_t}(\mathbf{z}_0^{t-1})$ and $m_{h_t}(\mathbf{z}_0^{t-1})$ defined as

$$R_{h_t h_t}\left(\mathbf{z}_0^{t-1}\right) \triangleq \sum_{l=0}^{L_I - 1} \mathcal{P}\big(i_t = l \big| \mathbf{z}_0^{t-1}\big)\, h_t\big(l, \mathbf{z}_0^{t-1}\big)\, h_t\big(l, \mathbf{z}_0^{t-1}\big)', \tag{6.20}$$

$$m_{h_t}\left(\mathbf{z}_0^{t-1}\right) \triangleq \sum_{l=0}^{L_I - 1} \mathcal{P}\big(i_t = l \big| \mathbf{z}_0^{t-1}\big)\, h_t\big(l, \mathbf{z}_0^{t-1}\big). \tag{6.21}$$

Finally, it is straightforward to verify that $\hat{p}_t(r_t, \mathbf{z}_0^{t-1})$ can still be computed only using $\hat{b}_t^{[k]}(r_t^{[k]})$, as shown by (6.17)–(6.18). In practice, a candidate of z_t is the index j_t given by the binary decisions $\{\hat{r}_t^{[k]}\}$, $k \in \{1, \ldots, R_I\}$, which will be studied by numerical examples in Section 6.5.

In the extreme case that there is no SI at the encoder, a simple practical controller is

$$g_t(r_t) = \ell_t \mathbf{E}\left\{ x_t \middle| r_t \right\} = \ell_t \bar{T}_t \mathbf{E}\left\{ h_t \middle| r_t \right\},$$

where $\bar{T}_t$ satisfies the equation $\bar{C}_t = \bar{T}_t H_t$. The matrix $\bar{C}_t$ is constructed by the conditional centroid $\mathbf{E}\left\{ x_t \middle| i_t = l \right\}$, $l \in \mathcal{L}_I$, and the matrix H_t has its l^{th} column $h_t(l)$ formed by $b_t(l)$, the binary codeword associated to $i_t = l$. Observe that, at each t there is only one encoding matrix, $\bar{T}_t$. Finally, we compute the vector $\mathbf{E}\left\{ h_t \middle| r_t \right\}$ according to

$$\mathbf{E}\left\{ h_t \middle| r_t \right\} = \left[m'_{h_t} \cdot \hat{p}_t\left(r_t\right) \right]^{-1} R_{h_t h_t} \hat{p}_t\left(r_t\right),$$

where $R_{h_t h_t}$ and m_{h_t} are given by

$$\begin{aligned} R_{h_t h_t} &\triangleq \sum_{l=0}^{L_I - 1} \mathcal{P}\left(i_t = l\right) h_t(l) h_t(l)', \\ m_{h_t} &\triangleq \sum_{l=0}^{L_I - 1} \mathcal{P}\left(i_t = l\right) h_t(l), \end{aligned}$$

and $\hat{p}_t(r_t)$ is still fully determined by $\hat{b}_t^{[k]}(r_t^{[k]})$, as given (6.18).

6.4 System Design

In Section 6.3, we have introduced a number of practical controllers among which the major difference lies in their accessible information. Regarding the optimization of encoder mappings, it is generally not possible to implement a similar version

Algorithm 6.4.1 Encoder–Controller Design Algorithm for BGC

1. Initialize the hard-decision-based encoder–controller mappings $\mathbf{f}_0^{T-1}$ and $\mathbf{g}_0^{T-1}$.
2. Optimize hard-decision-based encoder–controller pair according to Figure 3.3.
3. Fixing the trained encoder, replace the hard-decision-based controller with a combined controller, according to (6.19).

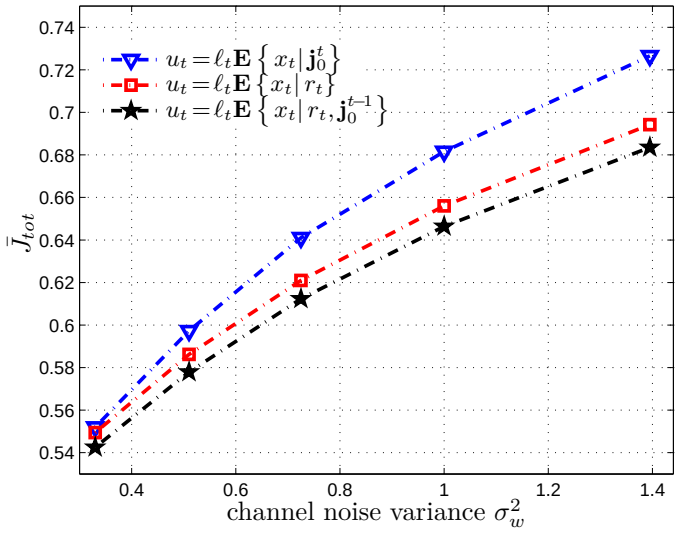
of (3.8) from Chapter 3, mainly because of the facts: (*i*) it is impractical to feedback to the encoder full SI due to the unrealistic bandwidth requirement of the SI feedback channel, and (*ii*), it is impractical to optimize an encoder exploiting soft information due to the complexity limitation. Given the above difficulties, we consider only practical encoders which receive hard-information as SI about past channel outputs.

Based on the above discussion concerning the encoder and controller, we propose Algorithm 6.4.1, which provides practical designs of encoder–controllers for the BGC. In short, the optimization is performed in two steps: First, train a hard-information-based encoder–controller according to Figure 3.3 from Chapter 3. Then, replace the hard-information-based controller by a combined controller (6.19) which exploits both the soft and hard information about the past channel outputs. Certainly, training according to Algorithm 6.4.1 can improve upon the performance with respect to the hard-information-based controller. Although we are not able to quantify the optimality of Algorithm 6.4.1, the algorithm works satisfactorily in practice, as shown in the next section.

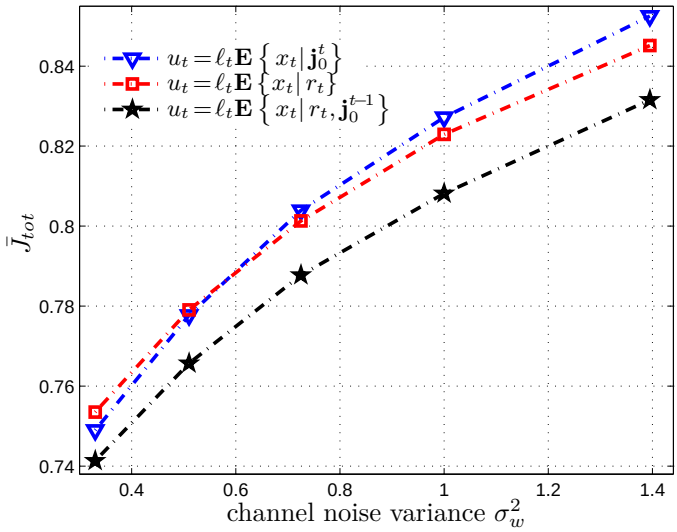
6.5 Numerical Examples

This section presents the numerical experiments conducted to study the performance of the various controllers described in Section 6.3. The impact of soft and hard information on the overall system performance is investigated. We also study the controllers in terms of σ_w^2 and P_t, the noise variance and the weighting factor of the control input, respectively. To focus on demonstrating the basic concept, in the experiments a scalar plant is considered where $M_e = 0$, i.e., at each t only the latest measurement is encoded and transmitted to the controller over the binary Gaussian channel. The linear plant has $A = 0.9$ and $B = C = 1$. The initial-state, process noise and measurement noise are zero-mean Gaussian with variances $\sigma_{x_0}^2 = 5$, $\sigma_{v_t}^2 = 1$ and $\sigma_e^2 = 1$, respectively. In the objective function, $T = 4$ and $V_t = 1$. Finally, the rate R_I is 2, i.e., 2 binary bits per state measurement.

In Figure 6.3, performance of three controllers are depicted with respect to the increasing noise variance σ_w^2, for two different P_t values, namely $P_t = 1$ and $P_t = 5$. The first controller, $u_t = \ell_t \mathbf{E}\{x_t | \mathbf{j}_0^t\}$, referred to as the hard-information-based controller, is restricted to only knowing the hard-information. The second controller,



(a) $P_t = 1$



(b) $P_t = 5$

Figure 6.3: The system performance by using the trained controllers: $g_t(\mathbf{j}_0^t)$, $g_t(r_t)$, $g_t\big(r_t, \mathbf{j}_0^{t-1}\big)$, for $P_t = 1$ and $P_t = 5$. The cost $\bar{J}_{tot}$ is derived by normalizing $\mathbf{E}\{J_{tot}\}$ with the cost obtained without any control.

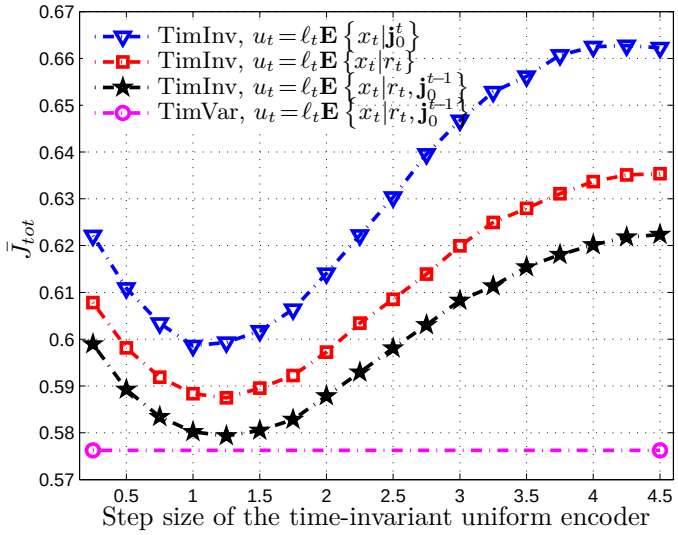
$u_t = \ell_t \mathbf{E}\{x_t | r_t\}$, referred to as the memoryless soft-information-based controller, is limited by not having access to memory. The third controller, $u_t = \ell_t \mathbf{E}\{x_t | r_t, \mathbf{j}_0^{t-1}\}$, referred to as the combined controller, exploits both the soft information of the current measurement and hard information of old measurements. For each P_t, there is a common encoder for all the three controllers. This encoder, designed according to CCS 4.6.4, as described in Chapter 4, is memoryless time-varying. Since the globally optimal solution is unknown, we take the special case of no control action ($u_t = 0$) as a reference system. The performance measure $\bar{J}_{tot}$ is obtained by normalizing $\mathbf{E}\{J_{tot}\}$ with the reference case where $u_t = 0$.

First, we see in Figure 6.3 that the combined controller always outperforms the other two controllers, evidently. Second, when the noise variance σ_w^2 is small and the weighting factor P_t is large, the hard-information-based controller is superior to the memoryless soft-information-based controller. For low-level channel noise, there is a minor difference between the soft bit and the hard bit. Hence, the hard-information-based controller can gain by having access to memory. On the other hand, when the channel becomes noisy, soft outputs carry considerably more information than hard bits, which explains why the memoryless soft-information-based controller is superior for large-valued σ_w^2. A large-valued P_t indicates a demanding power constraint on control inputs, which results in small-valued controls and slow converge to the steady state. Owing to the access to memory, there is an evident performance gain by using the hard-information-based controller, compared with using a memoryless soft-information-based controller. While, when P_t is small, the steady state is reached quickly that the significance of memory is reduced.

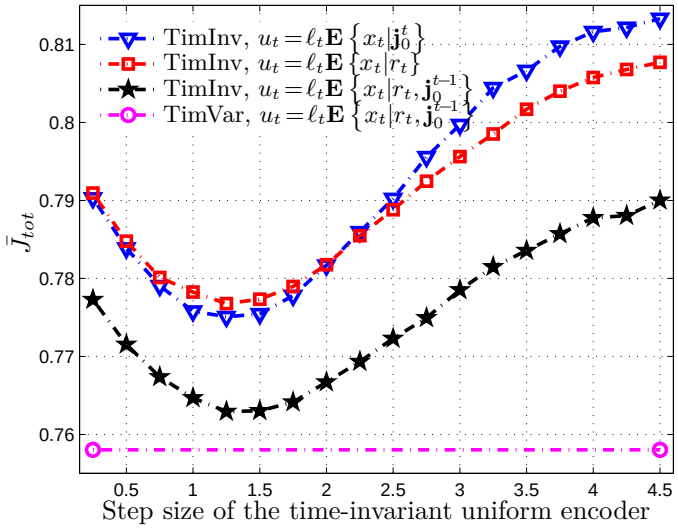
Figure 6.4 compares the performance of a trained encoder with time-invariant uniform encoders, for $P_t = 1$ and $P_t = 5$. The variance of channel noise is $\sigma_w^2 = 0.5$. The other system parameters are the same as in Figure 6.3. The horizonal line is obtained by training the encoder according to Figure 3.3. The other three curves in the figure are obtained by employing time-invariant uniform encoders where the step length is shown on the x-axis. As expected, the time-varying encoder outperforms the uniform encoders. Especially, if the uniform encoder is chosen improperly, it may cause severe consequences.

6.6 Summary

This chapter studied how the generalization to infinite-output channels affected the optimization of the encoder–controller theoretically and practically. From a practical point of view, the impact appeared not only in the training stage, but also in how to implement the trained encoder–controller pair. The challenge is that the trained encoder–controller can no longer be implemented as a simple look-up table. To get more insight to the optimal controller, the Hadamard-based soft controller which fully exploited the channel outputs was introduced. However, we can not implement the soft-information-based controller in practice due to the complexity and memory demands. Given this difficulty, a combined encoder–controller which exploited both



(a) $P_t = 1$



(b) $P_t = 5$

Figure 6.4: The system performance by using a time-varying (TimVar) encoder, trained according to Algorithm 6.4.1, or time-invariant (TimInv) uniform encoders, for $P_t = 1$ and $P_t = 5$. The cost $\bar{J}_{tot}$ is derived by normalizing $\mathbf{E}\{J_{tot}\}$ with the cost obtained without any control.

the hard and soft information of the channel outputs were proposed. Monte Carlo simulations showed that the proposed scheme has good performance compared to the controllers which only used hard-information or ignored the information carried in the memory.

6.A Proof of Lemma 6.3.2

Proof. Here, we briefly show the derivation of (6.14)–(6.16). According to (6.7), $\hat{h}_t(\mathbf{r}_0^t)$ is given by

$$\hat{h}_t\big(\mathbf{r}_0^t\big) = \frac{\displaystyle\sum_{l=0}^{L_I-1} p(\,r_t|\,i_t=l)\,\mathcal{P}\big(\,i_t=l|\,\mathbf{r}_0^{t-1}\big)\,h_t\big(l,\mathbf{r}_0^{t-1}\big)}{\displaystyle\sum_{k=0}^{L_I-1} p(\,r_t|\,i_t=k)\,\mathcal{P}\big(\,i_t=k|\,\mathbf{r}_0^{t-1}\big)}. \tag{6.22}$$

In what follows, we show that $\hat{h}_t(\mathbf{r}_0^t)$ can also be written as

$$\hat{h}_t\big(\mathbf{r}_0^t\big) = \frac{\displaystyle\sum_{l=0}^{L_I-1}\mathcal{P}\big(\,i_t=l|\,\mathbf{r}_0^{t-1}\big)h_t\big(l,\mathbf{r}_0^{t-1}\big)\,h_t\big(l,\mathbf{r}_0^{t-1}\big)'\frac{1}{L_I}\sum_{n=0}^{L_I-1} h_t\big(n,\mathbf{r}_0^{t-1}\big)\,p(\,r_t|\,i_t=n)}{\displaystyle\sum_{k=0}^{L_I-1}\mathcal{P}\big(\,i_t=k|\,\mathbf{r}_0^{t-1}\big)\,h_t\big(k,\mathbf{r}_0^{t-1}\big)'\,\frac{1}{L_I}\sum_{m=0}^{L_I-1} h_t\big(m,\mathbf{r}_0^{t-1}\big)\,p(\,r_t|\,i_t=m)}. \tag{6.23}$$

In the first place, consider the numerator of (6.23),

$$\begin{aligned}
&\sum_{l=0}^{L_I-1}\mathcal{P}\big(\,i_t=l|\,\mathbf{r}_0^{t-1}\big)\,h_t\big(l,\mathbf{r}_0^{t-1}\big)\,h_t\big(l,\mathbf{r}_0^{t-1}\big)'\,\frac{1}{L_I}\sum_{n=0}^{L_I-1} h_t\big(n,\mathbf{r}_0^{t-1}\big)p(\,r_t|\,i_t=n)\\
=\;&\sum_{l=0}^{L_I-1}\sum_{n=0}^{L_I-1}\frac{1}{L_I}\big(\mathcal{P}\,\big(\,i_t=l|\,\mathbf{r}_0^{t-1}\big)\,h_t\big(l,\mathbf{r}_0^{t-1}\big)\big)\Big(h_t\big(l,\mathbf{r}_0^{t-1}\big)'\,h_t\big(n,\mathbf{r}_0^{t-1}\big)\Big)\,p(\,r_t|\,i_t=n)\\
=\;&\sum_{l=0}^{L_I-1}\sum_{n=0}^{L_I-1}\frac{1}{L_I}\big(\mathcal{P}\big(\,i_t=l|\,\mathbf{r}_0^{t-1}\big)\,h_t\big(l,\mathbf{r}_0^{t-1}\big)\big)\,L_I\delta_{l,n}p(\,r_t|\,i_t=n)\\
=\;&\sum_{l=0}^{L_I-1}\big(\mathcal{P}\big(\,i_t=l|\,\mathbf{r}_0^{t-1}\big)\,h_t\big(l,\mathbf{r}_0^{t-1}\big)\big)p(\,r_t|\,i_t=l)\,.
\end{aligned} \tag{6.24}$$

where $\delta_{l,n}$ is the delta function, defined as

$$\delta_{l,n} \triangleq \begin{cases} 1, & l=n,\\ 0, & l\neq n.\end{cases}$$

Here, we have used the special property of the Hadamard matrix [Sko99a]:

$$h_t\big(l,\mathbf{r}_0^{t-1}\big)'\,h_t\big(n,\mathbf{r}_0^{t-1}\big) = \begin{cases} L_I, & l=n,\\ 0, & l\neq n.\end{cases}$$

In the second place, we treat the denominator of (6.23) in a similar way,

$$\begin{aligned}
&\sum_{k=0}^{L_I-1} \mathcal{P}\big(i_t=k|\,\mathbf{r}_0^{t-1}\big)\, h_t\big(k,\mathbf{r}_0^{t-1}\big)'\, \frac{1}{L_I} \sum_{m=0}^{L_I-1} h_t\big(m,\mathbf{r}_0^{t-1}\big)\, p(r_t|\,i_t=m) \\
= &\sum_{k=0}^{L_I-1}\sum_{m=0}^{L_I-1} \frac{1}{L_I}\mathcal{P}\big(i_t=k|\,\mathbf{r}_0^{t-1}\big)\Big(h_t\big(k,\mathbf{r}_0^{t-1}\big)'\, h_t\big(m,\mathbf{r}_0^{t-1}\big)\Big)\, p(r_t|\,i_t=m) \\
= &\sum_{k=0}^{L_I-1}\sum_{m=0}^{L_I-1} \frac{1}{L_I}\mathcal{P}\big(i_t=k|\,\mathbf{r}_0^{t-1}\big)\, L_I\delta_{k,m} p(r_t|\,i_t=m) \\
= &\sum_{k=0}^{L_I-1} \big(\mathcal{P}\big(i_t=k|\,\mathbf{r}_0^{t-1}\big)\big)\, p(r_t|\,i_t=k)\,.
\end{aligned}$$

Finally, dividing both the numerator and denominator of (6.23) by the normalization factor $\sum_{n=0}^{L_I-1} p(r_t|\,i_t=n)$, yields

$$\begin{aligned}
\hat{h}_t\big(\mathbf{r}_0^t\big) = &\frac{\displaystyle\sum_{l=0}^{L_I-1} \mathcal{P}\big(i_t=l|\,\mathbf{r}_0^{t-1}\big)\, h_t\big(l,\mathbf{r}_0^{t-1}\big)\, h_t\big(l,\mathbf{r}_0^{t-1}\big)'}{\displaystyle\sum_{k=0}^{L_I-1} \mathcal{P}\big(i_t=k|\,\mathbf{r}_0^{t-1}\big)\, h_t\big(k,\mathbf{r}_0^{t-1}\big)'} \\
&\times\left(\frac{\dfrac{1}{L_I}\displaystyle\sum_{n=0}^{L_I-1} h_t\big(n,\mathbf{r}_0^{t-1}\big)\, p(r_t|\,i_t=n)}{\dfrac{1}{L_I}\displaystyle\sum_{n=0}^{L_I-1} p(r_t|\,i_t=n)}\right) \\
&\times\left(\frac{\dfrac{1}{L_I}\displaystyle\sum_{m=0}^{L_I-1} h_t\big(m,\mathbf{r}_0^{t-1}\big)\, p(r_t|\,i_t=m)}{\dfrac{1}{L_I}\displaystyle\sum_{n=0}^{L_I-1} p(r_t|\,i_t=n)}\right).
\end{aligned}$$

Now we can easily identify the terms from (6.13)–(6.16). □

Chapter 7

Conclusions and Future Research

7.1 Concluding Remarks

In this thesis we have discussed various fundamental aspects of wireless networked control systems. Because of the importance and popularity of industrial wireless networking, research on designing closed-loop systems for control using measurement feedback over imperfect communication channels has received increasing attention. It can be claimed, however, that the research area is still in its infancy, and it evolves rapidly. As explained in the introduction chapter, considerable efforts have been devoted to various stability issues for quantized control systems. This thesis, on the other hand, focuses mainly on the optimization of the overall system performance. In particular, a stochastic control problem was formulated and several aspects of the design and analysis of encoder–controllers for control over low-rate noisy channels were studied. To deal with one of the most fundamental problems in control systems, namely how to make the best use of the feedback information for the future evolution, the encoder and controller are required to perform efficient estimation and control using a few bits per sensor measurement. Because of the complex relation to all past and future events, optimal estimation and control are difficult tasks.

In the main part of the thesis we studied the problem of optimizing the encoder–controller jointly, by using an iterative training approach. The basic principle is to alternate between the optimization of the encoder mappings and the controller mappings. For the special case of full SI, we showed that the optimal controller is a CE controller. For the general case of partial SI, we proposed an iterative design approach based on constraining the controller to be a CE controller. The major drawback of the proposed training approach is the computational complexity. To seek a good compromise between computational effort and efficient controls, various low-complexity encoder–controllers were proposed.

In the second part of the thesis, we studied the optimization of the rate allocation over time, another method to overcome the limited communication resources and to achieve better performance. We investigated how to allocate communication

rates both in the scenario of state estimation as well as for state feedback control. A solution based on high-rate quantization theory and Lagrangian duality was proposed, which gives a good performance compared to arbitrarily selected rate allocations.

In the final part of the thesis, we extended the design of encoder–controllers for channels of finite alphabets to include infinite-output channels. Optimizing for this class of channels increases the controller complexity substantially. The recursive structure of the optimal controller was exposed and a combined encoder–controller was proposed which exploited both the soft and hard information of the channel outputs.

7.2 Future Research

Below we discuss a few specific problems suggested for future research.

Iterative Design

The iterative encoder–controller design proposed in Chapter 3 has been shown to result in satisfactory performance. However, there is still much work to be done before the iterative design method is fully understood. First of all, theoretical results on the convergence properties and the optimality of the training method are important questions which need to be further investigated. The trade-off relation between the overall performance and the computational effort, such as the size of the sample set, and the number of iterations, should to be clarified.

Low-Complexity High-Performance Encoder–Controllers

The major drawback of the training-based approach is the heavy computational burden. Based on the discussions from Chapter 3, it is clear that the separation property is a desired feature, since it can reduce the overall optimization complexity significantly. There is no doubt that for most systems, the separation principle does not apply. Thus, we believe that it is important to find useful suboptimal solutions, which may or may not exploit the separation property. Seeking suboptimal solutions opens up a broad spectra of research themes, both theoretical and practical. In Chapter 4, an empirical study of several low-complexity encoder–controllers was conducted. Analysis of the performance degradation of those schemes still needs to be carried out.

As a theoretical foundation, it is of great importance to understand the significance of the variety of the system parameters, such as SI, encoder–controller memories etc. Especially there is a need to clarify the interactive relations among those factors. Moreover, from a practical point of view, encoder–controllers with finite memories should be considered. The synthesis and analysis of complexity-reduced coding–control schemes are certainly topics deserving considerable research efforts.

Unstable Plant

Without any control action, unstable poles will result in unbounded state trajectories and lead to unbounded estimation errors. In the literature, concerning stabilizing unstable plants over noisy channels, relatively little work has been performed so far. It can be claimed that we in this thesis considered both *stable* and *unstable* plants, since our performance criterion uses a finite horizon. It should be observed, however, that even for stable plants, many conventional controllers will fail to stabilize the closed-loop system over noisy channels. In the case of strictly unstable plants, the situation becomes more critical. In the thesis, we have demonstrated numerically that it is possible to stabilize unstable plants over noisy channels. However, there is of course a need to perform a more thorough stability analysis in the case of control over noisy links. How to optimize the overall system performance, while imposing a stability requirement over an infinite horizon is a challenging problem.

Distributed Network

Our study of quantized control systems is motivated by the challenges of control over wireless networked systems, given limited communication resources. Clearly, besides the problem of optimizing the performance for each individual plant, another major challenge is to coordinate all distributed control nodes to provide a satisfactory overall performance. In a networked system, it can happen that several plants are communicating simultaneously which may give congestion and interference. How to design communication protocols that minimize congestion and interference problems, and how to handle the congestion and interference when they are present, are issues which deserve thorough investigations. Regarding the resource allocation issues, how to distribute communication resources among multiple control nodes, and how to adopt to the varying demands of the control nodes, are also interesting problems. Finally, it has been shown that combining measurements from several distributed nodes has the potential to provide more accurate estimation about the objects a network is monitoring. How to apply the advances in distributed sensing and data fusion techniques to networked control problems also need to be carefully investigated.

Bibliography

[ÅL92] K. J. Åström and M. Lundh. Lund control program combines theory with hands-on experience. In *IEEE Control Systems Magazine*, volume 12, pages 22–30, 1992.

[Aok67] M. Aoki. *Optimization of Stochastic Systems - Topics in Discrete-Time Systems*. Academic Press, 1967.

[Åst70] K. J. Åström. *Introduction to Stochastic Control Theory*. Academic Press, INC., 1970.

[Ath71] M. Athans. The role and use of the stochastic linear-quadratic-Gaussian problem in control system design. *IEEE Transactions on Automatic Control*, pages 529–552, December 1971.

[Bai02] J. Baillieul. Feedback coding for information-based control: operating near the data-rate limit. In *Proc. of the 41st IEEE Conference on Decision and Control*, volume 3, pages 3229–3236, December 2002.

[Bai04] J. Baillieul. Data-rate requirements for nonlinear feedback control. In *IFAC Symp. Nonlinear Control Systems*, volume 1, September 2004.

[BBCea00] A. Berlin, D. Biegelsen, P. Cheung, and et al. Motion control of planar objects using large-area arrays of MEMS-like distributed manipulators. 2000.

[Bel61] R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[Ber76] D. Bertsekas. *Dynamic programming and stochastic control*. Academic Press, 1976.

[Ber95] D. Bertsekas. *Dynamic programming and optimal control*. Prentice-Hall, Inc., 1995.

[BL00] R. W. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, 45(7):1279–1289, July 2000.

[BL06] F. Bullo and D. Liberzon. Quantized control via locational optimization. *IEEE Transactions on Automatic Control*, 51(1):2–13, January 2006.

[BM95] V. S. Borkar and S. K. Mitter. LQG control with communication constraints. Technical report, MIT, 1995.

[BMT01] V. S. Borkar, S. K. Mitter, and S. Tatikonda. Markov control problems under communication constraints. *Communications in Information and Systems*, 1(1):15–32, January 2001.

[BS96] D. Bertsekas and S. E. Shreve. *Stochastic optimal control: the discrete-time case*. Athena Scientific, Belmont, Massachusetts, 1996.

[BSFJ09a] L. Bao, M. Skoglund, C. Fischione, and K. H. Johansson. Optimized rate allocation for closed-loop control over noisy channels. *Manuscript in preparation for journal publication*, 2009.

[BSFJ09b] L. Bao, M. Skoglund, C. Fischione, and K. H. Johansson. Optimized rate allocation for state estimation over noisy channels. In *Proc. of 2009 IEEE International Symposium on Information Theory*, 2009.

[BSFJ09c] L. Bao, M. Skoglund, C. Fischione, and K. H. Johansson. Optimized rate allocation for state feedback control over noisy channels. In *Proc. of the 48nd IEEE Conference on Decision and Control, submitted*, 2009.

[BSFJ09d] L. Bao, M. Skoglund, C. Fischione, and K. H. Johansson. Rate allocation for quantized control over noisy channels. In *Proc. of 2009 IEEE workshop on Control over communication channels*, 2009.

[BSJ06a] L. Bao, M. Skoglund, and K. H. Johansson. Encoder–decoder design for event-trigged feedback control. In *Proc. of 2006 American Control Conference*, pages 4183–4188, 2006.

[BSJ06b] L. Bao, M. Skoglund, and K. H. Johansson. Encoder–decoder design for feedback control over the binary symmetric channel. In *Proc. of 2006 IEEE International Symposium on Information Theory*, pages 2481–2485, 2006.

[BSJ07a] L. Bao, M. Skoglund, and K. H. Johansson. On optimal system design for feedback control noisy channels. In *Proc. of 2007 IEEE International Symposium on Information Theory*, 2007.

[BSJ07b] L. Bao, M. Skoglund, and K. H. Johansson. A scheme for joint quantization, error protection and control over noisy channels. In *Proc. of 2007 American Control Conference*, pages 4905–4910, 2007.

[BSJ08] L. Bao, M. Skoglund, and K. H. Johansson. On the separation principle in optimal control over noisy channels. In *IFAC World Congress 2008*, July 2008.

[BSJ09] L. Bao, M. Skoglund, and K. H. Johansson. Iterative encoder–controller design for feedback control over noisy channels. *Submitted for journal publication September 2007. Revised June 2009*, 2009.

[BST74] Y. Bar-Shalom and E. Tse. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, AC-19(5):494–500, October 1974.

[BT97] D. Bertsekas and J.N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, Massachusetts, 1997.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[CD02] M. Chiani and D. Dardari. Improved exponential bounds and approximation for the q-function with application to average error probability computation. In *Global Telecommunications Conference*, pages 1399–1402, 2002.

[CF08] C. D. Charalambous and A. Farhadi. LQG optimality and separation principle for general discrete time partially observed stochastic systems over finite capacity communication channels. *Automatica*, 44:3181–3188, 2008.

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc, 1991.

[Cur72] R. E. Curry. *Estimation and control with quantized measurements*. The M.I.T Press, 1972.

[Del88] D. F. Delchamps. The 'stabilization' of linear systems with quantized feedback. In *Proc. of the* 27*th IEEE Conference on Decision and Control*, December 1988.

[Del89] D. F. Delchamps. Controlling the flow of information in feedback systems with measurement quantization. In *Proc. of the* 28*th IEEE Conference on Decision and Control*, December 1989.

[Del90] D. F. Delchamps. Stabilizing a linear system with quantized state feedback. *IEEE Transactions on Automatic Control*, 35(8):916–924, August 1990.

[Del06] J. C. Delvenne. An optimal quantized feedback strategy for scalar linear systems. *IEEE Transactions on Automatic Control*, 51(2), February 2006.

[DLGH05] M. Drew, X. Liu, A. Goldsmith, and K. Hedrick. Networked control system design over a wireless LAN. In *Proc. of the 44st IEEE Conference on Decision and Control*, pages 6704–6709, December 2005.

[DS06] S. C. Draper and A. Sahai. Noisy feedback improves communication reliability. In *Proc. of 2006 IEEE International Symposium on Information Theory*, pages 69–73, July 2006.

[EF02] N. Elia and E. Frazzoli. *Quantized stabilization of two-input linear systems: a lower bound on the minimal quantization density.* Springer, 2002. Hybrid Systems Computation and Control, Lecture Notes in Computer Science.

[EHSW87] A. A. El Gamal, L. A. Hemachandra, I. Shperling, and V. K. Wei. Using simulated annealing to design good codes. *IEEE Transactions on Information Theory*, 33(1), January 1987.

[EM99] N. Elia and S. K. Mitter. Quantization of linear systems. In *Proc. of the 38th Conference on Decision and Control*, volume 4, pages 3428–3433, December 1999.

[EM01] N. Elia and S. K. Mitter. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 46(9), September 2001.

[EWRF07] F. G. Estern, C. Canudas De Wit, F. Rubio, and J. Fornes. Adaptive delta-modulation coding for networked controlled systems. In *Proc. of the 26th American Control Conference, ACC'2007)*, July 2007.

[Far90] N. Farvardin. A study of vector quantization for noisy channels. *IEEE Transactions on Information Theory*, 36(4):799–809, 1990.

[FM84] N. Farvardin and J. W. Modestino. Optimum quantizer performance for a class of non-Gaussian memoryless sources. *IEEE Transactions on Information Theory*, 30(3):485–497, May 1984.

[FV87] N. Farvardin and V. Vaishampayan. Optimum quantizer design for noisy channels: An approach to combined source-channel coding. *IEEE Transactions on Information Theory*, 33(6):827–838, November 1987.

[FV91] N. Farvardin and V. Vaishampayan. On the performance and complexity of channel-optimized vector quantizers. *IEEE Transactions on Information Theory*, 37(1):155–160, January 1991.

[FX05] M. Fu and L. Xie. The sector bound approach to quantized feedback control. *IEEE Transactions on Automatic Control*, 50(11):1698–1711, November 2005.

[FX06] M. Fu and L. Xie. Finite-level quantized feedback control for linear systems. In *Proc. IEEE Conf. Decision and Control*, December 2006.

[FX07] M. Fu and L. Xie. Dynamic quantized feedback control design for uncertain linear systems. In *Proc. IEEE Conf. Control and Automation*, pages 3004–3008, May 2007.

[FZ03] F. Fagnani and S. Zampieri. Stability analysis and synthesis for scalar linear systems with a quantized feedback. *IEEE Transactions on Automatic Control*, 48(9):1569–1584, September 2003.

[FZ05a] F. Fagnani and S. Zampieri. A symbolic approach to performance analysis of quantized feedback systems: the scalar case. *SIAM J. Control Optim.*, 44(3):818–866, 2005.

[FZ05b] B. Farber and K. Zeger. Quantization of multiple sources using integer bit allocation. In *Proc. of* 2005 *IEEE Data Compression Conference*, 2005.

[FZ06] B. Farber and K. Zeger. Quantization of multiple sources using nonnegative integer bit allocation. *IEEE Transactions on Information Theory*, 52:4945–4964, January 2006.

[Gal68] R. G. Gallager. *Information theory and reliable communication.* John Wiley and Sons, Inc., 1968.

[Ger79] A. Gersho. Asymptotically optimal block quantization. *IEEE Transactions on Information Theory*, 25(4):373–380, July 1979.

[GG92] A. Gersho and R. M. Gray. *Vector quantization and signal compression.* Kluwer, 1992.

[GN08] A. Gurt and G. N. Nair. Internal stability of dynamically quantised control for stochastic scalar plants. In *IFAC World Congress 2008*, pages 5197–5202, July 2008.

[GR95] W. R. Gardner and B. D. Rao. Theoretical analysis of the high-rate vector quantization of LPC parameters. 3(5):367–381, September 1995.

[GW02] A. Goldsmith and S. B. Wicker. Design challenges for energy-constrained ad hoc wireless networks. *IEEE Wireless Communications Magazine*, 9(4):8–27, August 2002.

[Ho80] Y. Ho. Team decision theory and information structures. *Proceedings of the IEEE*, 68(6), June 1980.

[Hor95] *Handbook of global optimization.* Kluwer Academic Publishers, 1995.

[HOV02] J. P. Hespanha, A. Ortega, and L. Vasudevan. Towards the control of linear systems with minimum bit-rate. In *Proc. of the 15th International Symposium on MTNS*, 2002.

[Joh00] K. H. Johansson. The quadruple-tank process: A multivariable laboratory process with an adjustable zero. *IEEE Transactions on Control Systems Technology*, 8(3):456–465, May 2000.

[Jon01] P. H. Jones. *Environmental Controls - Florida Greenhouse Vegetable Production Handbook*. Institute of Food and Agricultural Sciences, University of Florida, 2001.

[KA96] P. Knagenhjelm and E. Agrell. The Hadamard transform - a tool for index assignment. *IEEE Transactions on Information Theory*, 42(4):1139–1151, July 1996.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598), May 1983.

[Kha96] *Nonlinear Systems*. Prenice-Hall. Inc., 1996.

[Kie82] J. C. Kieffer. Exponential rate of convergence for Lloyd's method i. *IEEE Transactions on Information Theory*, 28(2):205–210, March 1982.

[KL99] P. Koppang and R. Leland. Linear quadratic stochastic control of atomic hydrogen masers. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 46(3):517–522, May 1999.

[Kle04] B. Kleijn. *A Basis for Source Coding*. KTH (Royal Institute of Technology), 2004.

[Kna93] P. Knagenhjelm. How good is your index assignment? In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume II, pages 423–426, December 1993.

[LE04] J. Liu and N. Elia. Quantized feedback stabilization of non-linear affine systems. *Internat. J. Control*, 73(3):239–249, 2004.

[LG04] X. Liu and A. Goldsmith. Wireless medium access control in networked control systems. In *American Control Conference*, volume 4, pages 3605–3610, June 2004.

[LH05] D. Liberzon and J. P. Hespanha. Stabilization of nonlinear systems with limited information feedback. *IEEE Transactions on Automatic Control*, 50(6):910–915, 2005.

[Lib02a] D. Liberzon. A note on stabilization of linear systems using coding and limited communication. In *Proc. of the* 41*st IEEE Conference on Decision and Control*, December 2002.

[Lib02b] D. Liberzon. Stabilization by quantized state or output feedback: A hybrid control approach. In *Proc. IFAC* 15*th Triennial World Congress*, 2002.

[Lib03] D. Liberzon. Hybrid feedback stabilization of systems with quantized signals. *Automatica*, 39:1543–1554, 2003.

[Lim05] J. Lim. Optimal bit allocation for noisy channels. *Electronics Letters*, 41(7), March 2005.

[Lin98] J. Lindèn. *Interframe quantization for noisy channels*. PhD thesis, Chalmers University of Technology, 1998.

[LL05a] Q. Ling and M. D. Lemmon. Optimal dynamic bit assignment in noise-free quantized linear control systems. In *Proc. of the* 44*th IEEE Conference on Decision and Control, and the European Control Conference 2005*, pages 8191–8196, December 2005.

[LL05b] Q. Ling and M. D. Lemmon. Stability of quantized control systems under dynamic bit assignment. *IEEE Transactions on Automatic Control*, 50(5):734–740, May 2005.

[LN07] D. Liberzon and D. Nešić. Input-to-state stabilization of linear systems with quantized state measurements. *IEEE Transactions on Automatic Control*, 52(5):767–781, 2007.

[Loo09] D. P. Looze. Linear-quadratic-Gaussian control for adaptive optics systems using a hybrid model. *Journal of the Optical society of America*, 26(1), January 2009.

[Mas90] J. Massey. Causality, feedback and directed information. In *Proc. of* 1990 *International Symposium on Information theory and its application*, pages 303–305, 1990.

[MFDN07] P. Minero, M. Franceschetti, S. Dey, and G. Nair. Towards feedback stabilization over fading channels. In *Proc. of the* 5*th International symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops*, pages 1–5, April 2007.

[MN93] S. W. Mclaughlin and D. L. Neuhoff. Asymptotic quantization for noisy channels. In *Proc. of 1993 IEEE International Symposium on Information Theory*, page 442, January 1993.

[MR06] C. R. Murthy and B. D. Rao. High-rate analysis of source coding for symmetric error channels. In *Data Compression Conf. (DCC)*, March 2006.

[MS98] F. J. Macwilliams and N. J. A. Sloane. *The theory of error-correcting codes*. North Holland, 1998.

[MS02] A. S. Matveev and A. V. Savkin. Optimal LQG control via limited capacity communication networks. In *Proc. of the* 41*st IEEE Conference on Decision and Control*, December 2002.

[MS04a] A. S. Matveev and A. V. Savkin. An analogue of Shannon information theory for networked control systems: Stabilization via a noisy discrete channel. In *Proc. of the* 43*nd IEEE Conference on Decision and Control*, volume 4, pages 4485–4490, December 2004.

[MS04b] 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.

[MS05] A. S. Matveev and A. V. Savkin. Shannon zero error capacity and the problem of almost sure observability over noisy communication channels. In *Proc. of the* 44*th IEEE Conference on Decision and Control*, pages 3219–3224, December 2005.

[MS06a] 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 *Proc. of the 2006 IEEE International Conference on Control Applications*, pages 2072–2077, October 2006.

[MS06b] A. S. Matveev and A. V. Savkin. Stabilization of stochastic linear plants via limited capacity stochastic communication channels. In *Proc. of the* 45*th IEEE Conference on Decision and Control*, pages 484–489, December 2006.

[MS07] A. S. Matveev and A. V. Savkin. An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels. *SIAM J. Control Optm.*, 46(4):1323–1367, 2007.

[MT06a] A. Mahajan and D. Teneketzis. Fixed delay optimal joint source–channel coding for finite–memory systems. In *Proc. of 2006 IEEE International Symposium on Information Theory*, pages 2319–2323, July 2006.

[MT06b] A. Mahajan and D. Teneketzis. Optimal performance of feedback control systems with limited communication over noisy channels. In *Proc. of the* 45*th IEEE Conference on Decision and Control*, pages 3228–3235, July 2006.

[NB06] G. N. Nair and J. Baillieul. Time to failure of quantized control via a binary symmetric channel. In *Proc. of the* 45*th IEEE Conference on Decision and Control*, pages 2883–2888, December 2006.

[NE02] G. N. Nair and R. J. Evans. Mean square stabilisability of stochastic linear systems with data rate constraints. In *Proc. of the* 41*st IEEE Conference on Decision and Control*, December 2002.

[NE03] G. N. Nair and R. J. Evans. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM Jour. Contr. Opt.*, February 2003.

[NEMM04] 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.

[NFZE07] G. N. Nair, F. Fagnani, S. Zampieri, and R. Evans. Feedback control under data rate constraints: an overview. In *Proc. of the IEEE*, pages 108–137, January 2007.

[NN95] S. Na and D. L. Neuhoff. Bennett's integral for vector quantizers. *IEEE Transactions on Information Theory*, 41(4):886–900, 1995.

[PB05] B. Picasso and A. Bicchi. Control synthesis for practical stabilization of quantized linear systems. *Rend. Sem. Mat. Univ. Pol. Torino, Control Theory and Stabil., I*, 63(4):397–410, 2005.

[PB07] B. Picasso and A. Bicchi. On the stabilization of linear systems under assigned I/O quantization. *IEEE Transactions on Automatic Control*, 52(10):1994–2000, October 2007.

[PB08] B. Picasso and A. Bicchi. Hypercubes are minimal controlled invariants for discrete-time linear systems with quantized scalar input. *Int. J. Nonlinear Analysis*, (2):706–720, 2008.

[PBC+07] S. D. Patek, M. D. Breton, Y. Chen, C. Solomon, and B. Kovatchev. Linear quadratic Gaussian-based closed-loop control of type 1 diabetes. *Journal of Diabetes and Technology*, 2007.

[Per04] C. De Persis. Results on stabilization on nonlinear systems under finite data-rate constraints. In *Proc. of the* 43*rd IEEE Conference on Decision and Control*, pages 3986–3991, December 2004.

[PGB02] B. Picasso, F. Gouaisbaut, and A. Bicchi. Construction of invariant and attractive sets for quantized-input linear system. In *Proc. of the 41th IEEE Conference on Decision and Control*, volume 1, pages 824–829, December 2002.

[PKA04] N. J. Ploplys, P. A. Kawka, and A. G. Alleyne. Closed-loop control over wireless networks. *IEEE Control Systems Magazine*, pages 58–71, June 2004.

[PPBJ04] B. Picasso, L. Palopoli, A. Bicchi, and K. H. Johansson. Control of distributed embedded systems in the presence of unknown-but-bounded noise. In *Proc. of the* 43*nd IEEE Conference on Decision and Control*, December 2004.

[Pro95] J. G. Proakis. *Digital Communications,* 3*rd ed.* McGrawHill, 1995.

[PS98] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover, 1998.

[RRM03] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, 2003.

[RS76] N. Rydbeck and C. E. Sundberg. Analysis of digital errors in nonlinear PCM systems. *IEEE Transactions on Communications*, 24(1):59–65, 1976.

[RU02] T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2002.

[Sah00] A. Sahai. Evaluating channels for control: Capacity reconsidered. In *Proc. of the American Control Conference*, June 2000.

[Sah04] A. Sahai. The necessity and sufficiency of anytime capacity for control over a noisy communication link. In *Proc. of the* 43*rd IEEE Conference on Decision and Control*, volume 2, pages 1896–1901, December 2004.

[Sav06] A. V. Savkin. Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control. *Automatica*, 42:51–62, 2006.

[Seg76] A. Segall. Bit allocation and encoding for vector sources. *IEEE Transactions on Information Theory*, 22(2):162–169, 1976.

[SG86] M. J. Sabin and R. M. Gray. Global convergence and empirical consistency of the generalized Lloyd algorithm. *IEEE Transactions on information theory*, 32(2):148–155, March 1986.

[SH87] H. H. Szu and R. L. Hartley. Nonconvex optimization by fast simulated annealing. In *Proc. of the IEEE*, volume 75, pages 1538–1540, June 1987.

[SH94] M. Skoglund and P. Hedelin. Vector quantization over a noisy channel using soft decision decoding. In *Proc. of IEEE Int. Conf. of Acoustics, Speech and Signal Processing*, volume V, April 1994.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, July 1948.

[Sha59] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.,*, pages 142–163, 1959.

[Sim04] T. Simsek. *Anytime Capacity Coding with Feedback*. PhD thesis, University of California at Berkeley, 2004.

[SJV04] T. Simsek, R. Jain, and P. Varaiya. Scalar estimation and control with noisy binary observation. *IEEE Transactions on Automatic Control*, 49(9):1598–1603, September 2004.

[Sko97] M. Skoglund. *On soft decoding and robust vector quantization*. PhD thesis, Chalmers University of Technology, 1997.

[Sko99a] M. Skoglund. Hadamard-based soft decoding for vector quantization over noisy channels. *IEEE Transactions on Information Theory*, 45(2):515–532, March 1999.

[Sko99b] M. Skoglund. Soft decoding for vector quantization over noisy channels with memory. *IEEE Transactions on Information Theory*, 45(4):1293–1307, May 1999.

[SLT04] S. Seth, J. P. Lynch, and D. M. Tilbury. Feasibility of real-time distributed structural control upon a wireless sensor network. In *Proc. of the 42nd Allerton Conference on Communication, Control, and Computing*, September 2004.

[SM06] A. Sahai and S. Mitter. The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link - part I, scalar systems. *IEEE Transactions on Information Theory*, 52(8):3369–3395, 2006.

[Söd02] T. Söderström. *Discrete-time Stochastic Systems*. Springer, 2002.

[SV03] T. Simsek and P. Varaiya. Noisy data-rate limited estimation: Renewal codes. In *Proc. of the 42nd IEEE Conference on Decision and Control*, December 2003.

[Tat00] S. Tatikonda. *Control under communication constraints*. PhD thesis, Massachusetts Institute of Technology, 2000.

[Tat08] S. Tatikonda. The structure of encoders for channels with feedback. In *Proc. of 2008 IEEE International Symposium on Information Theory*, pages 722–726, July 2008.

[TBS75] E. Tse and Y. Bar-Shalom. Generalized certainty equivalence and dual effect in stochastic control. *IEEE Transactions on Automatic Control*, 20:817–819, 1975.

[Ten06] D. Teneketzis. On the structure of optimal real-time encoders and decoders in noisy communication. *IEEE Transactions on Information Theory*, 52(9), September 2006.

[The57] H. Theil. A note on certainty equivalence in dynamic planning. *Econometrica*, 25:346–349, 1957.

[TM04a] S. Tatikonda and S. Mitter. Control over noisy channels. *IEEE Transactions on Automatic Control*, 49(7):1196–1201, September 2004.

[TM04b] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7):1056–1068, July 2004.

[TM09] S. Tatikonda and S. Mitter. The capacity of channels with feedback. *IEEE Transactions on Information Theory*, 55(1):323–349, 2009.

[TSM04] S. Tatikonda, A. Sahai, and S. Mitter. Stochastic linear control over a communication channel. *IEEE Transactions on Automatic Control*, 49(9):1549–1561, September 2004.

[VV95] S. Vembu and S. Verdú. The source-channel separation theorem revisited. *IEEE Transactions on Information Theory*, vol.41(1), January 1995.

[WB97] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints - part I: state estimation problem. *IEEE Transactions on Automatic Control*, 42(9), September 1997.

[WB99] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints - part II: Stabilization with limited information feedback. *IEEE Transactions on Automatic Control*, 44(9), September 1999.

[Wic95] S. B. Wicker. *Error control systems for digital communication and storage*. Prentice-Hall, Inc., 1995.

[Wid61] B. Widrow. Statistical analysis of amplitude-quantized sampled-data systems. *Amer. Inst. Elect. Eng., Pt.II (Appl. Ind.)*, pages 555–568, 1961.

[Wit71] H. S. Witsenhausen. Separation of estimation and control for discrete time systems. In *Proc. of the IEEE*, volume 59, pages 1557–1566, November 1971.

[WKL96] B. Widrow, I. Kollár, and M. Liu. Statistical theory of quantization. *IEEE Transactions on Instrumentation and Measurement*, 45(2), April 1996.

[WPJ+07] E. Witrant, G. P. Park, M. Johansson, C. Fischione, and K. H. Johansson. Predictive control over wireless multi-hop networks. In *IEEE International Conference on Control Applications*, pages 1037–1042, 2007.

[WW81] H. Van De Water and J. C. Willems. The certainty equivalence property in stochastic control theory. *IEEE Transactions on Automatic Control*, vol.26(5), October 1981.

[XJH+05] L. Xiao, M. Johansson, H. Hindi, S. Boyd, and A. Goldsmith. Joint optimization of wireless communication and networked control systems. *Chapter in Switching and Learning, Springer Lecture Notes in Computer Science 3355*, pages 248–272, September 2005.

[Xu05] Q. Xu. Anytime capacity of the AWGN+erasure channel with feedback. Technical report, University of California at Berkeley, Berkeley, 2005.

[YB04] S. Yüksel and T. Basar. Minimum rate coding for state estimation over noiseless channels. In *Proc. of the* 43*rd IEEE Conference on Decision and Control*, December 2004.

[YKT07] S. Yang, A. Kavčić, and S. Tatikonda. On the feedback capacity of power-constrained Gaussian noise channels with memory. *IEEE Transactions on Information Theory*, 53(3):929–954, 2007.

[ZG90] K. Zeger and A. Gersho. Pseudo-Gray coding. *IEEE Transactions on Communications*, vol.38(12):2147–2158, 1990.

[ZL08] H. Zhu and K. J. R. Liu. *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*. Cambridge University Press., 2008.

[ZM94] K. Zeger and V. Manzella. Asymptotic bounds on optimal noisy channel quantization via random coding. *IEEE Transactions on Information Theory*, 40(6), November 1994.