ROYAL INSTITUTE
OF TECHNOLOGY

Source–Channel Coding for Closed-Loop Control

Lei Bao

TRITA–EE 2006:013
ISSN 1653-5146

Communication Theory Lab
School of Electrical Engineering
Royal Institute of Technology
Stockholm, Sweden, 2006

Licentiate Thesis
Stockholm, Sweden, 2006

Source–Channel Coding for Closed-Loop Control

Communication Theory Laboratory
School of Electrical Engineering
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, Sweden

Tel. +46 8 790 8441, Fax. +46 8 790 7260
http://www.ee.kth.se

Abstract

Networked embedded control systems are present almost everywhere. A recent trend is to introduce wireless sensor networks in these systems, to take advantage of the added mobility and flexibility offered by wireless solutions. In such networks, the sensor observations are typically quantized and transmitted over noisy links. Concerning the problem of closed-loop control over such non-ideal communication channels, relatively few works have appeared so far. This thesis contributes to this field, by studying some fundamentally important problems in the design of joint source–channel coding and optimal control.

The main part of the thesis is devoted to joint design of the coding and control for scalar linear plants, whose state feedbacks are transmitted over binary symmetric channels. The performance is measured by a finite-horizon linear quadratic cost function. The certainty equivalence property of the studied systems is utilized, since it simplifies the overall design by separating the estimation and the control problems. An iterative optimization algorithm for training the encoder–decoder pairs, taking channel errors into account in the quantizer design, is proposed. Monte Carlo simulations demonstrate promising improvements in performance compared to traditional approaches.

Event-triggered control strategies are a promising solution to the problem of efficient utilization of communication resources. The basic idea is to let each control loop communicate only when necessary. Event-triggered and quantized control are combined for plants affected by rarely occurring disturbances. Numerical experiments show that it is possible to achieve good control performance with limited control actuation and sensor communication.

Keywords: optimal control, joint source–channel coding, joint coding and control, certainty equivalence, linear quadratic control, stochastic control, wireless sensor network.

Acknowledgments

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

First and foremost, I would like to express my sincere gratitude to my supervisors, Professor Mikael Skoglund and Associate Professor Karl Henrik Johansson. I own many thanks to Mikael for your guidance throughout the years. So much I have learnt from you. I greatly appreciate your generosity in sharing your expertise and time, your patience and pedagogic abilities. I'm grateful to Kalle for introducing me to the field of automatic control, for all of the inspiring discussions and the continuous encouragement. I am thoroughly impressed by your endless energy and great enthusiasm for scientific research.

I would also like to thank all my current and former colleagues at Communication Theory, Signal Processing and Automatic Control Labs for the pleasant and warm atmosphere. It has been a lot of fun to be here. Especially, I want to express my appreciation to Niklas Wernersson for all the valuable comments on the first draft of this manuscript and to Thanh Tùng Kim for being an excellent office neighbor who always provides answers to my questions. I'm also grateful to Karin Demin, Karin Karlsson Eklund and the ladies at STEX for helping with administrative issues, to Nina Unkuri and Andreas Stenhall for assisting computer issues.

I'm indebted to all my friends, especially to Yi, Xi and Kai. You make my life in Stockholm enjoyable and memorable. Warmest thanks go to my dear mom and sister for your great love throughout the years.

Last but not least, I want to thank Dr. Fredrik Gunnarsson who acts as the opponent on this thesis.

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 in 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 systems is the increasing demands 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 the available information at each distributed node and exchanging 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. 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 no delays and no transmission errors in the links between the plant

and the controller. Advanced methods and 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. In such a network, the sensor observations are typically quantized and transmitted over noisy links. Also difficulties, such as data delays and data drops, are encountered. Concerning control over non-ideal communication links, little work has been performed so far. Development of methods and tools for the analysis and synthesis of feedbacks 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, especially the quantization granularity, deteriorates the signals transmitted between the plant and the controller. This can potentially degrade the overall system performance. 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. This, since they are easy to implement. However, for applications with low rate and high communication costs, it is natural to study closer-to-optimal solutions.

Transmission errors are unavoidable in communications over unreliable media, for example wireless networks. Therefore, robustness to transmission errors is one of the fundamental requirements of all modern communication systems. Concerning control applications, 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. A joint design, with a constraint on the codeword length, which combines the source compression and the channel protection can therefore be suitable. In this thesis we will study the joint design of coding and control for an efficient use of the available communication resources.

Event-triggered control strategies are considered as another promising solution to the problem of efficient utilization of communication resources. The basic idea is to let each control loop communicate as infrequently as possible. It is believed that event-triggered controls can potentially be more efficient than the conventional time-triggered control in many

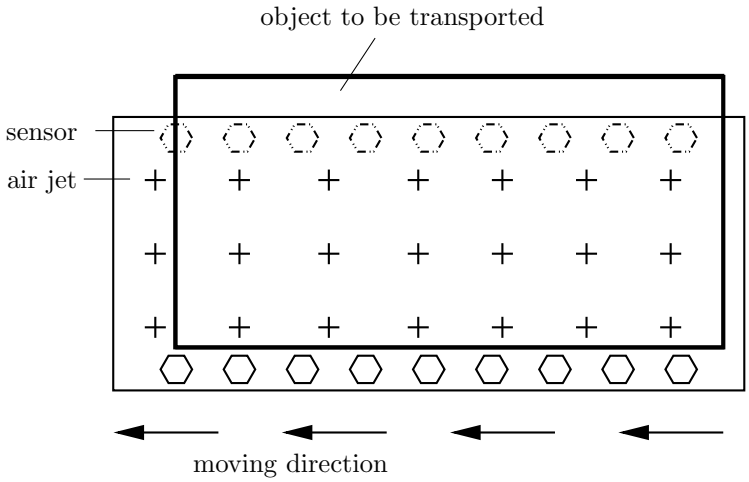


Figure 1.1: An example of a sensor network in motion control of a non-contact objective in industrial manufacturing.

cases. However, how to schedule a number of distributed subsystems is still an open problem. As an application of our proposed joint coding–control strategy, we combine event-triggered and quantized control for plants affected by rarely occurring disturbances.

1.2 Motivating Examples

Before presenting the particular control problem studied in this thesis, let us first consider two examples of sensor networks in control applications.

Example 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. Fig. 1.1 illustrates an example, inspired by [BBea00], where a planar object is transported by a number of air jets. In order 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

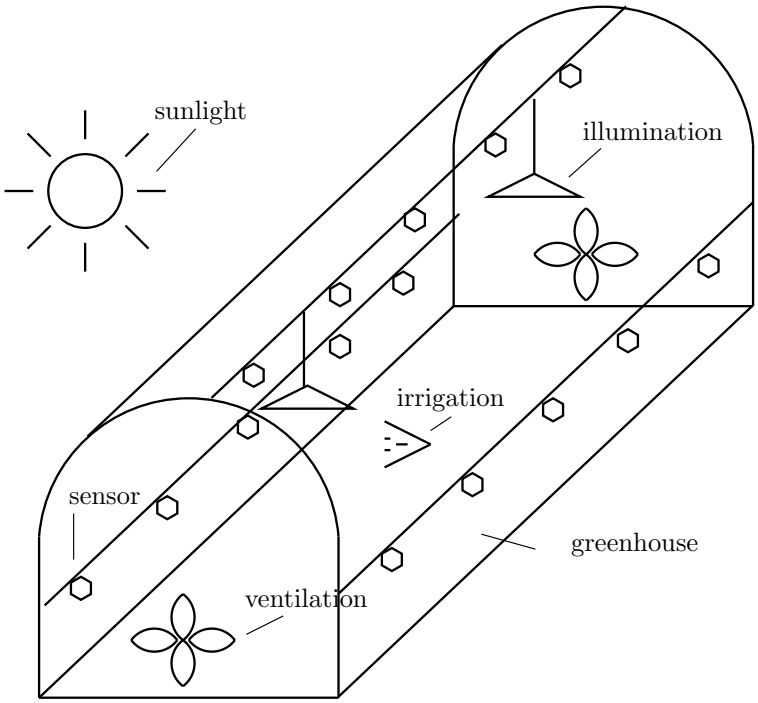


Figure 1.2: An example of a sensor network in environmental control of greenhouse vegetable production.

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. In this case, 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 2 Environmental Control of Greenhouses
The following example, inspired by [Jon01], illustrates the future use of control over sensor networks in agriculture applications. The new generation of greenhouses will support automatic environmental regulations. Fig. 1.2 depicts such a greenhouse, which exploits a feedback control system to maintain a perfect growing environment for plants. More precisely,

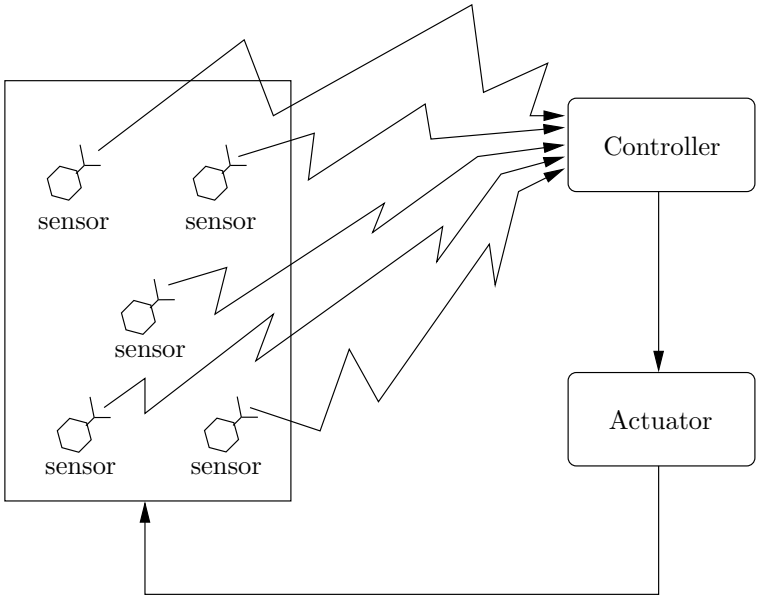


Figure 1.3: Control system utilizes data from wireless sensor network. The results of this thesis can be applied to optimize the use of the wireless medium.

the "perfect" environment is specified by a number of so called primary variables, such as temperature, humidity, light, CO_2 levels etc. The control task is to keep the primary variables within some 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. In order to have a general picture of the problems particularly studied in this thesis, let us consider the wireless networked control system shown in Fig. 1.3. The system consists of a number of sensor nodes that are connected 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 rarely occurring local disturbances. The control commands for keeping the states around the equilibrium working points are executed through a common actuator, therefore all the sensors report their state measurements to a common control node. 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. In order to efficiently utilize communication resources, it is especially interesting to study the case where each sensor transmits only few symbols and each symbol consists of few bits.

The goal is to design the encoding (compression and/or quantization, and error control), the decoding and the control strategy jointly. A decentralized approach is as follows: let the control command corresponding to a specific sensor be zero as long as no disturbance is detected. Once the sensor detects a disturbance, the sensor measurement is encoded and transmitted to the control node. The message is decoded at the control node and a control command is derived and actuated to counteract the specific disturbance observed at the transmitting sensor.

In this thesis, a stochastic control problem for the system described in Fig. 1.3 is formulated. Before moving on to the problem formulation, a short introduction to the topics of stochastic control theory and coding over noisy channels is given in the remaining part of this chapter.

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 technique 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 equivalency property*, which describes the separation of state estimation and control. There exists a rich literature in stochastic control theory, e.g., [Åst70, Ber95, BS96, Söd03].

1.3.1 Dynamic Programming

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

A simple stochastic control problem for a scala plant (of dimension one) is formulated below. The state space model of the plant is given by

$$\begin{cases} x_{t+1} &= \theta_t(x_t, u_t, v_t), \\ y_t &= h_t(x_t, e_t), \end{cases} \tag{1.1}$$

where θ_t is the system function and h_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 the cost function,

$$J = \mathbf{E}\left\{\sum_{t=0}^{T} l(x_t, y_t, u_t)\right\}, \tag{1.2}$$

where T denotes the finite time horizon and l denotes a function, which measures the instantaneous system performance. Finally, $\mathbf{E}$ is the expectation operator.

Let us review some results [Åst70,Söd03], which will simplify the studied optimization problem. For brevity, the time index t will be ignored for a moment. When having the complete state information, i.e., both x and y are available, the following result has been proven: assuming the function $l(x, y, u)$ has a unique minimum as a function of the control signal, at $u^\star(x, y)$, the minimization and the expectation are commutative, i.e.,

$$\min_{u(x,y)} \mathbf{E}\left\{l(x, y, u)\right\} = \mathbf{E}\left\{l(x, y, u^\star(x, y))\right\} = \mathbf{E}\left\{\min_u l(x, y, u)\right\}. \tag{1.3}$$

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

$$\begin{aligned} \min_{u(y)} \mathbf{E}\left\{l(x, y, u)\right\} &= \mathbf{E}\left\{l(x, y, u^\star(y))\right\} \\ &= \mathbf{E}_y\left\{\min_u \mathbf{E}\left\{l(x, y, u)|y\right\}\right\}, \end{aligned} \tag{1.4}$$

by assuming $\mathbf{E}\{l(x,y,u)|y\}$ has 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 strategy for the truncated problem whose cost is a summation from time s to T. Let the notation $\mathbf{z}_a^b=\{z_a,\ldots,z_b\}$ describe the evolution of a discrete-time signal z_t from $t=a$ to $t=b$. Based on the principle of optimality, the optimal u_t is the one that minimizes a sum of the future costs:

$$u_t^\star=\arg\min_{u_t}\mathbf{E}\left\{\sum_{s=t}^{T}l(x_s,y_s,u_s)|\mathbf{y}_0^t\right\}, \tag{1.5}$$

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

$$V^\star(\xi(t),t)=\min_{\mathbf{u}_t^T}\mathbf{E}\left\{\sum_{s=t}^{T}l(x_s,y_s,u_s)|\mathbf{y}_0^t\right\}, \tag{1.6}$$

where $\xi(t)$ denotes the conditional probability $p(x_t|\mathbf{y}_0^{t-1})$. $V^\star(\xi(t),t)$ is the result of the optimal control sequence $\{u_0^\star,\ldots,u_T^\star\}$. Then, the optimal control function $u_t^\star$ at time t can be derived by solving the following functional equation,

$$V^\star(\xi(t),t)=\min_{u_t}\mathbf{E}\left\{l(x_t,y_t,u_t)+V^\star(\xi(t+1),t+1)|\mathbf{y}_0^t\right\}. \tag{1.7}$$

1.3.2 Certainty Equivalence

A certainty equivalence (CE) controller [BST74, The57, TBS75, WW81], is obtained by replacing the full state observation x_t in the optimal deterministic solution, where the process noise is absent and the perfect state observations are available, with the state estimate $\hat{x}_{t|t}$, that is,

$$u_t^{OpD}=\phi_t(x_t),\quad u_t^{CE}=\phi_t(\hat{x}_{t|t}), \tag{1.8}$$

where $\phi_t(x_t)$ denotes the optimal deterministic solution (OpD) and $\hat{x}_{t|t}$ is the conditional mean estimate $\mathbf{E}\{x_t|\mathbf{y}_0^t,\mathbf{u}_0^{t-1}\}$. The equation (1.8) depicts a clear separation between the estimation $\hat{x}_{t|t}$ and the control ϕ_t. The CE controller is favored in stochastic control for its practicability.

Sometimes, a CE controller is also the optimal control strategy, which is often termed *certainty equivalency property*. However, a CE controller is in general only a suboptimum solution with few exceptions. The most well known exception is linear systems with quadratic costs and Gaussian distributed random variables (i.e., initial state x_0, process noise v_t and measurement noise e_t). Moreover, the variables are independent and a so-called *classical information pattern* is required, which means all past measurements are known to the controller. For this example, the separation in (1.8) applies and the optimal control is a linear function of the conditional mean estimate:

$$u_t^\star = -\ell_t \mathbf{E}\left\{x_t | \mathbf{y}_0^t\right\}, \tag{1.9}$$

where $\mathbf{E}\{x_t|\mathbf{y}_0^t\}$ denotes the expected x_t conditioned on all past measurements $\mathbf{y}_0^t$ and ℓ_t is the feedback gain.

Finally, we introduce the *separation property*, by which $\hat{x}_{t|t}$ is a *sufficient statistic* to derive the optimal control, such as

$$u_t^\star = \varphi_t(\hat{x}_{t|t}). \tag{1.10}$$

Separation property is a weak notion of the CE property, since φ_t is not necessary the optimal deterministic solution ϕ_t.

1.4 Elements of Source and Channel Coding

Recall the example in Fig. 1.3. The signal path from a sensor to the controller can be modeled as a point-to-point communication link. In Fig. 1.4, a block diagram of the elemental building blocks in a traditional communication system is given. *Source coding* and *channel coding* are presented briefly in Section 1.4.1 and Section 1.4.2. In Section 1.4.3, we give an introduction to the *source–channel separation theorem*.

1.4.1 Source Coding

Source coding deals with the compression of source data by removing the redundancy and irrelevancy in the data sequence. Based on whether or not the original data sequence can be reconstructed, the source coding methods can be divided into two categories. The first, referred to as *lossless coding*, requires the data processing to be reversible. For an independent and identically distributed (i.i.d.) random variable X, the

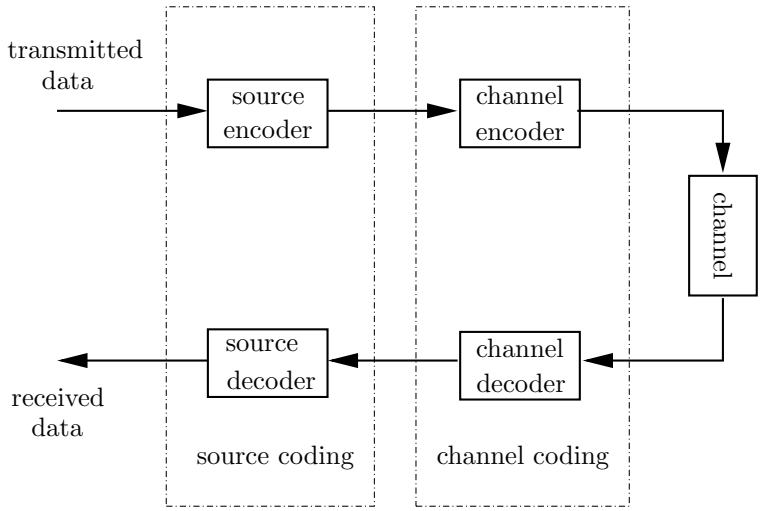


Figure 1.4: The functional diagram of a digital communication system.

average length of a uniquely decodable code is bounded by the *entropy rate* [CT91] of the source, which is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x), \tag{1.11}$$

where $p(x)$ is the probability mass function, $p(x) = \Pr\{X = x\}$. By using base 2 logarithm, $H(X)$ is measured in bits. In *lossy coding* (the second category), the restriction of the reversibility is relaxed. A measure of the distortion between the original data and the reconstructed data, $\varrho(X, \hat{X})$, is required for the optimization in the coder design. The distortion function ϱ will vary from application to application. In general, the main features of a proper distortion function are non negativity, physical meaningfulness and easy to calculate. The conflicting relation between a given distortion and a minimum rate is stated in a rate distortion function, see e.g., [CT91, Gal68, 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 possible values into a small set of discrete symbols. Scalar quantization operates

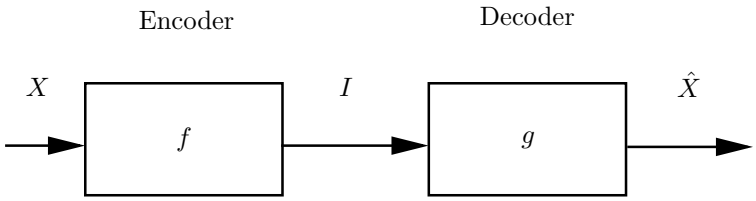


Figure 1.5: A source coding diagram. X is the random source symbol, I is the integer index and $\hat{X}$ is the estimate of X.

on scalar data, while vector quantization operates on multidimensional vectors. Fig. 1.5 depicts the source coding over a noise-free channel. Let X denote the source symbol, in particular the sample of a stationary ergodic stochastic process. The encoder function f performs a mapping of each source symbol (scalar) to an integer index: $I = f(X)$. The integer set $\mathcal{I} = \{0, 1, \ldots, 2^R - 1\}$, where R denotes the rate, contains all possible indices. The *encoder region* S_i is the set that contains all source symbols assigned index i:

$$S_i = \{x | f(x) = i\}, \quad i \in \mathcal{I}, \tag{1.12}$$

i.e., $X \in S_i \Leftrightarrow I = i$. The task of a decoding function g is to reconstruct an estimate $\hat{X}$ of the source symbol X, based on the received index, in particular, $\hat{X} = g(I)$. We define the *reconstruction point* q_i, which is associated with index i through

$$I = i \Rightarrow \hat{X} = g(i) = q_i. \tag{1.13}$$

The set of all reconstruction points is termed a *codebook*.

A pdf-optimized quantizer provides a set of reconstruction points and their corresponding encoder regions based on the statistics of the source symbol. The optimality is typically measured by minimizing a distortion D (or a cost), e.g., the *mean squared error* (MSE):

$$D_{MSE} = \mathbf{E}\left\{(X - \hat{X})^2\right\}. \tag{1.14}$$

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. Generally speaking, a discrete memoryless channel can be described by a conditional probability function $p(j|i)$, where i and j are input and output indices to the channel. The channel capacity for a stationary discrete memoryless channel is obtained as

$$\begin{aligned} C &= \max_{p(i)} \mathrm{I}(I;J) \\ &= \max_{p(i)} \left\{ \sum_{i\in\mathcal{I}, j\in\mathcal{J}} p(j|i)p(i) \frac{p(j|i)}{\sum_{i\in\mathcal{I}} p(j|i)p(i)} \right\}, \end{aligned} \tag{1.15}$$

where $\mathrm{I}(I;J)$ 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 [MS98, RU02, Wic95] are good introductions to the subject of channel coding.

1.4.3 The Source–Channel Separation Theorem

The source–channel separation theorem states that under certain conditions the combining of separately designed source and channel codes will still achieve the optimal performance for transmission over noisy channels. For lossless coding, it particularly states that there exists a source–channel code to transmit a stationary ergodic source with arbitrary low probability of error if the source satisfies the *asymptotic equipartition property*, e.g., [CT91], and 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. The asymptotic equipartition property concerns the probability of so-called "typical sequences". Due to the asymptotic equipartition property, 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). The output of the source encoder can be viewed as a new source sequence, where all source symbols are nearly equiprobable. Thereafter, by using channel codes of very long block lengths, the probability of error in the transmission approaches to zero. 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 wide classes of sources and channels can be found in e.g., [CT91, VV95].

There are obvious advantages to separating the source coding and channel coding problems. The separation theorem has served as a motivation 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 particular channel coding related topics: non-redundant channel coding (Section 1.4.4) and channel optimized quantization (Section 1.4.5).

1.4.4 Non-redundant Channel Coding

Index Assignment

As mentioned above, 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 zero redundant bit, where the designer is confronted with the problem of carefully labelling the codewords, referred to as the *index assignment* (IA). The index assignment deals with the combinatorial optimization problem of assigning non-redundant codewords to quantizer reconstruction points. Throughout this chapter, we assume that the reconstruction points are labelled by the indices $i \in \mathcal{I}$. The index assignment in our case describes the mapping between the reconstruction points and a set of binary codewords. Let each reconstruction point q_i have an associated integer index i and each index i is mapped to a codeword through a deterministic function $c(i)$. For brevity, $c(i)$ is also used as the notation for the resulting codeword, i.e., the reconstruction point q_i is assigned the codeword $c(i)$. The codewords are different for different indices, i.e., $c(i) = c(j)$ if $i = j$, and $c(i) \neq c(j)$, $\forall i \neq j$, $i, j \in \mathcal{I}$. In Fig. 1.6 we illustrate the signal path from the source encoder input to the channel decoder output, when the index assignment is used. Since a binary symmetric channel is considered in this example, $c(i)$ and $c(j)$ denote the binary transmitted and received codewords, respectively. Index assignment has been shown to be important for transmission over noisy channels. It is

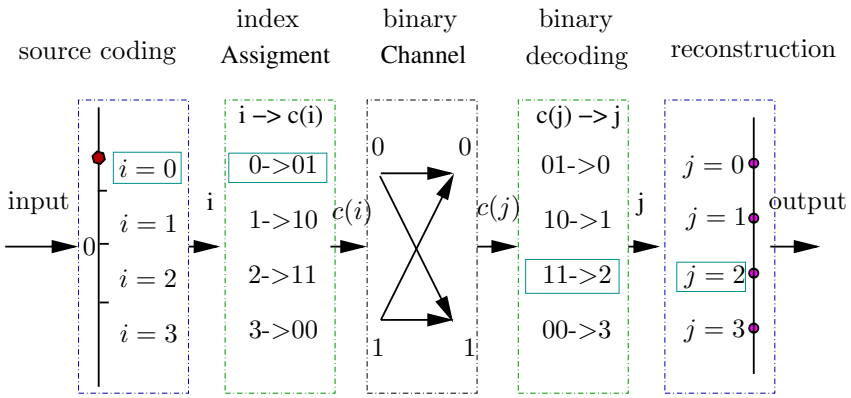


Figure 1.6: A diagram of data transmission over a binary symmetric channel. As an example, the index $i = 0$ renders to transmit the binary codeword 01, whereas the codeword 11 is received due to the noise in the channel. The decoded index is $j = 2$.

widely accepted that a carelessly designed index assignment will reduce the system performance severely. The varied ability to combat the channel errors is attributed to the different conditional transition probabilities among the binary codewords. For more research results on this topic, see e.g., [Kna93, KA96, RS76, SH94, ZG90]. We illustrate the importance of a carefully designed index assignment by giving a numerical example.

Example 3 Index Assignment
The input to the quantizer is drawn according to the generalized Gaussian distribution (GGD), see Appendix A, with parameters $\alpha = 4$ and $\beta = 2$. The quantizer has a rate $R = 2$ bits per source symbol, which induces 4 reconstruction points $\{q_0, \ldots, q_3\}$, 4 codewords and 24 different index assignments. A binary symmetric channel (BSC) (see Section 2.2) is adopted and the transition probabilities are $\epsilon = 0.1$ and $\epsilon = 0.2$. Due to the symmetries in the channel, the quantizer and the source distribution, some index assignments are given identical transition probabilities $p(q_j|q_i)$, where q_i and q_j are the transmitted and received reconstruction points, respectively. It turns out that the 24 different assignments can be divided into three groups, see Fig. 1.7, that each group will have the same channel transition probability. In Fig. 1.8 the bit error patterns of all three groups are listed, which tell the number of errors needed in a

$G1$	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
q_0 :	11	11	10	10	01	01	00	00
q_1 :	00	00	01	01	10	10	11	11
q_2 :	01	10	00	11	00	11	01	10
q_3 :	10	01	11	00	11	00	10	01
$G2$	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
q_0 :	10	01	11	00	11	00	10	01
q_1 :	00	00	01	01	10	10	11	11
q_2 :	01	10	00	11	00	11	01	10
q_3 :	11	11	10	10	01	01	00	00
$G3$	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
q_0 :	10	01	11	00	11	00	10	01
q_1 :	00	00	01	01	10	10	11	11
q_2 :	11	11	10	10	01	01	00	00
q_3 :	01	10	00	11	00	11	01	10

Figure 1.7: The 24 different index assignments are divided into three groups G1 - G3, based on the transition probability $p(q_j|q_i)$. In each group there are 8 assignments, denoted No.1 - No.8. As examples, the index assignment No.1 in G1 will assign the binary codeword 11 to the reconstruction point q_0 and the binary codeword 00 to the reconstruction point q_1 etc.

certain codeword to result in a certain decision error. Finally, the performance comparison of the 24 index assignments are shown in Fig. 1.9, where on the y-axis, the mean square error between the original source symbol and the decoded symbol is presented. The simulation shows that the assignments are divided into three groups based on the performance, which agrees well with the grouping in Section 2.2. Index assignments in Group 3 outperform those of the other two groups. By studying the bit error patterns, one can have a rough idea about why the index assignments in Group 3 perform better than the others. The merit of Group 3 is probably the strong protection, i.e., 2 error bits induce a decision error, against a decision error between q_0 (the smallest reconstruction point) and q_3 (the largest reconstruction point). This type of transmission errors give large valued quantization errors.

Group 1

q_i \ q_j	q_0	q_1	q_2	q_3
q_0	0	2	1	1
q_1	2	0	1	1
q_2	1	1	0	2
q_3	1	1	2	0

Group 2

q_i \ q_j	q_0	q_1	q_2	q_3
q_0	0	1	2	1
q_1	1	0	1	2
q_2	2	1	0	1
q_3	1	2	1	0

Group 3

q_i \ q_j	q_0	q_1	q_2	q_3
q_0	0	1	1	2
q_1	1	0	2	1
q_2	1	2	0	1
q_3	2	1	1	0

Figure 1.8: The number of bit errors in the codeword when there is a decoding error between the reconstruction points. As examples, using any index assignment in Group 1, transmitting q_1 and receiving q_0 is caused by 2 bit errors, while transmitting q_1 and receiving q_3 is caused by 1 bit error.

Simulated Annealing

In a convex optimization problem, given any initial point, it is straightforward to arrive to the global optimum by successively moving in the negative direction of the local gradient. When optimizing a nonconvex function, which may have multiple local minima, the gradient search method encounters the issue of convergence to a local minimum.

Index assignment is a combinatorial optimization problem, which belongs to the family of nonconvex optimization problems. An exhaustive search for a global optimum is often exceedingly computationally expensive. A fairly good local optimum can 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 so the defects in the material can be reduced. By observing the similarities between this metallurgic process and the optimization problem with multiple local optima, Kirkpatrick [KGV83] has developed an efficient algorithm to solve nonconvex optimization problems, see also [EHSW87]. About simulated annealing is used for the index assignment problem can be read in e.g., [Far90].

Before presenting the algorithm for using simulated annealing to design the index assignment, let us first introduce the permutation vector π, which is used to specify a certain index assignment. Let c_k be the binary presentation of the integer value $k \in \mathcal{I}$, i.e., c_k is the natural binary code. We define a state π as a permutation vector that specifies the mapping

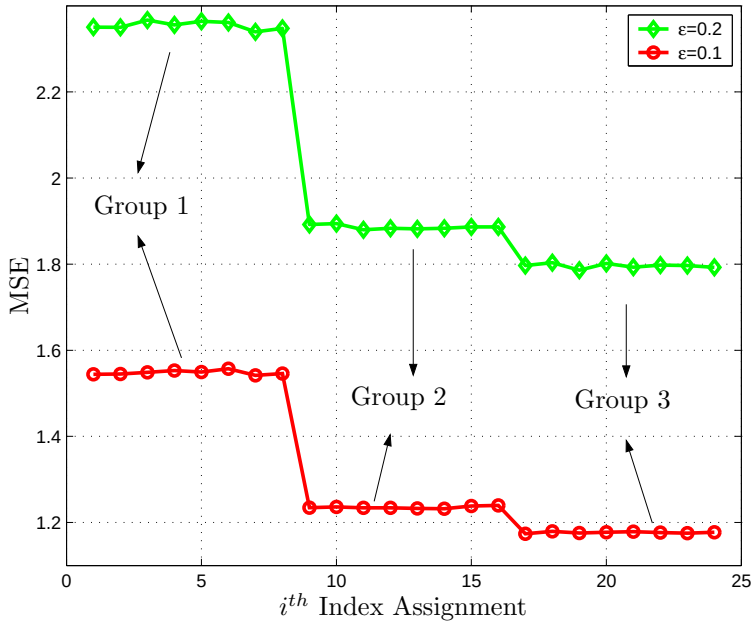


Figure 1.9: Comparison of costs among various index assignments.

from index i (as well the reconstruction point q_i) to the binary codeword c_k through $k = \pi(i)$. Fixing the permutation vector π, the deterministic mapping $c(i)$ is

$$c(i) = c_{\pi(i)}. \tag{1.16}$$

An example of the use of the permutation vector π is given in Fig. 1.10.

Returning to the simulated annealing algorithm, we define a cost J, which is a function of the state vector π. The goal of the optimization is to find the best state π, which minimizes the cost J. The overall optimization procedure is summarized in Algorithm 1. Note that at high temperatures where $e^{-\frac{\Delta J}{T}}$ approaches 1, the probability to accept a new index assignment is high. When the temperature decreases, it is more and more unlikely to adopt a new index assignment. Finally, we illustrate the perturbation in Algorithm 1 with Example 4.

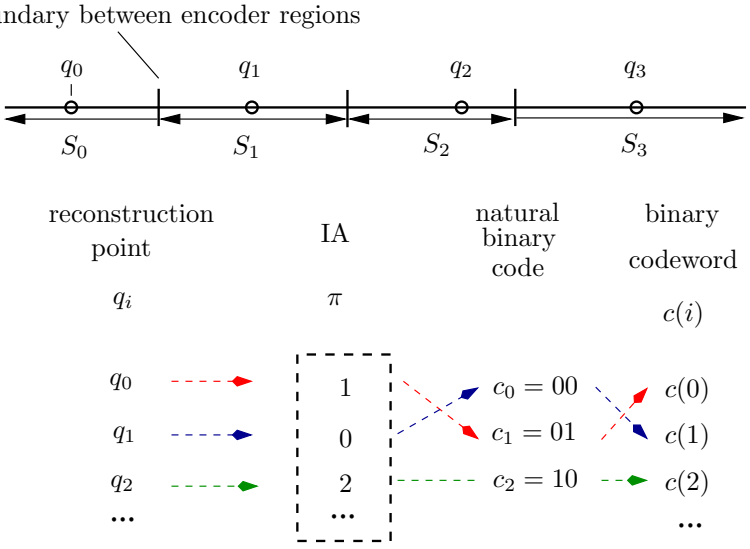


Figure 1.10: An illustration of the index assignment. q_i and S_i are the reconstruction point and the encoder region associated with the index i. c_i is the natural binary code for the integer i. $c(i)$ is the binary mapping of the index i. As an example, q_0 is the reconstruction point for $i = 0$, and it is coded to the binary codeword 01, which can also be described by the mapping $c(0) = 01$.

Example 4 Perturbation in Simulated Annealing
Given a quantizer with $R = 2$, the index i can take one of four values, $i \in \{0, 1, 2, 3\}$. The natural binary code is assigned as, $c_0 = 00$, $c_1 = 01$, $c_2 = 10$ and $c_3 = 11$. The permutation vector π can take one of 24 combinations. As an example, assume $\pi = (1023)$. The smallest reconstruction point will be encoded to $c(0) = c_1 = 01$. In an increasing order, the remaining reconstruction points will be coded to $c(1) = c_0 = 00$,$c(2) = c_2 = 10$, and $c(3) = c_3 = 11$. A perturbation $\bar{\pi}$ to π is a new permutation vector obtained by a position exchange between two entries in π. As an example, $\bar{\pi} = (1203)$ will result in an exchange between $c(1)$ and $c(2)$, i.e., $c(1) = c_2 = 10$ and $c(2) = c_0 = 00$.

Algorithm 1: Simulated Annealing

1. Select a cost function J. Define an effective temperature τ. Set $\tau = \tau_0$, where τ_0 is a high initial temperature.
2. Randomly choose an initial state π and calculate $J(\pi)$.
3. Select a new state $\bar{\pi}$ by randomly changing two entries in π.
4. Let $\Delta J = J(\bar{\pi}) - J(\pi)$,
   4a. if $\Delta J < 0$, replace π by $\bar{\pi}$;
   4b. if $\Delta J \geq 0$, replace π by $\bar{\pi}$ with probability $e^{-\frac{\Delta J}{\tau}}$.
5. If the number of cost drops exceeds a prescribed number or if too many unsuccessful perturbations occur, go to Step 6. Otherwise return to Step 3.
6. Lower the temperature. If the temperature is below some prescribed freezing temperature τ_f, terminate the iterations. Otherwise return to Step 3.

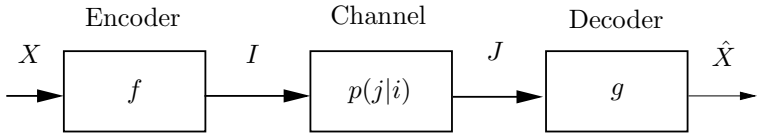


Figure 1.11: Quantization over a noisy channel. X and $\hat{X}$ are the source symbol and its estimate. I and J are the transmitted and received indices.

1.4.5 Channel Optimized Scalar Quantization

In the traditional separate design, the source coding block does not take the channel properties into account. It is well recognized that, in cases of short codewords, it is beneficial that a quantizer design takes the channel properties into consideration [FV87,Far90,FV90,Lin98,Sko97]. When the characteristics of the channel ($p(j|i)$) and the source ($p(x)$) are perfectly known, a design method referred to as *channel optimized quantization* (COQ) [FV87, Far90, FV90, Lin98, Sko97] can be used. The terminology quantizer is used here since the joint source–channel coder completes a mapping from an infinite set of continuous values to a finite set of discrete symbols. In Fig. 1.11 quantization over a noisy channel is depicted. Let X be the i.i.d. source symbol and the encoder function f performs a mapping of X to an integer index $I \in \mathcal{I}$. The encoder region associated with the index $i \in \mathcal{I}$ is defined as $S_i = \{x | f(x) = i\}$. Comparing to Fig. 1.5, I,

the output of the source encoder, is here fed into a channel described by the transition probability $p(j|i)$, where i and j are the transmitted and the received indices, respectively. The channel output $J \in \mathcal{I}$ will be used of the decoder to produce the estimate $\hat{X}$, in particular, $J = j \Leftrightarrow \hat{X} = q_j$, where q_j is the reconstruction point in the codebook.

In short, the aim of channel optimized quantization is to find the optimal encoder regions and reconstruction points that together minimize a certain distortion function, e.g., the MSE distortion,

$$\begin{aligned} D &= \mathbf{E}\left\{(X - \hat{X})^2\right\} \\ &= \sum_{i=0}^{2^R-1} \int_{S_i} p(x) \left(\sum_{j=0}^{2^R-1} p(j|i)(x - q_j)^2 \right) dx. \end{aligned} \tag{1.17}$$

It turns out that this type of optimization problems are difficult to solve. Up to now, the solution is often obtained by using a method similar to the Lloyd–Max algorithm, see [GG92]. The basic idea is to first fix the encoder and find the optimal decoder, then fix the decoder and find the best encoder. A local optimum is then obtained iteratively by alternating between the above two optimization procedures. As an example, we look at the necessary conditions for the solution to the minimum mean squared error (MMSE) problem. For a fixed encoder, the best reconstruction point $q_l^\star$ is

$$\begin{aligned} q_l^\star &= \mathbf{E}\{X|j = l\} \\ &= \frac{\sum_{k=0}^{2^R-1} p(j = l|i = k) \int_{S_i} xp(x)dx}{\sum_{k=0}^{2^R-1} p(j = l|i = k) \int_{S_i} p(x)dx}, \end{aligned} \tag{1.18}$$

while the best encoder region for fixed reconstruction points, is

$$\begin{aligned} S_l^\star = \Bigg\{ x : & \sum_{k=0}^{2^R-1} p(j = k|i = l)(x - q_k)^2 \\ & \leq \sum_{k=0}^{2^R-1} p(j = k|i = m)(x - q_k)^2, \quad \forall m \Bigg\}. \end{aligned} \tag{1.19}$$

For related topics in scalar and vector quantizer design for noisy channels (COSQ and COVQ), the interested readers are referred to [FM84, FV87, Far90, FV90, GG92, Lin98, Sko97].

Finally, we give a comment on the general use of the terms quantizer and encoder–decoder in this thesis. As shown above, in a communication context, an encoder–decoder pair is not necessarily a quantizer. For example a channel code typically will add redundant bits to the input codewords. However, when an encoder–decoder denotes a mapping from a large set of values to a small set of discrete symbols, it has functionality similar to a quantizer. In those cases, the encoder–decoder can be identified as a quantizer.

1.5 Control with Quantized Feedback

In this section, a brief introduction to the emerging research area of control with feedbacks over communication channels is given, by discussing a number of influential publications within this field.

The effect of a quantized feedback on the overall performance of a control system is not a new topic in the literature. However, until the end of the 1980's, quantization errors have mainly been modeled as additive white noise. Tools and methods from traditional stochastic control theory were applied to the quantized systems [Cur70].

A broad research interest has been evoked by the seminal paper of Delchamps [Del90]. In that paper the author has shown that, even for the simplest dynamic system (linear, scalar and noiseless), whenever the plant has its eigenvalue greater than 2, i.e., an unstable system, it is not possible to asymptotically stabilize the system with a finite data rate. Other interesting phenomena at low rates are limit cycles and chaotic behavior. By these observations, it has been realized 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 impacts of quantization effects in dynamic systems. Since then, topics of control designs for data-rate limited systems have 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. Before moving to a state-of-art survey, a few commonly used nomenclatures are first introduced.

The importance of *information patterns* in the analysis of system behavior and design of control strategies has been well recognized, see e.g., [Wit71]. The concept of information pattern is introduced to specify what information is available at each component in a connected system.

Several authors have in their work, see e.g., [Tat00], shown that information patterns play indeed a key role in the achievable performance for quantized control systems. Regarding the information patterns at the encoder, decoder and controller, the term *static* refers to a time-invariant function, while *dynamic* refers to a time-varying function. A static quantizer is memoryless, while for a dynamic quantizer, the memory access pattern varies, e.g., from *finite memory* to *infinite memory*.

The study of the asymptotic behavior 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 for estimations of the state; and *asymptotic stability* which describes the asymptotic behavior of the controlled state signals, see e.g., [Tat00]. The result of Delchamps has exposed that asymptotic stability cannot be achieved by using a static (memoryless) quantizer. Instead, *practical stability* [WB97] is formulated, where the state trajectories are only required to be bounded within a certain region.

In the remaining part of this section, several well studied problems are presented. Some recent results concerning stabilities are addressed in Section 1.5.1. A static quantizer is an example of an entity, which is easy to build, but not transparent to understand. The study of their asymptotic behavior involves advanced mathematics topics, such as ergodic theory, chaotic theory, symbolic theory etc. In Section 1.5.2, an introduction to the research of static (memoryless) quantizers will be given. In Section 1.5.3, a few issues regarding quantization in linear quadratic (LQ) control are discussed. Finally, a couple of information theoretic results are revealed in Section 1.5.4.

1.5.1 Stability and State Estimation

Since the main objective for control is to deal with uncertainty, the majority of control problems deal with various stability issues. Regarding control with quantized feedback, a fundamental problem that has aroused a lot of attention during the last decades is to find the smallest feedback data rate necessary to asymptotically stabilize an unstable dynamic system. In the sequel, the expression *minimum rate* refers to as the abovementioned smallest data rate. It has been shown that, in order to attain the asymptotic stability, the quantizer must be dynamic and have infinite reconstruction levels [Del90].

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 be addressed as uncertainty "volume"), due to unstable poles, should be counteracted by the "volume" reduction along the coding. 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 scalar linear plant ($x_{t+1} = ax_t + u_t$, where a describes the linear dynamics) can be found in e.g., [Tat00]. The author has proved that, 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., [NE02, Tat00]. The most common way (not necessarily the best) to tackle a multidimensional system is to transform the system matrix $\mathbf{A}$ (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 $R > \sum_s \log_2 |\lambda_s^u|$, where λ^u denotes the unstable eigenvalue of the system matrix $\mathbf{A}$. A realistic data rate should take on integer values. In [LL05a, LL05b], the authors have proposed a practical coding scheme by using 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, with fairly mild assumptions on process noise, a new lower bound of minimum rate has been derived in [NE03], based on differential entropy power.

Extensions to nonlinear systems can be found in [Bai04, LE04, Lib02b, Lib03, LH05, NEMM04, Per04]. In [Lib02b], the author has developed a zooming strategy to a nonlinear time-invariant system with an unknown initial value. Actually, the author has investigated two quantization problems. In the first, the quantizer is located at the observation link. In the second, the quantizer is located at the actuation link. The conditions for the global asymptotic stability are derived for both cases.

The synthesis of a feasible controller, which achieves the minimum rate, is as important as the derivation of the theoretical bound. The minimum rate achievable control strategy, proposed by Liberzon and

coworkers [BL00, BL06, Lib02a, Lib02b, Lib03], has significance in practical implementations. The basic principle behind their solution is, 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 decrease the quantizer resolution. However, even though at each time t the quantizer is memoryless and has finite reconstruction levels, a memoryless dynamic quantizer will have infinite quantization levels asymptotically.

The state estimation problem and the stabilization problem have always been studied side by side, see e.g., [SV03, Tat00, WB97]. In fact, estimation problems explore the observability of dynamic systems. It has been shown that [Tat00], even with a rate at which achieving asymptotic observability is impossible, it may still be possible to achieve asymptotic stability. In [SV03] a state estimation problem over a binary symmetric channel is considered. The encoder–decoder adopts the zooming idea to capture the state trajectory and bound estimation errors. In their setup, the quantizer is uniform and its range is adaptively adjusted according to the state evolution. At each time instant only one binary bit is transmitted over the channel, that is to say for a codeword of length 2^R the decoder needs to wait 2^R time units before the decoding. During the transmission of a codeword, the worst case situations are considered.

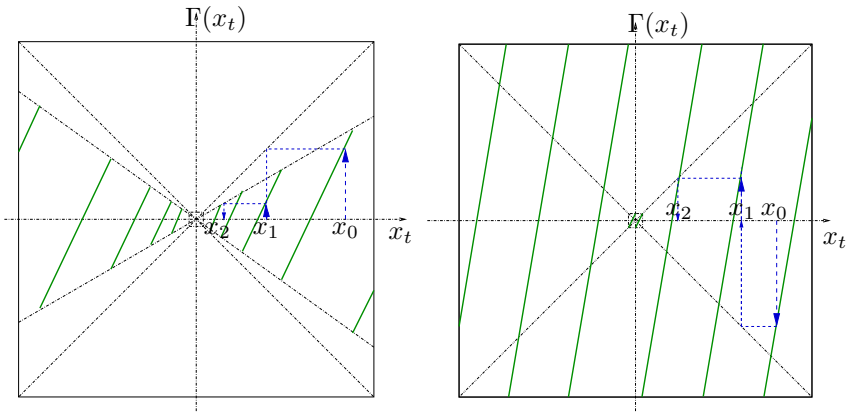
1.5.2 Static Quantizer

From the implementation point of view, it is more relevant to study a static quantizer. The recent research has shown an increasing interest in the fundamental properties of this type of quantizers. Problem formulations involve mostly a rigorous analysis of the long term behavior. The attention has also been devoted to problems such as proper performance measures for static quantizers; and impacts on system performance due to memory restriction.

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, see e.g., [Del88, Del89, Del90, Del06, FZ03, FZ05, PGB02, PPBJ04]. Together, they establish theoretical fundamentals for the practical design of quantized feedback control.

In the remaining part of this section, some features of static memoryless quantizers are presented. In this setup, the system evolution becomes

$$x_{t+1} = ax_t + u_t = ax_t + F(q(x_t)) = \Gamma(x_t). \tag{1.20}$$



(a) Logarithmic quantizer (b) Chaotic quantizer

Figure 1.12: Examples of piecewise affine maps for quantized control systems. a) A piecewise affine map of a logarithmic quantizer. The dashed arrows illustrate the system evolutions from x_0 to x_2. b) A piecewise affine map of a chaotic quantizer.

where q denotes static quantization and the feedback F is static as well. Since u_t is completely determined by x_t, it leads to the fact that x_{t+1} depends only on x_t. The system evolution can be described by a piecewise affine map Γ, see Fig. 1.12. In Fig. 1.12, the current state x_t is presented on the x-axis and the state one-step-ahead, i.e., $\Gamma(x_t)$, is presented on the y-axis. In particular, the map illustrates the two most important features of the type of control strategies. The first one is attractivity, referring to as the attraction from an initial region (large) to a target region (small). The second feature 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 suitable steady-state and transient properties. Fagnani and Zampieri have shown that there is a conflicting relation between the steady-state and the transient behavior for quantized systems [FZ03, FZ05]. The authors were particularly interested in quantitative analysis and comparison among different quantizers. They have suggested *contraction rate* for the steady state performance

and *expected time* for the transient behavior. Generally speaking, a large contraction rate is desired, 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. Fagnani and Zampieri have evaluated this conflicting relation for uniform quantizers, logarithmic quantizers and chaotic quantizers. More about the uniform quantizer and the chaotic quantizer can be found in the next paragraph. Here, we present a special property of the logarithmic quantizer. Shown by Elia and Mitter in [EM99, EM01], a logarithmic quantizer is special in the way that when the number of the reconstruction levels is a prior entity, the logarithmic quantizer is shown to be the most efficient quantization scheme, in the sense of quadratic Lyapunov stability. Lyapunov stability is a stability measurement commonly used in nonlinear 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.

For most researchers, the design has been part of the quantizer problem. Unlike those, Picasso and coworkers have considered the quantizer as a given building block and investigated its capabilities [PGB02, PPBJ04]. In particular, they worked with the construction of attractive and invariant sets for uniform quantizers. Due to the simplicity in implementation, the uniform quantizer is the most widely used quantization scheme. The authors have in particular derived several theoretical results concerning the use of a uniform quantizer to stabilize unstable linear systems, e.g., the construction of the target set for given an initial set and a quantizer, as well as the number of steps needed to enter a target set. Besides uniform quantizers and logarithmic quantizers, a special class of quantizers called chaotic quantizers have also been thoroughly studied, see e.g., [FZ03, FZ05]. The chaotic behavior is exploited to mitigate the conflicting relations between attractivity and invariance, done by combining two feedback strategies. 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 map. All the trajectories will eventually enter the target region. When the trajectory is within the target region, it switches to the second controller, which has a task to keep the trajectories within the target region ever after, cf. Fig. 1.12(b).

A recent paper of Delvenne [Del06] has shown a promising improvement in data rate by relaxing the restriction imposed on the structure

of quantization regions. This work has brightened the potential of new quantization sets compared to the traditional view of connected intervals.

1.5.3 Linear Quadratic Control

Control performance for quantized systems can be developed in the context of stochastic control, e.g., [BM99, MS02, MS04, Tat00]. Borkar and Mitter have in the early 1990's explored the linear stochastic control of partially observed Gaussian systems. They suggested to quantize the state innovations, out of a minimum variance filter. They have particularly studied a scenario of stable plants and noiseless channel. In [Tat00], the research on the separation principle and CE controllers has been carried forward. Analyses of systems with different information patterns are 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.

In [MS02,MS04], generalizations to multidimensional plants have been brought out. In their problem formulation, process noise and measurement noise are both Gaussian distributed. The transmission is performed over a noiseless channel. The exact information pattern is as follows: the encoder has access to all past observations, while the memoryless decoder has only access to the current received symbol. The authors have investigated a controller-coder separated design, where an optimal control signal is derived assuming no data rate constraint is imposed. Then, a coded version of the derived control signal is applied to the system. It has been proven that the controller-coder separated design is not the optimal solution among all searched control strategies.

1.5.4 Information Theoretic Results

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 the rate distortion theory, are in their traditional forms not useful for feedback systems [Sah00, Sah04, Tat00, TSM04, TM04]. The reasons are multiple. It is a common fact that control applications are much more time critical than

conventional communication applications. Moreover, in communication applications the performance of the current transmission has usually no impact on what information to transmit in the future, whereas the objective of a feedback control is to affect the future states.

In [Tat00, TSM04, TM04] a sequential rate distortion theory for feedback systems has been developed. The author has emphasized the difference between statistical dependence in traditional information theory and causality in a control context. The quantity of *directed mutual information* [Mas90] is pointed out to be instrumental in dealing with sequential rate distortion problems. Concerning the channel coding issues, in [Sah00, Sah04], the author has shown that the Shannon capacity is not a proper entity for characterizing feedback systems, instead, *anytime capacity* has been proposed. This new capacity resembles Gallager's error exponent [Gal68] with a crucial difference that the encoder is not allowed to have an arbitrary delay as the decoder does.

1.5.5 Summary

In this section, a brief overview of recent advances in control using quantized feedback has been given. The existing work has been mainly devoted to various stability analysis. Uniform quantizers are considered in most contributions, since they are easy to implement. However, for the applications with a high cost on the communications, it is appropriate to study optimal quantizers. There is a rich literature addressed to the optimal quantizer design for memoryless sources as well as causal coding problems [NG82, WV83, Wit79]. The main objective of this thesis is to jointly optimizing the encoder–decoder together with a controller..

1.6 Contributions and Outline

This thesis first gives a brief overview of the recent developments in the area of control with feedback over communication channels. In Chapter 2, a state space model of the overall system is presented and the information pattern at each component is specified. A quadratic stochastic linear control problem is then formulated. Iterative algorithms for optimizing coder–controllers are proposed for transmission over noise-free channels (Chapter 3) and noisy channels (Chapter 4), respectively. In Chapter 5, the proposed encoder–decoder design is applied to an event-triggered control problem. Finally, in Chapter 6, some topics for future research

are suggested. Below, the contributions are described in detail for each chapter.

Chapter 2

In this chapter, a general system model is presented. The state space presentation is employed. The model consists of a system equation and an observation equation. By varying the information pattern at each component, this model can cover a wide range of scenarios. A generalization to multidimensional systems is straightforward, by replacing scalar signals with vector signals.

Chapter 3

This chapter is devoted to the data rate restriction imposed by a non-ideal communication channel. In particular, we design the control feedback transmitted over error-free and band-limited channels. In the first part, the focus is on the simplest system, in which the only uncertainty is the initial state. Later on, process noise is included into the system description. For both scenarios, the conditions for the separation principle are derived and optimal control strategies together with coding schemes are proposed. This chapter is based on the paper [BSJ06a].

Chapter 4

Given a quantizer designed for a noise-free channel (as in Chapter 3), the system performance deteriorates severely if the real channel produces transmission errors. In the first part of this chapter, two remedies to this problem are studied. The first enhancement is an approach to design a good index assignment by using simulated annealing. Since, when communicating over a noisy channel, the mappings between the integer indices and the binary codewords can not be overlooked. We show that the index assignment plays an important role in the overall system performance. In the second approach, on top of the source code we will exploit a channel code with a short code length.

In the main part of this chapter we present a joint source–channel code where the channel characteristic has been taken into account in the quantizer design. The condition for the CE property is derived, by which the estimation problem and the control problem are separated. The control problem can be solved by applying the fundamentals of stochastic control theory, while the solution to the estimation problem resembles

channel optimized quantization. This chapter is based on the papers [BSJ06b, BSJ06c].

Chapter 5

In this chapter, a new controller is developed for a system perturbed by some infrequently occurring local disturbances. The disturbances are treated as events, which trigger the control actuation. Results from Chapter 3 are used to develop the event-triggered control strategy which has shown to be bandwidth-efficient. This chapter is based on the paper [BSJ06a].

Chapter 6

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 will be used:

a	:	Dynamics of a scalar linear system.
$\mathbf{A}$	:	Dynamics of a multidimensional linear system.
$\mathcal{C}_t$	:	The set of information at the controller.
$c(i)$	:	Binary transmitted codeword.
$c(j)$	:	Binary received codeword.
$\mathcal{D}_t$	:	The set of information at the decoder.
d_H	:	Hamming distance between binary codewords.
d_t	:	Output from the decoder.
$\mathcal{E}_t$	:	The set of information at the encoder.
$\mathbf{E}\{X\}$	:	The expectation of X.
$\mathbf{E}\{X\|y\}$	:	The expectation of X given $Y = y$.
$\mathbf{E}_X\{f(X)\}$	:	The expectation of $f(X)$ over the random variable X.
e_t	:	Measurement noise at each time instant.
f_t	:	Encoder function.
$GGD(\alpha, \beta)$	:	Generalized Gaussian distribution with parameters

		α and β.
g_t	:	Decoder function.
h_t	:	Measurement function in the state space model.
i_t	:	Index input to the discrete channel.
j_t	:	Index output of the discrete channel.
k_t	:	Channel function.
M_c	:	Channel memory.
M_t	:	System memory.
$p(x)$	:	Probability mass function for a discrete random variable. Probability density function for a continuous random variable.
$q_{\mathbf{j}_0^t}$	:	Reconstruction point in the codebook, based on the received indices $\mathbf{j}_0^t$.
R	:	Channel Rate.
S_i	:	Encoder region associated with index i
u_t	:	Control signal at time t.
v_t	:	Process noise at time t.
x_t	:	System state at time t.
$\hat{x}_t$	:	An estimate (arbitrary) of the state signal x_t.
$\hat{x}_{t\|t}$	:	Conditional expectation estimate of x_t : $\hat{x}_{s\|t} = \mathbf{E}\{x_s\|\mathbf{i}_0^t\}$ for a noiseless channel; $\hat{x}_{s\|t} = \mathbf{E}\{x_s\|\mathbf{j}_0^t\}$ for a noisy channel.
$\bar{x}_t$	:	$\bar{x}_t = x_t - \sum_{s=0}^{t-1} a^{t-1-s} u_t$.
$\tilde{x}_t$	:	$\tilde{x}_t = x_t - \hat{x}_{t\|t}$.
$\acute{x}_t$	:	$\acute{x}_t = \hat{x}_{t\|t} - \hat{x}_{t\|t-1}$.
y_t	:	State observation at time t.
$\mathbf{z}_a^b$	:	Evolution of a discrete signal z_t from $t = a$ to $t = b$: $\{z_a, \ldots, z_b\}$.
ϵ	:	The transition probability of a binary symmetric channel.
δ	:	The trigger level.

θ_t : System function in the state space model.

1.8 Acronyms

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

BSC Binary symmetric channel.

CE Certainty equivalence.

COSQ Channel optimized scalar quantizer.

COVQ Channel optimized vector quantizer.

GGD Generalized Gaussian distribution.

i.i.d. Independent identically-distributed.

IA Index assignment.

MMSE Minimum mean squared error.

MSE Mean squared error.

pdf Probability density function.

pmf Probability mass function.

SMCM Sequential Monte Carlo method.

VQ Vector quantization.

Chapter 2

Problem Statement

2.1 System Model and Information Pattern

The information pattern specifies the information available at each building block in a connected system. It plays a key role in achieving a certain system performance. In this thesis, we design the encoder, decoder and controller given the plant and channel, see Fig. 2.1. Listed below are the assumed information patters at each component. Let $\mathbf{z}_a^t$ denote the evolution of a discrete time signal z_t from $z = a$ to $z = b$. The system dynamics is then given by

$$\begin{cases} x_{t+1} &= \theta_t(\mathbf{x}_{t-M_x}^t, u_t, v_t), \\ y_t &= h_t(x_t, e_t), \end{cases} \tag{2.1}$$

which has a memory of order $M_x \geq 0$. The terms θ_t denotes a time-varying system function and h_t denotes a time-varying measurement function. The variables $x_t, u_t, y_t \in \mathbb{R}$ are the state, the control, and the measurement, respectively. The signal v_t refers to as the process noise and e_t is the measurement noise. In stochastic control, v_t is often associated with the uncertainty in the system model while e_t represents the uncertainty in state measurements.

Let $\mathcal{E}_t$ denote the set of information available at the encoder at time instant t, i.e., the set of variables whose values are known to the encoder. In particular, an encoder causally utilizes *full information*, when $\mathcal{E}_t = \{\mathbf{y}_0^t, \mathbf{i}_0^{t-1}, \mathbf{u}_0^{t-1}, \mathbf{j}_0^{t-1}\}$ is assumed. The *encoder* is a mapping from $\mathcal{E}_t$ to a discrete set of symbols. We take each symbol to be represented by an integer index. At time t, the index is $i_t \in \mathcal{I}_L = \{0, 1, \ldots, L-1\}$, where

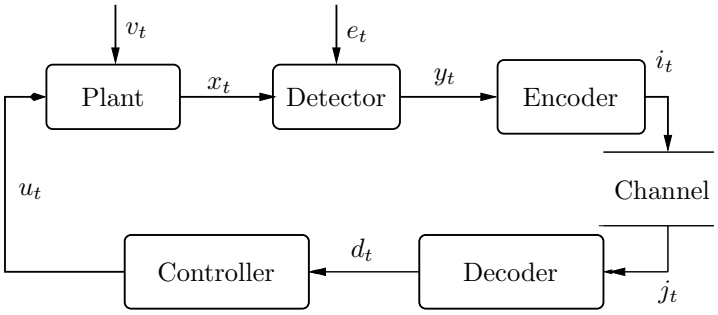


Figure 2.1: A general picture of the feedback control system.

$L = 2^R$ with R denoting the *rate* of the transmission, in bits per state measurement. Hence, the encoder is described by a mapping

$$i_t = f_t(\mathbf{y}_0^t, \mathbf{i}_0^{t-1}, \mathbf{j}_0^{t-1}, \mathbf{u}_0^{t-1}). \tag{2.2}$$

Let the *discrete channel* have input variable i_t and output variable $j_t \in \mathcal{I}_L$, with one channel use defined by

$$j_t = k_t(\mathbf{i}_{t-M_c}^t), \tag{2.3}$$

where $k_t : \mathcal{I}_L^{M_c+1} \to \mathcal{I}_L$ is a stochastic mapping, and $M_c \geq 0$ indicates the (potential) channel memory.

At the receiver, the information available at the decoder is denoted by $\mathcal{D}_t$. The decoder in consideration causally utilizes full information, i.e., $\mathcal{D}_t = \{\mathbf{j}_0^t, \mathbf{u}_0^{t-1}\}$. The *decoder* is a deterministic mapping

$$d_t = g_t(\mathbf{j}_0^t, \mathbf{u}_0^{t-1}) \tag{2.4}$$

from $\mathcal{D}_t$ to $\mathbb{R}$.

Let $\mathcal{C}_t = \{\mathbf{d}_0^t, \mathbf{u}_0^{t-1}\}$ denote the (full) information available at the controller. The *controller* is then defined by the mapping

$$u_t = z_t(\mathbf{d}_0^t, \mathbf{u}_0^{t-1}) \tag{2.5}$$

from $\mathcal{C}_t$ to $\mathbb{R}$. We also define

$$\hat{x}_{s|t} = \mathbf{E}\{x_s | \mathbf{j}_0^t, \mathbf{u}_0^{t-1}\} \tag{2.6}$$

as the MMSE estimator of the state x_s, based on $\mathbf{j}_0^t$ and $\mathbf{u}_0^{t-1}$.

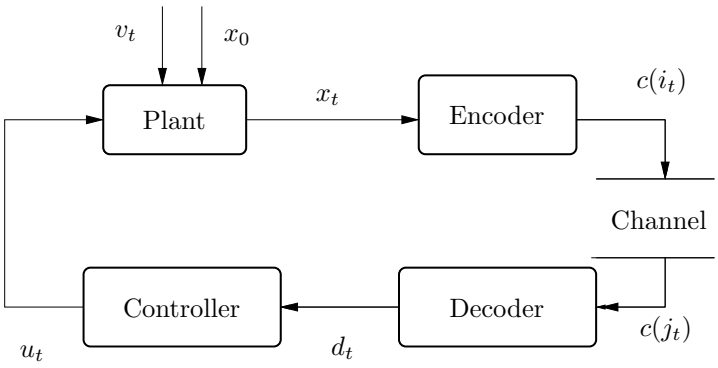


Figure 2.2: The feedback control system in (2.7), with channel constraints.

2.2 Linear System Model

It is well known in stochastic control that for most systems the optimal performance is achievable only when the estimation and the control are designed jointly. A closed form solution of the optimal control is in general prohibitive. However, for traditional control systems with infinite data rates, there are widely known examples where the separation principle applies. Since the main objective of this thesis is devoted to the study of the impact of bandlimited noisy channels, our attention will be restricted to a simple linear system whose exact solution to the optimal stochastic control problem, in absence of the rate limitation, is known. In particular, we consider the linear time-invariant plant illustrated in Fig. 2.2. The system can be described as

$$\begin{cases} x_{t+1} &= ax_t + u_t + v_t, \\ y_t &= x_t, \end{cases} \tag{2.7}$$

where v_t is white noise with the pdf $p(v_t)$. The linear dynamics is a, where $|a| < 1$ so that the plant is stable. Note that full state information is assumed available at the encoder, i.e., $y_t = x_t$ and hence $\mathcal{E}_t = \{\mathbf{x}_0^t, \mathbf{i}_0^{t-1}, \mathbf{u}_0^{t-1}, \mathbf{j}_0^{t-1}\}$. In case there is an infinite bandwidth for the feedback signal, it is straightforward to show that the separation principle applies to this system and an analytical solution can be derived for the optimal controller under a quadratic cost function.

The goal of this thesis is henceforth to design the coding and control for the system in (2.7)*, subject to a data-rate limited and potentially noisy channel.*

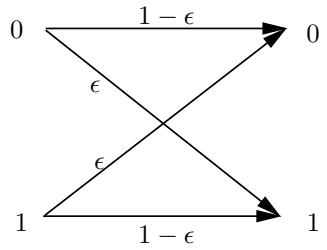


Figure 2.3: A binary symmetric channel. The variable ϵ is the channel transition probability.

Binary Symmetric Channel

A particular class of memoryless discrete channels is considered in this thesis, the so called binary symmetric channels (BSC). A binary symmetric channel is simply described by a channel transition probability, $\epsilon = p(0|1) = p(1|0)$, see Fig. 2.3. Let $c(i)$ and $c(j)$ refer to the binary transmitted and received codeword associated to index i and j. Assuming an independent transmission for each binary bit, the conditional probability $p(c(j)|c(i))$ is a function of ϵ,

$$p(c(j)|c(i)) = (1-\epsilon)^{R-d_H(c(i),c(j))}\epsilon^{d_H(c(i),c(j))}, \tag{2.8}$$

where $d_H(c(i), c(j))$ is the *Hamming distance* between the binary codewords $c(i)$ and $c(j)$, i.e., the number of bits by which they differ.

2.3 Encoder–Decoder Structure

We consider the encoder–decoder pairs as functions with memory. They map an infinite set of continuous real values to a finite set of discrete symbols. The finite set refers to as the codebook, whose entities are reconstruction points labelled with memory-based index sequences.

The presence of a noisy channel raises a synchronization problem for memory-based encoder–decoder pairs. In case of noiseless communications, it is true that $\mathbf{i}_0^t = \mathbf{j}_0^t$, where $i_t \in \mathcal{I}_L$ denotes the transmitted index at t and $j_t \in \mathcal{I}_L$ denotes the received index at t. There is no uncertainty in the information the decoder receives. In those cases, the encoder and the decoder have identical information about the previous events. Such a synchronization between the encoder and the decoder is not present

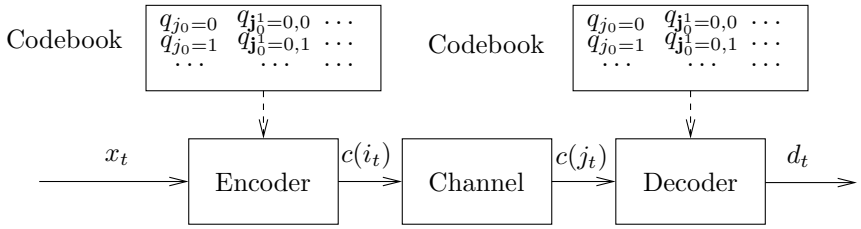


Figure 2.4: The encoder, the channel and the decoder. There is a common codebook at the encoder and the decoder based on the received indices $\mathbf{j}_0^t$.

when the channel contains transmission errors. The encoder cannot in advance make an error-free prediction about the index j_t that the decoder will receive. Even worse, the decoder has lost the ability to deduce the transmitted indices $\mathbf{i}_0^t$ by examining the received $\mathbf{j}_0^t$. However, by assuming the encoder has access to all past control signals, the received indices can be deduced at the encoder. Thus, the sequence $\mathbf{j}_0^t$ can define the common information about the previous events, referred to as the *common memory*. Fig. 2.4 depicts the common codebook at the encoder and the decoder for communications over noisy channels. Note that the reconstruction point $q_{\mathbf{j}_0^t}$ in the codebook is based on the received index sequence $\mathbf{j}_0^t$.

Let $\mathbf{C}$ denote the binary mapping from the index sequence $\{\mathbf{j}_0^{t-1}, i_t\}$ to a binary codeword $c(i_t)$ through $c(i_t) = \mathbf{C}(j_0^{t-1}, i_t)$, where $c(i_t) \in \{0,1\}^R$, with $i_t \in \mathcal{I}_L$, is the binary transmitted codeword at time t, see Fig. 2.7. In a similar way, $\mathbf{D}$ describes the inverse mapping $\mathbf{j}_0^t = \mathbf{D}(c(j_t), \mathbf{j}_0^{t-1})$, where $c(j_t) \in \{0,1\}^R$, with $j_t \in \mathcal{I}_L$, is the received binary codeword at time t. We comment here, for a joint source–channel code, the binary mapping is included in the encoding and decoding functions $\{f_t, g_t\}_{t=0}^{T-1}$. On the other hand, concerning the setup with separated source and channel codes, the encoding function f_t only takes care of the selection of the integer index. The binary mapping needs to be separately designed.

2.4 Performance Measure

The optimality of system performance depends on the design criterion. The goal of the design in this thesis is to solve an integrated coding–

control problem for the system in (2.7) measured by the following cost function

$$J_T = \mathbf{E}\left\{ x_T^2 + \sum_{t=0}^{T-1} [x_t^2 + \rho u_t^2] \right\}. \tag{2.9}$$

The above linear quadratic cost function is well established in the scope of stochastic control, cf. Section 1.3. There is often a practical relevance in that x_t^2 and u_t^2 are related to the energy of the state and the control at t. The parameter $\rho \geq 0$ specifies the relative weight assigned to the control signal u_t. The aim of optimizing this cost function is then to minimize the energy of the state at all time instants, with a power constraint on the control signal.

The overall optimization problem can be specified as

$$J_T^{opt} = \min_{\mathbf{f}_0^{T-1}, \mathbf{g}_0^{T-1}, \mathbf{z}_0^{T-1}} \mathbf{E}_{x_0, v_t} \left\{ x_T^2 + \sum_{t=0}^{T-1} [x_t^2 + \rho u_t^2] \right\}, \tag{2.10}$$

where $\mathbf{f}_0^{T-1}$, $\mathbf{g}_0^{T-1}$, $\mathbf{z}_0^{T-1}$ are the memory-based encoder, decoder and controller functions, respectively. The expectation is taken over the initial state x_0, the process noise v_t and the channel errors. Due to the large number of design parameters involved, even for a simple plant, such as the one in (2.7), the optimization problem is hard to solve.

Chapter 3

Feedback Control over Noise-Free Channels

Imperfect communication channels have several kinds of impacts on performance of control systems. In this thesis, we have restricted our attention to two particular types of channel imperfections. The first one is a limited channel rate and the second one is bit errors. This chapter is devoted to encoder–decoder designs for data rate limited channels. An underlying assumption of the coding strategy throughout is that the predefined coding rule is known to both the encoder and the decoder. In this chapter, the link between the sensor and the controller is error free, hence the encoder can calculate in advance the exact control signal the controller will produce. As will be shown later, at each time instant t, the decoding function is

$$d_t = g_t(\mathbf{i}_0^t), \quad t = 0, \ldots, T-1, \tag{3.1}$$

i.e., the common memory is the transmitted indices $\mathbf{i}_0^t$. The optimal state estimator based on $\mathcal{D}_t$ is

$$\hat{x}_{s|t} = \mathbf{E}\{x_s|\mathbf{i}_0^t\}, \quad t = 0, \ldots, T-1. \tag{3.2}$$

Fig. 3.1 depicts the reconstruction points $q_{\mathbf{i}_0^t}$ in the common codebook for the encoder and the decoder. Observe that Fig. 3.1 is almost the same as Fig. 2.4 except here the transmitted index i_t is available at the decoder.

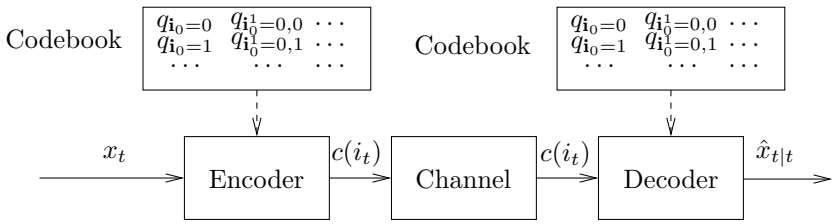


Figure 3.1: The encoder, the channel and the decoder. There is a common codebook at the encoder and the decoder based on the transmitted indices $\mathbf{i}_0^t$, since the channel is noise-free.

3.1 Certainty Equivalence

In general, the optimal performance is achievable only when the coder and the controller are designed jointly. Such an optimization problem involves a large number of design parameters, see (2.10). Finding the optimal solution is often computationally prohibitive. It is therefore attractive if the overall problem can be decomposed into several complexity-reduced subproblems. One fundamental question is hence whether or not there will be any loss in separating the coding and the control. In this section we explore the conditions when the separation principle applies for some classes of linear systems subjected to a bandlimited channel.

3.1.1 Absence of Process Noise

Let us first consider the simplest system, where the only uncertainty is the initial state. The system is modelled as

$$\begin{cases} x_{t+1} &= ax_t + u_t, \\ y_t &= x_t, \\ j_t &= i_t, \end{cases} \tag{3.3}$$

where the initial state x_0 is drawn according to a pdf $p(x_0)$ and the linear dynamics is given by a, $|a| < 1$.

To apply the separation principle to the system (3.3), the design of optimal sequences of decoders $\{g_t\}_{t=0}^{T-1}$ and controllers $\{z_t\}_t^{T-1}$ is considered. The derivation of the optimal control, which turns out to be the CE controller, is in accordance to [BST74]. That is, at each time t, the optimal control minimizes the "cost-to-go" V_t, which is the expected sum

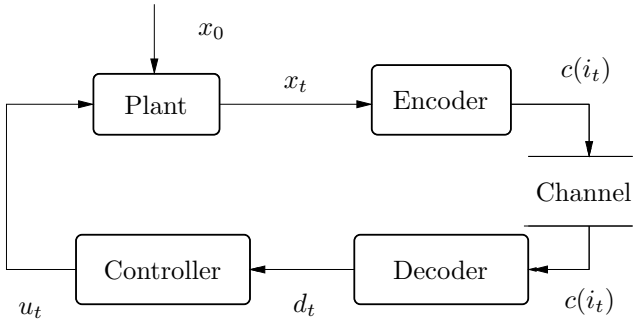


Figure 3.2: A closed-loop control system with data rate constraints, in absence of the process noise.

of $x_s^2 + \rho u_s^2$ from time $s = t$ to $s = T$. The proof exploits mathematical induction over V_t, which states that the optimal "cost-to-go" at time $t+1$ has the form

$$V_{t+1}^\star = \mathbf{E}\left\{P_{t+1}x_{t+1}^2|\mathbf{i}_0^{t+1}, \mathbf{u}_0^t\right\} + \alpha_{t+1}, \tag{3.4}$$

where the variable P_{t+1} is independent of $\mathbf{u}_0^{t-1}$, and the variable α_{t+1} is independent of $\mathbf{u}_0^t$. The terms P_t and α_t will be derived later.

Let us first take a look at the time instant $t + 1 = T$ (recall that T is the time horizon in (2.9)). By choosing $P_T = 1$ and $\alpha_T = 0$, the statement (3.4) is obviously true. Henceforth, assuming the statement is valid for time $t+1$, we derive the conditions for the statement to be true for time t. It gives

$$\begin{aligned}
V_t^\star &= \min_{u_t} \mathbf{E}\left\{x_t^2 + \rho u_t^2 + V_{t+1}^\star|\mathbf{i}_0^t, \mathbf{u}_0^{t-1}\right\} \\
&= \min_{u_t} \mathbf{E}\left\{x_t^2 + \rho u_t^2 + P_{t+1}x_{t+1}^2|\mathbf{i}_0^t, \mathbf{u}_0^{t-1}\right\} + \mathbf{E}\left\{\alpha_{t+1}|\mathbf{i}_0^t, \mathbf{u}_0^{t-1}\right\} \\
&= \min_{u_t} \mathbf{E}\left\{x_t^2 + \rho u_t^2 + P_{t+1}(ax_t + u_t)^2|\mathbf{i}_0^t, \mathbf{u}_0^{t-1}\right\} \\
&\qquad + \mathbf{E}\left\{\alpha_{t+1}|\mathbf{i}_0^t, \mathbf{u}_0^{t-1}\right\} \\
&= \min_{u_t} \mathbf{E}\{(a^2P_{t+1} + 1)x_t^2 + 2aP_{t+1}x_tu_t \\
&\qquad + (P_{t+1} + \rho)u_t^2|\mathbf{i}_0^t, \mathbf{u}_0^{t-1}\} + \mathbf{E}\left\{\alpha_{t+1}|\mathbf{i}_0^t, \mathbf{u}_0^{t-1}\right\} \\
&= \min_{u_t} \mathbf{E}\{(a^2P_{t+1} + 1)x_t^2 + 2aP_{t+1}x_tu_t + (P_{t+1} + \rho)u_t^2|\mathbf{i}_0^t\} \\
&\qquad + \mathbf{E}\left\{\alpha_{t+1}|\mathbf{i}_0^t\right\}.
\end{aligned} \tag{3.5}$$

The last step is since $\mathbf{u}_0^{t-1}$ is completely determined by $\mathbf{i}_0^{t-1}$. It is straightforward to verify that the optimal control, which minimizes (3.5), is

$$u_t^\star = -\frac{aP_{t+1}}{P_{t+1}+\rho}\mathbf{E}\left\{x_t|\mathbf{i}_0^t\right\} = -\ell_t\hat{x}_{t|t}, \tag{3.6}$$

where ℓ_t denotes the linear feedback. Substituting $u_t^\star$ into (3.5), implies

$$\begin{aligned} V_t^\star &= \mathbf{E}\{(a^2P_{t+1}+1)x_t^2 - \frac{2a^2P_{t+1}^2x_t\hat{x}_{t|t}}{P_{t+1}+\rho} + \frac{a^2P_{t+1}^2\hat{x}_{t|t}^2}{(P_{t+1}+\rho)^2}|\mathbf{i}_0^t\} + \mathbf{E}\{a_{t+1}|\mathbf{i}_0^t\} \\ &= \mathbf{E}\{(1+\frac{a^2P_{t+1}\rho}{\rho+P_{t+1}})x_t^2 + \frac{a^2P_{t+1}^2}{\rho+P_{t+1}}(x_t-\hat{x}_{t|t})^2|\mathbf{i}_0^t\} + \mathbf{E}\left\{\alpha_{t+1}|\mathbf{i}_0^t\right\}. \end{aligned} \tag{3.7}$$

Evaluate (3.4) at time t, and compare it with (3.7). In order to make (3.4) valid for time t, P_t and α_t need to fulfil

$$P_t = 1 + \frac{a^2P_{t+1}\rho}{P_{t+1}+\rho}, \tag{3.8}$$

$$\alpha_t = \frac{a^2P_{t+1}^2}{\rho+P_{t+1}}\mathbf{E}\left\{(x_t-\hat{x}_{t|t})^2|\mathbf{i}_0^t\right\} + \mathbf{E}\left\{\alpha_{t+1}|\mathbf{i}_0^t\right\}. \tag{3.9}$$

In addition, it must also hold true that

$$\mathbf{E}\left\{(x_t-\hat{x}_{t|t})^2|\mathbf{i}_0^t\right\} \tag{3.10}$$

is not a function of $\mathbf{u}_0^{t-1}$. If that is the case, $u_t = -\ell_t\,\hat{x}_{t|t}$ is the optimal control strategy and the separation principle applies. The constraint in (3.10) can be interpreted as that the estimation errors are not dependent on the past control signals.

To summarize, in case the separation principle applies, the optimal decoder and controller sequences are given by

$$d_t = \hat{x}_{t|t}, \quad \text{and} \quad u_t = -\ell_t d_t \tag{3.11}$$

(i.e., $z_t(\mathbf{d}_0^t) = -\ell_t\,d_t$) with

$$\ell_t = \frac{a\,P_{t+1}}{P_{t+1}+\rho}, \quad P_t = 1 + \frac{a^2P_{t+1}\rho}{P_{t+1}+\rho}, \tag{3.12}$$

for $t = T-1,\ldots,0$ and P_t is initialized with $P_T = 1$. Notice that since the separation principle holds, there is no loss in separating the decoder–controller into two entities, cf. Fig. 3.2. Henceforth, the decoder mapping is fixed to $d_t = \hat{x}_{t|t}$, so $\hat{x}_{t|t}$ will be utilized by the controller to produce the control output u_t.

Optimal total cost

In Section 3.1.1, it has been shown that the optimal control can be derived by minimizing the "cost-to-go" at each t. In this section, we revisit the control strategy by elaborating on the total cost (2.9). In particular, we now want to show that the controller in (3.11) is in fact a CE controller that the CE property applies, cf. Section 1.3.1. According to (3.12), the relationship between P_t and P_{t+1} can be written as

$$P_t = a^2 P_{t+1} + 1 - \frac{a^2 P_{t+1}^2}{P_{t+1} + \rho}, \tag{3.13}$$

where P_t and ℓ_t, defined in (3.12), are related by

$$P_t = a^2 P_{t+1} + 1 - \ell_t^2 (P_{t+1} + \rho). \tag{3.14}$$

Observe that

$$x_{t+1}^2 P_{t+1} = (ax_t + u_t)^2 P_{t+1}, \tag{3.15}$$

$$x_t^2 P_t = x_t^2 (a^2 P_{t+1} + 1 - \ell_t^2 (P_{t+1} + \rho)). \tag{3.16}$$

By using $P_T = 1$, together with (3.12), (3.15) and (3.16), x_T^2 can be written as

$$\begin{aligned} x_T^2 &= x_0^2 P_0 + \sum_{t=0}^{T-1} \left[x_{t+1}^2 P_{t+1} - x_t^2 P_t \right] \\ &= x_0^2 P_0 + \sum_{t=0}^{T-1} \left[(ax_t + u_t)^2 P_{t+1} - x_t^2 (a^2 + P_{t+1} + 1 - \ell_t^2 (P_{t+1} + \rho)) \right] \\ &= x_0^2 P_0 + \sum_{t=0}^{T-1} \left[(\ell_t^2 (P_{t+1} + \rho) - 1) x_t^2 + 2a x_t u_t P_{t+1} + u_t^2 P_{t+1} \right]. \end{aligned} \tag{3.17}$$

Substituting (3.17) into the cost function J_T in (2.9), gives

$$\begin{aligned} J_T &= \mathbf{E} \left\{ x_T^2 + \sum_{t=0}^{T-1} \left[x_t^2 + \rho u_t^2 \right] \right\} \\ &= \mathbf{E} \left\{ P_0 x_0^2 + \sum_{t=0}^{T-1} \left[\ell_t^2 (P_{t+1} + \rho) x_t^2 + 2a P_{t+1} x_t u_t + (P_{t+1} + \rho) u_t^2 \right] \right\} \\ &= \mathbf{E} \left\{ P_0 x_0^2 + \sum_{t=0}^{T-1} (P_{t+1} + \rho)(x_t \ell_t + u_t)^2 \right\}. \end{aligned} \tag{3.18}$$

Note that, (3.18) reveals that when the state signal x_t is available at the controller, the optimal deterministic control is $u_t = -\ell_t x_t$. Hence, the control in (3.11) is a CE controller.

However, under the assumption of this section, the control signal u_t is a function of the quantized x_t, in particular, a function of $\mathbf{i}_0^t$. Hence, the optimization problem becomes

$$\begin{aligned}
&\min_{\mathbf{u}_0^{T-1}} \mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t x_t)^2\right\}\\
&\quad=\mathbf{E}_{\mathbf{i}_0^t}\left\{\min_{\mathbf{u}_0^{T-1}}\mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t x_t)^2\Big|\mathbf{i}_0^t\right\}\right\}\\
&\quad=\mathbf{E}_{\mathbf{i}_0^t}\left\{\min_{\mathbf{u}_0^{T-1}}\mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t(x_t-\hat{x}_{t|t}+\hat{x}_{t|t}))^2\Big|\mathbf{i}_0^t\right\}\right\}\\
&\quad=\mathbf{E}_{\mathbf{i}_0^t}\left\{\min_{\mathbf{u}_0^{T-1}}\mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t\hat{x}_{t|t}+\ell_t(\hat{x}_{t|t}-x_t))^2\Big|\mathbf{i}_0^t\right\}\right\}\\
&\quad=\mathbf{E}_{\mathbf{i}_0^t}\left\{\min_{\mathbf{u}_0^{T-1}}\left(\mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t\hat{x}_{t|t})^2\Big|\mathbf{i}_0^t\right\}\right.\right.\\
&\qquad\left.\left.+\sum_{t=0}^{T-1}\mathbf{E}\{\ell_t^2(\rho+P_{t+1})(x_t-\hat{x}_{t|t})^2|\mathbf{i}_0^t\}\right)\right\},
\end{aligned}\tag{3.19}$$

where $\mathbf{E}_{\mathbf{i}_0^t}$ denotes the expectation over $\mathbf{i}_0^t$. Note that, if the last term in (3.19),

$$\mathbf{E}_{\mathbf{i}_0^t}\left\{\sum_{t=0}^{T-1}\ell_t^2(\rho+P_{t+1})\mathbf{E}\{(\hat{x}_{t|t}-x_t)^2|\mathbf{i}_0^t\}\right\}\tag{3.20}$$

is not a function of $\mathbf{u}_0^{t-1}$, then the same decoder and controller as in (3.11) will minimize (3.19), accordingly the total cost (2.9).

3.1.2 Presence of Process Noise

In this section, we generalize the system in (3.3) to one including process noise:

$$\left\{\begin{array}{rcl} x_{t+1} &=& ax_t+u_t+v_t,\\ y_t &=& x_t,\\ j_t &=& i_t,\end{array}\right.\tag{3.21}$$

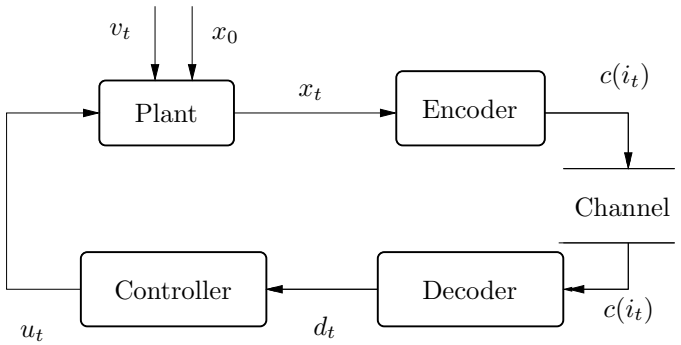


Figure 3.3: A closed-loop control system with data rate constraints, in presence of the process noise v_t.

where x_t, u_t and v_t are the state, the control signal and the process noise, respectively. The pdf's $p(x_0)$ and $p(v_t)$ are given, as well as the linear dynamics $a, |a| < 1$. The state x_t is fully observed at the encoder ($y_t = x_t$), and there is no transmission error in the communication channel ($j_t = i_t$). Note that the linear quadratic cost function in (2.9) should here be interpreted as an average over x_0 and v_t.

We derive the optimal control for fixed encoder–decoder pairs $\{f_t, g_t\}_{t=0}^{T-1}$. The approach is similar to the one presented in Section 3.1.1. Observe that, in presence of the process noise, (3.15) and (3.16) turn out to be

$$x_{t+1}^2 P_{t+1} = (ax_t + u_t + v_t)^2 P_{t+1}, \tag{3.22}$$

$$x_t^2 P_t = x_t^2 (a^2 P_{t+1} + 1 - \ell_t^2 (P_{t+1} + \rho)). \tag{3.23}$$

By using $P_T = 1$, together with (3.12), (3.22) and (3.23), x_T^2 becomes

$$\begin{aligned} x_T^2 &= x_0^2 P_0 + \sum_{t=0}^{T-1} \Big[(ax_t + u_t + v_t)^2 P_{t+1} - x_t^2 (a^2 P_{t+1} + 1 - \ell_t^2 (P_{t+1} + \rho)) \Big] \\ &= x_0^2 P_0 + \sum_{t=0}^{T-1} \Big[2v_t (ax_t + u_t) P_{t+1} + v_t^2 P_{t+1} \\ &\qquad + (\ell_t^2 (P_{t+1} + \rho) - 1) x_t^2 + 2ax_t u_t P_{t+1} + u_t^2 P_{t+1} \Big]. \end{aligned} \tag{3.24}$$

Substituting (3.24) into the expression for J_T in (2.9), gives

$$\begin{aligned} J_T &= \mathbf{E}\Bigg\{ x_0^2 P_0 + \sum_{t=0}^{T-1} \left[2v_t P_{t+1}(ax_t + u_t) + v_t^2 P_{t+1}\right] \\ &\qquad + \sum_{t=0}^{T-1} \left[x_t^2 \ell_t^2 (P_{t+1} + \rho) + 2a x_t P_{t+1} u_t + u_t^2 (P_{t+1} + \rho)\right] \Bigg\} \\ &= \mathbf{E}\Bigg\{ x_0^2 P_0 + \sum_{t=0}^{T-1} \left[2v_t P_{t+1}(ax_t + u_t) + v_t^2 P_{t+1}\right] \\ &\qquad + \sum_{t=0}^{T-1} (P_{t+1} + \rho)(x_t \ell_t + u_t)^2 \Bigg\}. \end{aligned} \tag{3.25}$$

By assumption, v_t is white noise, uncorrelated with x_t and u_t, so J_T can hence be simplified to

$$J_T = \mathbf{E}\left\{ x_0^2 P_0 + v_t^2 P_{t+1} + \sum_{t=0}^{T-1} (P_{t+1} + \rho)(x_t \ell_t + u_t)^2 \right\}. \tag{3.26}$$

Note that we have obtained an expression similar to (3.18). The difference lies in the additional term $\mathbf{E}\{v_t^2 P_{t+1}\}$, which is not relevant to the minimization problem. Accordingly, we will reach the same result as in (3.19). In that case, the condition of the CE property requires that the following term (same as (3.20))

$$\mathbf{E}_{\mathbf{i}_0^t}\left\{ \sum_{t=0}^{T-1} \ell_t^2 (\rho + P_{t+1}) \mathbf{E}\{(\hat{x}_{t|t} - x_t)^2 | \mathbf{i}_0^t\} \right\} \tag{3.27}$$

is independent of $\mathbf{u}_0^{t-1}$. We conclude that, if (3.27) is independent of $\mathbf{u}_0^{t-1}$, the optimal control for the system in (3.21) has the form (3.11). Note that, $\hat{x}_{t|t}$ here depends now on $p(x_0)$, $\{f_t, g_t\}_{t=0}^{T-1}$ and $p(v_t)$. Obviously, the system (3.3) is a special case of the system (3.21) where process noise always equals to zero.

3.2 Encoder–Decoder Structure

The condition of the CE property imposes certain structure on the encoder–decoder pairs. In this section, we investigate encoder–decoder

structures for two classes of linear plants, namely, the system (3.3) and the system (3.21). It is worth noting that, since the channel is error-free, the system performance is independent of the mappings between the indices and the binary codewords, see Section 2.3.

3.2.1 Absence of Process Noise

Let us first restrict the system model to (3.3), i.e., the initial state is the only uncertainty. At time $t = 0$, once x_0 is available at the encoder, the encoder is able to calculate all future states based on the system equation (3.3). The encoder–decoder design is therefore essentially equivalent to the problem of constructing an optimal strategy for the encoder to successively inform the decoder about the initial state over a discrete noiseless channel. Note that $\mathcal{E}_t = \{\mathbf{x}_0^t, \mathbf{i}_0^{t-1}, \mathbf{u}_0^{t-1}\}$ is equivalent to $\mathcal{E}_t = \{x_0, \mathbf{i}_0^{t-1}\}$. This, since the encoder knows $\mathbf{u}_0^{t-1}$ based on $\mathbf{i}_0^{t-1}$ (remember, $j_t = i_t$), and since x_1^t can be computed from $\{x_0, \mathbf{u}_0^{t-1}\}$. Hence, we can assume from now on, that i_t is generated as

$$i_t = f_t(x_0, \mathbf{i}_0^{t-1}), \quad t = 0, \ldots, T-1, \tag{3.28}$$

based on x_0 and the known previous indices $\mathbf{i}_0^{t-1}$.

For a fixed sequence of encoders $\{f_t\}_{t=0}^{T-1}$ with $f_t = f_t(x_0, \mathbf{i}_0^{t-1})$, we notice that the CE property requires that the estimation error $\mathbf{E}\{(x_0 - \hat{x}_{0|t})^2\}$, $t = 0, \ldots, T$, is not a function of $\mathbf{u}_0^s$, $s = 0, \ldots, t-1$, for any sequence $\mathbf{u}_0^s$ of controls. It is straightforward to verify that this condition holds true in our case, since $i_0 = f_0(x_0)$, $i_1 = f_1(x_0, i_0)$, $\ldots$, $i_t = f_t(x_0, \mathbf{i}_0^{t-1})$. Hence the values taken on by $\mathbf{i}_0^t$, given a fixed set of encoders, depend only on x_0 and not on $\mathbf{u}_0^s$ for any $s < t$. Thus, the separation principle holds under our assumptions.

At the receiver, full information is equivalent to $\mathcal{D}_t = \{\mathbf{i}_0^t\}$, since $j_t = i_t$ and $\mathbf{u}_0^t$ can be computed from $\mathbf{i}_0^t$, as stated previously. Therefore the decoder mapping is

$$d_t = g_t(\mathbf{i}_0^t), \quad t = 0, \ldots, T-1. \tag{3.29}$$

3.2.2 Presence of Process Noise

Unlike the situation in Section 3.2.1, for the system (3.21), all future states can not be calculated at time $t = 0$, since at each t, new information about $v(t)$ will arrive at the encoder.

The encoder–decoder design is then essentially equivalent to the problem of constructing an optimal strategy for the encoder to successively inform the decoder about x_0 and $\mathbf{v}_0^{t-1}$. Note that the encoder knows $\mathbf{v}_0^{t-1}$, based on $\mathbf{x}_0^t$ and $\mathbf{u}_0^{t-1}$. Moreover, $\mathbf{u}_0^{t-1}$ is completely determined by $\mathbf{i}_0^{t-1}$. Hence i_t is generated as

$$i_t = f_t(x_0, \mathbf{v}_0^{t-1}, \mathbf{i}_0^{t-1}), \quad t = 0, \ldots, T-1. \tag{3.30}$$

Now, since $i_0 = f_0(x_0), i_1 = f_1(x_0, v_0, i_0), \ldots, i_t = f_t(x_0, \mathbf{v}_0^{t-1}, \mathbf{i}_0^{t-1})$, $\mathbf{i}_0^t$ depends only on x_0 and $\mathbf{v}_0^{t-1}$, and not on $\mathbf{u}_0^s$ for any $s < t$. The CE controller in (3.11) is optimal for fixed encoders $\{f_t\}_0^{T-1}$ with the structure $i_t = f_t(x_0, \mathbf{v}_0^{t-1}, \mathbf{i}_0^{t-1})$.

For convenience, let us define $\bar{x}_t$ as follows

$$\bar{x}_t = a^t x_0 + \sum_{s=0}^{t-1} a^{t-1-s} v_s. \tag{3.31}$$

Let us assume that the encoder has the structure $i_t = f_t(\bar{x}_t, \mathbf{i}_0^{t-1})$ and the reconstruction point in the codebook is $q_{\mathbf{i}_0^t} = \mathbf{E}\{\bar{x}_t | \mathbf{i}_0^t\}$. Since $x_t - \hat{x}_{t|t} = \bar{x}_t - q_{\mathbf{i}_0^t}$ is not a function of $\mathbf{u}_0^s$, the CE controller in (3.11) is thus also optimal for fixed encoders $\{f_t\}_0^{T-1}$ with the structure $i_t = f_t(\bar{x}_t, \mathbf{i}_0^{t-1})$. The term $\bar{x}_t$ is a sufficient statistic and there is no loss in performance by enforcing the encoder structure $f_t(\bar{x}_t, \mathbf{i}_0^{t-1})$.

Finally, the decoder mapping at the receiver is the same one as in (3.29).

3.3 Iterative Design Algorithm

A method for jointly optimizing the encoder, decoder and controller functions is proposed in this section. Similar to traditional iterative algorithms for optimal source coder designs [GG92], the basic idea is to search for locally optimal encoder–decoder pairs by alternating between updating the decoders for fixed encoders and vice versa, until convergence. The convergence is monitored based on updating the value of J_T in each step. Unfortunately, this design does not guarantee global optimality. However, the algorithm converges to a local minimum, and in this sense produces a "good" solution.

Algorithm 2: Encoder–Decoder Design for the System (3.3)

1. Initialize the encoder and decoder mappings $\{f_t\}_0^{T-1}$ and $\{g_t\}_0^{T-1}$. Compute the controller parameters $\{\ell_t\}$ by using (3.12).
2. For each $t = 0, \ldots, T-1$:
   2a. Update the encoder mapping f_t by using (3.35).
   2b. Update the decoder mapping d_t by using (3.32).
   2c. Set $u_t = -\ell_t d_t$.
3. If J_T has not converged, return to Step 2, otherwise terminate the iteration.

3.3.1 Absence of Process Noise

For any given sequence $\{f_t\}_{t=0}^{T-1}$ of encoder mappings, with $f_t = f_t(x_0, \mathbf{i}_0^{t-1})$, the optimal decoder and controller mappings are given by (3.11), i.e.,

$$g_t(\mathbf{i}_0^t) = \mathbf{E}\{x_t|\mathbf{i}_0^t\} = a^t q_{\mathbf{i}_0^t} + \sum_{s=0}^{t-1} a^{t-1-s} u_s, \tag{3.32}$$

where $q_{\mathbf{i}_0^t} = \mathbf{E}\{x_0|\mathbf{i}_0^t\}$ is the reconstruction point at the time t, stored in the common codebook at the decoder and the encoder, cf. Fig. 3.1.

Given the sequence of decoder $\{g_t\}_0^{T-1}$ and controller mappings $\{z_t\}_0^{T-1}$, and assuming in addition that f_t, $t = 0, \ldots, T-2$, are fixed and known, it is straightforward to realize that the optimal $f_{T-1} = f_{T-1}(x_0, \mathbf{i}_0^{T-2})$ is described by

$$i_{T-1} = \arg\min_{i \in \mathcal{I}_L} \mathbf{E}\left\{ [x_T^2 + \rho u_{T-1}^2] \Big| x_0, \mathbf{i}_0^{T-2}, i_{T-1} = i \right\}. \tag{3.33}$$

Note that $x_1, \ldots, x_{T-1}$ and $u_0, \ldots, u_{T-2}$ are known deterministically conditioned on x_0 and $\mathbf{i}_0^{T-2}$, given f_t for $t = 0, \ldots, T-2$. The reason for this is that the decoder and controller mappings are fixed. Hence, testing different values for i_{T-1} influences only u_{T-1} and $x_T = ax_{T-1} + u_{T-1}$. For $t < T-1$, the impact of i_t on all future terms should be taken into consideration. Hence, for given decoders and controllers, and given the encoder mappings f_t for $t = 0, \ldots, t-1, t+1, \ldots, T-1$, the optimal $f_t = f_t(x_0, \mathbf{i}_0^{t-1})$ is

$$i_t = \arg\min_{i \in \mathcal{I}_L} \mathbf{E}\left\{ \sum_{k=t}^{T} [x_k^2 + \rho u_k^2] \Big| x_0, \mathbf{i}_0^{t-1}, i_t = i \right\}. \tag{3.34}$$

Note that $\sum_{k=t}^{T}[x_k^2+\rho u_k^2]$ is a deterministic function of i_t, thus (3.34) is equal to

$$i_t = \arg\min_{i\in\mathcal{I}_L}\left(\sum_{k=t}^{T}[x_k^2+\rho u_k^2]\Big|x_0,\mathbf{i}_0^{t-1},i_t=i\right). \tag{3.35}$$

The optimal mapping f_t indeed has the form $i_t = f_t(x_0,\mathbf{i}_0^{t-1})$. Although straightforward in principle, the expression in (3.35) has not appeared in our context before.

Based on (3.12), (3.32) and (3.35), an encoder–decoder design algorithm is formulated in Algorithm 2.

3.3.2 Presence of Process Noise

For systems as (3.21), given the encoder, the optimal decoder is

$$g_t(\mathbf{i}_0^t) = \mathbf{E}\{x_t|\mathbf{i}_0^t\} = q_{\mathbf{i}_0^t} + \sum_{s=0}^{t-1} a^{t-1-s}u_s, \tag{3.36}$$

where $q_{\mathbf{i}_0^t} = \mathbf{E}\{\bar{x}_t|\mathbf{i}_0^t\}$ is the reconstruction point for $\mathbf{i}_0^t$. We see that the main task for the decoder is to estimate $\bar{x}_t$ based on $\mathbf{i}_0^t$, and it is not interesting to know x_0 and $\mathbf{v}_0^{t-1}$ separately.

The optimal encoder needs to take the impact of the future state evolutions into account. Hence, for fixed decoders and controllers, and given the encoder mappings f_t for $t=0,\ldots,t-1,t+1,\ldots,T-1$, the optimal $f_t = f_t(x_0,\mathbf{v}_0^{t-1},\mathbf{i}_0^{t-1})$ is

$$i_t = \arg\min_{i\in\mathcal{I}_L}\mathbf{E}\left\{\sum_{k=t}^{T}[x_k^2+\rho u_k^2]\Big|x_0,\mathbf{v}_0^{t-1},\mathbf{i}_0^{t-1},i_t=i\right\}. \tag{3.37}$$

According to (3.37), the encoding rule is updated once the reproduction points are recalculated. Therefore, both the encoder and decoder are specified by the set of reconstruction points $\{q_{\mathbf{i}_0^t}\}_{t=0}^{T-1}$.

Based on (3.12), (3.36) and (3.37), an encoder–decoder design algorithm is summarized in Algorithm 3.

3.4 Practical Implementation Issues

Before presenting the numerical experiments, we comment on some practical issues. Section 3.4.1 addresses the implementation of the conditional

Algorithm 3: Encoder–Decoder Design for the as System (3.21)

1. Initialize the encoder and decoder mappings $\{f_t\}_0^{T-1}$ and $\{g_t\}_0^{T-1}$. Compute the controller parameters $\{\ell_t\}$ by using (3.12).
2. For each $t = 0, \ldots, T-1$:
   2a. Update the encoder mapping f_t by using (3.37).
   2b. Update the decoder mapping d_t by using (3.36).
   2c. Set $u_t = -\ell_t d_t$.
3. If J_T has not converged, return to Step 2, otherwise terminate the iteration.

mean estimator, while in Section 3.4.2, a couple of initialization methods are discussed.

3.4.1 Conditional Mean Estimator

The most computationally intensive part in the proposed iterative algorithms is the computation of

$$\mathbf{E}\{\bar{x}_t|\mathbf{i}_0^t\} \quad \text{and} \quad \mathbf{E}\{x_s^2 + \rho u_s^2|x_0, \mathbf{v}_0^{t-1}, \mathbf{i}_0^{t-1}, i_t\}, \quad t < s \leq T-1.$$

When the encoder–decoder pairs as well as the pdf's $p(x_0)$ and $p(v_t)$ are known, the conditional pdf's $p(\bar{x}_t|\mathbf{i}_0^t)$, $t < T$, can be derived. Similarly, given the pdf of the current state, the pdf's of future estimates can also be calculated. Unfortunately, these estimation problems are computationally demanding. As an example, the state update from x_t to x_{t+1} alone is a function involving diverse linear and nonlinear operations. To derive the exact closed-form estimators is therefore not possible in general. In this thesis, the simulation based sequential Monte Carlo method (SMCM) is exploited to handle the nonlinear filtering problems. Here "sequential" refers to the fact that the filtering is accomplished for each new observation, which arrives sequentially in time. We illustrate the filtering procedure by taking the example of calculating $\mathbf{E}\{\bar{x}_t|\mathbf{i}_0^t\}$ for the system (3.3). At each time t, the Monte Carlo simulation involves the steps listed in Algorithm 4.

3.4.2 Initialization

One major weakness of the proposed algorithms is that the result rests heavily on the initial conditions. Here, we give comments on some

Algorithm 4: Simulation of $\mathbf{E}\{\bar{x}_t|\mathbf{i}_0^t\}$ using SMCM

1. Generate a set of points according to $p(\bar{x}_{t-1}|\mathbf{i}_0^{t-1})$.
2. Generate a set of process noise samples v_{t-1} according to $p(v_t)$.
3. By using the samples from Step 1 and Step 2, obtain a sequence of $\bar{x}_t$ based on (3.3).
4. Encode the sequence $\bar{x}_t$ according to (3.35).
5. Estimate $\mathbf{E}\{\bar{x}_t|\mathbf{i}_0^t\}$ based on the samples in Step 4.

encoder–decoder initialization methods used in later simulation experiments (Section 3.5 and Section 5.5). Given $p(x_0)$ for x_0 and $\mathbf{i}_0^{t-1}$ (and hence the corresponding $\mathbf{u}_0^{t-1}$), the pdf $p(\bar{x}_t|\mathbf{i}_0^{t-1})$ can be derived (estimated from a training set). A natural choice to initialize the proposed algorithms is to use the encoder–decoder of scalar Lloyd–Max quantizers [GG92], designed for the pdf's $p(\bar{x}_t|\mathbf{i}_0^{t-1}), t = 0, \ldots, T-1$. Starting by designing f_0 for $p(x_0)$, and using the resulting reconstruction points as initial estimates for $\hat{x}_{0|0}$ as well as ℓ_0 from (3.12), the conditional pdf's $p(\bar{x}_1|i_0)$, $i_0 \in \mathcal{I}_L$, can be determined. Base on these, L different Lloyd–Max quantizers can be trained to determine $f_1(x_0, i_0)$, and so on. This initialization will be particularly evaluated in Section 5.5. Alternatively, the encoder–decoders can be initialized as properly scaled uniform quantizers. By "properly scaled" we mean that the quantizer has memory and the time varying range is adapted to the range of $p(\bar{x}_t|\mathbf{i}_0^{t-1})$.

3.5 Numerical Experiments

In Section 3.5.1, Algorithm 2 developed for systems as (3.3) is evaluated by numerical experiments. While in Section 3.5.2, simulations based on Algorithm 3, developed for systems as (3.21), are presented.

3.5.1 Absence of Process Noise

Fig. 3.4 gives comparisons of the system performances for different initial pdf's. The setup of the numerical experiment is as follows. The initial state x_0 is drawn according to a generalized Gaussian distribution, see Appendix A. The notation $GGD(\alpha, \beta)$ is used, where α describes the exponential rate of decay, and β is the standard deviation. By varying α, the $GGD(\alpha, \beta)$ distribution provides a wide coverage from narrow-

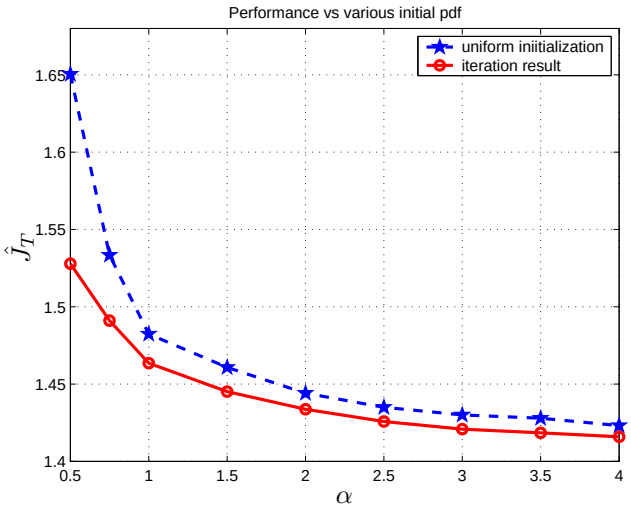


Figure 3.4: Performances vs various $GGD(\alpha, \beta)$ initial distributions. The normalized performance measure $\hat{J}_T$ is defined in (3.38).

tailed to broad-tailed pdf's. The variance $\beta = 2$ is used throughout the experiment, while α varies. The plant dynamics is given by $a = 0.8$. Regarding the cost function (2.9), the time horizon T is 3 and the weight ρ is 1, which together give $J_T = x_3^2 + \sum_{t=0}^{2}[x_t^2 + u_t^2]$. Since the initial states have the same variance, it is convenient to normalize the cost according to

$$\hat{J}_T = \frac{J_T}{\mathbf{E}\{x_0^2\}}. \tag{3.38}$$

Finally, the channel rate is $R = 2$ bits per time unit (per channel use).

Fig. 3.4 depicts that for an initial state with a broad pdf (a large α value) the performance is better than those of peaked pdf's (small α values). This result is expected, since a peaked pdf produces more often large-valued quantization errors. The dashed curve is obtained by employing uniform quantizers, which have optimized step-sizes with respect to the conditional pdf's $p(x_0|\mathbf{i}_0^{t-1})$. The solid curve is obtained by applying the optimized encoder–decoder using Algorithm 2 and initialized with the above-mentioned uniform quantizers. A clear performance improvement is observed, especially at small α-values.

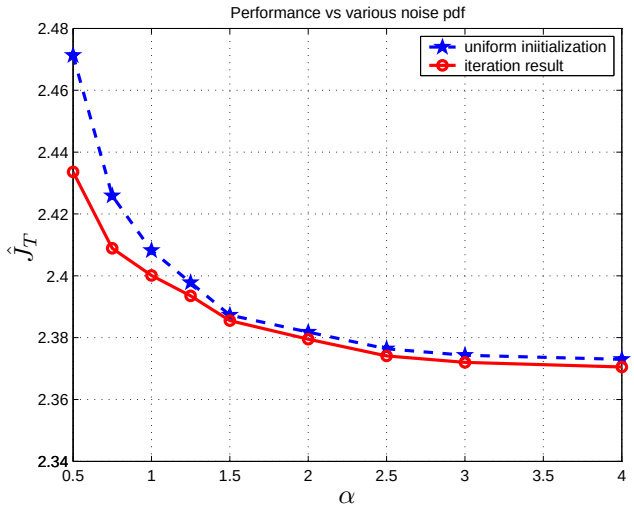


Figure 3.5: Performances comparison between different $GGD(\alpha, \beta)$ noise pdf's. The normalized performance measure $\hat{J}_T$ is defined in (3.38).

3.5.2 Presence of Process Noise

In this section the encoder–decoder pairs presented in Section 3.3.1 are evaluated by numerical experiments. Fig. 3.5 provides comparisons of system performances for process noise with different pdf's. The variances of the GGD process noise are held constant by setting $\beta = 0.25$ throughout, while α varies. The distribution of the initial state x_0 is restricted to $GGD(2, 1)$. The other system parameters are $a = 0.8$, $T = 3$, $\rho = 1$ and $R = 2$. The cost function is given by $\hat{J}_T$ in (3.38). As expected, the noise with a peaked pdf induces more harm to the entire cost function than the noise with a broad pdf. The dashed curve employs uniform quantizers whose step-sizes are optimized for each $p(\bar{x}_t|\mathbf{i}_0^{t-1})$. The solid curve is obtained by exploiting the optimized encoder–decoder pairs according to Algorithm 3 and initialized with the above-mentioned uniform quantizers. The simulations show a clear improvement in the system performance, especially for noise with peaked pdf's.

Chapter 4

Feedback Control over Binary Symmetric Channels

4.1 Background

In Chapter 3 we have considered noise-free channels. When communicating in hostile environments, for example over radio channels, transmission errors are however unavoidable. A well designed system should be capable to tolerate transmission errors. For control applications with strict delay constraints and low transmission rates, a source and channel separated strategy appears to be inefficient. This will be explored in the first part of this chapter. In particular, two alternatives of source and channel separated approaches are experimentally investigated in Section 4.2 and Section 4.3. Then, in Section 4.4, a joint design of source–channel code is proposed to further improve the overall system performance.

A fundamental problem in the system design involving noisy channels is how to handle the impairment caused by channel errors. The two basic strategies are either to prevent in advance or handle errors after they have occurred. In this thesis, the first strategy is considered, that is to design the reconstruction points taking the probability of bit errors into account.

As described previously in Section 2.3, the noise channel raises a "synchronization" problem to memory-based encoders and decoders. In case of error-free communications, $\mathbf{i}_0^t$ always equals $\mathbf{j}_0^t$, thus, in advance, the

encoder knows exactly the symbol the decoder will receive. Such a synchronization between the encoder and the decoder is not present when the channel produces transmission errors. In advance, the encoder cannot make error-free estimation of the received index j_t. Even worse, the decoder has lost the ability to track the transmitted symbol $\mathbf{i}_0^t$ after receiving $\mathbf{j}_0^t$. However, if the past control signals are available at both the encoder and the decoder, the received sequence $\mathbf{j}_0^t$ can be deduced at the encoder. Thus, the encoder and the decoder can have access to the same codebook (see Fig. 2.4) of reconstruction points $\{q_{\mathbf{j}_0^t}\}_0^{T-1}$. Since, at each time instant t, j_t is available to the decoder, the decoding function is

$$d_t = g_t(\mathbf{j}_0^t), t = 0, \ldots, T-1. \tag{4.1}$$

The optimal state estimator, based on $\mathcal{D}_t$ is

$$\hat{x}_{s|t} = \mathbf{E}\{x_s|\mathbf{j}_0^t\}, \quad t = 0, \ldots, \ T-1. \tag{4.2}$$

As will be explained later, the decoder mapping will be specialized to $d_t = \hat{x}_{t|t}$, so $\hat{x}_{t|t}$ will be utilized by the controller to produce the control output u_t.

4.2 Non-redundant Channel Coding

Consider the situation where a source coder (quantizer) designed for a noiseless channel (according to the methods in Chapter 3), encounters a binary symmetric channel. When the error probability ϵ is large, the overall system performance is deteriorated severely. In order to improve the system performance, we will here study how to design a good index assignment. This problem belongs to the class of non-redundant channel coding problems, see Section 1.4.4. Note that, since the encoder and the decoder are memory-based, the assignment involves a large number of different indices. A full search for the best combination is therefore not feasible. Simulated annealing, introduced in Section 1.4.4, is a practical method to obtain a fairly good index assignment with a low computational complexity.

Recall the settings in Section 1.4.4 that each index assignment is specified by a state vector π. For the memory-based problem, we let the state vector π consist of a number of sub-vectors $\pi_{\mathbf{j}_0^t}$, sorted according to the increasing (length) sequence $\mathbf{j}_0^t, t = 0, \ldots, T-1$. A sub-vector $\pi_{\mathbf{j}_0^t}$ is a function of the received indices $\mathbf{j}_0^t$ and it specifies the index assignment

Algorithm 5: Simulated Annealing

1. Select a cost function J_{IA}. Define an effective temperature τ and set $\tau = \tau_0$, where τ_0 is a high initial temperature.
2. Randomly choose an initial state π, which is a set of permutation sub-vectors specifying the index assignments for $\{f_t, g_t\}_{t=0}^{T-1}$. Calculate J_{IA} for the given π.
3. Randomly choose a new state $\bar{\pi}$, which performs a perturbation on one of the sub-vectors $\pi_{\mathbf{j}_0^t}$.
4. Let $\Delta J_{IA} = J_{IA}(\bar{\pi}) - J_{IA}(\pi)$,
   4.1. if $\Delta J_{IA} < 0$, replace π by $\bar{\pi}$;
   4.2. if $\Delta J_{IA} \geq 0$, replace π by $\bar{\pi}$ with a probability $e^{-\frac{\Delta J_{IA}}{\tau}}$.
5. If the number of cost drops exceeds a prescribed number or if too many unsuccessful perturbations occur, go to Step 6. Otherwise return to Step 3.
6. Lower the temperature. If the temperature is below some prescribed freezing temperature τ_f, terminate the iteration. Otherwise return to Step 3.

for the transmission of i_{t+1} when $\mathbf{j}_0^t$ has been received. The index assignment at each t consists of the mapping $\mathbf{C}$ from the index sequence $\{\mathbf{j}_0^{t-1}, i_t\}$ to a binary codeword $c(i_t)$:

$$c(i_t) = \mathbf{C}(\mathbf{j}_0^{t-1}, i_t), \tag{4.3}$$

and the inverse mapping $\mathbf{D}$ from $\{\mathbf{j}_0^{t-1}, c(j_t)\}$ to $\mathbf{j}_0^t$:

$$\mathbf{j}_0^t = \mathbf{D}(\mathbf{j}_0^{t-1}, c(j_t)), \tag{4.4}$$

see also Section 2.3. A perturbation $\bar{\pi}$ of the state π will select only one of the sub-vectors and exchange two entities in that sub-vector. The selection of the sub-vector and the change of the positions are pursued randomly. We present Algorithm 5 as a general algorithm for using simulated annealing to design the index assignment. The goal of the design is to minimize an overall cost J_{IA}, which is a function of index assignment π.

4.2.1 Cost Function

This section is devoted to a discussion of several issues about the cost function used in simulated annealing. The cost function J_T in (2.9) is

denoted as J_{IA} in Algorithm 5. Let us first take a look at computation J_T. Note that, given the distribution of the initial state, the encoder–decoder functions and the index assignment, J_T can be calculated. When the pdf's are described by a training set, J_T can be obtained numerically by sequential Monte Carlo simulations. The sequential Monte Carlo simulation starts by generating a sufficiently large set of samples of initial states x_0. A summary of the steps for the system evolution from the time instant t to $t+1$ is described in Algorithm 6. As was stated earlier, an exhaustive search for the globally optimal index assignment is computationally expensive. In the worst case, $(2^R)^T!$ combinations need to be tested, with $n!$ denoting n factorial. Hence the number of comparisons is infeasible for large valued R and T. By using simulated annealing, the number of comparisons can be reduced. Still, the amount of computation is prohibitive if J_T is re-calculated for each comparison. Note that, the complexity for simulating each J_T also increases exponentially with R and T. In each calculation of J_T, all the variables $\{x_t, i_t, j_t, u_t\}_{t=0}^{T-1}$ are involved. These variables are produced sequentially according to the below ordering:

$$\begin{aligned} &x_0 \to i_0 \to j_0 \to u_0 \to x_1 \to i_1 \to j_1 \to u_1 \to x_2 \dots \\ &\to x_{T-1} \to i_{T-1} \to j_{T-1} \to u_{T-1} \to x_T. \end{aligned}$$

In order to reduce the complexity of calculating J_T, we should avoid recalculating a full J_T by comparing only those terms affected by the index assignment. It is worth noting that, for each new state $\bar{\pi}$, the permutation of a certain sub-vector $\bar{\pi}_{\mathbf{j}_0^t}$ not only affects the cost $\mathbf{E}\{x_t^2 + \rho u_t^2 | \mathbf{j}_0^t\}$, but also the costs of all future time instants. On the other hand, the change in $\bar{\pi}_{\mathbf{j}_0^t}$ does not affect the costs contributed by the early time instants $s = 0, \dots, t-1$. Obviously, the index assignment at time t has most effect on u_t, and hence also x_{t+1}. Note that, the contributions to J_T, from x_t and u_t, decrease with time, so that the influence of the current index assignment on the future states is limited. Therefore, in Example 5 we will only have a look at the effect of the index assignment on the term $\mathbf{E}\{x_{t+1}^2 + \rho u_t^2 | \mathbf{j}_0^t\}$.

Example 5 Index assignment related terms in $\mathbf{E}\{x_{t+1}^2 + \rho u_t^2 | \mathbf{j}_0^t\}$
A unit step updating from x_t to x_{t+1} involves

$$x_t \to i_t \to j_t \to u_t \to x_{t+1}.$$

Let $\bar{u}_t$ denote the intended control signal based on $\{\mathbf{j}_0^{t-1}, i_t\}$ and u_t denote the actual control signal based on the received symbol $\mathbf{j}_0^t$. Observe that

Algorithm 6: Unit Step Update of x_t to x_{t+1}

1. Appling the encoding mapping f_t to x_t, gives i_t.
2. Appling the binary mapping $\mathbf{C}\{\mathbf{j}_0^{t-1}, i_t\}$, gives $c(i_t)$.
3. Let the coded binary symbol $c(i_t)$ be fed into a BSC.
4. Appling the inverse binary mapping $\mathbf{D}\{\mathbf{j}_0^{t-1}, c(j_t)\}$, gives $\mathbf{j}_0^t$.
5. Appling the decoding mapping g_t to $\mathbf{j}_0^t$, gives d_t.
6. Calculate the control signal u_t based on $\mathbf{d}_0^t$.
7. Evolving the system according to the system equation, gives x_{t+1}.

the transition probability $p(u_t|\bar{u}_t)$ depends on the index assignment. Let us make an elaboration on the term $\mathbf{E}\{x_{t+1}^2 + \rho u_t^2|\mathbf{j}_0^t\}$:

$$\begin{aligned}
&\mathbf{E}\{x_{t+1}^2 + \rho u_t^2|\mathbf{j}_0^t\} \\
=\ &\mathbf{E}\{(ax_t + u_t)^2 + \rho u_t^2|\mathbf{j}_0^t\} \\
=\ &\mathbf{E}\{(ax_t + \bar{u}_t + u_t - \bar{u}_t)^2 + \rho(u_t - \bar{u}_t + \bar{u}_t)^2|\mathbf{j}_0^t\} \\
=\ &\mathbf{E}\{(ax_t + \bar{u}_t)^2 + 2(ax_t + \bar{u}_t)(u_t - \bar{u}_t) + (u_t - \bar{u}_t)^2 \\
&\qquad +\rho(u_t - \bar{u}_t)^2 + 2\rho\bar{u}_t(u_t - \bar{u}_t) + \rho\bar{u}_t^2|\mathbf{j}_0^t\} \\
=\ &\underbrace{\mathbf{E}\{(ax_t + \bar{u}_t)^2 + \rho\bar{u}_t^2|\mathbf{j}_0^t\}}_{\textbf{(I)}} + \underbrace{\mathbf{E}\{2(ax_t + \bar{u}_t + \rho)(u_t - \bar{u}_t)|\mathbf{j}_0^t\}}_{\textbf{(II)}} \\
&\qquad \underbrace{+\mathbf{E}\{(\rho+1)(u_t - \bar{u}_t)^2|\mathbf{j}_0^t\}}_{\textbf{(III)}}. \qquad (4.5)
\end{aligned}$$

Observe that, the term **(I)** is independent of the index assignment at time t, but **(II)** and **(III)** are. Thus in the comparison, **(I)** can be neglected. Moreover, the values of u_t and $\bar{u}_t$ are determined by the reconstruction points $q_{\mathbf{j}_0^t}$ and there are a finite number of them. Hence, all possible $u_t - \bar{u}_t$ and $(u_t - \bar{u}_t)^2$ can be calculated once and stored in a look-up table.

Besides the identification of the terms relevant to the index assignment, the computational complexity can be further reduced if there are some symmetry properties to exploit. As an example, a binary symmetric channel can result in the same transition probability $p(u_t|\bar{u}_t)$ for different index assignments. Revisiting the example in Section 1.4.4, there are totally 24 ways to assign the binary codewords $\{00,\ 01,\ 10,\ 11\}$ to the integer indices $\{0, 1, 2, 3\}$. However due to the symmetry in the channel,

the quantizer and the distribution of the input signal, there are only 3 different values of the cost. How to divide the bit assignments into the three groups is illustrated in Fig. 1.7.

4.2.2 Numerical Example

In this section we present a comparison of system performances between an average of randomly selected index assignments and an index assignment designed by using Algorithm 5.

The encoder–decoder pairs employed in this example are designed for a system as (3.3) with $a = 0.8$. The channel is assumed to be error-free and has a rate $R = 2$. The initial state is drawn according to $p(x_0) = GGD(4, 2)$. The system performance is evaluated as the cost $\tilde{J}_3$, which is defined as

$$\tilde{J}_3 = \mathbf{E}\left\{\sum_{t=1}^{3}[x_t^2 + u_{t-1}^2]\right\}. \tag{4.6}$$

Compared to the cost function in (2.9), the contribution from the initial state x_0 is omitted in (4.6). When the channel produces errors the system performance is degraded, since the reconstruction points are optimal only when the correct indices are received. In (3.3), the dashed line is the average performance of randomly selected index assignments, whereas the solid line shows the performance by exploiting the index assignment resulted from the simulated annealing. An improvement in performance is obvious, but the improvement is marginal. A further improvement can be achieved when exploiting joint source–channel codes, which will be presented in Section 4.4.

4.3 Separate Source–Channel Coding

In the previous section, non-redundant channel coding was experimentally examined. In this section, the restriction that the whole channel rate is occupied by the source code is relaxed. The trade-off between source coding and separate channel coding becomes part of the overall design. This means that a greater protection against channel noise can be obtained at the cost of a coarser quantization. In general, fundamental results in coding theory, such as the source–channel separation theorem, hinge entirely on an infinitely large codeword length. Concerning control applications, the real-time nature may require extraordinarily short

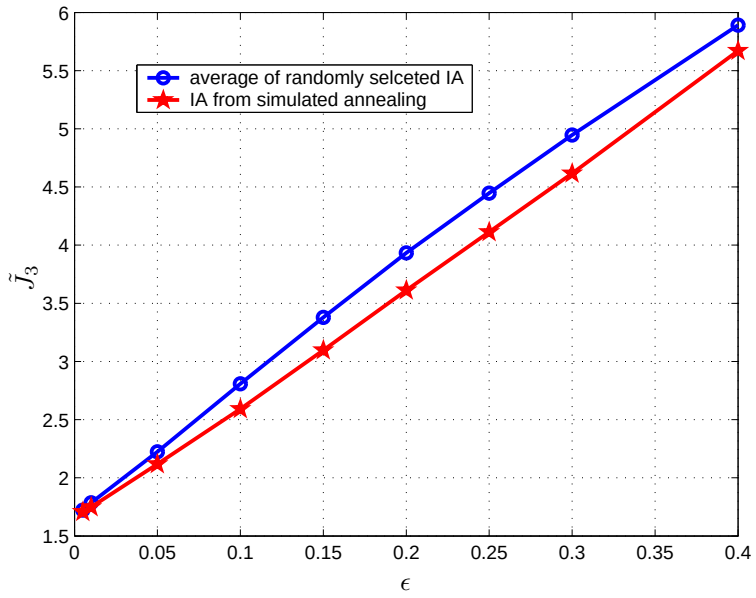


Figure 4.1: The performance comparison between average of randomly selected index assignments and a designed index assignment using Algorithm 5.

codewords. In the following section, a numerical example is given to demonstrate the poor performance when source and channel codes are designed separately.

4.3.1 Numerical Example

The system is the one in (3.3) with $a = 0.8$ and $p(x_0) = GGD(2, 2)$. The cost function is a modification of the one in (2.9) with $T = 3$ and $\rho = 1$, namely (4.6).

Now first assuming that the channel rate is $R = 3$, which implies three ways to allocate the three bits between the source code and the channel code. As an example, we allot two bits on the source coding. Only one bit is left for the channel protection and a simple channel code is proposed as below. The codewords that the encoder will choose are $\{000, 110, 101, 011\}$ so that the minimum distance of this code is 2. At the decoder, received $\{000, 001\}$ are decoded to $\{000\}$; $\{110, 100\}$ are

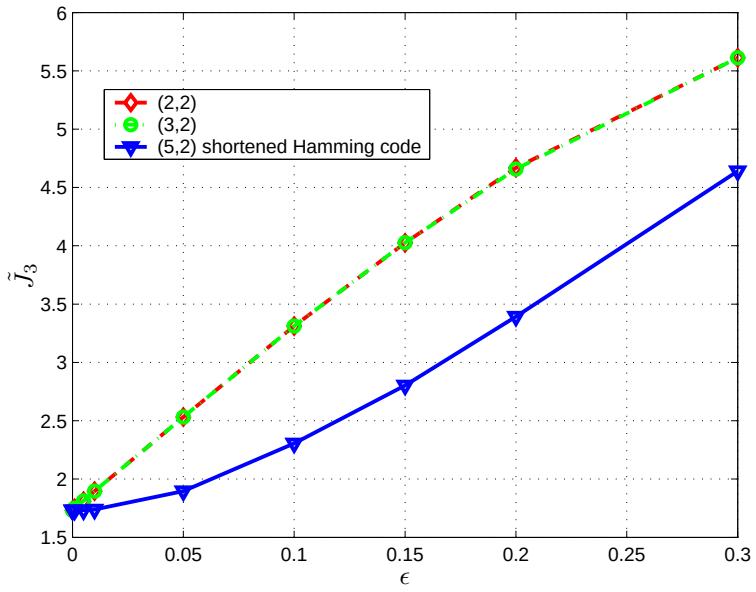


Figure 4.2: The improvement in the system performance when using separate channel codes. By (n, k) we mean k information bits will be coded to a codeword of n bits.

decoded to $\{110\}$; $\{101, 001\}$ are decoded to $\{101\}$ and finally, $\{011, 111\}$ are decoded to $\{011\}$. According to channel coding theory, a code that can correct single-bit errors requires at least a minimum distance of three, which is not offered by the proposed code. Obviously, there is no channel code of length three, which has a minimum distance of three and a code rate of $2/3$. The code rate is defined as the ratio between the number of input bits and the number of output bits. Based on the minimum distance, we expect a remarkably poor performance of this code. Even though at a low error probability, for example only one bit error per codeword, there is only a chance of $1/3$ for this code can correct the error. At high error probabilities, the channel code may do more harm than good.

Fig. 4.2 provides a simulated performance comparison among the cases with and without channel codes for an error probability region between $\epsilon = 0$ to $\epsilon = 0.3$. In all experiments, the same quantizer with

$R = 2$ is exploited. We have implemented two channel codes. Besides the three-bits code described in the previous paragraph, a $(5, 2)$ shortened Hamming code is adopted, which has 2 information bits as the input and produces a codeword of 5 bits. The codewords are namely $\{00000, 10110, 01011, 11101\}$ with a minimum distance of 3. This codeword can correct all single-bit errors and even some 2-bits errors. Fig. 4.2 shows that the improvement using the $(3, 2)$ block code is invisible, while the shortened Hamming code performs much better. However, it is worth noting that the $(5, 2)$ block code has used 3 bits more to transmit the state information compared to the solution without any channel code. That means the good performance is paid by a considerably inefficient code rate. With this example we have illustrated that to design good separate source and channel codes of extraordinarily short lengths is very difficult. In the next section, a design of joint source–channel code is presented that will improve the control performance significantly.

4.4 Channel Optimized Encoder–Decoder

The two coding strategies presented in the previous sections belong to the category where source and channel codes are designed separately. The first method was referred to as non-redundant channel coding, while the second approach employed channel codes with short block lengths. Numerical experiments shows that for both strategies, the capabilities of protecting against channel errors are not impressive. For control applications with strict delay constants and bandwidth limitations, we stress that a combined source–channel coding scheme can improve upon these traditional approaches. In this section, we propose a joint design, suitable for short codewords, accomplishing the source compression and the channel protection simultaneously.

In particular, we will examine the following time-invariant linear system,

$$\begin{cases} x_{t+1} &= ax_t + u_t + v_t, \\ y_t &= x_t, \end{cases} \tag{4.7}$$

where the pdf's $p(x_0)$ and $p(v_t)$ are known, as well as the linear dynamics $a, |a| < 1$, see Fig. 2.2. Compared to the system in (3.21), here, the received index j_t is not necessarily equal to the transmitted index i_t, due to the potential transmission errors produced by a binary symmetric channel. Again, we emphasize that the encoder and the decoder have access

to the same codebook (see Fig. 2.4), which consists of the reconstruction points $\{q_{\mathbf{j}_0^t}\}_0^{T-1}$. As will be explained later, the reconstruction point $q_{\mathbf{j}_0^t}$ is namely

$$q_{\mathbf{j}_0^t} = \mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\}, \tag{4.8}$$

with $\bar{x}_t$ defined in (3.31), i.e.,

$$\bar{x}_t = x_t - \sum_{s=0}^{t-1} a^{t-1-s}u_s = a^t x_0 + \sum_{s=0}^{t-1} a^{t-1-s}v_s.$$

4.4.1 Certainty Equivalence

Let us first take a look at the conditions for the CE property, given fixed encoder–decoder pairs. According to Section 3.1.1, by the assumption that v_t is uncorrelated with x_t, and u_t is a function of $\mathbf{j}_0^t$, the optimization of the cost function J_T in (3.18) is simplified to

$$\begin{aligned}
&\min_{\mathbf{u}_0^{T-1}} \mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t x_t)^2\right\} \\
&\quad= \mathbf{E}_{\mathbf{j}_0^t}\left\{\min_{\mathbf{u}_0^{T-1}} \mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t x_t)^2\Big|\mathbf{j}_0^t\right\}\right\} \\
&\quad= \mathbf{E}_{\mathbf{j}_0^t}\left\{\min_{\mathbf{u}_0^{T-1}} \mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t(x_t-\hat{x}_{t|t}+\hat{x}_{t|t}))^2\Big|\mathbf{j}_0^t\right\}\right\} \\
&\quad= \mathbf{E}_{\mathbf{j}_0^t}\left\{\min_{\mathbf{u}_0^{T-1}} \left(\mathbf{E}\left\{\sum_{t=0}^{T-1}(\rho+P_{t+1})(u_t+\ell_t\hat{x}_{t|t})^2\Big|\mathbf{j}_0^t\right\}\right.\right. \\
&\qquad\left.\left.+\sum_{t=0}^{T-1}\mathbf{E}\{\ell_t^2(\rho+P_{t+1})(x_t-\hat{x}_{t|t})^2|\mathbf{j}_0^t\}\right)\right\},
\end{aligned} \tag{4.9}$$

where P_t and ℓ_t are defined in (3.12), and $\hat{x}_{t|t} = \mathbf{E}\{x_t|\mathbf{j}_0^t\}$. With a similar reasoning as the one in Section 3.1, the condition for the CE property can be shown to be to require that

$$\mathbf{E}_{\mathbf{j}_0^t}\left\{\sum_{t=0}^{T-1}\ell_t^2(\rho+P_{t+1})\mathbf{E}\{(\hat{x}_{t|t}-x_t)^2|\mathbf{j}_0^t\}\right\} \tag{4.10}$$

is not a function of $\mathbf{u}_0^{t-1}$, which means the mean squared estimation errors are not dependent on the past control signals. If that is the case, the CE controller is optimal and is given by

$$d_t = \hat{x}_{t|t}, \quad \text{and} \quad u_t = -\ell_t d_t \tag{4.11}$$

(i.e., $z_t(\mathbf{d}_0^t) = -\ell_t d_t$) with

$$\ell_t = \frac{aP_{t+1}}{P_{t+1}+\rho}, \quad P_t = 1 + \frac{a^2 P_{t+1}\rho}{P_{t+1}+\rho}, \tag{4.12}$$

for $t = 0, \ldots, T-1$, and P_t is initialized with $P_T = 1$. The equations in (4.11) and (4.12) are almost identical with those in (3.11) and (3.12). The difference lies in the term $\hat{x}_{t|t}$. Here, we have $\hat{x}_{t|t} = \mathbf{E}\{x_t|\mathbf{j}_0^t\}$ instead of $\hat{x}_{t|t} = \mathbf{E}\{x_t|\mathbf{i}_0^t\}$ as in Chapter 3.

4.4.2 Encoder–Decoder Structure

By using a similar reasoning as the one in Section 3.2.1, the encoder–decoder structure for the system (4.7) is as follows. Note that the encoder knows $\mathbf{v}_0^{t-1}$, based on $\mathbf{x}_0^t$ and $\mathbf{u}_0^{t-1}$. Moreover, $\mathbf{u}_0^{t-1}$ is completely determined by $\mathbf{j}_0^{t-1}$. Let i_t be generated as

$$i_t = f_t(x_0, \mathbf{v}_0^{t-1}, \mathbf{j}_0^{t-1}), \quad t = 0, \ldots, T-1. \tag{4.13}$$

Since $i_0 = f_0(x_0), i_1 = f_1(x_0, v_0, j_0), \ldots, i_t = f_t(x_0, \mathbf{v}_0^{t-1}, \mathbf{j}_0^{t-1})$, the values $\mathbf{j}_0^t$ depend only on $x_0, \mathbf{v}_0^{t-1}$ and the potential channel errors, and not on $\mathbf{u}_0^s$ for any $s < t$. The CE controller in (4.11) is optimal for fixed encoders $\{f_t\}_0^{T-1}$ with the structure $i_t = f_t(x_0, \mathbf{v}_0^{t-1}, \mathbf{j}_0^{t-1})$.

Observe that, $x_t - \hat{x}_{t|t}$ equals to $\bar{x}_t - \mathbf{E}\{\bar{x}|\mathbf{j}_0^{t-1}\}$ and therefore the estimation error is not a function of $\mathbf{u}_0^s$. We see that $\bar{x}_t$ is a sufficient statistic to derive the optimal control. There is no loss in performance by enforcing the encoder structure $f_t(\bar{x}_t, \mathbf{j}_0^{t-1})$ and the reconstruction point $q_{\mathbf{j}_0^t} = \mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\}$. The decoder mapping at the decoder will be

$$d_t = g_t(\mathbf{j}_0^t) = \hat{x}_{t|t}, \quad t = 0, \ldots, T-1, \tag{4.14}$$

as shown in Section 4.1. In Example 6, we will study the encoder–decoder structure for a special case of the general model (4.7), in particular when $v_t = 0, \forall t$.

Example 6 Absence of Process Noise
In this example, we consider a special case of the general model (4.7), namely, when $v_t = 0, \forall t$. The system model is

$$\begin{cases} x_{t+1} &= ax_t + u_t, \\ y_t &= x_t, \end{cases} \tag{4.15}$$

where the pdf $p(x_0)$ is known, as well as the linear dynamics $a, |a| < 1$. In this case, the necessary and sufficient condition for the separation derived in Section 4.4.1, can be shown to correspond to the requirement that the average estimation error,

$$\mathbf{E}\{(x_0 - \hat{x}_{0|t})^2\}, \quad t = 0, \ldots, T-1, \tag{4.16}$$

is not a function of $\mathbf{u}_0^s$, $s = 0, \ldots, t-1$. By enforcing the encoder to have the structure $f_t(x_0, \mathbf{j}_0^{t-1})$, it is quite straightforward to verify that the above condition holds true, based on the fact that $i_0 = f_0(x_0), i_1 = f_1(x_0, j_0), \ldots, i_t = f_t(x_0, \mathbf{j}_0^{t-1})$. The reconstruction point can therefore be $q_{\mathbf{j}_0^t} = \mathbf{E}\{x_0|\mathbf{j}_0^t\}$. Observe that $\bar{x}_t = a^t x_0$ when $v_t = 0, \forall t$. The encoder $f_t(\bar{x}_t, \mathbf{j}_0^{t-1})$ from Section 4.4.2 in this case will become $f_t(ax_0, \mathbf{j}_0^{t-1})$, which agrees well with the proposed structure $f_t(x_0, \mathbf{j}_0^{t-1})$.

4.4.3 Iterative Training Algorithm

An iterative training method is proposed in this section to optimize the encoder–decoder pairs. Similar as in Section 3.4, this iterative approach starts with any given initial setup, updates the reconstruction points and the encoding rules considering not only the past but also the future state evolution.

Let us first look at the optimal decoder for the fixed $\{f_t\}_{t=0}^{T-1}$. Recall that the optimal controller is $u_t = -\ell_t d_t$, where $d_t = g_t(\mathbf{j}_0^t)$, then the optimal decoder is

$$g_t(\mathbf{j}_0^t) = \hat{x}_{t|t} = q_{\mathbf{j}_0^t} + \sum_{s=0}^{t-1} a^{t-1-s} u_s, \tag{4.17}$$

where $q_{\mathbf{j}_0^t} = \mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\}$ denotes the reconstruction point at time t, stored in a codebook at the decoder (and the encoder, c.f. Fig. 2.4).

The optimal encoder needs to take the impact of future state evolutions into account. Hence, for fixed decoders and controllers, and given

Algorithm 7: Encoder–Decoder Design for the System (4.7)

1. Initialize the encoder and decoder mappings $\{f_t, g_t\}_{t=0}^{T-1}$. Compute the controller parameters $\{\ell_t\}$ by using (4.12).
2. For each $t = 0, \ldots, T-1$:
   2a. Update the encoder mapping f_t by using (4.18).
   2b. Update the decoder mapping d_t by using (4.17).
   2c. Set $u_t = -\ell_t d_t$.
3. If J_T has not converged, return to Step 2, otherwise terminate the iterations.

the encoder mappings f_t for $t = 0, \ldots, t-1, t+1, \ldots, T-1$, the optimal $f_t = f_t(\bar{x}_t, \mathbf{j}_0^{t-1})$ is

$$i_t = \arg\min_{i \in \mathcal{I}_L} \mathbf{E}\left\{ \sum_{k=t}^{T} x_k^2 + \rho u_k^2 \Big| \bar{x}_t, \mathbf{j}_0^{t-1}, i_t = i \right\}. \tag{4.18}$$

According to (4.18), the encoding rule is updated once the reproduction points are recalculated. Therefore the coder is specified by the set of reconstruction points $\{q_{\mathbf{j}_0^t}\}_{t=0}^{T-1}$.

Based on (4.12), (4.17) and (4.18), an encoder–decoder design algorithm is formulated in Algorithm 7. The convergence is monitored based on updating the value of J_T in each step.

4.4.4 Evolution of the Estimation Error

For control applications, the imperfections of the communication link are in general not critical as long as they do not deteriorate the closed-loop performance. However, the idea of separation suggests to optimize the estimation and the control separately. Hence the optimum estimation problem is still of interest. In Chapter 3, we have shown that for fixed the encoder–decoder pairs, the controller in (3.11) is optimal in the senses of both "cost-to-go" (3.4) and the total cost (2.9). In this section, the optimal control is revisited from an estimation perspective. In particular, we are going to explore the evolution of the estimation error.

Let the encoder have the structure $f_t(\bar{x}_t, \mathbf{j}_0^{t-1})$ and define the estimation error $\tilde{x}_t$ as

$$\tilde{x}_t = x_t - \hat{x}_{t|t}. \tag{4.19}$$

Observe that $\mathbf{E}\{x_t^2|\mathbf{j}_0^t\}$ can be written as

$$\begin{aligned}\mathbf{E}\{x_t^2|\mathbf{j}_0^t\} &= \mathbf{E}\{(x_t-\hat{x}_{t|t}+\hat{x}_{t|t})^2|\mathbf{j}_0^t\}\\ &= \mathbf{E}\{(x_t-\hat{x}_{t|t})^2|\mathbf{j}_0^t\} - 2\mathbf{E}\{(x_t-\hat{x}_{t|t})\hat{x}_{t|t}|\mathbf{j}_0^t\} + \hat{x}_{t|t}^2\\ &= \mathbf{E}\{(x_t-\hat{x}_{t|t})^2|\mathbf{j}_0^t\} - 2\hat{x}_{t|t}(\hat{x}_{t|t}-\hat{x}_{t|t}) + \hat{x}_{t|t}^2\\ &= \mathbf{E}\{(x_t-\hat{x}_{t|t})^2|\mathbf{j}_0^t\} + \hat{x}_{t|t}^2, \end{aligned} \tag{4.20}$$

and accordingly, the cost function J_T can be written as

$$\begin{aligned}J_T &= \mathbf{E}\left\{x_{T+1}^2 + \sum_{t=0}^{T}[x_t^2+\rho u_t^2]\right\}\\ &= \mathbf{E}\Bigg\{(x_{T+1}-\hat{x}_{T+1|T+1})^2 + \hat{x}_{T+1|T+1}^2\\ &\qquad + \sum_{t=0}^{T}\left[(x_t-\hat{x}_{t|t})^2+\hat{x}_{t|t}^2+\rho u_t^2\right]\Bigg\}\\ &= \mathbf{E}\left\{\hat{x}_{T+1|T+1}^2 + \sum_{t=0}^{T}[\hat{x}_{t|t}^2+\rho u_t^2] + \sum_{t=0}^{T+1}(x_t-\hat{x}_{t|t})^2\right\}. \end{aligned} \tag{4.21}$$

Regarding the term $x_t-\hat{x}_{t|t}$ in (4.21), it can be rewritten as

$$\begin{aligned}x_t-\hat{x}_{t|t} &= \bar{x}_t + \sum_{s=0}^{t-1}a^{t-1-s}u_s - \mathbf{E}\{\bar{x}_t + \sum_{s=0}^{t-1}a^{t-1-s}u_s|\mathbf{j}_0^t\}\\ &= \bar{x}_t - \mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\} = \tilde{x}_t, \end{aligned} \tag{4.22}$$

where $\tilde{x}_t$ is a function of $\bar{x}_t$ and hence does not depend on the past control sequence $\mathbf{u}_0^{t-1}$. For convenience, we define a new linear quadratic cost function $\bar{J}$ that is

$$\begin{aligned}\bar{J} &= \mathbf{E}\left\{\hat{x}_{T+1|T+1}^2 + \sum_{t=0}^{T}\left[\hat{x}_{t|t}^2+\rho u_t^2\right]\right\}\\ &= J_T - \sum_{t=0}^{T+1}\mathbf{E}\{\tilde{x}_t^2\}. \end{aligned} \tag{4.23}$$

Taking a close look at the term $\hat{x}_{t+1|t+1}$, we can write it as

$$\begin{aligned}\hat{x}_{t+1|t+1} &= \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t+1}\} \\ &= \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t}\} + \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t+1}\} - \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t}\} \\ &= \mathbf{E}\{ax_t + u_t + v_t|\mathbf{j}_0^{t}\} + \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t+1}\} - \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t}\} \\ &= a\mathbf{E}\{x_t|\mathbf{j}_0^{t}\} + u_t + \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t+1}\} - \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t}\}, \end{aligned} \tag{4.24}$$

since v_t is independent of $\mathbf{j}_0^t$. In particular, j_t is related to x_0 and $\mathbf{v}_0^{t-1}$ (recall $x_t = ax_{t-1} + u_{t-1} + v_{t-1}$). Observe that, the equation (4.24) describes the evolution of $\hat{x}_{t|t}$ with time, i.e,

$$\hat{x}_{t+1|t+1} = a\hat{x}_{t|t} + u_t + \acute{x}_{t+1}, \tag{4.25}$$

where $\acute{x}_{t+1} = \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t+1}\} - \mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t}\}$. The term $\acute{x}_{t+1}$ can be interpreted as the difference in estimating x_{t+1} before and after j_{t+1} is observed.

By exploiting the sequence P_t specified in (3.8), the term $\bar{J} + \mathbf{E}\{\hat{x}_{0|0}^2 P_0\}$ can be written as

$$\begin{aligned}\bar{J} + \mathbf{E}\{\hat{x}_{0|0}^2 P_0\} &= \mathbf{E}\left\{\hat{x}_{T+1|T+1}^2 + \sum_{t=0}^{T}\left[\hat{x}_{t|t}^2 + \rho u_t^2\right]\right\} + \mathbf{E}\{\hat{x}_{0|0}^2 P_0\} \\ &= \sum_{t=0}^{T}\mathbf{E}\{\hat{x}_{t|t}^2 + \rho u_t^2 + \hat{x}_{t+1|t+1}^2 P_{t+1} - \hat{x}_{t|t}^2 P_t\} \\ &= \sum_{t=0}^{T}\mathbf{E}\{(1-P_t)\hat{x}_{t|t}^2 + \rho u_t^2 + P_{t+1}(a\hat{x}_{t|t} + u_t + \acute{x}_{t+1})^2\} \\ &= \sum_{t=0}^{T}\mathbf{E}\{(1-P_t)\hat{x}_{t|t} + \rho u_t^2 + P_{t+1}(a\hat{x}_{t|t} + u_t)^2 \\ &\qquad + 2P_{t+1}(a\hat{x}_{t|t} + u_t)\acute{x}_{t+1} + P_{t+1}\acute{x}_{t+1}^2\}. \end{aligned} \tag{4.26}$$

Regarding the term $\acute{x}_{t+1}$, one can show that the following holds true,

$$\begin{aligned}\mathbf{E}\{\acute{x}_{t+1}|\mathbf{j}_0^t\} &= \mathbf{E}\{\mathbf{E}\{x_{t+1}|\mathbf{j}_0^{t+1}\} - \mathbf{E}\{x_{t+1}|\mathbf{j}_0^t\}\ |\mathbf{j}_0^t\} \\ &= \mathbf{E}\{x_{t+1}|\mathbf{j}_0^t\} - \mathbf{E}\{x_{t+1}|\mathbf{j}_0^t\} = 0. \end{aligned} \tag{4.27}$$

Hence, the cost function $\bar{J} + \mathbf{E}\{\hat{x}_{0|0} P_0\}$ can be written as

$$\bar{J} + \mathbf{E}\{\hat{x}_{0|0} P_0\} = \sum_{t=0}^{T} \Big[\mathbf{E}\{(1-P_t)\hat{x}_{t|t}^2 + \rho u_t^2 + P_{t+1}(a\hat{x}_{t|t} + u_t)^2\} + \mathbf{E}\{P_{t+1}\acute{x}_{t+1}^2\} \Big]. \tag{4.28}$$

Substituting P_t into (4.28), gives

$$\begin{aligned}
\bar{J} + \mathbf{E}\{\hat{x}_{0|0}^2 P_0\} &= \sum_{t=0}^{T} \mathbf{E}\Big\{ \left(-\frac{a^2 \rho P_{t+1}}{\rho + P_{t+1}} \right) \hat{x}_{t|t} + \rho u_t^2 + P_{t+1} a^2 \hat{x}_{t|t}^2 \\
&\qquad + P_{t+1} u_t^2 + 2a P_{t+1} u_t \hat{x}_{t|t} \Big\} + \mathbf{E}\{P_{t+1}\acute{x}_{t+1}^2\} \\
&= \sum_{t=0}^{T} \mathbf{E}\Big\{ \frac{1}{\rho + P_{t+1}} ((\rho + P_{t+1})^2 u_t^2 + P_{t+1}^2 a^2 \hat{x}_{t|t}^2 \\
&\qquad + 2a P_{t+1} \hat{x}_{t|t} (\rho + P_{t+1}) u_t \Big\} + \mathbf{E}\{P_{t+1}\acute{x}_{t+1}^2\} \\
&= \sum_{t=0}^{T} \mathbf{E}\Big\{ \frac{1}{\rho + P_{t+1}} ((\rho + P_{t+1}) u_t + a P_{t+1} \hat{x}_{t|t})^2 \Big\} \\
&\qquad + \mathbf{E}\{P_{t+1}\acute{x}_{t+1}^2\}.
\end{aligned} \tag{4.29}$$

We see again that, the control that minimizes (4.29) is the CE controller in (4.11), i.e.,

$$u_t^\star = -\frac{a P_{t+1}}{\rho + P_{t+1}} \hat{x}_{t|t} = -\ell_t \hat{x}_{t|t}.$$

Substituting $u_t^\star$ into J_T, the resulting $J_T^\star$ is the minimum cost for the given encoder–decoder pairs,

$$J_T^\star = \mathbf{E}\{\hat{x}_{0|0}^2 P_0\} + \sum_{t=0}^{T} \mathbf{E}\{P_{t+1}\acute{x}_{t+1}^2\} + \sum_{t=0}^{T+1} \mathbf{E}\{\tilde{x}_t^2\}. \tag{4.30}$$

Note that, the cost $J_T^\star$ is a function of the encoder–decoder pairs $\{f_t, g_t\}_{t=0}^{T-1}$. The overall optimization problem is thus equivalent to finding the encoder–decoder pairs, among all the encoder–decoder pairs, which result in the minimum $J_T^\star$:

$$J_T^{opt} = \min_{f_0^{T-1}, g_0^{T-1}} J_T^\star. \tag{4.31}$$

Algorithm 8: Calculation of $\mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\}$ using SMCM

1. Generate a set of points according to $p(\bar{x}_{t-1}|\mathbf{j}_0^{t-1})$.
2. Generate a set of process noise v_{t-1} according to $p(v_{t-1})$.
3. Using the samples from Step 1 and Step 2, obtain a sequence of $\bar{x}_t$ based on (4.7).
4. Encode the sequence $\bar{x}_t$, according to (4.18).
5. Simulate the transmission over the channel and create a sequence of received symbols j_t.
6. Decode the received symbols j_t and derive $\mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\}$.

This size-reduced optimization problem (4.31), cf. (2.10), however still involves a large number of design parameters.

4.4.5 Practical Implementation Issues

In this section we consider some issues regarding the practical implementation of Algorithm 7.

Conditional Mean Estimator

Comparing with Section 3.4.1, here, the most computationally intensive part of Algorithm 7 is the computation of

$$\mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\} \text{ and } \mathbf{E}\{x_s^2 + \rho u_s^2|x_0, \mathbf{v}_0^{t-1}, \mathbf{j}_0^{t-1}, i_t\}, \quad s > t.$$

When the encoder–decoder pairs are known, as well as the pdf's $p(x_0)$, $p(v_t)$ and ϵ, the conditional pdf $p(\bar{x}_t|\mathbf{j}_0^t)$, $t < T$, can be derived. Similarly, given the pdf of the current state, the pdf's of future estimates can also be derived. However these estimation problems are computationally demanding. As in Section 3.4, a sequential Monte Carlo approach is adopted to handle the nonlinear filtering problems.

At each time t, to obtain $\mathbf{E}\{\bar{x}_t|\mathbf{j}_0^t\}$, a Monte Carlo simulation involves the steps listed in Algorithm 8. Since the transition probability of a BSC has a closed-form expression, Step 4 and Step 5 can be replaced by

$$\mathbf{E}\{\bar{x}_t|\mathbf{j}_0^{t-1}, j_t = k\} = \sum_{i_t} \mathbf{E}\{\bar{x}_t|\mathbf{j}_0^{t-1}, i_t\} \frac{p(j_t = k|i_t)p(i_t)}{p(j_t = k)}, \tag{4.32}$$

where $p(i_t)$ and $p(j_t)$ are simulated.

Initialization

Different initialization methods will be compared in Section 4.4.6. Similar as in Section 3.4, all of them exploit memory-based quantizers. The least computationally intensive initialization employs a time varying uniform quantizer. The range of the quantizer has only taken the pdf of the initial state and the dynamics of the plant into account. The second initialization applies Lloyd–Max quantizer, designed for each pdf $p(\bar{x}_t|\mathbf{j}_0^t),\ t = 0, \ldots, T-1$. That is to say, starting by designing f_0 for $p(x_0)$, and using the resulting reconstruction points as initial estimates for $\hat{x}_{0|0}$ as well as ℓ_0 from (4.9), the conditional pdf's $p(x_0|j_0)$, $j_0 \in \mathcal{I}_L$, can be determined, and so on for increasing t. The last initialization applies channel optimized quantizers, designed also for each pdf $p(\bar{x}_t|\mathbf{j}_0^t)$. It is worth noting that $p(\bar{x}_t|\mathbf{j}_0^t)$ is dependent on $\{f_s, g_s\}_{s=0}^t$.

4.4.6 Numerical Experiments

To evaluate the advantage of a joint design of source–channel coding, Algorithm 7 is applied first to a system without process noise, as (4.15), and thereafter to a system with process noise, as (4.7) .

Absence of Process Noise

Numerical experiments are pursued for a plant with $a = 0.8$. The performance is measured by J_T in (2.9) with $\rho = 1$ and $T = 3$. The random magnitude of the initial state is modelled using the generalized Gaussian distribution. Fig. 4.3 shows simulation results for $p(x_0) = GGD(4, 2)$, where the encoder–decoders for $R = 2$ and $R = 3$ have been designed. The channel transition probability ϵ is shown on the x-axis, while on the y-axis, we plot the costs J_T's. The horizontal dashed line in Fig. 4.3 represents the cost if no control action is taken. We can read from the figure that there is still some advantage to feedback the state observation over a noisy channel even with 40% bit errors. An interesting problem would be to examine the trade-off between ϵ and establish how coarse the feedback information can be, in order to meet a certain performance requirement.

As described earlier, the iterative design requires an initial setup of reconstruction points. Two initialization methods are compared. The first method, referred to as $\mathbf{ini}_{uni}$, employs properly scaled uniform quantizers. The second method, referred to as $\mathbf{ini}_{co}$, applies channel optimized

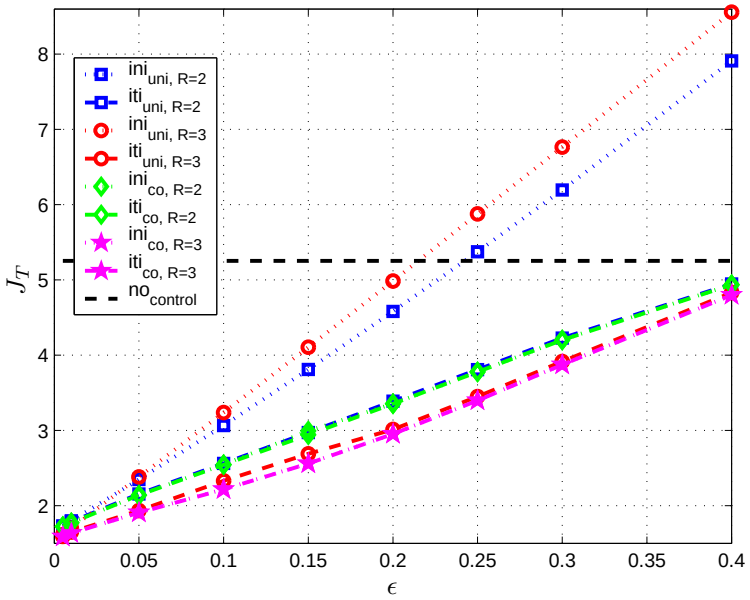


Figure 4.3: Performance comparisons among encoder–decoder pairs. $\mathbf{ini}_{uni}$: uniform initialization; $\mathbf{ini}_{co}$: channel optimized initialization; $\mathbf{iti}_{uni}$: iteratively trained with $\mathbf{ini}_{uni}$ as initialization; $\mathbf{iti}_{co}$: iteratively trained with $\mathbf{ini}_{co}$ as initialization.

quantizers designed for each pdf $p(x_0|\mathbf{j}_0^t)$. Fig. 4.3 shows that with iterative improvements, both initializations converge to quite similar final results, $\mathbf{iti}_{uni}$ and $\mathbf{iti}_{co}$, although the performance of $\mathbf{ini}_{uni}$ is noticeably worse than the performance of $\mathbf{ini}_{co}$. The reason for this phenomenon deserves a thorough investigation. Recall that for noisy channel scenarios, a good index assignment is essential for the system performance. The training in Algorithm 7 has the advantage that it produces a good index assignment simultaneously as generating the encoder–decoder pairs.

For stable systems, the contributions of x_t and u_t to the total cost J_T decrease with time. Hence, the reconstruction points at $t = 0$, i.e., $\{q_{j_0=k}\}_{k=0}^{L-1}$, are the most important design parameters. Fig. 4.4 shows the resulting encoder regions and the reconstruction points at $t = 0$ versus ϵ. On the x-axis there is the channel transition probability ϵ, while on the y-axis we plot the reconstruction points, together with the

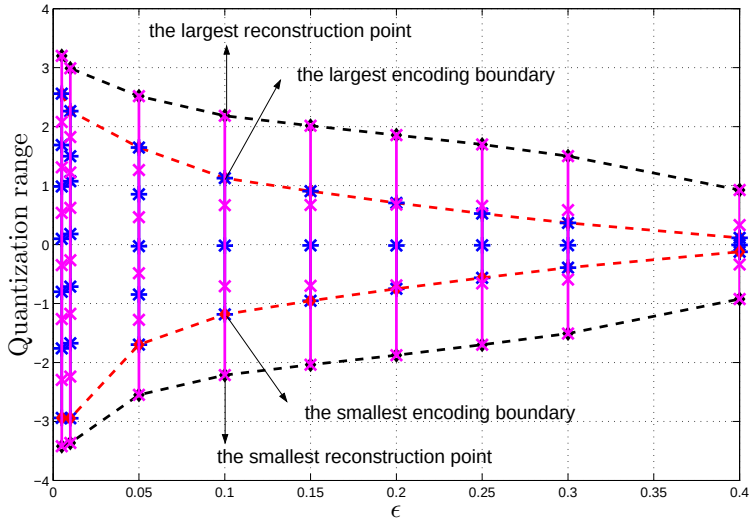


Figure 4.4: Reconstruction points q_{j_0} and encoding boundaries versus ϵ (at time $t = 0$). The mark $*$ denotes an encoding boundary. The mark $\times$ denotes a reconstruction point.

decision boundaries between each encoder regions. We note that the number of reconstruction points chosen by the encoder decreases with increasing ϵ, but the number of reconstruction points remains the same. This phenomenon is well known in quantization for noisy channels and is attributed to the varying abilities of binary codewords in combating channel errors. For very noisy channels, it is beneficial to transmit only the "stronger" codewords, providing true redundancy for error protection. Another impact of increasing ϵ is that the ranges of the reconstruction points decrease so that the boundaries and the reconstruction points are all moved closer to zero, indicating that only small-valued control actions are allowed.

The proposed encoder–decoder is designed for a certain ϵ. The system performance is evidently degraded if the true ϵ deviates from the design value. In Fig. 4.5, the robustness of the system against a mismatch in the transition probability ϵ is studied. Each dashed line is obtained by employing fixed encoder–decoder pairs designed for a certain ϵ, to channels with different ϵ's. The solid line is obtained by using

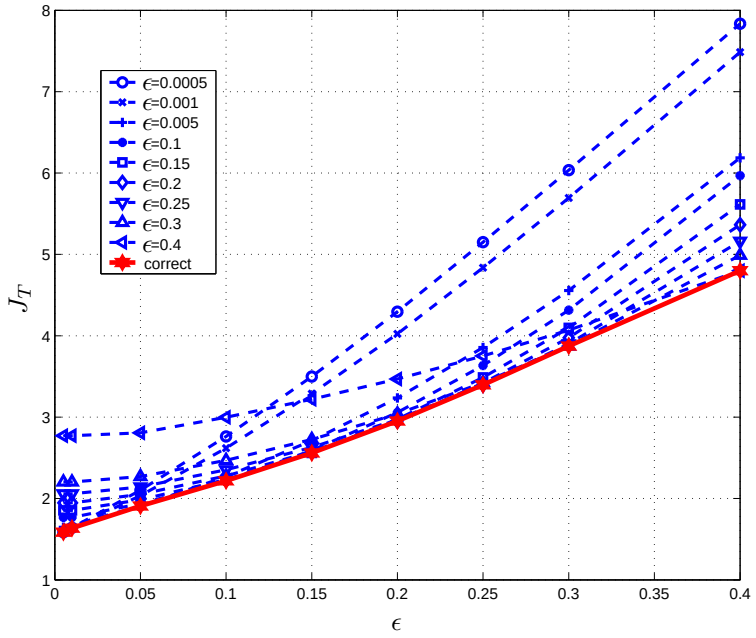


Figure 4.5: The robustness test against various ϵ.

the encoder–decoder designed for the true ϵ. Obviously, the performance is deteriorated if there is a mismatch between the designed and true ϵ. We also observe that underestimating the errors is more harmful than overestimating it.

Presence of Process Noise

In this section we present the simulation results of applying Algorithm 7 to a system as (4.7). The performance is measured with $\tilde{J}_3$, as defined in (4.6). The random magnitudes of the initial state and the process noise are both modelled using the generalized Gaussian distribution. We consider an experiment with following design parameters: $a = 0.8$, $\rho = 1$, $T = 3$, $R = 2$, $p(x_0) = GGD(2, 1)$ and $p(v_t) = GGD(2, 0.25)$. In Fig. 4.6, the cost $\tilde{J}_3$ is plotted against the channel transition probability ϵ. The horizontal dashed line represents the cost if no control action is taken.

We see that it is still beneficial to feeding back the state observation over a noisy channel even with 35% bit errors.

Three initialization methods are compared in this setup. The notations are the same as used in the process noise free scenario. The first method, referred to as $\mathbf{ini}_{uni}$, exploits properly scaled uniform quantizers. The second method, referred to as $\mathbf{ini}_{max}$, applies Lloyd-Max quantizers, designed for each pdf $p(\bar{x}_t|\mathbf{j}_0^t)$. The last method, referred to as $\mathbf{ini}_{co}$, applies channel optimized quantizers, designed for each pdf $p(\bar{x}_t|\mathbf{j}_0^t)$. Fig. 4.6 shows that with iterative improvements, all three initializations converge to quite similar final results, i.e., $\mathbf{iti}_{uni}$, $\mathbf{iti}_{max}$, and $\mathbf{iti}_{co}$, although the performances of $\mathbf{ini}_{uni}$ and $\mathbf{ini}_{max}$ are notably worse than the performance of $\mathbf{ini}_{co}$.

As discussed before, the reconstruction points at $t = 0$, i.e., $\{q_{j_0=k}\}_{k=0}^{L-1}$, are important design parameters. Fig. 4.7 shows the encoder regions and the reconstruction points at $t = 0$ versus ϵ. While in Fig. 4.8 we demonstrate how the encoder regions at $t = 0$ are changed after iterations. Similar results as for the process noise free scenario are also observed here. For example, the number of reconstruction points chosen by the encoder decreases with increasing ϵ, but not the number of reconstruction points. In Fig. 4.7, for $\epsilon = 0.3$ and $\epsilon = 0.35$, the two middle reconstruction points are close, so that additional zooming would be needed to distinguish them. Additionally, an increasing of ϵ renders also the moving of boundaries and reconstruction values closer to zero that only small-valued control actions are allowed.

If the encoding regions are intervals, like in the experiments, the encoder complexity can be significantly reduced by exploiting the knowledge about decision boundaries. This, since the tedious estimation of the future states is replaced by a simple comparison with a scalar value. Note also, the boundaries can be calculated once the reconstruction points are fixed, and then stored in a look-up table. We illustrate the calculation of the boundaries by a simple example.

Example 7 Boundaries for $\mathbf{ini}_{co}$ at $t = 0$
In this example, we calculate the boundary for $\mathbf{ini}_{co}$ at $t = 0$. Recall that, $\mathbf{ini}_{co}$ applies $COSQ$ to each conditional distribution of the state. In particular, the coding functions $\{f_0, g_0\}$ for the time instant $t = 0$ do not take the future evolution into account. For brevity, let x denote the state signal x_0, while i and j denote the transmitted and received indices i_0 and j_0, respectively. Also, q_i denotes the reconstruction point at $t = 0$ with $i \in \mathcal{I}_L$. Finally, $p(j|i)$ denotes $p(c(j_0)|c(i_0))$ where $c(i_0)$ and $c(j_0)$

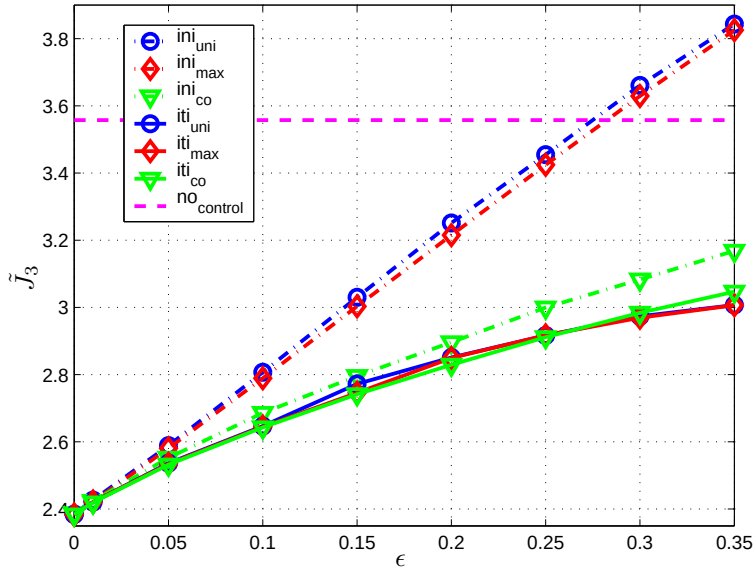


Figure 4.6: Performance comparisons among encoder–decoder pairs. $\mathbf{ini}_{uni}$: uniform initialization; $\mathbf{ini}_{max}$: Lloyd–Max initialization; $\mathbf{ini}_{co}$: Channel Optimized initialization; $\mathbf{iti}_{un}$: iteratively trained with $\mathbf{ini}_{uni}$ as initialization; $\mathbf{iti}_{max}$: iteratively trained with $\mathbf{ini}_{max}$ as initialization; $\mathbf{iti}_{co}$: iteratively trained with $\mathbf{ini}_{co}$ as initialization.

are the binary received and transmitted codewords.

Let us define:

$$\alpha_k(x) = \sum_{j=0}^{L-1} p(j|i=k)(x-q_j)^2. \tag{4.33}$$

It turns out that $\alpha_k(x)$ can be written as

$$\alpha_k(x) = (x^2 - m_k)^2 + s_k - (m_k)^2, \tag{4.34}$$

where m_k and s_k are defined as

$$m_k = \sum_{j=0}^{L-1} p(j|i=k)q_j, \quad s_k = \sum_{j=0}^{L-1} p(j|i=k)q_j^2. \tag{4.35}$$

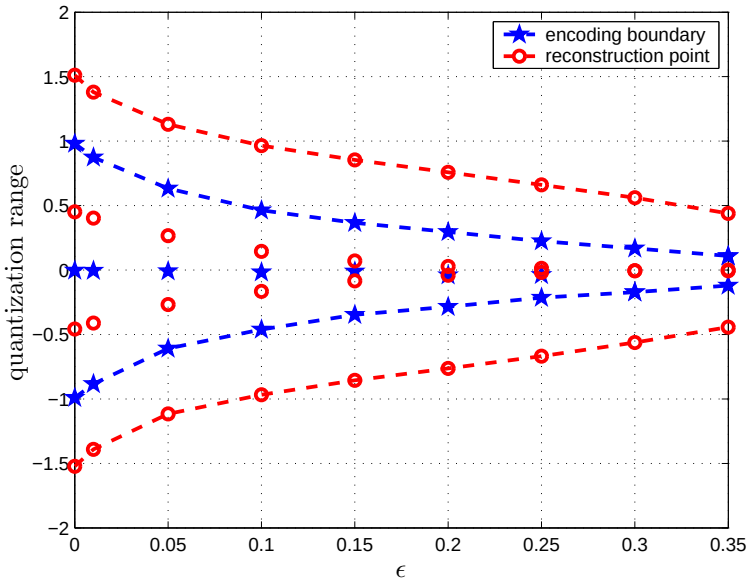


Figure 4.7: q_{j_0} and encoding boundaries versus ϵ (at $t = 0$). Stars: decision boundary; circles : reconstruction point.

This, since expanding the right term in (4.33) gives

$$\begin{aligned}\alpha_k(x) &= x^2 - 2x\sum_{j=0}^{L-1} p(j|i=k)q_j + \sum_{j=0}^{L-1} p(j|i=k)q_j^2 \\ &= (x^2 - \sum_{j=0}^{L-1} p(j|i=k)q_j)^2 + \sum_{j=0}^{L-1} p(j|i=k)q_j^2 \\ &\quad - (\sum_{j=0}^{L-1} p(j|i=k)q_j)^2. \end{aligned} \tag{4.36}$$

Let t_{kl} be the boundary between the encoder regions S_k and S_l, where $S_i = \{x|f_0(x) = i\}$, that is

$$t_{kl} = \left\{ x : \sum_{m=0}^{2^R-1} p(j=m|i=k)(x-q_m)^2 \right.$$

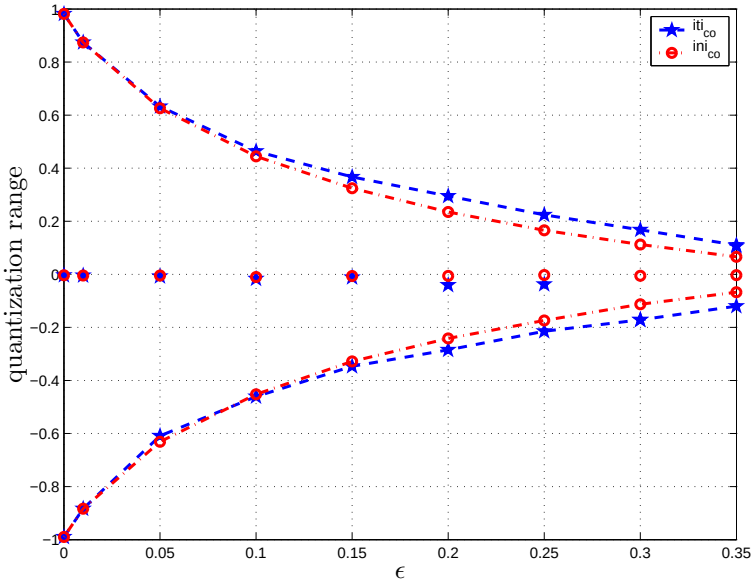


Figure 4.8: The change of the encoding boundaries after iterations, cf. Fig. 4.7.

$$= \sum_{m=0}^{2^R-1} p(j=m|i=l)(x-q_m)^2 \Bigg\}. \tag{4.37}$$

At t_{kl}, the following equation is valid:

$$\frac{(t_{kl}-m_k)^2 + s_k - m_k^2}{(t_{kl}-m_l)^2 + s_l - m_l^2} = 1. \tag{4.38}$$

Hence, t_{kl} can be calculated as

$$t_{kl} = \frac{s_l - s_k}{2(m_l - m_k)}. \tag{4.39}$$

For example, studying the case when $k=1$ and $l=2$, renders

$$t_{12} = \frac{s_2 - s_1}{2(m_2 - m_1)}$$

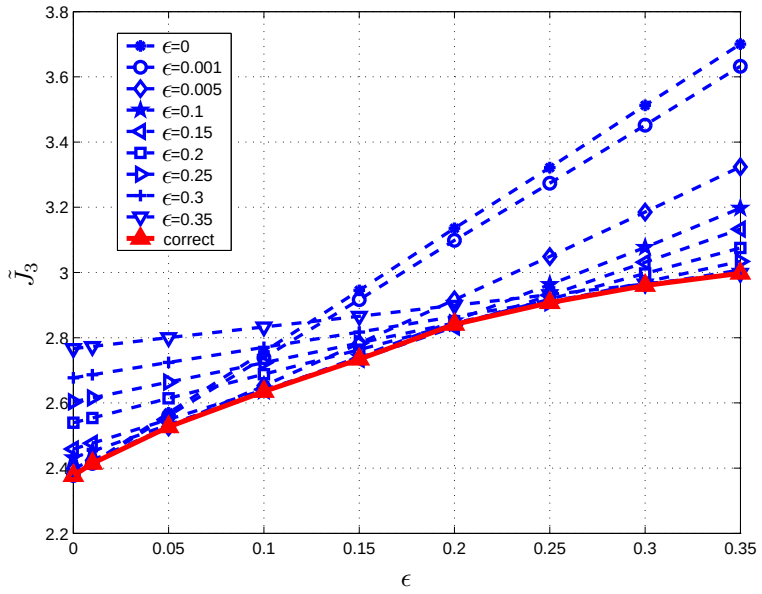


Figure 4.9: The robustness test against various ϵ.

$$= \frac{\sum_{j=0}^{L-1} p(j|i=2)q_j^2 - \sum_{j=0}^{L-1} p(j|i=1)q_j^2}{2(\sum_{j=0}^{L-1} p(j|i=2)q_j - \sum_{j=0}^{L-1} p(j|i=1)q_j)}.$$

The calculation of the boundaries for $\mathbf{ini}_{co}$ at $t = 0$ does not involve any memory as the one in (4.18). Although the memory-based encoding rule renders more tedious calculation, but the basic principle is similar.

Finally, shown in Fig. 4.9, experiments were pursued to evaluate the robustness of the proposed encoder–decoder to the variations in the channel error probability ϵ. Along the x-axis, we have the channel transition probability ϵ, while the cost $\tilde{J}_3$ is plotted on the y-axis. The dashed lines are obtained by using the encoder–decoder pairs designed for fixed ϵ-values, while the solid line is obtained by using the encoder–decoder pairs designed for the true ϵ values. As stated earlier, since each codebook is trained for a particular ϵ, the system performance is degraded if the true ϵ deviates from the design parameter. From the figure, one can verify that a robust coding scheme prefers an overestimate of the error probability to an underestimated one.

Chapter 5

Application to Event-Triggered Control

5.1 Background

Event-triggered control strategies have the potential in many cases to be more efficient than conventional time-triggered (or sampled) control [ÅB99]. To optimally utilize the communication resources, it is desirable to let each control loop communicate only when necessary. How this should be done in general is quite unexplored and only preliminary results are available, e.g., [ÅB99, SJ03]. In this chapter, a new control strategy combining the approaches from event-triggered with quantized control is proposed. It is shown that for an interesting class of systems, which are affected by rarely occurring disturbances drawn from a known probability distribution, it is possible to achieve a good control performance with limited control actuation and sensor communication. The focus here is on how to encode sensor data efficiently. Related problems on encoding control commands for motion control have been studied in [BMP03, Bro88, EB03]. In applications to communication systems, contributions can be found in e.g., [MCea05].

5.2 Wireless Sensor Network Example

Consider the example of wireless sensor networks presented in Section 1.2, shown again in Fig. 5.1. Suppose that the dynamics of each plant can be

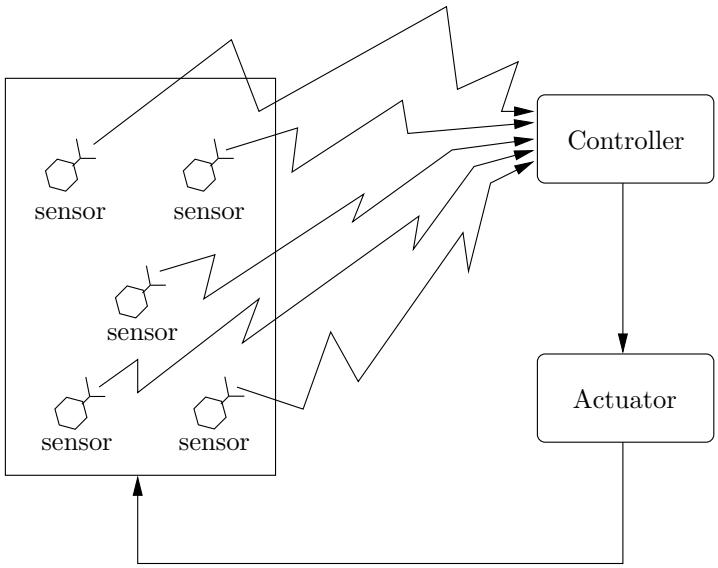


Figure 5.1: Control system utilizes data from wireless sensor network. The results of this thesis can be applied to optimize the use of the wireless medium. (Identical to Fig. 5.1).

represented as a scalar system

$$x_{t+1} = ax_t + u_t + \omega_t, \qquad t = 0, 1, \ldots, \tag{5.1}$$

where x_t is the state, u_t the control signal, and ω_t the event-triggered disturbance at time instant t. The disturbance is equal to zero except at time instants $t = t_i \geq 0$, $i = 1, 2, \ldots$. This disturbance is not measurable but estimates of its distribution in time and space are known.

The control objective is to keep the state of each plant close to the origin by using a small amount of control actuation. Since the communication resources between sensor and control nodes are shared, it is desirable to limit the amount of transmitted data. This is possible since the plants are stable and exposed to rarely occurring individual disturbances. A plausible decentralized control strategy is the following. Let u_t be equal to zero as long as x_t is close to the origin. Define a δ-ball around the origin. When the sensor detects that x_t is outside of the δ-ball, the sensor measurement is encoded and transmitted to the control node. The

message is decoded at the control node and a control command is derived and actuated. The controller and the encoder–decoder pairs are designed based on a finite-horizon linear quadratic criterion.

5.3 Problem Statement

The stable scalar linear time-invariant plant using an ideal channel is descried by

$$\left\{ \begin{array}{rcl} x_{t+1} & = & ax_t + u_t + \omega_t, \\ y_t & = & x_t, \\ j_t & = & i_t, \end{array} \right. \tag{5.2}$$

where the pdf's $p(x_0)$ and $p(\omega_t)$ are given, as well as the linear dynamics a, $|a| < 1$. The disturbance occurs (i.e., ω_t is non-zero) at random instants of time.

5.3.1 Performance Measure

The goal is to solve an integrated encoder–decoder design and optimal control problem for the system in (5.2), with a performance measure of the form

$$J_0 = \lim_{K\to\infty} \frac{1}{K} \sum_{t=0}^{K-1} \mathbf{E}\left\{x_t^2 + \rho u_t^2\right\}, \tag{5.3}$$

where $\rho \geq 0$ specifies the relative weight assigned to the cost (contribution to J_0) of the control signal u_t. As mentioned earlier, let $t_i,\ i = 0, 1, \ldots,$ $t_i > 0,\ \forall\ i > 0$ and $t_i \neq t_j,\ \forall\ i \neq j$, denote the time instant for the occurrence of the ith disturbance. We assume that the system is initialized at time $t = t_0 = -1$ by the disturbance $\omega_{-1} \neq 0$ ($x_{-1} = 0$ and $u_{-1} = 0$). Thus, at $t = 0$, we have the initial state $x_0 = \omega_{-1} \neq 0$. Then J_0 is

$$J_0 = \lim_{K\to\infty} \frac{1}{K} \sum_{i=0}^{K-1} \left[\frac{1}{t_{i+1} - t_i} \sum_{t=t_i+1}^{t_{i+1}} \mathbf{E}\left\{x_t^2 + \rho u_t^2\right\} \right], \tag{5.4}$$

conditioned on a fixed sequence $\{t_i\}$.

Suppose now that a substantial part of the contribution of a disturbance to the performance measure is concentrated to a (small) time-interval of length T time-instants following each disturbance, i.e., the

contribution in the interval $t_i + 1 + T \leq t < t_{i+1} + 1$ is small compared to the one in $t_i + 1 \leq t \leq t_i + T$. This is a reasonable assumption if the feedback control succeeds in attenuating the disturbance within time T, and the intensity of the disturbance is low enough such that, with high probability, there is only one disturbance per interval. Then, J_0 can be approximated by

$$\lim_{K\to\infty} \frac{1}{KT} \sum_{i=0}^{K-1} \left[\sum_{t=t_i+1}^{t_i+T} \mathbf{E}\{x_t^2 + \rho u_t^2\} \right]. \tag{5.5}$$

From now on, we assume that the plant is not controlled in the interval $t_i + 1 + T \leq t < t_{i+1} + 1$, i.e., $u_t = 0$ for these time instants. We also explicitly assume that $t_{i+1} - t_i > T, \ \forall i$.

In (5.5) the expectation is (implicitly) taken with respect to the pdf $p(\omega_t)$, while the random occurrences $\{t_i\}$ are averaged out by the time-average. Alternatively, by treating each disturbance separately and shift time to zero ("re-starting the clock") after the occurrence of each disturbance, we end up with the following criterion function

$$J_T = \sum_{t=0}^{T} \mathbf{E}\left\{x_t^2 + \rho u_t^2\right\}, \tag{5.6}$$

with the constraint $u_T = 0$.

In (5.6) the expectation is to be interpreted as averaging over both the values and time-instants of the disturbances, in the sense that the expectation $\mathbf{E}$ is taken with respect to a new pdf p_0 for x_0 defined as

$$p_0 = \lim_{K\to\infty} \frac{1}{K} \sum_{i=0}^{K-1} p_i, \tag{5.7}$$

where p_i is the pdf of x_{t_i-1}. As per our discussion, it should hold that $J_T \approx J_0$. In the remaining parts, any reference to optimality will refer to J_T and the implicit averaging over all the different disturbances in the sense of p_0.

5.3.2 Disturbance Detector

A detector is employed to mark instants when a disturbance event has taken place. Since no measurement error is present, estimates of the system evolution are readily calculated without error. Hence, a disturbance

Algorithm 9: The Overall System Design (SD)

1. Initiate p_0 with the disturbance pdf $p(\omega_t)$. Initialize the encoder–decoder $\{f_t, g_t\}_{t=0}^{T-1}$ using Lloyd–Max or uniform quantizers, as described in Section 3.3.2.
2. Set $\delta = 0$. Design the encoder, decoder and controller by using Algorithm 2 in Section 3.3.1.
3. Use the resulting encoder–decoder pairs and simulate the disturbance sequence according to the statistical properties of $\{\omega_t\}$. Find the value $\bar{\delta}$ for the threshold that minimizes J_T.
4. While J_T has not converged,
   4a. Apply the trained encoder–decoder and $\bar{\delta}$ to the disturbed control system. Collect a training set that describes p_0.
   4b. Train the encoder–decoder using Algorithm 2 in Section 3.3.2.
5. For the final encoder–decoder, find the threshold that minimizes J_T.

can be perfectly detected by comparing a predicted value for x_t with the observed value.

Even though there is a perfect disturbance detection mechanism, the quantization effect makes it non-trivial to decide when to act. For example, when the bit rate is low, it may be better to let the disturbance die out by the plant's own stabilizing dynamics, than to apply a control with relatively large estimation errors. We therefore introduce a threshold $\delta > 0$. The control action is triggered only when the magnitude of the observed state is larger than the threshold. The threshold is obviously an additional design parameter, which will be discussed more in Section 5.4.

5.4 System Design

We observe that the proposed control problem can partly be solved by exploiting Algorithm 2 in Chapter 3, if the initial pdf p_0 is known or can be accurately estimated. However, as discussed in Section 5.3, an important parameter, with respect to the overall performance, is the threshold δ. The choice of threshold is taken into account in the overall design algorithm summarized in Algorithm 9. In Step 4*b* the encoder–decoder pairs from the prior iteration is used to initialize the training.

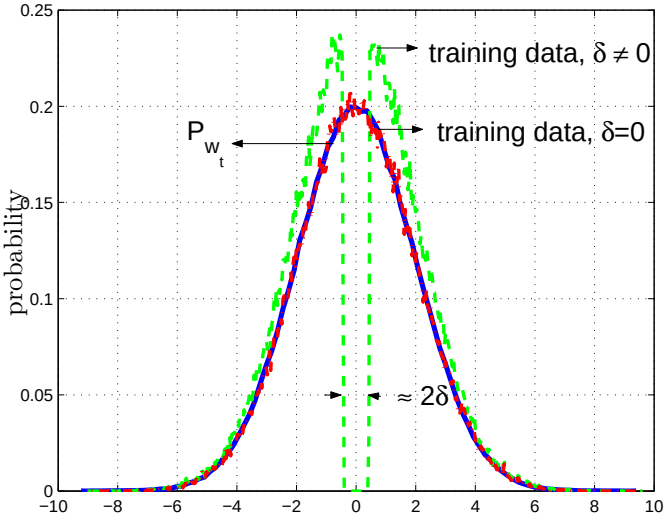
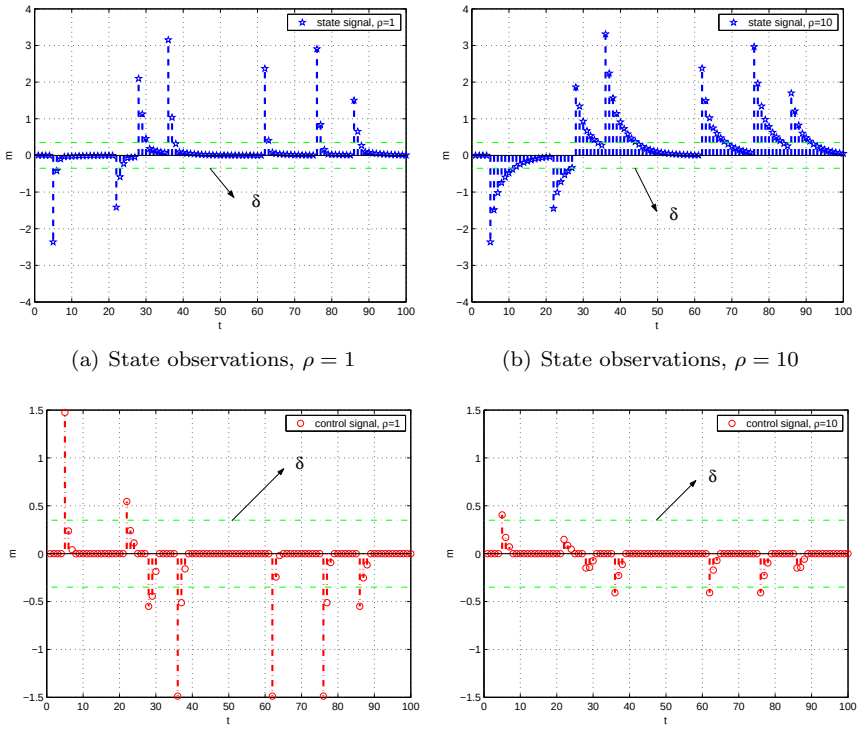


Figure 5.2: GGD distribution $p(\omega_t)$, empirical distributions for $\varepsilon = 0$ and $\varepsilon \neq 0$.

Note that selecting a threshold is performed once in Step 3, and for the second time in Step 5 to set a final value for δ. Obviously, the final δ is not the best one for the overall optimization problem, but a local optimum.

5.5 Numerical Experiments

The setup for the numerical experiments is given by $a = 0.8$, $\rho = 1$, $T = 3$ and $R = 2$. The disturbance events $\{t_i\}$ are modeled as a Poisson process, and the pdf $p(\omega_t)$ of the disturbance magnitude is modeled as a generalized Gaussian distribution. We use $\beta = 2$ throughout. Fig. 5.2 illustrates how the distribution of the training data changes when a threshold $\delta \neq 0$ is introduced. In Fig. 5.3, the impact of the parameter ρ on the system behavior is investigated. Recall that ρ is the penalty on the control signal in the performance measure. In Fig. 5.3(a) and Fig. 5.3(b), we show state evolutions of 100 samples for $\rho = 1$ and $\rho = 10$, respectively. While in Fig. 5.3(c) and Fig. 5.3(d), we shown the corresponding control signals



(a) State observations, $\rho = 1$

(b) State observations, $\rho = 10$

(c) Control signals, $\rho = 1$

(d) Control signals, $\rho = 10$

Figure 5.3: The state behavior and control signals for $\rho = 1$ and 10, respectively. The event threshold is $\delta = 0.35$.

at those periods. Note that when ρ is large, the optimal control signal has small magnitude, so consequently the response times to the pulse disturbances are longer.

To demonstrate the improvement in system performance when using optimized encoder–decoder pairs, a comparison with using uniform quantizers has been made. In the first case, referred to as $\mathbf{ini}_{uniF}$, a fixed optimum step-size uniform quantizer is employed for the entire period $0 \leq t \leq T-1$. The second approach, referred to as $\mathbf{ini}_{uniV}$, employs time-varying uniform quantizers with optimized step size (Section 3.4.2).

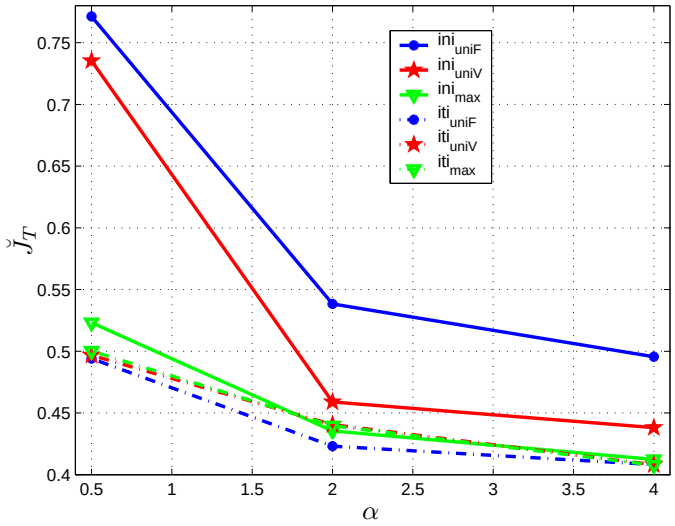


Figure 5.4: The improvement of the system performance by applying the optimized encoder–decoder. $\mathbf{ini}_{uniF}$: time-invariant uniform quantizer. $\mathbf{ini}_{uniV}$: time-varying uniform quantizer. $\mathbf{ini}_{max}$: Lloyd–Max quantizer. $\mathbf{iti}_{uniF}$: trained with $\mathbf{ini}_{uniF}$ as initialization. $\mathbf{iti}_{uniV}$: trained with $\mathbf{ini}_{uniV}$ as initialization. $\mathbf{iti}_{max}$: trained with $\mathbf{ini}_{max}$ as initialization.

Fig. 5.4 shows the performance improvement as

$$\breve{J}_T = \frac{\mathbf{E}\left\{\sum_0^T [x_t^2 + \rho u_t^2]\right\} - \mathbf{E}\left\{\sum_0^T \omega_t^2\right\}}{\mathbf{E}\left\{\sum_0^T \omega_t^2\right\}}, \tag{5.8}$$

versus α, which represents various GGD pdf's $p(\omega_t)$. Note that a small $\breve{J}_T$ indicates an effective control strategy. We observe that the proposed algorithm always results in an improvement over the initial encoder–decoder pairs. In particular, for disturbances with a peaked distribution, e.g., GGD$(0.5, 2)$, the performance after training is significantly improved. In this case, the optimal quantization should obviously not be uniform. Recalling that the parameter α describes the exponential rate of decay, the pdf is closer to a uniform distribution when α is large. Another interesting observation is that the training algorithm still provides similar results irrespective of the initialization of the training.

Chapter 6

Conclusions and Future Research

6.1 Concluding Remarks

Designing a control system with feedbacks over imperfect communication channels has received an increasing attention during the last decades. The research area is new and evolves rapidly. This thesis has studied approaches to joint source–channel coding and feedback control for scalar linear plants, subjected to data rate limited noisy channels.

A brief overview on recent publications of feedback quantization in control systems has been given. In this research field, a massive effort has been devoted to various stability issues. This thesis however focused on the performance optimization and solved a stochastic control problem. An iterative optimization algorithm for training the encoder–decoder pairs has been proposed and scenarios regarding various information patterns have been investigated. Sequential Monte Carlo simulations have shown promising improvements for the proposed coding–control schemes compared to traditional approaches.

6.2 Future Research

This thesis takes some steps towards future work on practical designs in joint control and source–channel coding. Suggesting a number of research directions, listed below are a few issues for the near-future research.

- **Multidimensional Systems**

  Although in this work, we have only considered linear scalar plants (of dimension one), the overall problem has been formulated under quite general assumptions. A generalization to multidimensional plants is straightforward, but the solutions are far from evident. For certain classes of systems, the CE controller may still be the optimal control strategy. In that case, it is interesting to know if traditional stochastic control theory is still useful for the controller design, and if vector quantization is the relevant counterpart to the scalar quantization studied in this thesis. However, compared to the scalar case, the information pattern for a multidimensional system has a larger variety. Witsenhausen has demonstrated with a famous counterexample [Wit71] that a small change in the information pattern can render completely different solutions. This counterexample has stressed that care must be taken when generalizing to multidimensional systems.

- **Distributed Systems**

  The main motivation for study of the quantized control and the event-triggered control is to efficiently utilize the limited communication resources. Besides the problem of optimizing the performance for each individual plant, another challenge is how to coordinate all distritbuted nodes as one system. In Chapter 5, we have studied the response of a single plant to triggered events. In a networked system, it may happen that several plants are triggered simultaneously resulting congestions in networks. How to design the trigger events that minimize the congestions in a network, and how to handle the congestions when they have occurred, are issues which deserve thorough investigations. Moreover, it has been shown that to combine the measurements from 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 network control problems needs also to be carefully investigated.

- **Measurement Noise**

  The assumption of the absence of observation noise is unrealistic. Once the measurement noise is considered in system models, the problem becomes complicated. In that case the encoder does not know the exact system state, but only has a copy of the corrupted

state measurement. In traditional stochastic control, the solution relies on the minimum variance estimate by exploiting a Kalman filter. Concerning quantized systems, how good is the approach when the minimum variance estimate out of a Kalman filter is quantized? The state estimation problems can certainly appeal to advances in nonlinear filtering theory.

- **Suboptimal Solutions**

  The separation property has been addressed as a desired feature, since it can reduce the overall optimization complexity. Even though the separation property is attractive, there is no doubt that for most systems, the separation principle does not apply. Also, it turns out that the memory-based CE controller is computationally demanding. Moreover, the proposed separation scheme has its weakness when generalized to unstable systems. This, since without any control action, the unstable poles may result in unbounded state trajectories and lead to unbounded estimation errors, although the plant probably will be stabilized. For unstable systems, the stability issues should be taken into account in the coding and the estimation. Thus, we believe that it is important to find feasible suboptimal solutions, which may not exploit the separation property. Seeking suboptimal solutions opens up a broad spectra of research themes, both theoretical and practical. From a practical aspect, encoder–decoders with finite memories should be considered. Another obvious obstacle to practical implementations is the prohibitive computational complexity. Finding complexity-reduced algorithms is also a topic deserving considerable research efforts.

Appendix A

Generalized Gaussian Distribution

The generalized Gaussian Distribution (GGD) is described by the following probability density function,

$$p(x) = \left(\frac{\alpha \eta(\alpha, \beta)}{2\Gamma(1/\alpha)} \right) \exp\{-(\eta(\alpha, \beta)|x|)^{\alpha}\}, \tag{A.1}$$

where Γ denotes the Gamma function and $\eta(\alpha, \beta)$ is defined as

$$\eta(\alpha, \beta) = \beta^{-1} \left(\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)} \right)^{1/2}. \tag{A.2}$$

Regarding the probability density function, two parameters should be specified, namely, α and β. In particular, α describes the exponential rate of decay, while β states the standard deviation.

In this thesis, the *inverse method* for continuous distributions is exploited to implement the random variable generator for the generalized Gaussian distribution. The inverse method is based on the fundamental theorem which states the relation between a uniform distribution compared to other continuous distributions. Suppose that X is a continuous random variable with a strictly increasing cumulative distribution function (cdf) F, whose inverse function is denoted as F^{-1}. Let U be a random variable uniformly distributed between 0 and 1, then F serves

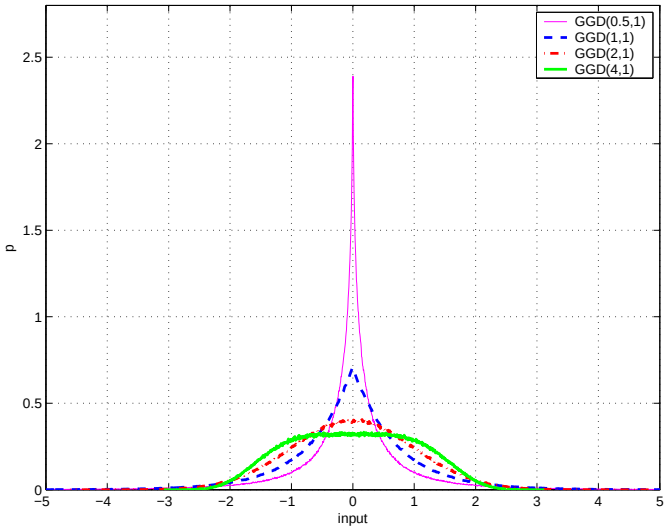


Figure A.1: Different generalized Gaussian distributions.

as the cdf for $F^{-1}(U)$ as well as for X. This means that we can generate a random variable by applying the inverse function of the intended cumulative distribution function to a uniform random variable.

The cdf of the generalized Gaussian distribution is given by

$$F(x) = \frac{1}{2} + \frac{\Gamma((\eta(\alpha,\beta)|x|)^{\alpha}, \frac{1}{\alpha})}{2\Gamma(\frac{1}{\alpha})}, \tag{A.3}$$

where the Gamma function and the incomplete Gamma function are defined as

$$\Gamma(a) = \int_0^{\infty} e^{-t}t^{a-1}dt, \quad \Gamma(a,x) = \int_0^{x} e^{-t}t^{a-1}dt. \tag{A.4}$$

Note that the inverse function F^{-1} is straightforward to calculate if the inverse of the incomplete Gamma function is available. In the remaining part of this chapter, based on [DM86], we provide a short summary of a numerical solution to the inverse problem.

Algorithm 10: Iteration of x_k, $k = 1, \ldots$

1. $x_{n+1} = x_n(1 - h_n)$
2. $h_n = t_n + w_n t_n^2$
3. $w_n = (a - 1 - x_n)/2$
4. $t_n = \begin{cases} \frac{p(a,x_n)-p(a,x)}{r(a,x_n)}, & p(a,x) \leq \frac{1}{2} \\ \frac{q(a,x)-q(a,x_n)}{r(a,x_n)}, & p(a,x) > \frac{1}{2} \end{cases}$
5. $r(a, x_n) = \frac{e^{-x} x^a}{\Gamma(a)}$.

Let us first define $p(a, x)$ and $q(a, x)$ as

$$p(a,x) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt, \tag{A.5}$$

$$q(a,x) = 1 - p(a,x) = \frac{1}{\Gamma(a)} \int_x^\infty e^{-t} t^{a-1} dt. \tag{A.6}$$

In particular, the algorithm is about the computation of x for given a and $p(a, x)$, when $a \geq 0$, $x \geq 0$ and $a + x \neq 0$. Listed in Algorithm 10 is the iterative procedure. Three iterations at most are required to obtain 10 significant digit accuracy for x. However, the initial state x_0 is not straightforward to acquire. We describe the calculation in details. Define $B = q(a, x)\Gamma(a)$ and let c denote Euler's constant. Consider two situations:

Case 1) For $a \leq 1$, x_0 is approximated by:

- Case 1.1: when ($B > 0.6$) or ($B \geq 0.45$ and $a \geq 0.3$):

$$x_0 = \frac{u}{1 - u/(a+1)}$$

$$u = \begin{cases} (p(a,x)\Gamma(a+1))^{1/a}, & Bq(a,x) > 10^{-6} \\ e^{-q(a,x)/(a-c)}, & \text{otherwise.} \end{cases}$$

- Case 1.2: when ($0.35 \leq B \leq 0.6$ and $a < 0.3$):

$$x_0 = te^u$$

$$t = e^{-c-B}$$

$$u = te^t.$$

- Case 1.3: when ($0.15 \leq B \leq 0.45$ and $a \geq 0.3$) or ($0.15 \leq B < 0.35$):

$$\begin{aligned} x_0 &= y - (1-a)\ln v - \ln\left(1 + \frac{1-a}{1+v}\right) \\ y &= -\ln B \\ v &= y - (1-a)\ln y. \end{aligned}$$

- Case 1.4: when ($0.01 < B < 0.15$):

$$\begin{aligned} x_0 &= y - (1-a)\ln v - \ln\frac{v^2 + 2(3-a)v + (2-a)(3-a)}{v^2 + (5-a)v + 2} \\ y &= -\ln B \\ v &= y - (1-a)\ln y. \end{aligned}$$

- Case 1.5: when ($B \leq 0.01$):

$$\begin{aligned} x_0 &= y + c_1 + c_2\, y + c_3\, y^2 + c_4\, y^3 + c_5\, y^4 \\ y &= \ln B \\ c_1 &= (a-1)\ln y \\ c_2 &= (a-1)(1+c_1) \\ c_3 &= (a-1)\left(-\frac{1}{2}c_1^2 + (a-2)c_1 + \frac{3a-5}{2}\right) \\ c_4 &= (a-1)\left(\frac{1}{3}c_1^3 - \frac{3a-5}{2}c_1^2 + (a^2-6a+7)c_1 \right. \\ &\quad \left. + \frac{11a^2 - 46a + 47}{6}\right) \\ c_5 &= (a-1)\left(-\frac{1}{4}c_1^4 + \frac{11a-17}{c}c_1^3 + (-3a^2+13a-13)c_1^2 \right. \\ &\quad \left. + \frac{2a^3 - 25a^2 + 72a - 61}{2}c_1 + \frac{25a^3 - 195a^2 + 477a - 379}{12}\right). \end{aligned}$$

Case2) For $a > 1$, denote the Cornish-Fisher 6-term approximation

for x_0 as

$$\begin{aligned} w &= a + s\sqrt{a} + \frac{s^2-1}{3} + \frac{s^3-7s}{36\sqrt{a}} - \frac{3s^4+7s^2-16}{810a} \\ &\quad + \frac{9s^5+256s^3-433s}{38880a\sqrt{a}} \qquad \text{(A.7)} \\ s &= (-1)^m \left(t - \frac{a_0 + a_1 t + a_2 t^2 + a_3 t^3}{1 + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4} \right) \qquad \text{(A.8)} \\ t &= \sqrt{-2\ln\tau} \\ m &= \begin{cases} 1, & 0 < p(a,x) < 1/2 \\ 0, & 1/2 \le p(a,x) < 1 \end{cases} \\ \tau &= \begin{cases} p(a,x), & 0 < p(a,x) < 1/2 \\ q(a,x), & 1/2 \le p(a,x) < 1, \end{cases} \end{aligned}$$

where

$$\begin{aligned} a_0 &= 3.31125922108741, \quad b_1 = 6.61053765625462 \\ a_1 &= 11.6616720288968, \quad b_2 = 6.40691597760039 \\ a_2 &= 4.28342155967104, \quad b_3 = 1.27364489782223 \\ a_3 &= 0.21362349371585, \quad b_4 = 0.036117081018842. \end{aligned}$$

Then following cases are considered

- Case 2.1: when ($a \ge 500$ and $|1 - w/a| < 10^{-6}$):
  $x = w$.
- Case 2.2: when ($a < 500$) or ($|1 - w/a| \ge 10^{-6}$):
  - Case 2.2.1: $p(a,x) > 1/2$:
    - Case 2.2.1.1: $w < 3a$, $x_0 = w$.
    - Case 2.2.1.2: $w \ge 3a$, $D = \max(2, a(a-1))$:
      - Case 2.2.1.2.1: $B > 10^{-D}$

$$\begin{aligned} u &= T(w) = -\ln B + (a-1)\ln w \\ &\quad - \ln\left(1 + \frac{1-a}{1+w}\right) \\ x_0 &= T(u). \end{aligned}$$

- Case 2.2.1.2.2: $B \leq 10^{-D}$:

$$x_0 = y + c_1 + c_2\ y + c_3\ y^2 + c_4\ y^3 + c_5\ y^4.$$

The coefficients are same as in Case 1.5.

- Case 2.2.2: $p(a,x) \leq 1/2$:
define the functions

$$\begin{aligned} F_n(x) &= e^{(v+x-\ln S_n(x))/a} \\ v &= \ln\left(p(a,x)\Gamma(a+1)\right) \\ s_0 &= 1 \\ s_n &= 1 + \sum_{i=1}^{n} x^i/((a+1)(a+2)\cdots(a+i)), \end{aligned}$$

calculate z according to

- Case: $w > 0.15(a+1) \Rightarrow z = w$.
- Case: $w \leq 0.15(a+1)$

$$\begin{aligned} u_1 &= F_0(w) \\ u_2 &= F_2(u_1) \\ u_3 &= F_2(u_2) \\ z &= F_3(u_3). \end{aligned}$$

Then

- Case 2.2.2.1: $(z \leq 0.002(a+1)) \Rightarrow x = z$.
- Case 2.2.2.2: $(0.002(a+1) < z \leq 0.01(a+1))$ or $(z > 0.7(a+1)) \Rightarrow x_0 = z$.
- Case 2.2.2.3: $(0.01(a+1) < z \leq 0.7(a+1))$

$$\begin{aligned} \bar{z} &= F_N(z) \\ x_0 &= \bar{z}\left(1 - \frac{a\ln\bar{z} - \bar{z} - v + \ln s_n(z)}{a - \bar{z}}\right), \end{aligned}$$

where N is the smallest integer such that

$$z^N/((a+1)\cdots(a+N)) \leq 10^{-4}.$$

Bibliography

[ÅB99] K. J Åström and B. Bernhardsson. Comparison of periodic and event based sampling for first-order stochastic system. In *Proc. of the* 14*th IFAC World Congress*, 1999.

[Åst70] K. J Åström. *Introduction to stochastic control*. Academic Press, Inc., 1970.

[Bai02] J. Baillieul. Feedback coding for information-based control: operating near the data-rate limit. In *Proc. of the* 41*st IEEE Conference on Decision and Control*, 2002.

[Bai04] J. Baillieul. Data-rate requirements for nonlinear feedback control. In *IFAC Symp. Nonlinear Control Systems*, 2004.

[BBea00] A. Berlin, D. Biegelsen, and P. Cheung et al. Motion control of planar objects using large-area arrays of MEMS-like distributed manipulators. In *Micromechatronics* 2000, June 2000.

[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*, vol.45(7), July 2000.

[BL06] F. Bullo and D. Liberzon. Quantized control via locational optimization. *IEEE Transactions on Automatic Control*, vol.51(1):2–13, January 2006.

[BM99] V. S. Borkar and S. K. Mitter. LQG control with communication contraints. In *Proc. of the* 38*th Conference on Decision and Control*, December 1999.

[BMP03] A. Bicchi, A. Marigo, and B. Piccoli. Encoding steering control with symbols. In *Proc. of the* 42*nd IEEE Conference on Decision and Control*, December 2003.

[Bro88] R. W. Brockett. On the computer control of movements. In *Proc.* 1988 *IEEE International Conference on Robotics and Automation*, April 1988.

[BS96] D. Bertsekas and S. E. Shreve. *Stochastic optimal control: the discrete-time case*. Athena Scientific, Belmont, Massachusetts, 1996.

[BSJ06a] L. Bao, M. Skoglund, and K. H. Johansson. Encoder-decoder design for event-trigged feedback control. In *Proc. of* 2006 *American Control Conference*, 2006. To appear.

[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.*, 2006. To appear.

[BSJ06c] L. Bao, M. Skoglund, and K. H. Johansson. Feedback control over the binary symmetric channel. In *Proc. of the* 45*th IEEE Conference on Decision and Control*, 2006. Submitted.

[BST74] Y. Bar-Shalom and E. Tse. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, vol.AC-19(5), Octorber 1974.

[CT91] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley and Sons, Inc, 1991.

[Cur70] R. E. Curry. *Estimation and control with quantized measurements*. The M.I.T Press, 1970.

[Del88] D. F. Delchamps. The stabilization of linear systems with quantized feedback. In *Proc. of the* 27*th IEEE Conference on Decision and Control*, 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*, 1989.

[Del90] D. F. Delchamps. Stabilizing a linear system with quantized state feedback. *IEEE Transactions on Automatic Control*, vol.35(8):916–924, August 1990.

[Del06] J. C. Delvenne. An optimal quantized feedback strategy for scalar linear systems. *IEEE Transactions on Automatic Control*, vol.51(2), Febrary 2006.

[DM86] A. R. Didonato and A. H. Morris. Computation of the incomplete Gamma function ratios and their inverse. *ACM Transactions on Mathematical Software*, vol.12(4):377–393, December 1986.

[EB03] M. Egerstedt and R. W. Brockett. Specification complexity of motor programs. *IEEE Transactions on Automatic Control*, vol.48(2), Febrary 2003.

[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*, vol.33(1), January 1987.

[EM99] N. Elia and S. K. Mitter. Quantization of linear systems. In *Proc. of the* 38*th Conference on Decision and Control*, December 1999.

[EM01] N. Elia and S. K. Mitter. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, vol.46(9), September 2001.

[Far90] N. Farvardin. A study of vector quantization for noisy channels. *IEEE Transactions on Information Theory*, vol.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*, vol.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*, vol.33(6):827–838, November 1987.

[FV90] N. Farvardin and V. Vaishampayan. On the performance and complexity of channel-optimized vector quantizers. *IEEE Transactions on Information Theory*, vol.37(1):155–160, January 1990.

[FZ03] F. Fagnani and S. Zampieri. Stability analysis and synthesis for scalar linear systems with a quantized feedback. *IEEE Transactions on Automatic Control*, vol.48(9), September 2003.

[FZ05] F. Fagnani and S. Zampieri. A symbolic approach to performance analysis of quantized feedback systems: the scalar case. *SIAM J. Control Optim.*, vol.44(3):818–866, 2005.

[Gal68] R. G. Gallager. *Information theory and reliable communication*. John Wiley and Sons, Inc., 1968.

[GG92] A. Gersho and R. M. Gray. *Vector quantization and signal compression*. Kluwer, 1992.

[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*, vol.42(4):1139–1151, July 1996.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simualted annealing. *Science*, vol.220(4598), May 1983.

[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 nonlinear affine systems. *Internat. J. Control*, vol.73(3):239–249, 2004.

[LH05] D. Liberzon and J. P. Hespanha. Stabilization of nonlinear systems with limited information feedback. *IEEE Transactions on Automatic Control*, vol.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*, vol.39:1543–1554, 2003.

[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*, December 2005.

[LL05b] Q. Ling and M. D. Lemmon. Stability of quantized control systems under dynamic bit assignment. *IEEE Transaction on Automatic Control*, vol.50(5):734–740, May 2005.

[Mas90] J. Massey. Causality, feedback, and directed information. In *Proc. of* 1990 *International Symposium on Information theory and its application*, pages 303–305, 1990.

[MCea05] N. Möller, I. Cabrera, and K. H. Johansson et al. Using radio network feedback to improve TCP performance over cellular networks. In *Proc. of the* 44*th IEEE Conference on Decision and Control, and the European Control Conference 2005*, 2005.

[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. Optmial LQG control via limited capacity communication networks. In *Proc. of the* 41*st IEEE Conference on Decision and Control*, December 2002.

[MS04] A. S. Matveev and A. V. Savkin. The problem of LQG optimal control via a limited capacity couuminication channel. *Systems and Control letters*, vol.53:51–64, 2004.

[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.*, Febrary 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*, vol.49(9):1585–1597, 2004.

[NG82] D. L. Neuhoff and R. K. Gilbert. Causal source codes. *IEEE Transactions on Information Theory*, vol.28(5), September 1982.

[Per04] C. De Persis. Results on stablization on nonlinear systems under finite data-rate constraints. In *Proc. of the* 43*rd IEEE Conference on Decision and Control*, 2004.

[PGB02] B. Picasso, F. Gouaisbaut, and A. Bicchi. Construction of invariant and attractive sets for quantized-input linear system. In *Proc. of the* 41*st IEEE Conference on Decision and Control*, December 2002.

[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* 42*nd IEEE Conference on Decision and Control*, 2004.

[RS76] N. Rydbeck and C. E. Sundberg. Analysis of digital errors in nonlinear PCM systems. *IEEE Transactions on Communications*, vol.24(1):59–65, 1976.

[RU02] T. Richardson and R. Urbanke. *Modern Coding Theory*. http://lthcwww.epfl.ch/mct/index.php, 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*, December 2004.

[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.

[SJ03] A. Speranzon and K. H. Johansson. On some communication schemes for distributed pursuit-evasion games. In *Proc. of the* 42*nd IEEE Conference on Decision and Control*, December 2003.

[Sko97] M. Skoglund. *On soft decoding and robust vector quantization*. PhD thesis, Chalmers University of Technology, 1997.

[Söd03] T. Söderström. *Discrete-time stochastic systems*. Springer, 2003.

[SV03] T. Simsek and P. Varaiya. Noisy data-rate limited estimation: Renewal codes. In *Proc. of the* 42*nd IEEE Conference on Decision and Control*, December 2003.

[Tat00] S. Tatikonda. *Control under communication constraints*. PhD thesis, Massachusetts Institute of Technology, 2000.

[TBS75] E. Tse and Y. Bar-Shalom. Generalized certainty equivalence and dual effect in stochastic control. *IEEE Transactions on Automatic Control*, vol.20:817–819, 1975.

[The57] H. Theil. A note on certainty equivalence in dynamic planning. *Econometrica*, vol.25:346–349, 1957.

[TM04] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, vol.49(7):1056–1068, July 2004.

[TSM04] S. Tatikonda, A. Sahai, and S. Mitter. Stochastic linear control over a communication channel. *IEEE Transactions on Automatic Control*, vol.49(9), September 2004.

[VV95] S. Vembu and S. Verd*ú*. 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*, vol.42(9), September 1997.

[Wic95] S. B. Wicker. *Error control systems for digital communication and storage*. Prentice-Hall, Inc., 1995.

[Wit71] H. Witsenhausen. Separation of estimation and control for discrete time systems. In *Proc. of THE IEEE*, volume vol.59, pages 1557–1566, November 1971.

[Wit79] H. Witsenhausen. On the structure of real-time source coders. *The Bell System Technical Journal*, vol.58(6), 1979.

[WV83] J. C. Walrand and P. Varaiya. Optimal causal coding–decoding problems. *IEEE Transactions on Information Theory*, vol.29(6), November 1983.

[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), Octorber 1981.

[ZG90] K. Zeger and A. Gersho. Pseudo-gray coding. *IEEE Transactions on Communications*, vol.38(12):2147–2158, 1990.