COMMUNICATION IN DECENTRALIZED CONTROL

by

Demosthenis Teokerzis
B.S. University of Patras
Patras, Greece
1974
S.M. Massachusetts Institute of Technology
1976
E.E. Massachusetts Institute of Technology
1977

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 1979
(Re. February 1980)

Signature of Author

Department of Electrical Engineering and
Computer Science, September 14, 1979

Certified by

Thesis Supervisor

Accepted by

Chairman, Departmental Committee on
Graduate Students

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY
MAR 25 1980
LIBRARIES
COMMUNICATION IN DECENTRALIZED CONTROL

by

Demosthenis Teneketzis

Submitted to the Department of Electrical Engineering and Computer Science

on September 14, 1979 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ABSTRACT

In the first part of the thesis the Real Time Feedforward Communication Problem (P) and the Real Time Noiseless Feedback Communication Problem (P') are considered. Two Real Time Distortion Rate Problems are presented whose solution provides lower bounds to (P) and (P'). These bounds are shown to lie above the lower bound provided by the classical Rate Distortion Problem.

In the second part, the Centralized Filtering Problem (N) is formulated using an Information Theoretic Approach, and a Real Time Distortion Rate Problem is presented whose solution provides a lower bound to (N). The similarities and Differences between the Real Time Communication Problem and the Centralized Filtering Problem are discussed.

Finally, two decentralized Filtering Problems are formulated, and lower bounds to these problems are provided via the solution of a Real Time Distortion Rate Problem and a Linear Filtering Problem.

Thesis supervisor: Nils R. Sandell, Jr.
Title: Associate Professor of Electrical Engineering and Computer Science
Dedicated to

My mother and father
Athina and Thomas Teneketzis,
who made it possible for me
to study at M.I.T.
ACKNOWLEDGEMENTS

I would like to express my deep appreciation to all those who contributed to my education at M.I.T.: Professors Sanjoy Mitter, Yu Chi Ho, James Massey, Robert Kennedy, Nils Sandell, Jeff Shapiro, Alan Willsky, Dr. Horace Yuen; and my fellow graduate students Steve Barta, Howard Chizeck, Ted Djaferis, David Flamm, Pooi Kam, Bob Klein, Wolf Kohn, Richard Ku, Greg Lauer, Ricky Lee, John Neese, Daniel Orhlac, Mahmood Shahjahan, Avram Tetewsky, Bob Washburn and Khaled Yared.

I would especially like to thank Professor Sanjoy Mitter for his interest in my studies and for his unfailingly encouraging attitude towards helping me explore possible employment opportunities.

I would like to thank Karen Glennon, Frantiska Frolik, and Deborah Buenz for their excellent typing of the thesis, as well as Arthur Giordani and Norman Darling for their elegant work on the figures.

Special thanks to my dear friends Barbara Loizou, Yannis Avgoustis and Christos Platsoucas for their support.

This research was carried at the M.I.T. Laboratory for Information and Decision Systems with support extended by the U.S. Department of Energy under contract EX-76-C-01-2087 and by the Office of Naval Research under contracts ONR-N00014-76-C-0346 and ONR-N00014-77-C-0532.
# CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>12</td>
</tr>
<tr>
<td>2.4</td>
<td>13</td>
</tr>
<tr>
<td>2.5</td>
<td>15</td>
</tr>
<tr>
<td>2.6</td>
<td>16</td>
</tr>
<tr>
<td>2.7</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>18</td>
</tr>
<tr>
<td>3.1.1</td>
<td>18</td>
</tr>
<tr>
<td>3.1.2</td>
<td>20</td>
</tr>
<tr>
<td>3.2</td>
<td>21</td>
</tr>
<tr>
<td>3.2.1</td>
<td>21</td>
</tr>
<tr>
<td>3.2.2</td>
<td>22</td>
</tr>
<tr>
<td>3.3</td>
<td>24</td>
</tr>
<tr>
<td>3.4</td>
<td>25</td>
</tr>
</tbody>
</table>
3.3.1 Statement of the Problem ................. 24
3.3.2 Properties and Characteristics of the Proposed Formulation ........... 25

3.4 Investigation of the Tightness of the Lower Bound .... 46
3.4.1 Two Special Forms of Real Time Communication Problems ............. 47
3.4.2 A Real Time Distortion Rate Problem for (R) and (R') .................... 49

3.5 Another Formulation of the Real Time Feedforward Communication Problem .......... 59
3.5.1 Statement of the Problem .................... 59
3.5.2 Formulation of a Discrete Finite State Real Time Distortion Rate Problem for (Q) .... 61

3.6 Summary ........................................ 64

4 The Real Time Distortion Rate Problem - Continuous Amplitude Sources .......... 65
4.1 The Real Time Communication Problem ............... 65
4.2 Formulation of a Real Time Distortion Rate Problem for Continuous Amplitude Sources .......... 65
4.2.1 Statement of the Problem .................... 65
4.2.2 Properties of (4.2.1)-(4.2.3) .................. 66
4.2.3 Difference Distortion Measures - Shannon-type Lower Bounds ............ 71
4.2.4 Example 4.2.1 .................. 74
4.2.5 Investigation of the Tightness of the Lower Bound of (4.2.1)-(4.2.3) ............ 78

4.3 Summary ........................................ 87

5 The Filtering Problem .................. 88
5.1 The Nonlinear Filtering Problem ............ 88
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Formulation of a Real Time Distortion Rate Problem for (N)</td>
<td>90</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Statement of the Problem</td>
<td>90</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Properties of the Formulation</td>
<td>91</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Example</td>
<td>93</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Comparison with the Results of Zakai-Ziv and Galdos</td>
<td>95</td>
</tr>
<tr>
<td>5.2.5</td>
<td>Relation to the Real Time Communication Problem</td>
<td>100</td>
</tr>
<tr>
<td>5.3</td>
<td>Formulation of a Decentralized Filtering Problem</td>
<td>102</td>
</tr>
<tr>
<td>5.4</td>
<td>Formulation of a Real Time Distortion Rate Problem for (DN)</td>
<td>105</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Statement of the Problem</td>
<td>105</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Properties of (5.4.1)-(5.4.4)</td>
<td>106</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Example 5.4.1</td>
<td>109</td>
</tr>
<tr>
<td>5.5</td>
<td>Another Formulation of the Decentralized Filtering Problem</td>
<td>112</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Statement of the Problem</td>
<td>112</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Formulation of a Real Time Distortion Rate Problem for (DNG)</td>
<td>112</td>
</tr>
<tr>
<td>5.6</td>
<td>Formulation of a Decentralized Filtering Problem (L) Involving Delays</td>
<td>115</td>
</tr>
<tr>
<td>5.6.1</td>
<td>Statement of the Problem</td>
<td>115</td>
</tr>
<tr>
<td>5.6.2</td>
<td>A Lower Bound to (L)</td>
<td>118</td>
</tr>
<tr>
<td>5.7</td>
<td>Summary</td>
<td>123</td>
</tr>
<tr>
<td>6</td>
<td>Summary, Conclusions and Suggestions for Further Research</td>
<td>124</td>
</tr>
<tr>
<td>6.1</td>
<td>Summary</td>
<td>124</td>
</tr>
<tr>
<td>6.2</td>
<td>Conclusions, Suggestions for Further Research</td>
<td>125</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>REFERENCES</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Solution of (3.3.53)-(3.3.59)</td>
<td>128</td>
</tr>
<tr>
<td>B</td>
<td>Solution of (3.4.19)-(3.4.28)</td>
<td>135</td>
</tr>
<tr>
<td>C</td>
<td>Solution of (3.4.19)-(3.4.28)</td>
<td>143</td>
</tr>
<tr>
<td>D</td>
<td>Solution of (4.2.47)-(4.2.53)</td>
<td>145</td>
</tr>
<tr>
<td>E</td>
<td>Solution of (4.2.92)-(4.2.98)</td>
<td>150</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>158</td>
</tr>
<tr>
<td>G</td>
<td>Solution of (5.2.22)-(5.2.26)</td>
<td>160</td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
<td>166</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

1.1 Preliminaries

A key characteristic of the design techniques used in conventional modern control theory is that optimal laws are used in which every sensor output affects every actuator input [1], [2]. This situation is called centralized control. The above characteristic is not possible for many systems, particularly large scale systems. Decentralized control theory has risen in response to the failure of modern control theory to deal with certain issues of concern in large scale systems. The basic characteristic of decentralized control is that there are restrictions on information transfer between certain groups of sensors or actuators.

Consider, for example, the system of Figure 1.1 (page 2). Assume that the observation \( y_1(y_2) \) and the information transferred from location 2(1) to location 1(2) through the communication link 2→1 (1→2) are used to form the estimate \( \hat{x}_1(\hat{x}_2) \). Assume also that the rate of information transfer is below that necessary to implement the centralized solution.

The problem that naturally arises in this situation is the following:

\[
(*) \begin{cases}
\text{What are the best estimates } \hat{x}_1(t) \text{ and } \hat{x}_2(t) \text{ that we can achieve with} \\
\text{a given set of observations and a given set of channel capacities?}
\end{cases}
\]

The problem (*) posed above, has characteristics similar to, but not identical to those of a distortion rate problem.\(^1\)

The basic differences between the distortion rate problem and problem (*), that make impossible the direct application of rate distortion theory to decentralized control problems are the following:

\(^1\) The distortion rate problem is the following: Given the channel of a communication system find the encoder and the decoder that achieve the minimum distortion between the input and the output of the system.
Figure 1.1
In control problems the sources are usually nonstationary and have memory. For this class of sources the rate distortion function has been defined asymptotically \([3], [5]\) and \([6]\).

Moreover, the solution of the distortion rate problem which is a mathematical programming problem, is attained asymptotically, i.e. when infinite delays are allowed and code words of block length which approaches infinity can be used. The fact that the solution of the distortion rate problem is asymptotically attained is guaranteed by the direct and converse coding theorems \([3], [4], [7] - [10], [21] - [22]\). The asymptotic nature of both the rate distortion function for sources with memory and the coding theorem, are in contrast with the real time character of control problems.

In control problems, no delays are allowed, i.e. the real time constraint is an important requirement. Consequently a real time reformulation of the distortion rate problem is required.

On the other hand, problem (*) can be viewed as a control problem where the encoder 2(1) and the estimator 1(2) are two controllers with different information at each instant of time. Because there is a restriction on information transfer between the two controllers, problem (*) is different from the classical centralized stochastic control problem. The fundamental difficulty of decentralized stochastic control was discovered by Witsenhausen [11]. He showed that a simple two stage scalar LQG problem had a nonlinear solution. The method of proof was to exhibit an ad-hoc nonlinear control law with performance exceeding that of the best linear control law. Analytical determination of the optimal control law has never been accomplished; numerical determination of the
control law is very difficult since the problem is equivalent to a non-convex infinite dimensional nonlinear programming problem. The non-linearities of the optimal control laws are due to the signaling phenomena* [12], [16]. There are special cases of stochastic control problems with nonclassical information pattern where the optimal solution is linear [20], [15]. There are also very few decentralized stochastic control problems where the structure of the information pattern is such that they can be solved using either dynamic programming [13], [14] or information theory [15]. In general however, analytical and numerical determination of the optimal control laws is very difficult.

Since the solution of control problems were signaling occurs is presently unknown, lower bounds are valuable since they help us evaluate linear suboptimal solutions. The development of lower bounds for stochastic control problems with nonclassical information pattern where signaling occurs will be the topic of this thesis. We will derive these bounds by using various real time reformulations of the distortion rate problem.

1.2 Organization

Since we take a rate distortion theoretic approach to nonclassical stochastic control problems where signaling occurs at first, in chapter 2, we briefly discuss the main concepts of rate distortion theory.

In chapter 3 we formulate the real time communication problems (P)

* Signaling is the act of transferring information from one decision maker to another through the use of decision variables. It is impossible (and perhaps even meaningless) to know what portion of a control law is used for signaling and what portion for control.
and (P') for finite state sources and then present real time distortion rate problems which provide lower bounds to problems (P) and (P').

In chapter 4 we extend the results of chapter 3 to continuous amplitude sources.

In chapter 5 we present an information theoretic approach to the filtering problem. First we rederive a lower bound to the classical filtering problem using a new approach. Then, we formulate a decentralized filtering problem (DN), and present a real time distortion rate problem which provides a lower bound to (DN). Finally, we formulate a decentralized filtering problem (L) involving delays and provide a lower bound to (L) by solving filtering problem.

Chapter 6 contains conclusions and suggestions for further research.

1.3 Contributions of the Thesis

1. The real time communication problem, the centralized filtering problem and decentralized filtering problems are formulated using an information theoretical approach. These formulations reveal the connections and differences among the above problems.

2. Various real time reformulations of the distortion rate problem are used to obtain the following results:

   (i) Lower bounds for the real time communication problem; these bounds lie above the lower bounds provided by the classical rate distortion problem.

   (ii) A lower bound to a decentralized filtering problem; this bound lies above the solution of the centralized filtering problem in the case of linear Gaussian sources linear observations and the squared
error criterion.

3. A lower bound to a decentralized filtering problem involving delays is obtained via the solution of a linear filtering problem.
Chapter 2

Summary of Rate Distortion Theory

In this chapter we briefly present the main concepts of Rate Distortion Theory. An extensive treatment of the Rate Distortion Problem can be found in [3].

2.1 The Rate Distortion Problem

Consider the communication system of Figure 2.1. In its simplest form the Rate Distortion Problem can be stated as follows:

Suppose we want to reconstruct the output of a discrete memoryless source\(^1\) \(\{X_t, P\}\) within a certain prescribed accuracy at some receiving point. Given the level of accuracy we want to find the channel with the minimum capacity that can achieve it. Now we proceed to give the mathematical formulation of the problem.

2.2. The Variational Problem Defining the Rate Distortion Function for Memoryless Sources.

Consider a discrete memoryless source \(\{X_t, P\}\) with a finite alphabet \(A_M = \{0, 1, 2, \ldots, M-1\}\), i.e. \(x_t \in A_M\). Elements \(X_n = (x_1, x_2, \ldots, x_n) \in A_M^n\), where \(A_M^n\) denotes the n-fold cartesian product of \(A_M\) with itself, are called source words of length n. In order to determine whether or not the output of the source is reconstructed within a certain accuracy, we need to assign a nonnegative cost function \(\rho(x_n', y_n)\) to every possible

\(^1\) A discrete memoryless source \(X_t, P\) is a sequence of independent discrete random variables identically distributed with probability distribution \(P\).
Figure 2.1
approximation \( \mathcal{Y}_n \) of a source word. This cost function \( \rho(x_n, y_n) \) is called a word distortion measure and describes the penalty charged for reproducing the source word \( x_n \) by \( y_n \). A sequence of word distortion measures

\[
F = \{ \rho(x_n, y_n) = \frac{1}{n} \sum_{i=1}^{n} \rho(x_i, y_i); \ 1 \leq n < \infty \}
\]

is called a fidelity criterion. A fidelity criterion composed of word distortion measures of the form

\[
\rho(x_n, y_n) = \frac{1}{n} \sum_{i=1}^{n} \rho(x_i, y_i)
\]

is called a single letter fidelity criterion. That is, with a single letter fidelity criterion the distortion between a source word and a reproducing word is the arithmetic average of the distortions between their corresponding letters. Assume that the discrete memoryless source \( \{X_t, P\} \) described above, and a single letter fidelity criterion generated by a single letter distortion measure \( \rho(\cdot) \) in accordance with (2.2.2) are given. A discrete memoryless channel with input alphabet

\[
A_M = \{m_0, m_1, \ldots, m_{M-1}\}
\]

and output alphabet \( A_N = \{n_0, n_1, \ldots, n_{N-1}\} \) is described completely by specifying for every ordered pair \((j,k)\), in the product space \(A_M \times A_N\), the conditional probability \( \tilde{Q}(n_k | m_j) \) that the letter \( n_k \) appears at the channel output when the input letter is \( m_j \). Assume that information about the source is transmitted through a discrete memoryless channel and that the reproducing alphabet is

\[
A_N = \{0, 1, 2, \ldots, N-1\} .
\]

With each conditional probability assignment, (c.p.a.), \( \tilde{Q}(k | j) \), \( \{k \in A_N, \ j \in A_M\} \) there is a joint distribution

\[
p(j, k) = p(j) \tilde{Q}(k | j)
\]

associated with the product space \( A_M \times A_N \). Then, the single letter distortion measure \( \rho(\cdot) \) that generates \( F \) becomes a random
variable over the joint ensemble. Since its expected value depends on 
\( \mathbb{Q}(k|j) \), we denote it by \( d(\mathbb{Q}) \) and

\[
d(\mathbb{Q}) = \sum_{j,k} p(j) \mathbb{Q}(k|j) \rho(j,k) \tag{2.2.3}
\]

\( d(\mathbb{Q}) \) is called the average distortion associated with \( \mathbb{Q} \). A c.p.a. \( \mathbb{Q}(k|j) \)
is said to be D-admissible if and only if

\[ d(\mathbb{Q}) \leq D. \]

Denote by

\[
\mathcal{Q}_D \overset{\Delta}{=} \{ \mathbb{Q}(k|j) | d(\mathbb{Q}) \leq D \} \tag{2.2.4}
\]

the set of all D-admissible c.p.a. Each c.p.a. gives rise to a mutual information

\[
I(\mathbb{Q}) = \sum_{j,k} p(j) \mathbb{Q}(k|j) \log \frac{\mathbb{Q}(k|j)}{\sum_{l} \mathbb{Q}(l|l)} \tag{2.2.5}
\]

For a fixed D define

\[
R(D) = \min_{\mathbb{Q} \in \mathcal{Q}_D} I(\mathbb{Q}). \tag{2.2.6}
\]

\( R(D) \) viewed as a function of \( D \), \( \mathbb{D} \in [0,\infty) \), is called the rate distortion function of the discrete memoryless source \( \{X_t, P\} \).

The rate distortion function for continuous amplitude and discrete time sources can be defined in a similar way. Consider the source

\( \{X_t, P\} \) that produces independent identically distributed outputs governed by an absolutely continuous probability distribution \( P(x) \) with density \( p(x) \). Assume that both the source alphabet \( X \) and the reproducing alphabet \( Y \) are the entire real time. We measure the accuracy of reproduction by a nonnegative function \( \rho(x,y) \) called a distortion measure or cost function. The distortion that results when the source word

\[ x = (x_1, x_2, \ldots, x_n) \]
is reproduced as \( y = (y_1, y_2, \ldots, y_n) \) is
\[ \rho(x_n, y_n) \overset{\Delta}{=} \rho_n(x, y) = \frac{1}{n} \sum_{k=1}^{n} \rho(x_k, y_k) \]  

(2.2.7)

The family

\[ F_\rho = \{ \rho_n(x, y); \ 1 \leq n < \infty \} \]

generated by \( \rho(*) \) according to (2.2.7) is called the single letter fidelity criterion.

To every conditional probability density \( q(y|x) \) we assign both an average distortion

\[ d(q) = \iint dx dy \ p(x) \ q(y|x) \ \rho(x, y) \]  

(2.2.8)

and an average mutual information

\[ I(q) = \iint dx dy \ p(x) \ q(y|x) \ \log \frac{q(y|x)}{\int dx \ p(x) \ q(y|x)} \]  

(2.2.9)

The rate distortion function of \( \{ x_t, p \} \) with respect to \( F_\rho \) is then defined by

\[ R(D) = \inf_{q \in \mathcal{Q}_D} I(q) \]  

(2.2.10)

where

\[ \mathcal{Q}_D \overset{\Delta}{=} \{ q(y|x) \ \bigg| \ d(q) < D \} \]  

(2.2.11)

Because of its definition, \( R(D) \) specifies the minimum rate at which one must receive information concerning the source output in order to be able to reproduce it within an average distortion not exceeding \( D \). In other words, the rate distortion function determines the channel with the minimum capacity needed for a communication system so that a prescribed fidelity criterion is achieved. There is an alternative interpretation of the definition of \( R(D) \). Since \( R(D) \) is monotonic decreasing, (all the properties of \( R(D) \) will be listed in section 2.4)
the same curve would have resulted if we had fixed the average mutual information and then chose \( Q(k|j) \) so as to minimize the average distortion. The only difference would be an interchange in the roles of the dependent and independent random variables. With this approach we get a "distortion rate function" which determines the minimum distortion that can be achieved by a communication system when the source and the channel over which it is transmitted are given. This interpretation of \( R(D) \) is intuitively more appealing from the control and filtering point of view. Both interpretations will be used and the symmetry between the rate distortion and the distortion rate problems will prove very useful in our development.

In control systems the sources are continuous amplitude systems that have memory. The definition of the rate distortion function for stationary sources with memory will be the topic of the next section.

2.3 The Rate Distortion Function for Sources with Memory

Let \( F_\rho \) be a single letter fidelity criterion and let \( p(x) = p(x_1, x_2, \ldots, x_n) \) be the joint probability density governing \( n \) successive letters produced by the discrete time stationary source \( \{X_t\} \). Consider all conditional probability densities \( q(y|x) \) for reproducing words \( y = (y_1, y_2, \ldots, y_n) \) given source words \( x = (x_1, x_2, \ldots, x_n) \). The rate distortion function \( R(D) \) of \( \{X_t\} \) with respect to \( F_\rho \) is defined as

\[
R(D) = \lim_{n \to \infty} R_n(D) \tag{2.3.1}
\]

where

\[
R_n(D) = n^{-1} \inf_{q \in Q_D} I(q) \tag{2.3.2}
\]
\[ Q_D = \{ q(y|x) \mid d(q) = \int \int dx \ dy \ q(y|x) \ p(x) \ \rho_n(x, y) \leq D \} \]  

(2.3.3)

\[ \rho_n(x, y) = \frac{1}{n} \sum_{t=1}^{n} \rho(x_t, y_t) \]  

(2.3.4)

\[ I(q) = \int \int dx \ dy \ p(x) \ q(y|x) \ \log \ \frac{q(y|x)}{\int q(y|x') \ p(x') dx'} \]  

(2.3.5)

Even for a single letter fidelity criterion, the rate distortion function for sources with memory is defined asymptotically. This happens because the successive outputs of the source are statistically dependent, and one has to take into account the inherent statistical dependence to find the minimum information per letter needed to achieve a specific fidelity.

In the next section we will list some important properties of the rate distortion function.

2.4 Properties of the Rate Distortion Function

We list the properties of the rate distortion function for memoryless sources. Similar properties hold for the rate distortion function for sources with memory.

The first property of the rate distortion curve that must be established is the existence. The existence of \( R(D) \) is established in section 2.4 of [3]. It is shown that for

\[ 0 \leq D \leq \min_k \sum_j P_j \ \rho_{jk} \leq D_{\max} \]

\( R(D) \) is positive, and vanishes for \( D \geq D_{\max} \). The most important
property of the rate distortion function is its convexity. This property is expressed by the following theorem:

**Theorem 2.4.1 (Theorem 2.4.2 [3])**

\[ R(D') \text{ is a convex function of } D. \text{ That is, for any pair of distortion values } D' \text{ and } D'' \text{ and any number of } \lambda \in [0,1], \text{ we have the inequality} \]
\[ R(\lambda D' + (1-\lambda)D'') \leq \lambda R(D') + (1-\lambda)R(D'') \]

Since \( R(D) \) is convex it is continuous in the interval \((0, D_{\text{max}})\). It is shown in Theorem 2.5.4 of [3] that \( R(D) \) is also continuous at \( D = 0 \). Convexity of \( R(D) \) also implies that for \( D < D_{\text{max}} \), \( R(D) \) is strictly decreasing. Therefore, the minimizing c.p.a. \( Q(k|j) \) always lies on the boundary of the set \( Q_D \) as defined by (2.2.4). Usually each point on the rate distortion curve, except perhaps the point \((D_{\text{max}}, 0)\) is achieved by a unique c.p.a. In certain cases, however, there can be many choices of c.p.a. \( Q \in Q_D \) for which a point on the rate distortion curve is achieved. This fact is expressed by the following theorem.

**Theorem 2.4.2 (Theorem 2.4.2, [3])**

If \( Q' \) and \( Q'' \) both achieve the point \((D, R(D))\) then
\[ \bar{Q} = \lambda Q' + (1 - \lambda)Q'' \]
for any \( \lambda \in [0,1] \) also achieves the same point. \[ \square \]

The properties of the rate distortion function guarantee that for every distortion level there exists at least one c.p.a. \( Q \) for which \( R(D) \) is achieved. The coding theorems, which will be stated in the next section, specify in what sense the rate distortion function represents the effective rate at which a source generates information.
subject to the requirement that its output must be reproduced with
fidelity D.

2.5 Coding Theorems

The coding theorems are the fundamental theorems of rate distortion
type. They are very important because they essentially establish the
concept of the "rate of the source". One cannot talk about the rate of
a source unless the direct coding theorem is proved. The direct and
converse coding theorems can be stated as follows:

Theorem 2.5.1 Source Coding Theorem (Theorem 3.2.1 [3]

Let the d.m.s. \( \{X_t, P\} \) and the single letter fidelity criterion \( F_\rho \) be
given, and let \( R(\cdot) \) denote the rate distortion function of \( \{X_t, P\} \) with
respect to \( F_\rho \). Then, given any \( \epsilon > 0 \) and any \( D > 0 \), an integer \( n \) can
be found such that there exists a \( D + \epsilon \) - admissible code, (i.e. a
code that achieves an average distortion less than \( D + \epsilon \)) of block
length \( n \) with rate* \( R < R(D) + \epsilon \).

Theorem 2.5.2 Converse Coding Theorem (Theorem 3.2.2 [3]

No \( D \)-admissible code has a rate less than \( R(D) \). □

The source coding theorems have very important consequences in
communications theory. These consequences will be discussed in the
next section.

* A source code with size \( \Lambda \) (i.e. \( \Lambda \) reproducing words) and block length
n is said to have rate \( R = n^{-1} \log \Lambda \).
2.6 Discussion of the Fundamental Results of Rate Distortion Theory

The most important characteristic of the source coding theorem is its asymptotic nature. The source encoder of an ideal system must operated on long source words even though the letters comprising these words are generated independently of one another. An intuitive explana-
tion of the asymptotic nature of the coding theorem can be given as follows: Consider the communication system of Figure 2.1 in which the source encoder effects a random transformation on successive letters in accordance with the conditional probability assignment $Q(k|j)$ which solves the variational problem defining the rate distortion function. The output of such an encoder reproduces the source with an average distortion $D$, and the average mutual information between the input and the output of the encoder is $R(D)$ bits per source letter. However, since the source encoder operates in a probabilistic way, the condi-
tional entropy of its output given its input is positive. Consequently, the entropy rate of the source encoder exceeds $R(D)$ the mutual informa-
tion between its input and output. Since for error-free transmission the information transmission theorem (Theorem 3.3.1 [3], also Theorem 5.6.2 [4]) requires the capacity of the channel to be greater than or equal to the entropy of the source encoder, a channel with capacity strictly greater than $R(D)$ bits per source letter is required. The additional capacity required is equal to the conditional entropy of the random transformation applied to the source encoder. No capacity additional to $R(D)$ would be required if the transformation applied by the source coder were deterministic. This is exactly what a block length source code does. A good source code takes advantage of the
statistical regularity of the source over long periods and results in a
deterministic mapping from source words to code words. If in the above
code we consider the letters individually, then the probability that
the letter m of a source code will be mapped into the letter l of a
code word approaches the solution Q(l|m) of the variational problem
defining R(D) as the blocklength of the code approaches infinity.

Up to this point we presented the main concepts of Rate Distortion
Theory. In the next section we will compare this theory to the real time
distortion theory required for control.

2.7 Comparison to the Real Time Rate Distortion Theory Required for
Control

The asymptotic character of the rate distortion function for
sources with memory and of the coding theorem are the main reasons why
classical rate distortion theory is not appropriate for control purposes.

One of the main characteristics of control problems is that no
delays are allowed, i.e. the real time constraint is an important
requirement. In order to have a rate distortion theory appropriate for
control purposes we should be able to define a real time rate distortion
function for sources with memory and, if possible, prove a real time
coding theorem, (i.e. a coding theorem for a delay dependent fidelity
criterion, where the penalty on any delay is infinite*). This will be
the topic of the next chapter.

* Krch and Berger [19], derived coding theorems for a delay dependent
fidelity criterion where there is a finite penalty on the delay. However,
because of the real time constraint of our problem, we must have an
infinite penalty on any delay.
CHAPTER 3

The Real Time Distortion Rate Problem - Finite State Sources

At the beginning of this chapter we will state the feedforward real time communication problem (P) and the noiseless feedback real time communication problem (P'). Then we will formulate two real time distortion rate problems whose solutions provide lower bounds on (P) and (P'). We will discuss in detail the properties and characteristics of the formulations. Finally we will present and briefly discuss other formulations of the feedforward real time communication problem.

3.1 The Real Time Communication Problem

3.1.1 The Noiseless Feedback Real Time Communication Problem

Consider the communication system shown in figure 3.1. The noiseless feedback real time communication problem can be stated as follows:

\[
\begin{align*}
\text{Minimize} & \quad \frac{1}{T} \sum_{t=1}^{T} \rho(x_t, \hat{x}_t), \\
\text{subject to} & \quad I(y_t; z_t) \leq C \quad \forall t, \\
& \quad z_t = f'(x_t, y_{t-1}) \quad \forall t, \\
& \quad \hat{x}_t = g'_t(y_t) \quad \forall t.
\end{align*}
\]

(3.1.1) (P') (3.1.2) (3.1.3) (3.1.4)
Figure 3.1

Figure 3.2
In (3.1.2) $C$ is the capacity of the channel, and in (3.1.1) $\rho(\cdot)$ is a distortion measure; it can be $\rho = (x_t - \hat{x}_t)^2$ or $\rho = |x_t - \hat{x}_t|$ for example.

In this thesis we will be mostly concerned with the feedforward real time communication problem. We state the feedback problem here because we will relate a real time distortion rate problem to it.

3.1.2 The Feedforward Real Time Communication Problem

Consider the communication system of figure 3.2. The feedforward real time communication problem can be stated as follows:

Minimize \[
\frac{1}{T} E \sum_{t=1}^{T} \rho(x_t, \hat{x}_t)
\]

subject to \[
I(y_t; z_t) \leq C \forall t \quad (P)
\]
\[
z_t = f_t(x_t) \forall t \quad (3.1.7)
\]
\[
\hat{x}_t = g_t(y_t) \forall t \quad (3.1.8)
\]

$C$ is again the capacity of the channel and $\rho(\cdot)$ is a distortion measure. The real time constraint makes problem (P) different from a classical communication problem where arbitrary delays are allowed. On the other hand, the real time communication problem can be looked at as a control problem where the encoder and the decoder are viewed as two controllers which at every instant of time form their decisions based upon two different information sets: the outputs of the source and the outputs of the channel. Moreover the information of the decoder depends on the
decisions of the encoder. Thus, the feedforward real time communication problem, looked at from the control point of view, has all the difficulties associated with stochastic control problems with nonclassical information patterns where signaling occurs. It is an infinite dimensional, nonconvex, nonlinear programming problem. Since the solution of \( P \) remains unknown up to the present, we will try to formulate a real time distortion rate problem whose solution will provide a lower bound on \( P \), and which will avoid difficulties due to the signaling phenomena.

3.2 **Formulation of a Discrete Finite State Real Time Distortion Rate Problem**

3.2.1 **Statement of the Problem**

Consider the system of figure 3.2. Assume that the source is described by a finite state process with known initial probability \( \pi(0) \) and known transition probability matrix \( P_t \) for all \( t, t=1,2,\ldots \) .

At every instant of time \( t \), the encoder output \( z_t \) depends on all previous outputs of the source and the decoder output \( \hat{x}_t \) depends on all previous outputs of the channel, i.e.,

\[
z_t = f_t(x_t) \quad \forall t \tag{3.2.1}
\]

\[
\hat{x}_t = g_t(y_t) \quad \forall t \tag{3.2.2}
\]

The channel is assumed to be memoryless and have capacity \( C \).

Under the above assumptions the real time distortion rate problem can be stated as follows:
Minimize \( \mathcal{J}_T = \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \rho(x_t, \hat{x}_t) \)

\[
= \frac{1}{T} \sum_{t=1}^{T} \sum_{x_t, x_t', y_{t-1}} p(x_t, y_{t-1}) q(\hat{x}_t | x_t, y_{t-1}) \rho(x_t, \hat{x}_t)
\]

subject to

\[
I(\hat{x}_t; x_t | y_{t-1}) = \sum_{x_t, \hat{x}_t, y_{t-1}} p(x_t, y_{t-1}) q(\hat{x}_t | x_t, y_{t-1}) \cdot \\
\cdot \log \frac{q(\hat{x}_t | x_t, y_{t-1})}{Q(\hat{x}_t | y_{t-1})} \leq C \quad \forall t
\]

\( q(\hat{x}_t | x_t, y_{t-1}) \geq 0 \quad \forall t, \forall (\hat{x}_t, x_t, y_{t-1}) \)

\( \sum_{\hat{x}_t} q(\hat{x}_t | x_t, y_{t-1}) = 1 \quad \forall t, \forall (x_t, y_{t-1}) \)

\[
Q(\hat{x}_t | y_{t-1}) = \sum_{x_t'} q(\hat{x}_t | x_t', y_{t-1}) p(x_t' | y_{t-1})
\]

In the sequel we will study the properties of the above formulation and discuss its characteristics.

3.2.2 Properties and Characteristics of the Proposed Formulation

At first we will show that the solution of (3.2.3)-(3.2.6) is a lower bound on \((P)\) and then we will try to solve (3.2.3)-(3.2.6).
Theorem 3.2.1

The solution of (3.2.3)-(3.2.6) provides a lower bound on (P).

Proof:

Consider $I(\hat{x}_t; x_t | y_{t-1})$, use (3.2.2) and the memoryless property of the channel, and apply the data processing theorem ([4], pg. 80). Then

$$I(\hat{x}_t; x_t | y_{t-1}) \leq I(y_t; x_t | y_{t-1}) = I(y_t; x_t | y_{t-1}) =$$

$$= H(y_t | y_{t-1}) - H(y_t | x_t, y_{t-1}) \leq H(y_t) - H(y_t | z_t, x_t, y_{t-1}) =$$

$$= H(y_t) - H(y_t | z_t) = I(y_t; z_t) \quad (3.2.7)$$

Because of (3.2.7), the set of feasible solutions of problem (P) is a subset of the set of feasible solutions of the problem (3.2.3)-(3.2.6).

Therefore, the solution of (3.2.3)-(3.2.6) provides a lower bound on (P). QED □.

Let us consider the solution of (3.2.3)-(3.2.6). We note that for all $t$, the information constraint (3.2.4) as well as the cost (3.2.3) require knowledge of the probability distribution of $y_{t-1}$. The probability distribution of $y_{t-1}$ depends on the decisions of the encoder up to time $t-1$. Consequently determining the probability distribution of $y_{t-1}$ is equivalent to solving the signaling problem. Thus, the proposed formulation has all the undesirable characteristics of stochastic control problems with nonclassical information patterns where signaling occurs. Since the solution of (3.2.3)-(3.2.6) remains unknown,
we will try another formulation of the real time distortion rate problem that will provide a lower bound on \( (P) \). This will be the topic of the next section.

3.3 Another Formulation of a Discrete Finite State Distortion Rate Problem

3.3.1 Statement of the Problem

We formulate the real time distortion rate problem as follows:

Minimize \( J'_T = \frac{1}{T} E \sum_{t=1}^{T} \rho(x_t, \hat{x}_t) \)

\( \{q(\hat{x}_t | x_t)\}_{t=1}^{T} \)

\[
= \frac{1}{T} \sum_{t=1}^{T} \sum_{x_t, \hat{x}_t} p(x_t) q(\hat{x}_t | x_t) \rho(x_t, \hat{x}_t)
\] (3.3.1)

subject to

\[
I(\hat{x}_t; x_t | x_{t-1}) = \sum_{\hat{x}_t, x_t} p(x_t) q(\hat{x}_t | x_t) \log \frac{q(\hat{x}_t | x_t)}{Q(\hat{x}_t | x_{t-1})} \leq C \ \forall t
\] (3.3.2)

\( q(\hat{x}_t | x_t) \geq 0 \ \forall t, \forall(\hat{x}_t, x_t) \) (3.3.3)

\[
\sum_{\hat{x}_t} q(\hat{x}_t | x_t) = 1 \ \forall t, \forall x_t
\] (3.3.4)

\[
\left( Q(\hat{x}_t | x_{t-1}) = \sum_{x'_t} p(x'_t | x_{t-1}) q(\hat{x}_t | x'_t x_{t-1}) \right)
\]

We will study the properties of the above formulation and discuss its advantages and disadvantages.
3.3.2 Properties and Characteristics of the Proposed Formulation

First we will show that the solution of (3.3.1)-(3.3.4) provides a lower bound on (P) and (P').

Theorem 3.3.1

The solution of (3.3.1)-(3.3.4) lies below the solution of (3.2.3)-(3.2.6), hence it provides a lower bound on (P). The solution of (3.3.1)-(3.3.4) also provides a lower bound on (P').

Discussion

Intuitively we expect that the solution of (3.3.1)-(3.3.4) lies below that of (3.2.3)-(3.2.6). In problem (3.3.1)-(3.3.4) we assume that after (t-1) transitions the decoder knows \( x_{t-1} \). Actually the decoder knows \( y_{t-1} \), less than \( x_{t-1} \). Hence according to (3.3.1)-(3.3.4) the amount of information required at time \( t \) in order to reproduce \( x_t \) perfectly (or up to some level of fidelity) is less than the information actually needed in order to have a perfect reproduction (or a reproduction of certain fidelity) at the output of the decoder. Thus, the rate distortion curve of problem (3.3.1)-(3.3.4) should lie below that of (3.2.3)-(3.2.6).

Proof:

Under the assumptions

\[
  z_t = f(x_t)
\]

\[
  \hat{x}_t = g(y_t)
\]
we have,

\[ q(\hat{x}_t|x_t) = \sum_{\hat{y}_{t-1}} q(\hat{x}_t, \hat{y}_{t-1}|x_t) = \]

\[ = \sum_{\hat{y}_{t-1}} q(\hat{x}_t|\hat{y}_{t-1}, x_t) p(y_{t-1}|x_t, x_{t-1}) \]  \hspace{1cm} (3.3.5)

But

\[ p(y_{t-1}|x_t, x_{t-1}) = \frac{p(x_t, x_{t-1}, y_{t-1})}{p(x_t, x_{t-1})} = \]

\[ = \frac{p(x_t|x_{t-1}) p(y_{t-1}|x_{t-1}) p(x_{t-1})}{p(x_t|x_{t-1}) p(x_{t-1})} = p(y_{t-1}|x_{t-1}) \]  \hspace{1cm} (3.3.6)

and

\[ q(\hat{x}_t|y_{t-1}, x_t, x_{t-1}) = \frac{p(x_t, \hat{x}_t, x_{t-1}, y_{t-1})}{p(x_t, \hat{x}_t, x_{t-1}, y_{t-1})} = \]

\[ = \frac{p(x_{t-1}|x_t, y_{t-1}) q(\hat{x}_t|y_{t-1}, x_t) p(x_t, y_{t-1})}{p(x_{t-1}|x_t, y_{t-1}) p(x_t, y_{t-1})} = \]

\[ = q(\hat{x}_t|x_t, y_{t-1}) \]  \hspace{1cm} (3.3.7)

Because of (3.3.6) and (3.3.7), (3.3.5) gives

\[ q(\hat{x}_t|x_t) = \sum_{y_{t-1}} p(y_{t-1}|x_{t-1}) q(\hat{x}_t|x_t, y_{t-1}) \]  \hspace{1cm} (3.3.8)

The cost to minimize in problems (3.3.1)-(3.3.4) and (3.2.3)-(3.2.6) can be written as
\[ \mathcal{T}_T = \mathcal{T}_T^* = \frac{1}{T} \sum_{t=1}^{T} \left[ \sum_{\hat{y}_{t-1}, \hat{y}_t} \sum_{\hat{x}_t} p(y_{t-1} | x_{t-1}) q(\hat{x}_t | x_t, y_{t-1}) \cdot \right. \\
\left. \cdot p(x_t) \rho(x_t, \hat{x}_t) \right] \cdot (3.3.9) \]

Each of the information constraints of problem (3.3.1)-(3.3.4) can be written as

\[ I(\hat{x}_t; x_t | x_t) \Delta I(q(\hat{x}_t | x_t)) = I\left( \sum_{\hat{y}_{t-1}} p(y_{t-1} | x_{t-1}) q(\hat{x}_t | x_t, y_{t-1}) \right) = \]

\[ = \sum_{x_t, \hat{x}_t} p(x_t) q(\hat{x}_t | x_t) \log \frac{q(\hat{x}_t | x_t)}{Q(\hat{x}_t | x_{t-1})} \quad (3.3.10) \]

Each of the information constraints of problem (3.2.3)-(3.2.6) can be written as

\[ I(\hat{x}_t; x_t | y_{t-1}) \Delta I(q(\hat{x}_t | x_t, y_{t-1})) = \]

\[ = \sum_{\hat{y}_{t-1}} p(y_{t-1} | x_{t-1}) \left[ \sum_{\hat{x}_t} q(\hat{x}_t | x_t, y_{t-1}) p(x_t) \cdot \right. \\
\left. \cdot \log \frac{q(\hat{x}_t | x_t, y_{t-1})}{Q(\hat{x}_t | y_{t-1})} \right] = \]

\[ = \sum_{x_t} p(x_t) \sum_{\hat{y}_{t-1}} p(y_{t-1} | x_{t-1}) \sum_{\hat{x}_t} q(\hat{x}_t | x_t, y_{t-1}) \cdot \log \frac{q(\hat{x}_t | x_t, y_{t-1})}{Q(\hat{x}_t | y_{t-1})} \quad (3.3.11) \]
Using Jensen's inequality

\[ f \left( \sum_j a_j x_j \right) \leq \sum_j a_j f(x_j) \]  \hspace{1cm} (3.3.12)

(where \( f \) is a convex function of \( x = \sum_i a_j x_j \) and \( \sum_i a_j = 1 \)) we get that

\[
I(\hat{x}_t; x_t | x_{t-1}) \leq I \left( \sum_{y_{t-1}} p(y_{t-1} | x_{t-1}) q(\hat{x}_t | x_t, y_{t-1}) \right)
\]

\[
\leq \sum_{y_{t-1}} p(y_{t-1} | x_{t-1}) \sum_{x_t} p(x_t) \sum_{\hat{x}_t} q(\hat{x}_t | x_t, y_{t-1}) \cdot 
\]

\[
\log \frac{q(\hat{x}_t | x_t, y_{t-1})}{Q(\hat{x}_t | y_{t-1})} = I(\hat{x}_t; x_t | y_{t-1}) \]  \hspace{1cm} (3.3.13)

Because of (3.3.13) the set of feasible solutions \( \{ q(\hat{x}_t | x_t, y_{t-1}) \} \) of the problem (3.2.3)-(3.2.6) is a subset of the set of feasible solutions of the problem (3.3.1)-(3.3.4). Therefore the solution of (3.3.1)-(3.3.4) lies below that of (3.2.3)-(3.2.6). From Theorem 3.2.1, the solution of (3.2.3)-(3.2.6) is a lower bound on (P). Consequently the solution of (3.3.1)-(3.3.4) provides a lower bound on (P).

To prove that the solution of (3.3.1)-(3.3.4) lies below (P') we consider 
\( I(\hat{x}_t; x_t | x_{t-1}) \), apply the data processing theorem and the fact that the channel is memoryless.

Then we get

\[
I(\hat{x}_t; x_t | x_{t-1}) \leq I(y_t; x_t | x_{t-1}) = 
\]

\[
= \sum_{x_t, y_t} p(x_t) p(y_t | x_t) \log \frac{p(y_t | x_t)}{p(y_t | x_{t-1})} \]  \hspace{1cm} (3.3.14)
But
\[ p(y_t | x_t) = p(y_t | x_t, y_{t-1}) p(y_{t-1} | x_t) \]  \hspace{1cm} (3.3.15) \\
\[ p(y_t | x_{t-1}) = p(y_t | x_{t-1}, y_{t-1}) p(y_{t-1} | x_{t-1}) \]  \hspace{1cm} (3.3.16) \\
and
\[ p(y_{t-1} | x_t) = \frac{p(x_t, y_{t-1})}{p(x_t)} = \frac{p(x_t | x_{t-1}) p(y_{t-1} | x_{t-1}) p(x_{t-1})}{p(x_t | x_{t-1}) p(x_{t-1})} \]
\[ = p(y_{t-1} | x_{t-1}) . \]  \hspace{1cm} (3.3.17)

Because of (3.3.15)-(3.3.17), (3.3.14) gives
\[ I(\hat{x}_t; x_t | x_{t-1}) \leq I(y_t; x_t | x_{t-1}) = I(y_t; x_t | x_{t-1}, y_{t-1}) \]
\[ = H(y_t | x_{t-1}, y_{t-1}) - H(y_t | x_t, y_{t-1}) \leq \]
\[ \leq H(y_t) - H(y_t | z_t | x_t, y_{t-1}) = H(y_t) - H(y_t | z_t) = \]
\[ = I(y_t; z_t) . \]  \hspace{1cm} (3.3.18)

(3.3.18) shows that the set of feasible solutions of (P') is a subset of the set of feasible solutions of (3.3.1)-(3.3.4). Therefore, the solution of (3.3.1)-(3.3.4) provides a lower bound on (P') QED □

Let us consider now the solution of (3.3.1)-(3.3.4) and its properties. Since a decision taken at a specific time does not affect the decisions taken in the future, the problem (3.3.1)-(3.3.4) is equivalent to the following series of optimization problems:

\[ \text{Min } \mathcal{J}_T = \frac{1}{T} \sum_{t=1}^{T} \text{Min } \mathbb{E} \rho(x_t, \hat{x}_t) , \]
\[ \{ q(\hat{x}_t | x_t) \}^{T}_{t=1} q(\hat{x}_t | x_t) \]
\[ = \frac{1}{T} \sum_{t=1}^{T} \text{Min } \sum_{x_t, \hat{x}_t} p(x_t) q(\hat{x}_t | x_t) \rho(x_t, \hat{x}_t) , \]  \hspace{1cm} (3.3.19)
subject to

\[ I(\hat{x}_t; x_t | \hat{x}_{t-1}) = \sum_{\hat{x}_t, x_t} p(x_t) q(\hat{x}_t | x_t) \log \frac{q(\hat{x}_t | x_t)}{Q(\hat{x}_t | x_{t-1})} \leq C \quad \forall t, \]

(3.3.20)

\[ q(\hat{x}_t | x_t) \geq 0 \quad \forall t, \quad \forall (x_t, \hat{x}_t) \]

(3.3.21)

\[ \sum_{\hat{x}_t} q(\hat{x}_t | x_t) = 1 \quad \forall x_t, \quad \forall t \]

(3.3.22)

Each one of the above static optimization problems has the following form:

Minimize \[ E \rho(x_t, \hat{x}_t) \]

\[ q(\hat{x}_t | x_t) \]

subject to

\[ I(\hat{x}_t; x_t | x_{t-1}) \leq C \]

(3.3.24)

\[ q(\hat{x}_t | x_t) \geq 0 \quad \forall (\hat{x}_t, x_t) \]

(3.3.25)

\[ \sum_{\hat{x}_t} q(\hat{x}_t | x_t) = 1 \quad \forall x_t \]

(3.3.26)

The problem symmetric to (3.3.23)-(3.3.26) is

Minimize \[ I(q(\hat{x}_t | x_t)) \Delta I(\hat{x}_t; x_t | x_{t-1}) \]

\[ q(\hat{x}_t | x_t) \]

subject to

\[ E \rho(x_t, \hat{x}_t) \leq D(C) \]

(3.3.28)

\[ q(\hat{x}_t | x_t) \geq 0 \quad \forall (x_t, \hat{x}_t) \]

(3.3.29)
\[ \sum_{\hat{x}_t} q(\hat{x}_t \mid x_t) = 1 \quad \forall x_t \quad (3.3.30) \]

The problems (3.3.23)-(3.3.26) and (3.3.27)-(3.3.30) are similar to a classical distortion rate problem and a classical rate distortion problem, respectively. As we have already discussed in Chapter 2, the curves resulting from (3.3.23)-(3.3.26) and (3.3.27)-(3.3.30) by varying \( C \) are identical.

The quantity \( I(\hat{x}_t; x_t \mid x_{t-1}) \) gives the new information that can be provided to the user by the transmission at time \( t \). (3.3.22) defines the maximum rate at which the new information can be transmitted so that the error does not exceed on the average a distortion level \( D \).

Apart from this difference, the rate distortion function defined by (3.3.27)-(3.3.30) has properties similar to those of the classical rate distortion function. These properties are derived in Chapter 2.4 of [3] and have been stated in Section 2.4 of this report. Moreover, the equations and theorems derived by the variational problem corresponding to (3.3.27)-(3.3.30) are similar to the results derived by the variational problem defining the classical rate distortion function in [3]. Therefore, a number of results follow immediately and are simply stated below.

The solution of (3.3.27)-(3.3.30) is derived by forming the augmented function

\[ \mathcal{F}(q(\hat{x}_t \mid x_t)) = I(q(\hat{x}_t \mid x_t)) - \sum_{\hat{x}_t} \mu_{\hat{x}_t} \sum_{\hat{x}_t} q(\hat{x}_t \mid x_t) - \\
- s \sum_{x_t, \hat{x}_t} p(x_t) q(\hat{x}_t \mid x_t) \rho(x_t, \hat{x}_t) \quad (3.3.31) \]
(where $\mu$ and $s$ are Lagrange multipliers), and then determining its stationary point from the requirements

$$\frac{d \mathcal{R}(q(\hat{x}_t | x_t))}{d q(\hat{x}_t | x_t)} = 0. \quad (3.3.32)$$

(3.3.32) gives, (following a procedure similar to that of Section 2.5 in [3]),

$$q(\hat{x}_t | x_t) = \lambda_{x_t | x_{t-1}} Q(\hat{x}_t | x_{t-1}) e^{s\rho(x_t, \hat{x}_t)}, \quad (3.3.33)$$

where

$$\lambda_{x_t | x_{t-1}}^{-1} = \sum_{\hat{x}_t} Q(\hat{x}_t | x_{t-1}) e^{s\rho(x_t, \hat{x}_t)}, \quad (3.3.34)$$

$$Q(\hat{x}_t | x_{t-1}) = \sum_{x_t} q(\hat{x}_t | x_t) p(x_t | x_{t-1}), \quad (3.3.35)$$

and $\{p(x_t | x_{t-1})\}$ is the transition probability matrix for the source at time $t$.

The rate distortion curve can then be expressed parametrically in terms of $Q(\hat{x}_t | x_{t-1})$ and $s$ as

$$D = \sum_{x_{t-1}} p(x_{t-1}) \sum_{x_t, \hat{x}_t} p(x_t | x_{t-1}) \lambda_{x_t | x_{t-1}} Q(\hat{x}_t | x_{t-1}) \cdot e^{s\rho(x_t, \hat{x}_t)}, \quad (3.3.36)$$

and

$$R = sD + \sum_{x_{t-1}} p(x_{t-1}) \sum_{x_t} p(x_t | x_{t-1}) \log \lambda_{x_t | x_{t-1}}. \quad (3.3.37)$$
The properties of the optimal solution are given by the following theorems:

**Theorem 3.3.2 (Similar to Lemma 1, pg. 32, [3])**

For every \( s > \infty \), the optimum \( q(\hat{x}_t|\hat{x}_t) \) is such that if \( q(\hat{x}_t|x_t) = 0 \) for some \( x_t \), given \( x_{t-1} \), then \( q(\hat{x}_t|x_t) = 0 \) for all \( x_t \), i.e., if the transition probability from any source letter \( x_t \), given the previous outputs \( x_{t-1} \), to the reproducing letter \( \hat{x}_t \) is zero, then all transition probabilities to that letter vanish. Therefore, conditioned on the previous source outputs \( x_{t-1} \), the letter \( \hat{x}_t \) may be deleted from the reproducing alphabet without affecting the solution for this value of \( s \).

**Theorem 3.3.3 (Similar to Theorem 2.5.1, [3])**

The parameter \( s \) represents the slope of the rate distortion function at the point \( (D_s, R_s) \), that is parametrically generated by (3.3.36), (3.3.37). That is

\[
\frac{dR(D_s)}{dD_s} = s \quad \text{\( s \in \mathbb{R} \)}
\]

(3.3.38)

Theorem 3.3.2 means that for each \( D > 0 \) there is a nonzero conditional probability connecting every source letter \( x_t \) to every reproducing letter \( \hat{x}_t \) that is used at all. At \( D = 0 \) every source letter is connected with the reproducing letter with which it has zero distortion. Because of the result of Theorem 3.3.2, the reproducing alphabet, conditioned on the previous source outputs \( x_{t-1} \), can be partitioned into two subsets, the boundary set \( B_{\hat{x}_{t-1}} \) and the interior set \( V_{\hat{x}_{t-1}} \). If the optimum \( q(\hat{x}_t|x_t) = 0 \), and \( \hat{x}_t \in V_{\hat{x}_{t-1}} \). Given any vector
\[ Q \triangleq (Q(x^1_t|x_{t-1}), Q(x^2_t|x_{t-1}), \ldots, Q(x^m_t|x_{t-1})) , \]

(where \( m \) is the size of the reproducing alphabet), consider the boundary set \( \mathcal{B}_{Q,x_{t-1}} \) and the interior set \( \mathcal{V}_{Q,x_{t-1}} \). Associated with \( Q \) is a conditional probability assignment \( q(Q) \) given by (3.3.33)-(3.3.35).

\( q(Q) \), given by (3.3.33)-(3.3.35) is said to be a tentative solution for that point (or those points) on the rate distortion curve where

\[ dR(D)|dD = s , \]

if

\[ \sum_{x_t} \lambda x_t|x_{t-1} p(x_t|x_{t-1}) e^{sp(x_t, \hat{x}_t)} = 1 , \quad \hat{x}_t \in \mathcal{V}_{Q,x_{t-1}} . \] (3.3.39)

The notion of the tentative solution can be used to give a necessary and sufficient condition for a conditional probability assignment to be optimum. This condition is given in the following theorem.

**Theorem 3.3.4** (Similar to Theorem 2.5.2 [3])

A necessary and sufficient condition for a conditional probability assignment to yield a point on the rate distortion curve is that it be a tentative solution \( q(Q) \) that satisfies the added condition

\[ \sum_{x_t} \lambda x_t|x_{t-1} p(x_t|x_{t-1}) e^{sp(x_t, \hat{x}_t)} \leq 1 , \quad \forall \hat{x}_t \in \mathcal{B}_{Q,x_{t-1}} \] (3.3.40)

The result of Theorem 3.3.4 allows us to calculate a lower bound to the rate distortion function. The lower bound is given by the following theorem:
**Theorem 3.3.5** (Similar to Theorem 2.5.3 [3])

An alternative definition of the rate distortion function is

\[ R(D) = \max_{s \leq 0, \lambda \in \Lambda_s} \left( sD + \sum_{x_{t-1}} p(x_{t-1}) \sum_{x_t} p(x_t | x_{t-1}) \log \frac{\lambda_{x_t | x_{t-1}}}{\lambda_{\hat{x}_t | x_{t-1}}} \right), \]

(3.3.41)

where \( \Lambda_s \) is the set of all vectors

\[ \lambda_{x_t | x_{t-1}} = \left( \lambda_1^{x_t | x_{t-1}}, \lambda_2^{x_t | x_{t-1}}, \ldots, \lambda_M^{x_t | x_{t-1}} \right), \]

(where \( M \) is the size of the input alphabet), with nonnegative components that satisfy the inequality constraints

\[ \sum_{x_t} p(x_t | x_{t-1}) \lambda_{x_t | x_{t-1}} e^{s \rho(x_t, \hat{x}_t)} \leq 1. \]

(3.3.42)

Moreover for each \( s \leq 0 \) a necessary and sufficient condition for \( \lambda_{x_t | x_{t-1}} \) to achieve a maximum in (3.3.41) is that its components be given by (3.3.33)-(3.3.35) for a probability vector

\[ Q = (Q(\hat{x}_1^{x_t | x_{t-1}}), Q(\hat{x}_2^{x_t | x_{t-1}}), \ldots, Q(\hat{x}_M^{x_t | x_{t-1}})), \]

for which (3.3.39) and (3.3.40) hold.

Theorem 3.3.5 can be used to establish the continuity of the rate distortion function at \( D = 0 \).

**Theorem 3.3.6** (Similar to Theorem 2.5.4 [3])

The rate distortion function \( R(D) \) is continuous at \( D = 0 \).

Finally the behavior of the slope \( R(D) \) is described by the following theorem.
Theorem 3.3.7 (Similar to Theorem 2.5.5 [3])

The slope of the rate distortion function \( R(D) \) is continuous in the interval \( 0 < D < D_{\text{max}} \) and tends to \(-\infty\) as \( D \to 0 \). A discontinuity of \( dR(D)/dD \) can occur only at \( D = D_{\text{max}} \).

From Theorems 3.3.2-3.3.7 we note that according to the proposed formulation, for each instant of time we have to solve a family of static problems (3.3.27)-(3.3.30), each one of which is similar to a classical rate distortion problem for a stationary memoryless source. Thus, the algorithms developed for the solution of the classical rate distortion problem, [29],[30], can be used to solve each of the static problems (3.3.27)-(3.3.30). A disadvantage of the formulation is that the number of computations grows with time. This happens because the number of possible previous source outputs increases with time; therefore, the number of rate distortion problems we have to solve at every instant of time increases with time. We can overcome this problem by considering finite memory encoders and decoders. We can assume that the decisions of the encoder and the decoder depend on the last \( N \) outputs of the sources and the channel, respectively, i.e., \( \forall t \geq N \)

\[
z_t = f_t(x_t, x_{t-1}, \ldots, x_{t-N}), \tag{3.3.43}
\]

\[
\hat{x}_t = g_t(y_t, y_{t-1}, \ldots, y_{t-N}). \tag{3.3.44}
\]

Assuming again that we have a memoryless channel of capacity \( C \), we can formulate the real time distortion rate problem as follows:
Minimize \( \{q(\hat{x}_t|x_t,x_{t-1},\ldots,x_{t-N})\} \) \\
\( \mathcal{L}_T = \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \rho(x_t,\hat{x}_t) = \) \\
\( \frac{1}{T} \sum_{t=1}^{T} \sum_{x_t,\ldots,x_{t-N},\hat{x}_t} p(x_t,x_{t-1},\ldots,x_{t-N}) \cdot q(\hat{x}_t|x_t,\ldots,x_{t-N}) \cdot \rho(x_t,\hat{x}_t), \) \\
subject to \\
\( I(\hat{x}_t;x_t|x_{t-1},\ldots,x_{t-N}) = \sum_{x_t',x_{t-1}',\ldots,x_{t-N}',\hat{x}_t} p(x_t',\ldots,x_{t-N}') \cdot q(\hat{x}_t|x_t',\ldots,x_{t-N}') \cdot \frac{q(\hat{x}_t|x_t',\ldots,x_{t-N}'}{Q(\hat{x}_t|x_{t-1}',\ldots,x_{t-N}')}) \leq c, \forall t, \)

\( Q(\hat{x}_t|x_{t-1},\ldots,x_{t-N}) = \sum_{x_t'} p(x_t'|x_{t-1},\ldots,x_{t-N}) \cdot q(\hat{x}_t|x_t',\ldots,x_{t-N}'), \forall t, \) \hspace{1cm} (3.3.47) \\
\( q(\hat{x}_t|x_t,\ldots,x_{t-N}) \geq 0, \forall t \quad (\hat{x}_t,x_t,\ldots,x_{t-N}), \) \hspace{1cm} (3.3.48) \\
\( \sum_{\hat{x}_t} q(\hat{x}_t|x_t,\ldots,x_{t-N}) = 1, \forall t \quad (x_t,\ldots,x_{t-N}). \) \hspace{1cm} (3.3.49)

The properties of (3.3.45)-(3.3.49) are the same as those of (3.3.1)-(3.3.4). Following the proof of Theorem 3.3.1 it is straightforward to show that the solution of (3.3.45)-(3.3.49) provides a lower bound on the real time communication problems (P) and (P') with encoders and decoders which have finite memory. Thus, in (3.3.45)-(3.3.49) the
number of computations increases with time until \( t = N \) and afterwards remains constant.

The formulation described by (3.3.1)-(3.3.4) allows us to overcome some of the analytical, computational and conceptual difficulties due to the signaling phenomena. When signaling occurs in stochastic control problems with nonclassical information pattern, it gives rise to infinite dimensional, nonconvex, nonlinear programming problems with many local minima. With our formulation we get a series of convex nonlinear programming problems for the solution of which there are algorithms [29], [30]. Convexity results because we do not worry about the detailed way the encoder and the decoder process the data, but we are only interested in the transition probability from the input of the encoder to the output of the decoder. The result of the input-output description of the communication systems of figures 3.1 and 3.2 is a lower bound on the real time communication problems (P) and (P'). In order to show that this lower bound is attained we must either prove a real time coding theorem, or provide an example of a system which attains the bound. Since we have not been able to prove a real time coding theorem, we will provide an example of a feedforward system which attains the bound. The system we will present uses unavailable information. Hence the bound should be loose for feedforward systems that satisfy the information constraints.

**Example 3.3.1**

Consider the system of figure 3.3. Assume that the source is binary Markov, described by the equation

\[
x_{t+1} = x_t \oplus \eta_t ,
\]

(3.3.50)

(where \( \oplus \) means addition modulo 2).
Assume that initially we have

\[ \text{Prob}(x(1) = 0) = \pi_1 \]  \hspace{1cm} (3.3.51)

\( (\pi_1 < \frac{1}{2} \) without loss of generality) and that

\[ \text{Prob}(\eta_t = 1) = p_t = \pi_1 \ \forall t \geq 1 \]  \hspace{1cm} (3.3.52)

Moreover assume that the distortion measure is the Hamming distance between the input \( x_t \) and the output \( \hat{x}_t \) (for all \( t \)) and that a memoryless channel of capacity \( C \) is available.

We will consider the two shot problem (i.e., \( T = 2 \)). According to (3.3.1)-(3.3.4) we have to solve the following problem:

\[
\text{Minimize} \quad \frac{1}{2} \mathbb{E}[\rho(x_1, \hat{x}_1) + \rho(x_2, \hat{x}_2)] = \\
\{ q(\hat{x}_1 | x_1), q(\hat{x}_2 | x_1 x_2) \} = \\
= \frac{1}{2} \left[ \sum_{x_1, \hat{x}_1} p(x_1) q(\hat{x}_1 | x_1) \rho(x_1, \hat{x}_1) + \sum_{x_1, x_2, \hat{x}_2} p(x_1 x_2) \cdot \\
\cdot q(\hat{x}_2 | x_1 x_2) \rho(x_2, \hat{x}_2) \right], \hspace{1cm} (3.3.53)
\]

subject to

\[
I(\hat{x}_1; x_1) = \sum_{x_1, \hat{x}_1} p(x_1) q(\hat{x}_1 | x_1) \log \frac{q(\hat{x}_1 | x_1)}{Q(\hat{x}_1)} \leq C, \hspace{1cm} (3.3.54)
\]

\[
I(\hat{x}_2; x_2 | x_1) = \sum_{x_1, x_2, \hat{x}_2} p(x_1 x_2) q(\hat{x}_2 | x_1 x_2) \log \frac{q(\hat{x}_2 | x_1 x_2)}{Q(\hat{x}_2 | x_1)} \leq C \hspace{1cm} (3.3.55)
\]

\[
q(\hat{x}_1 | x_1) \geq 0 \ \forall (x_1, \hat{x}_1), \hspace{1cm} (3.3.56)
\]

\[
q(\hat{x}_2 | x_1 x_2) \geq 0 \ \forall (x_1, x_2, \hat{x}_2) \hspace{1cm} (3.3.57)
\]
\[ \sum_{\mathbf{x}_1} q(\mathbf{\hat{x}}_1 | \mathbf{x}_i) = 1 \quad \forall \mathbf{x}_1 , \quad (3.3.58) \]

\[ \sum_{\mathbf{x}_2} q(\mathbf{\hat{x}}_2 | \mathbf{x}_1 \mathbf{x}_2) = 1 \quad \forall (\mathbf{x}_1, \mathbf{x}_2) . \quad (3.3.59) \]

The details of the solution of (3.3.53)-(3.3.59) are given in Appendix A.

Here we give the results. They are:

**Time** \( t = 1 \):

\[ C = H_b(\pi_1) - H_b(D_1) , \quad (3.3.60) \]

**Time** \( t = 2 \)

\[ C = H_b(p_1) - H_b(D_2) = H_b(\pi_1) - H_b(D_2) , \quad (3.3.61) \]

(where \( H_b(\cdot) \) denotes binary entropy). (3.3.60) and (3.3.61) can be solved for \( D_1 \) and \( D_2 \), respectively; the minimum average distortion per shot is

\[ D = \frac{1}{2} (D_1 + D_2) . \quad (3.3.62) \]

The test channels \( \{ q(\mathbf{\hat{x}}_1 | \mathbf{x}_i) \} \) and \( \{ q(\mathbf{\hat{x}}_2 | \mathbf{x}_j ; \mathbf{x}_1 = k) \} \) \((k = 0, 1)\) are the same because \( p_1 = \pi_1 \). Thus

\[ D = D_1 = D_2 . \quad (3.3.63) \]

Consider the case where the channel of the system of figure 3.3 happens to be equal to the test channel \( \{ q(\mathbf{\hat{x}}_1 | \mathbf{x}_i) \} \). Each source output is available at the decoder with one unit of delay. At each instant of time \( t+1 \) \((t \geq 1)\), the decoder combines \( x_t \) with \( \mathbf{\hat{n}}_t \), the transmitted version of the innovations of the source \( \mathbf{n}_t \), and forms the estimate

41
\[ \hat{x}_{t+1} = x_t \Theta \hat{n}_t. \]  

(3.3.64)

The source output \( x_1 \) and the innovations \( n_t \) \((t > 1)\) are transmitted without any encoding, i.e.,

\[ z_1 = x_1, \]  

(3.3.65)

\[ z_t = n_{t-1} \quad \forall (t > 1). \]  

(3.3.66)

We show that under the above encoding and decoding strategies the system of figure 3.3 achieves the bound of (3.3.1)-(3.3.4).

At time \( t = 1 \) the mutual information \( I(\hat{x}_1; x_1) \) is

\[ I(\hat{x}_1; x_1) = \sum_{x_1, \hat{x}_1} p(x_1) q(\hat{x}_1|x_1) \log \frac{q(\hat{x}_1|x_1)}{q(\hat{x}_1)} = \]

\[ = \sum_{z_1, y_1} p(z_1) q(y_1|z_1) \log \frac{q(y_1|z_1)}{q(y_1)} = C, \]  

(3.3.67)

because

\[ x_1 = z_1, \quad \hat{x}_1 = y_1, \]

and the channel of the system is equal to the test channel \( \{q(\hat{x}_1|x_1)\} \).

(3.3.67) shows that the information constraint is not violated, hence the above transmission is allowable. The average distortion is

\[ D'_1 = E[p(x_1, \hat{x}_1) = E[p(y_1, z_1) = \]

\[ = \sum_{y_1, z_1} p(z_1) q(y_1|z_1) \rho(y_1, z_1) = \]

\[ = \sum_{x_1, \hat{x}_1} p(x_1) q(\hat{x}_1|x_1) \rho(x_1, \hat{x}_1) = D_1. \]  

(3.3.68)
At time \( t = 2 \) the mutual information \( I(\hat{x}_2; x_2 | x_1) \) is

\[
I(\hat{x}_2; x_2 | x_1) = I(x_1 \oplus \hat{n}_1; x_1 \oplus n_1 | x_1) = 
\]

\[
= H(x_1 \oplus \hat{n}_1 | x_1) - H(x_1 \oplus \hat{n}_1 | x_1 \oplus n_1, x_1) = 
\]

\[
= H(\hat{n}_1) - H(\hat{n}_1 | n_1) = I(\hat{n}_1 ; n_1) = I(y_2 ; z_2) , \tag{3.3.69}
\]

and

\[
I(y_2 ; z_2) = C \tag{3.3.70}
\]

because the probability distribution of \( n_1 \) matches the channel (i.e., capacity is achieved with an input distribution to the channel equal to the probability distribution of \( n_1 \)). (3.3.69) and (3.3.70) show that the information constraint at \( t = 2 \) is not violated; therefore, the encoding-decoding strategy

\[
z_2 = n_1 \\
\hat{x}_2 = x_1 \oplus \hat{n}_1 ,
\]

is allowable.

The average distortion at time \( t = 2 \) is

\[
D'_2 = \sum_{x_2, \hat{x}_2} p(x_2, \hat{x}_2) \rho(x_2, \hat{x}_2) = 
\]

\[
= \sum_{n_1, \hat{n}_1} p(x_1 \oplus n_1, x_1 \oplus \hat{n}_1) \rho(x_1 \oplus n_1, x_1 \oplus \hat{n}_1) = 
\]

\[
= \sum_{n_1, \hat{n}_1} p(n_1, \hat{n}_1) \rho(n_1, \hat{n}_1) = 
\]

\[
= \sum_{n_1, \hat{n}_1} p(n_1) q(\hat{n}_1 | n_1) \rho(n_1, \hat{n}_1) = D_1 = D_2 , \tag{3.3.71}
\]

43
because of
\[ p(x_1 = i) = p(n_1 = i) \]
\[ q(\hat{n}_1 = j | n_1 = k) = q(\hat{x}_1 = j | x_1 = k) \]
\[ \rho(x_1 = j \hat{x}_1 = k) = \rho(n_1 = j \hat{n}_1 = k) \]
and (3.3.63), (3.3.68).

The average distortion per shot is
\[ D' = \frac{1}{2} (D_1' + D_2') = \frac{1}{2} (D_1 + D_2) = D. \quad (3.3.72) \]

Thus, the system of figure 3.3 achieves the lower bound resulting from our formulation.

Let us compare now our lower bound to that provided by the classical rate distortion problem. For \( T = 2 \) the rate distortion function resulting from the classical rate distortion problem for our example is given by (see [27]),
\[ R(\bar{D}) = \frac{1}{2} \left[ \mathcal{H}_b(\pi_1) + \mathcal{H}_b(p_1) \right] - \mathcal{H}_b(\bar{D}), \quad (3.3.73) \]
for \( \bar{D} \leq D_C \left( D_C = \frac{1}{2} - \frac{1}{2} \left[ 1 - \left( \frac{p_1}{1 - p_1} \right)^2 \right]^{1/2} \right) \).

Let
\[ \pi_1 = 0.5, \quad (3.3.74) \]
\[ p_1 = 0.25, \quad (3.3.75) \]
\[ C = 0.75 \text{ bits}. \quad (3.3.76) \]

Then (3.3.73) gives
\[ 0.75 = \frac{1}{2} [\mathcal{H}_b(0.5) + \mathcal{H}_b(0.25)] - \mathcal{H}_b(\bar{D}), \quad (3.3.77) \]
for $\bar{D} \leq D_C$ and $D_C = 0.028$ for this problem.

Hence,

$$0.75 = \frac{1}{2} [1 + 0.8113] - \mathcal{H}_b(\bar{D}) \Rightarrow$$

$$\mathcal{H}_b(\bar{D}) = 0.1556 \Rightarrow$$

$$\bar{D} = 0.0226 . \quad (3.3.78)$$

(In fact, $\bar{D} < D_C$; therefore, we can use (3.3.73) to determine $\bar{D}$).

(3.3.60) and (3.3.61) give

$$0.75 = \mathcal{H}_b(0.5) - \mathcal{H}_b(D_1) , \quad (3.3.79)$$

$$0.75 = \mathcal{H}_b(0.25) - \mathcal{H}_b(D_2) , \quad (3.3.80)$$

or

$$D_1 = 0.0417 , \quad (3.3.81)$$

$$D_2 = 0.0072 \quad (3.3.82)$$

and

$$D_{\text{real time lower bound}} \triangleq D_{r.t.l.b.} = \frac{D_1 + D_2}{2} = 0.0244 . \quad (3.3.83)$$

(3.3.78) and (3.3.83) show that the lower bound on the per shot average error resulting from our formulation lies above the bound provided by the classical rate distortion problem $\mathcal{X}$.

The system presented in Example 3.3.1 has only theoretical value. We cannot implement such a system because it is impossible for the decoder to know perfectly the outputs of the source with one unit of delay. Thus, although we have shown that our lower bound is nontrivial (as it lies above the bound provided by the classical rate distortion
problem), we have been unable to find a practical system that attains the bound. It remains to investigate whether there exists a system whose performance approaches the bound. This will be the topic of the next section.

3.4 Investigation of the Tightness of the Lower Bound

The only case where the real time encoding-decoding problem is trivial is when the source is stationary and memoryless and has a probability distribution that matches the channel, i.e., capacity is achieved for the source's input distribution. Then the channel is equal to the test channel and the bound provided by the solution of (3.3.1)-(3.3.4) is attained without any encoding or decoding. In control problems the sources are usually nonstationary and have memory. For this class of sources the test channel for problem (3.3.1)-(3.3.4) is time varying. Thus, even if the channel happens to be equal to the test channel for problem (3.3.1)-(3.3.4) and \( t = 1 \), it will not be equal to the test channels for all instants of time afterwards. Then the encoder and the decoder will be, in general, deterministic functions of the previous outputs of the source and the channel, and such deterministic functions, which combined with the channel give (or even approach) the test channels for all \( t \), are very difficult to find. The difficulty in determining good encoders and decoders is not surprising. The encoding-decoding problem is nontrivial even in the case where infinite delays are allowed. The real time character of our problem introduces an additional difficulty which is not present in the classical coding problem. Since a systematic procedure for constructing good encoders and decoders does not exist, we have to restrict attention to special cases. We will
present communication problems where it is possible that the decoder sees a probability distribution of the source that matches the channel for all instants of time \( t \). For these problems we will formulate a real time distortion rate problem for which it will be possible to get a time invariant test channel. We will realize the solution by a real time communication system and compare the performance of the system to our lower bound.

3.4.1 Two Special Forms of Real Time Communication Problems

A. The feedforward problem

Consider the communication system of figure 3.2. A special form of the feedforward real time communication problem is the following:

\[
\begin{align*}
\min_{(f_T, g_T)} & \frac{1}{T} E \left\{ \rho(x_T, \hat{x}_T) + \min_{(f_{T-1}, g_{T-1})} E \left\{ \rho(x_{T-1}, \hat{x}_{T-1}) + \ldots \right. \right. \\
& \left. \left. \ldots + \min_{(f_1, g_1)} E \rho(x_1, \hat{x}_1) \right\} \ldots \right\},
\end{align*}
\]

subject to

\[
\begin{align*}
I(y_t; z_t) & \leq C \forall t, \quad \text{(R)} \\
z_t & = f_t(x_t) \forall t, \quad \text{(3.4.3)} \\
\hat{x}_t & = g_t(x_t) \forall t, \quad \text{(3.4.4)}
\end{align*}
\]

where \( C \) is the capacity of the channel and \( \rho(\cdot) \) a distortion measure.

B. The noiseless feedback problem

Consider the communication system of figure 3.4. A special form of the noiseless feedback real time communication problem can be stated as follows:
\[
\begin{aligned}
\min_{(\lambda_T, \mu_T)} & \quad \frac{1}{T} \mathbb{E}\left\{ \rho(x_T, \hat{x}_T) + \min_{(\lambda_{T-1}, \mu_{T-1})} \mathbb{E}\left\{ \rho(x_{T-1}, \hat{x}_{T-1}) + \right. \\
& \left. \quad \ldots + \min_{(\lambda_1, \mu_1)} \mathbb{E}\rho(x_1, \hat{x}_1) \ldots \right\} \right\} \\
\text{subject to} & \quad I(y_t; z_t) \leq C \quad \forall t,
\end{aligned}
\] (3.4.5)

subject to
\[
\begin{aligned}
I(y_t; z_t) & \leq C \quad \forall t, \\
\hat{x}_t & = \lambda_t (x_t, \hat{x}_{t-1}) \forall t, \\
\hat{x}_t & = \mu_t (y_t, \hat{x}_{t-1}).
\end{aligned}
\] (3.4.6-3.4.8)

A characteristic of the above problems (R) and (R') is that each instant of time the encoder and the decoder act without taking future transmissions into consideration. Hence, signaling phenomena are not present in problem (R). Another characteristic of problems (R) and (R') is that the decision of the decoder at any time \( t \) depends on all previous decisions up to time \( t-1 \).

We will formulate a real time distortion rate problem \((\pi)\), which provides a lower bound for problems (R) and (R'); we will find cases where the bound provided by \((\pi)\) is attained, we will realize the solution of \((\pi)\) by a real time communication system and compare the performance of that system to the bound of (3.3.1)-(3.3.4).

3.4.2 A Real Time Distortion Rate Problem for (R) and (R')

A real time distortion rate problem for (R) and (R') can be stated as follows:

\[
\begin{aligned}
\min_{q(\hat{x}_T \mid x_T, \hat{x}_{T-1})} \quad & \frac{1}{T} \mathbb{E}\left\{ \rho(x_T, \hat{x}_T) + \min_{q(\hat{x}_{T-1} \mid x_{T-1}, \hat{x}_{T-2})} \mathbb{E}\left\{ \rho(x_{T-1}, \hat{x}_{T-1}) + \right. \\
& \left. \quad \ldots + \min_{q(\hat{x}_1 \mid x_1)} \mathbb{E}\rho(x_1, \hat{x}_1) \right\} \ldots \right\} \\
& \quad + \ldots + \min_{q(\hat{x}_1 \mid x_1)} \mathbb{E}\rho(x_1, \hat{x}_1) \right\} \ldots \right\} \\
& \quad \right\} \ldots \right\} \\
\text{(3.4.9)}
\end{aligned}
\]
subject to
\[
\begin{align*}
I(\hat{x}_t; x_t | \hat{x}_{t-1}) & \leq C \quad \forall t \\
q(\hat{x}_t | x_t, \hat{x}_{t-1}) & \geq 0 \quad \forall t \quad \forall (\hat{x}_t, x_t) \\
\sum_{\hat{x}_t} q(\hat{x}_t | x_t, \hat{x}_{t-1}) &= 1 \quad \forall t \quad \forall (x_t, \hat{x}_{t-1}) .
\end{align*}
\]  

(3.4.10) \quad (3.4.11) \quad (3.4.12)

(3.4.9)-(3.4.12) is a series of optimization problems each one of which, conditioned on the previous decisions of the decoder, is a convex non-linear programming problem similar to a classical distortion rate problem. Problem \((\pi)\) has characteristics similar to those of \((R)\) and \((R')\), i.e.: (i) the solution at any time \(t\) depends on all previous solutions up to time \((t-1)\), and (ii) at any instant of time future transmissions are not taken into consideration. We will show that \((\pi)\) provides, in general, a lower bound on \((R)\) and \((R')\).

**Theorem 3.4.1**

The solution of (3.4.9)-(3.4.12) provides a lower bound on \((R)\) and \((R')\).

**Proof:**

Consider the quantity \(I(\hat{x}_t; x_t | \hat{x}_{t-1})\); using (3.4.4),(3.4.8), the fact that the channel is memoryless, and the data processing theorem we get

\[
I(\hat{x}_t; x_t | \hat{x}_{t-1}) \leq I(y_t; \hat{x}_{t-1}; x_t | \hat{x}_{t-1}) = H(y_t; \hat{x}_{t-1} | \hat{x}_{t-1}) - H(y_t ; \hat{x}_{t-1} | \hat{x}_{t-1} x_t) =
\]

\[
= H(y_t | \hat{x}_{t-1}) - H(y_t | \hat{x}_{t-1} x_t) = I(y_t; x_t | \hat{x}_{t-1}) \leq
\]

\[
= H(y_t) - H(y_t ; z_t) = H(y_t) - H(y_t | z_t) =
\]

\[
= I(y_t; z_t) .
\]  

(3.4.13)
(3.4.13) shows that the set of feasible solutions of problems (3.4.1)-(3.4.4) and (3.4.5)-(3.4.9) are subsets of the set of feasible solutions of problem (3.4.9)-(3.4.12). Hence, problem (π) provides a lower bound on (R) and (R').

We will present an example of a system which attains the bound provided by (π) and compare the performance of that system to the lower bound of (3.3.1)-(3.3.4).

**Example 3.4.1**

Consider a binary Markov source described by

\[ x_{t+1} = x_t \oplus n_t, \]  

(3.4.14)

(where \( \oplus \) means addition modulo 2).

Assume that

1. \( \text{Prob}(x_1 = 0) = \pi_1 = 0.3 \) .
2. \( \text{Prob}(n_t = 1) = 0.278 \quad \forall t \geq 1 \) .
3. A memoryless channel described by

\[ q(y = 0|z = 0) = 0.895, \]  

(3.4.17)

\[ q(y_1 = 1|z = 1) = 0.9513, \]  

(3.4.18)

and having capacity \( C = 0.6 \) bits is available.

4. The distortion measure is the Hamming distance between the input \( x_t \) and the output \( \hat{x}_t \).

Consider the three shot problem (i.e., \( T = 3 \)). According to (3.4.9)-(3.4.12) we have to solve the problem
\[
\begin{align*}
\text{Min} & \quad \frac{1}{3} \{ E[q(\hat{x}_3|x_3,\hat{x}_1,\hat{x}_2)|q(\hat{x}_1|x_1)] + \\
& \quad + \min_{q(\hat{x}_2|x_1)} \left\{ E[q(\hat{x}_2|x_2,\hat{x}_1)|q(\hat{x}_1|x_1)] \right\} \}
\end{align*}
\]

subject to

\[
I(\hat{x}_1;x_1) = \sum_{x_1,\hat{x}_1} p(x_1)q(\hat{x}_1|x_1) \log \frac{q(\hat{x}_1|x_1)}{Q(\hat{x}_1)} \leq C ,
\]

(3.4.20)

\[
I(\hat{x}_2;x_2|\hat{x}_1) = \sum_{\hat{x}_1,x_2,\hat{x}_2} p(x_2,\hat{x}_1)q(\hat{x}_2|x_2,\hat{x}_1) \log \frac{q(\hat{x}_2|x_2,\hat{x}_1)}{Q(\hat{x}_2|\hat{x}_1)} \leq C ,
\]

(3.4.21)

\[
I(\hat{x}_3;x_3|\hat{x}_1,\hat{x}_2) = \sum_{x_3,\hat{x}_3} p(x_3,\hat{x}_1,\hat{x}_2) \cdot \\
\cdot q(\hat{x}_3|x_3,\hat{x}_1,\hat{x}_2) \log \frac{q(\hat{x}_3|x_3,\hat{x}_1,\hat{x}_2)}{Q(\hat{x}_3|\hat{x}_1,\hat{x}_2)} \leq C ,
\]

(3.4.22)

\[
q(\hat{x}_1|x_1) \geq 0 \quad \forall (x_1,\hat{x}_1) ,
\]

(3.4.23)

\[
q(\hat{x}_2|x_2,\hat{x}_1) \geq 0 \quad \forall (\hat{x}_2,x_2,\hat{x}_1) ,
\]

(3.4.24)

\[
q(\hat{x}_3|x_3,\hat{x}_1,\hat{x}_2) \geq 0 \quad \forall (\hat{x}_3,x_3,\hat{x}_1,\hat{x}_2) ,
\]

(3.4.25)

\[
\sum_{\hat{x}_1} q(\hat{x}_1|x_1) = 1 \quad \forall x_1 ,
\]

(3.4.26)

\[
\sum_{\hat{x}_2} q(\hat{x}_2|x_2,\hat{x}_1) = 1 \quad \forall (x_2,\hat{x}_1) ,
\]

(3.4.27)

\[
\sum_{\hat{x}_3} q(\hat{x}_3|x_3,\hat{x}_1,\hat{x}_2) = 1 \quad \forall (x_3,\hat{x}_1,\hat{x}_2) .
\]

(3.4.28)
The details of the solution of (3.4.19)-(3.4.28) are given in Appendix B. Here we give the results.

**Time 
\( t = 1 \)**

The solution is

\[ C = 0.6 = \mathcal{H}_b(0.3) - \mathcal{H}_b(D_1) , \]

(where \( \mathcal{H}_b(\cdot) \) means binary entropy),

\[ D_1 = 0.0488 , \quad (3.4.29) \]

and

\[ q(\hat{x}_1 = 0 | x_1 = 0) = 0.895 , \quad (3.4.30) \]
\[ q(\hat{x}_1 = 1 | x_1 = 0) = 0.105 , \quad (3.4.31) \]
\[ q(\hat{x}_1 = 0 | x_1 = 1) = 0.0487 , \quad (3.4.32) \]
\[ q(\hat{x}_1 = 1 | x_1 = 1) = 0.9513 . \quad (3.4.33) \]

**Time 
\( t = 2 \)**

The probability distribution \( p(x_2 | \hat{x}_1) \) is

\[ p(x_2 = 1 | \hat{x}_1 = 1) = p(x_2 = 0 | \hat{x}_1 = 0) = 0.7 , \quad (3.4.34) \]
\[ p(x_2 = 1 | \hat{x}_1 = 0) = p(x_2 = 0 | \hat{x}_1 = 1) = 0.3 . \quad (3.4.35) \]

The solution is

\[ C = 0.6 = \mathcal{H}_b(x_2 | \hat{x}_1) - \mathcal{H}_b(D_2) = \mathcal{H}_b(0.3) - \mathcal{H}_b(D_2) , \quad (3.4.36) \]

or

\[ D_2 = 0.0488 . \quad (3.4.37) \]
The test channel is equal to

\[ q(\hat{x}_2 = 0|x_2 = 1 \hat{x}_1 = 1) = q(\hat{x}_2 = 1|x_2 = 0 \hat{x}_1 = 0) = 0.0487 , \quad (3.4.38) \]
\[ q(\hat{x}_2 = 1|x_2 = 0 \hat{x}_1 = 1) = q(\hat{x}_2 = 0|x_2 = 1 \hat{x}_1 = 0) = 0.105 , \quad (3.4.39) \]
\[ q(\hat{x}_2 = 0|x_2 = 0 \hat{x}_1 = 0) = q(\hat{x}_2 = 1|x_2 = 1 \hat{x}_1 = 1) = 0.9513 , \quad (3.4.40) \]
\[ q(\hat{x}_2 = 0|x_2 = 0 \hat{x}_1 = 1) = q(\hat{x}_2 = 1|x_2 = 1 \hat{x}_1 = 0) = 0.895 . \quad (3.4.41) \]

Time \( t = 3 \)

The probability distribution \( p(x_3|\hat{x}_2, \hat{x}_1) \) is

\[ p(x_3 = 0|\hat{x}_2 = 0 \hat{x}_1) = p(x_3 = 1|\hat{x}_2 = 1 \hat{x}_1) = 0.7 , \quad (3.4.42) \]
\[ p(x_3 = 0|\hat{x}_2 = 1 \hat{x}_1) = p(x_3 = 1|\hat{x}_2 = 0 \hat{x}_1) = 0.3 . \quad (3.4.43) \]

The solution is

\[ C = 0.6 = \mathcal{H}_b(x_3|x_1x_2) - \mathcal{H}_b(D_3) = \mathcal{H}_b(0.3) - \mathcal{H}_b(D_3) , \quad (3.4.44) \]

or

\[ D_3 = 0.0488 . \quad (3.4.45) \]

The test channel is equal to

\[ q(\hat{x}_3 = 1|x_3 = 0 \hat{x}_2 = 0 \hat{x}_1) = q(\hat{x}_3 = 0|x_3 = 1 \hat{x}_2 = 1 \hat{x}_1) = 0.0487 , \quad (3.4.46) \]
\[ q(\hat{x}_3 = 0|x_3 = 0 \hat{x}_2 = 0 \hat{x}_1) = q(\hat{x}_3 = 1|x_3 = 1 \hat{x}_2 = 1 \hat{x}_1) = 0.9513 , \quad (3.4.47) \]
\[ q(\hat{x}_3 = 0|x_3 = 0 \hat{x}_2 = 1 \hat{x}_1) = q(\hat{x}_3 = 1|x_3 = 1 \hat{x}_2 = 0 \hat{x}_1) = 0.895 , \quad (3.4.48) \]
\[ q(\hat{x}_3 = 1|x_3 = 0 \hat{x}_2 = 1 \hat{x}_1) = q(\hat{x}_3 = 0|x_3 = 1 \hat{x}_2 = 0 \hat{x}_1) = 0.105 . \quad (3.4.49) \]
The characteristics of the above solution are:

(i) Originally the source matches the channel (i.e., capacity is achieved for an input distribution equal to the source's initial probability distribution).

(ii) For all subsequent instants of time the decoder sees, as a result of his information, a probability distribution of the source that matches the channel. That is why the test channel for problem (3.4.19)-(3.4.28) is time invariant.

We will present a real time communication system which achieves the solution of (3.4.19)-(3.4.28). Consider the system of figure 3.5 (page ).

The encoding-decoding strategies are the following:

At time $t = 1$ we transmit the output of the source $x_1$ without any encoding and take $\hat{x}_1$ to be the output of the channel, i.e.,
\[
\hat{x}_1 = y_1.
\]

At time $t = 2$ the encoder compares $\hat{x}_1$ to $x_2$.
If $x_2 = \hat{x}_1$ then $x_2$ is transmitted from input 1 of the channel.
If $x_2 \neq \hat{x}_1$ then $x_2$ is transmitted from input 0 of the channel. $\hat{x}_2$ is taken to be equal to the output of the channel at time $t = 2$, i.e.,
\[
\hat{x}_2 = y_2.
\]

At time $t = 3$ the encoder compares $x_3$ to $\hat{x}_2$.
If $x_3 = \hat{x}_2$ then $x_3$ is transmitted from input 1 of the channel.
If $x_3 \neq \hat{x}_2$ then $x_3$ is transmitted from input 0 of the channel. 
$\hat{x}_3$ is taken to be equal to the output of the channel at time $t = 3$.

In Appendix C we show that with the above encoding-decoding strategies we achieve the solution of (3.4.19)-(3.4.28). We cannot achieve the solution of (3.4.19)-(3.4.28) with a feedforward system and encoding-decoding strategies similar to the above. The reason is that although the decoder sees the source having a probability distribution that always matches the channel, the probability of the next source output being 0 or 1 depends on the last decision of the decoder. Thus, the encoder cannot decide which channel input to use to transmit the next source output, unless he knows the last decision of the decoder. Since in a feedforward system the encoder can never know the decisions of the decoder, it is impossible to achieve the solution of (3.4.19)-(3.4.28) with a feedforward system and the simple encoding-decoding strategies described above. It may be possible to achieve the solution of (3.4.19)-(3.4.28) with a feedforward real time communication system, but the encoding-decoding strategies that achieve that solution remain unknown.

Let us compare now the performance of the noiseless feedback system we described above to the lower bound provided by the solution of the problem (3.3.1)-(3.3.4).

The average error of the system described above is

$$D_1 = D_2 = D_3 = 0.0488,$$  \hspace{1cm} (3.4.50)

and in general

$$D_t = 0.0488 \quad \forall t \geq 1.$$  \hspace{1cm} (3.4.51)
Hence
\[ D^3 \triangleq \frac{1}{3} (D_1^* + D_2^* + D_3^*) = 0.0488. \] (3.4.52)

The lower bound provided by the solution of (3.3.1)-(3.3.4) is

(\text{using Example 3.3.1}):
\[ D_1^* = 0.0488, \] (3.4.53)
\[ D_2^* = 0.0429, \] (3.4.54)
\[ D_3^* = 0.0429, \] (3.4.55)

and in general
\[ D_t^* = 0.0429 \quad \forall t > 1. \] (3.4.56)

Hence
\[ D^{*3} \triangleq \frac{1}{3} (D_1^* + D_2^* + D_3^*) = 0.0445, \] (3.4.57)

and
\[ \frac{D^{*3}}{D^3} = \frac{0.0445}{0.0488} = 91\%. \] (3.4.58)

As \( T \to \infty \) we get, because of (3.4.51) and (3.4.56),
\[ \frac{D^{*T}}{D^T} = \frac{0.0429}{0.0488} = 88\% \] (3.4.59)

The above example shows that with a noiseless feedback communication system we can achieve an average distortion very close to the lower bound provided by the solution of (3.3.1)-(3.3.4). It remains unknown how tight this bound is for feedforward real time communication systems.
Up to this point we presented a real time feedforward communication problem (P) and a real time noiseless feedback communication problem (P'). We formulated two distortion rate problems, namely (3.2.3)-(3.2.5) and (3.3.1)-(3.3.4), whose solutions provided lower bounds on (P) and (P'). We discussed the characteristics of the two real time distortion rate problems and found an example where the bound provided by the solution of (3.3.1)-(3.3.4) is tight. In the next section we will present another formulation of a real time communication problem and its corresponding real time distortion rate problem.

3.5 Another Formulation of the Real Time Feedforward Communication Problem

3.5.1 Statement of the Problem

Consider the system of figure 3.2. The real time communication problem can be formulated as follows:

\[
\begin{align*}
\text{Minimize} & & \frac{1}{T} E \sum_{t=1}^{T} \rho(x_t, \hat{x}_t), \\
\text{subject to} & & I(y_t; z_t) \leq tC \quad \forall t, \\
& & z_t = w_t(x_t) \quad \forall t, \\
& & \hat{x}_t = \mathcal{C}(y_t) \quad \forall t.
\end{align*}
\]  

(3.5.1) (Q) (3.5.2) (3.5.3) (3.5.4)

The difference between problems (P) and (Q) lies in the information constraint. (Q) always lies below (P) because (3.1.6) is a stronger constraint than (3.5.2). This is shown in the following theorem.
Theorem 3.5.1

The solution of (Q) lies below the solution of (P).

Proof:

(P) has the same solution as the problem

\[
\text{Minimize} \quad \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \rho(x_t, \hat{x}_t), \quad \{\delta_t, \gamma_t\}_{t=1}^{T}
\]

subject to

\[
\sum_{\tau=1}^{t} I(y_{\tau}; z_{\tau}) \leq tC \quad \forall t,
\]

\[
z_t = \delta_t(x_t) \quad \forall t,
\]

\[
\hat{x}_t = \gamma_t(y_t) \quad \forall t.
\]

To prove the theorem all we have to show is that

\[
I(y_t; z_t) \leq \sum_{\tau=1}^{t} I(y_{\tau}; z_{\tau}) \quad \forall t.
\]

Then, (3.5.9) will imply that the set of feasible solutions of (3.5.5)-(3.5.8) is a subset of the set of feasible solutions of (Q); therefore, (Q) will lie below (P).

To prove (3.5.9) we use basic entropy inequalities and the fact that the channel is memoryless.

By definition

\[
I(y_t; z_t) = H(y_t) - H(y_t | z_t).
\]

Since the channel is memoryless

\[
H(y_t | z_t) = \sum_{\tau=1}^{t} H(y_{\tau} | z_{\tau}) \quad \forall t.
\]
From [4]

\[ H(y_t) = \sum_{\tau=1}^{t} H(y_{\tau|y_{\tau-1}}) \leq \sum_{\tau=1}^{t} H(y_{\tau}) \quad \forall t \]  \hspace{1cm} (3.5.12)

(3.5.10)-(3.5.12) give

\[ I(y_t;z_t) \leq \sum_{\tau=1}^{t} H(y_{\tau}) - \sum_{\tau=1}^{t} H(y_{\tau|z_{\tau}}) = \]

\[ = \sum_{\tau=1}^{t} I(y_{\tau};z_{\tau}) \quad \forall t \quad \text{QED} \]

\hspace{1cm} (3.5.13)

The real time communication problem (Q) has the same characteristics as (P), i.e., it is a stochastic control problem with a nonclassical information pattern where signaling occurs. Since we cannot solve the problem itself, we will find a lower bound by formulating a real time distortion rate problem for (Q).

### 3.5.2 Formulation of a Discrete Finite State Real Time Distortion Rate Problem for (Q)

The real time distortion rate problem for (Q) can be stated as follows:

\[
\text{Minimize} \quad \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \rho(x_t, \hat{x}_t), \quad \{q(\hat{x}_t|x_t)\}_{t=1}^{T}
\]

subject to

\[ I(\hat{x}_t;x_t) \leq C \quad \forall t , \]  \hspace{1cm} (3.5.15)

\[ q(\hat{x}_t|x_t) \geq 0 \quad \forall t , \quad \forall (x_t,\hat{x}_t) , \]  \hspace{1cm} (3.5.16)

\[ \sum_{\hat{x}_t} q(\hat{x}_t|x_t) = 1 \quad \forall t , \forall x_t . \]  \hspace{1cm} (3.5.17)
(3.5.14)-(3.5.17) differs from (3.3.1)-(3.3.4) only in the information constraint. In (3.3.1)-(3.3.4), $I(\hat{x}_t; x_t | x_{t-1})$ gives the new information that is provided to the user by the transmission at time $t$. The information constraint of (3.5.14)-(3.5.17) has the following interpretation: When we decode $x_t$, we have already used the channel $t$ times and $\hat{x}_t$ is a function of all the previous outputs of the channel. Consequently, $\hat{x}_t$ cannot contain more information about the source than the amount that has been transmitted through the $t$ uses of the channel.

It is easy to show that the solution of (3.5.14)-(3.5.17) provides a lower bound on (Q).

**Theorem 3.5.2**

The solution of (3.5.14)-(3.5.17) is a lower bound on (Q).

**Proof:**

Consider $I(\hat{x}_t; x_t)$ and apply the data processing theorem

$$I(\hat{x}_t; x_t) \leq I(y_t; x_t) \leq I(y_t; z_t).$$

(3.5.18)

Because of (3.5.18), the set of feasible solutions of (3.5.14)-(3.5.17) contains the set of feasible solutions of (Q); hence, (3.5.14)-(3.5.17) provides a lower bound on (Q) \(\blacksquare\)

It is easy to solve (3.5.14)-(3.5.17) because it decomposes into a series of static optimization problems, each one of which, conditioned on the previous decisions, is similar to a classical distortion rate problem. The disadvantage of the formulation is that the lower bound becomes trivial as $t \to \infty$. Since we are dealing with finite state sources, $\hat{x}_t$ takes only a finite number of values (for each $t$), $H(\hat{x}_t)$ and

62
$H(\hat{x}_t|x_t)$ are finite for all $t$, $H(\hat{x}_t) - H(\hat{x}_t|x_t) = I(\hat{x}_t;x_t)$ is always finite; therefore,

$$\frac{1}{t} I(\hat{x}_t;x_t) \to 0 \quad \text{as} \quad t \to \infty.$$  

Consequently, the information constraint is trivially satisfied as $t \to \infty$, the optimal solution is

$$q(\hat{x}_t|x_t) = \begin{cases} 1 & \text{if } \hat{x}_t = x_t, \\ 0 & \text{if } \hat{x}_t \neq x_t, \end{cases} \quad (3.5.19)$$

and if $\rho(x_t,\hat{x}_t) = 0$ whenever $x_t = \hat{x}_t$, the lower bound on the average distortion tends to zero as $t \to \infty$.

The above result is intuitively expected for the following reason: For very large $t$, $x_t$ tends to be independent (or very weakly correlated) of $x_\tau$, for all $\tau << t$. Therefore, $\hat{x}_t$ contains no information (or very little information) about $x_\tau$. This fact, however, is not taken into account into the information constraint. On the contrary, the information constraint considers that $\hat{x}_t$ contains information about all $x_\tau$, even those for which $\tau << t$, and for each transmission this information can be no more than the channel capacity of the communication system. Thus, as $t \to \infty$ (3.5.15) is trivially satisfied and this fact leads to a trivial lower bound.

In (3.3.1)-(3.3.4) the information constraint is never trivially satisfied because at each instant of time $t$ we are concerned only with the additional information gained about $x_t$ from the transmission at time $t$. That is why (3.3.1)-(3.3.4) is a better formulation of a real time distortion rate problem than (3.5.14)-(3.5.17).
3.6 Summary

In this chapter we presented the real time feedforward communication problems (P) and (Q) and the real time noiseless feedback communication problem (P'). Problems (P) and (Q) are infinite dimensional, nonlinear, nonconvex and their solution is presently unknown. We formulated two real time distortion rate problems, namely (3.3.1)-(3.3.4) and (3.5.14)-(3.5.17) which provided lower bounds on (P), (P') and (Q), respectively. We showed through Example 3.3.1 that the lower bound of (3.3.1)-(3.3.4) is nontrivial as it lies above the solution of the classical distortion rate problem. We presented a noiseless feedback system whose performance is close to the solution of (3.3.1)-(3.3.4). Finally, we showed that the bound provided by (3.5.14)-(3.5.17) is trivial as $t \to \infty$.

In the next chapter we will extend the results of this chapter to continuous amplitude sources.
CHAPTER 4

The Real Time Distortion Rate Problem - Continuous Amplitude Sources

In this chapter we will extend the results of Chapter 3 to continuous amplitude sources.

4.1 The Real Time Communication Problem

The real time forward communication problem and the real time noiseless feedback communication problem for continuous amplitude sources can be stated in exactly the same way as problems (P) and (P') of the previous chapter respectively. Following the developments of Chapter 3, we will first formulate a real time distortion rate problem whose solution will provide a lower bound on (P) and (P'), and then we will discuss the properties of the formulation.

4.2 Formulation of a Real Time Distortion Rate Problem for Continuous Amplitude Sources

4.2.1 Statement of the problem

Consider the systems of figures 3.1 and 3.2 (p.19). Assume that the source is continuous amplitude with known initial probability density p(x(0)) and known transition probability density matrix \( p(x_t|x_{t-1}) \) for all \( t^* \), \( t=1,2,\ldots \). The channel of both systems is memoryless and has capacity \( C \). We formulate the real time distortion rate problem as follows:

*We assume that we are dealing with absolutely continuous probability distributions \( p(x(0)) \) and \( p(x_t|x_{t-1}) \) for \( t \), so that the corresponding densities exist.*
Minimize \[ J_T = \frac{1}{T} \sum_{t=1}^{T} \rho(x_t, \hat{x}_t) = \]
\[ \{q(\hat{x}_t | x_t)\}_{t=1}^{T} \]
\[ \frac{1}{T} \sum_{t=1}^{T} \int p(\hat{x}_t | x_t) q(\hat{x}_t | x_t) \rho(x_t, \hat{x}_t) \, dx_t \, d\hat{x}_t \]  \hspace{1cm} (4.2.1)

Subject to
\[ I(\hat{x}_t; x_t | x_{t-1}) = \int p(x_t) q(\hat{x}_t | x_t) \log \frac{q(\hat{x}_t | x_t)}{Q(\hat{x}_t | x_{t-1})} \, d\hat{x}_t \, dx_t \]
\[ \leq C \quad \forall t \]  \hspace{1cm} (4.2.2)
\[ \int q(\hat{x}_t | x_t) \, d\hat{x}_t = 1 \quad \forall t \]  \hspace{1cm} (4.2.3)
\[ Q(\hat{x}_t | x_{t-1}) = \int p(x_t | x_{t-1}) q(\hat{x}_t | x_t) \, dx_t. \]

(4.2.1)-(4.2.3) is analogous to (3.3.1)-(3.3.4). We briefly discuss the properties of the proposed formulation.

### 4.2.2 Properties of (4.2.1)-(4.2.3)

First we show that (4.2.1)-(4.2.3) provides a lower bound to problems (P) and (P') for continuous amplitude sources.

**Theorem 4.2.1**

The solution of (4.2.1)-(4.2.3) provides a lower bound to (P) and (P').

**Proof**

First we prove that the solution of (4.2.1)-(4.2.3) provides a lower bound to (P). Consider \( I(\hat{x}_t; x_t | x_{t-1}) \), and use the data processing Theorem and the fact that the channel is memoryless. Then

\[ I(\hat{x}_t; x_t | x_{t-1}) \leq I(y_t; x_t | x_{t-1}) = \]
\[ = \int p(x_t) p(y_t | x_t) \log \frac{p(y_t | x_t)}{p(y_t | x_{t-1})} \, dx_t \, dy_t. \]  \hspace{1cm} (4.2.4)
But
\[
p(y_t | x_t) = p(y_t | x_{t-1} x_t) \quad p(y_{t-1} | x_t) = \\
= p(y_t | x_t) \quad p(y_{t-1} | x_{t-1})
\] (4.2.5)

Because of (4.2.5), (4.2.4) gives
\[
I(\hat{x}_t ; x_t | x_{t-1}) \leq I(y_t ; x_t | x_{t-1}) = I(y_t ; x_t | x_{t-1}) \\
= H(y_t | x_{t-1}) - H(y_t | x_t) \leq H(y_t) - H(y_t | z_t, x_t) \\
= H(y_t) - H(y_t | z_t) = I(y_t ; z_t)
\] (4.2.6)

Because of (4.2.6) the set of feasible solutions of problem (P) is a subset of the set of feasible solutions of (4.2.1)-(4.2.3). Hence (4.2.1)-(4.2.3) provides a lower bound to (P).

To prove that (4.2.1)-(4.2.3) provides a lower bound to problem (P'), we consider again \( I(\hat{x}_t ; x_t | x_{t-1}) \) and apply the data processing Theorem.

\[
I(\hat{x}_t ; x_t | x_{t-1}) \leq I(y_t ; x_t | x_{t-1}) = \\
= \int p(x_t) \quad p(y_t | x_t) \quad \log \frac{p(y_t | x_t)}{p(y_t | x_{t-1})} \quad dx_t \quad dy_t
\] (4.2.7)

But
\[
p(y_t | x_t) = p(y_t | x_t, y_{t-1}) \quad p(y_{t-1} | x_{t-1}) \\
p(y_t | x_{t-1}) = p(y_t | x_{t-1}, y_{t-1}) \quad p(y_{t-1} | x_{t-1})
\] (4.2.8)

Because of (4.2.8), (4.2.7) gives
\[
I(\hat{x}_t ; x_t | x_{t-1}) \leq I(y_t ; x_t | y_{t-1} x_{t-1}) \leq \\
\leq H(y_t) - H(y_t | z_t, x_t) = H(y_t) - H(y_t | z_t) \\
= I(y_t ; z_t)
\] (4.2.8a)
Because of (4.2.8a) the set of feasible solutions of problem (P') is a subset of the set of feasible solutions of (4.2.1)-(4.2.3). Hence the solution of (4.2.1)-(4.2.3) provides a lower bound to problem (P').

The properties of the proposed formulation are very similar to those of the real time distortion rate problem (3.3.1)-(3.3.4) for finite state sources. (4.2.1)-(4.2.3) decomposes into a series of static optimization problems each one of which is of the form

\[
\begin{align*}
\text{Minimize} & \quad E \rho(x_t, \hat{x}_t) = \\
& \int p(x_t) q(\hat{x}_t | x_t) \rho(x_t, \hat{x}_t) \; dx_t \; d\hat{x}_t
\end{align*}
\]  

subject to

\[
I(\hat{x}_t; x_t | x_{t-1}) = \int p(x_t) q(\hat{x}_t | x_t) \log \frac{q(\hat{x}_t | x_t)}{Q(\hat{x}_t | x_{t-1})} \; dx_t \; d\hat{x}_t 
\]

\[
\leq C
\]

\[
\int q(\hat{x}_t | x_t) \; d\hat{x}_t = 1.
\]

(4.2.9)

(4.2.10)

(4.2.11)

The corresponding rate distortion form of (4.2.9)-(4.2.11) is

\[
\begin{align*}
\text{Minimize} & \quad I(q(\hat{x}_t | x_t)) + I(\hat{x}_t; x_t | x_{t-1}) \\
\text{subject to} & \quad E \rho(x_t, \hat{x}_t) \leq D(C) \\
& \int q(\hat{x}_t | x_t) \; d\hat{x}_t = 1.
\end{align*}
\]

(4.2.12)

(4.2.13)

(4.2.14)

(4.2.9)-(4.2.11) are infinite dimensional convex programming problems similar to the classical distortion rate problem and the classical rate distortion problems for a source with continuous amplitude. Results analogous to equations (3.3.33)-(3.3.37) and to Theorems 3.3.3-3.3.5, 3.3.7 can be easily derived, in the same way as in section 3.3, from the variational problem corresponding to (4.2.12)-(4.2.14). They are the following:
The rate distortion curve has the parametric representation

\[ D_t = \int_{x_{t-1}} p(x_{t-1}) \, dx_{t-1} \int_{x_t, \hat{x}_t} p(x_t | x_{t-1}) \, \lambda_{x_t | x_{t-1}} \cdot \]

\[ \cdot Q(\hat{x}_t | x_{t-1}) \, e^{s \rho(x_t, \hat{x}_t)} \, \rho(x_t, \hat{x}_t) \, dx_t \, d\hat{x}_t \]

(4.2.15)

\[ R = \delta D_t + \int_{x_{t-1}} dx_{t-1} p(x_{t-1}) \int_{x_t, \hat{x}_t} dx_t \, d\hat{x}_t \, p(x_t | x_{t-1}) \log \lambda_{x_t | x_t} \]

(4.2.16)

where

\[ \lambda_{x_t | x_{t-1}} = (\int d\hat{x}_t \, Q(\hat{x}_t | x_{t-1}) \, e^{s \rho(x_t, \hat{x}_t)}) \]

(4.2.17)

and \( \delta \) is the Lagrange multiplier defined in Section 3.3. The continuous amplitude analogs of Theorems 3.3.3-3.3.5 and 3.3.7 are:

**Theorem 4.2.2**

The parameter \( s \) represents the slope of the rate distortion function at the point \( D_s, R_s \), that is parametrically generated by (4.2.15) and (4.2.16)

**Theorem 4.2.3**

A transition probability density \( q(\hat{x}_t | x_t) \) yields a point on \( R(D) \) if there exists a probability density \( Q(\hat{x}_t | x_{t-1}) \), a value \( s \leq 0 \), and a subset \( V \) of the real time such that

(i) \( q(\hat{x}_t | x_t) = \lambda_{x_t | x_{t-1}} Q(\hat{x}_t | x_{t-1}) \, e^{s \rho(x_t, \hat{x}_t)} \)

(4.2.18)

(ii) \( Q(\hat{x}_t | x_{t-1}) > 0 \) for \( \hat{x}_t \in V \), \( Q(\hat{x}_t | x_{t-1}) = 0 \) for \( \hat{x}_t \notin V \)

(4.2.19)

(iii) \( \int d\hat{x}_t \lambda_{x_t | x_{t-1}} p(x_t | x_{t-1}) \, e^{s \rho(x_t, \hat{x}_t)} = 1 \)

for \( \hat{x}_t \in V \) and
\[
\int \text{d}x_t \lambda_{x_t \mid x_{t-1}} p(x_t \mid x_{t-1}) e^{sp(x_t, \hat{x}_t)} \leq 1
\]  
(4.2.21)

for \( \hat{x}_t \notin V \)

Theorem 4.2.4

Define \( \Lambda_s \) to be the set of all nonnegative functions \( \lambda_{x_t \mid x_{t-1}} \) that satisfy

\[
\int \text{d}x_t \lambda_{x_t \mid x_{t-1}} p(x_t \mid x_{t-1}) e^{sp(x_t, \hat{x}_t)} \leq 1
\]  
(4.2.22)

for all \( \hat{x}_t \). Then

\[
R(D) = \sup_{s \leq 0, \lambda_{x_t \mid x_{t-1}} \in \Lambda_s} \log \lambda_{x_t \mid x_{t-1}}
\]  
(4.2.23)

For every \( s \leq 0 \), \( \lambda_{x_t \mid x_{t-1}} \) realizes the supremum in (4.2.23) iff there exists a probability density \( Q(\hat{x}_t \mid x_{t-1}) \), (related to \( \lambda_{x_t \mid x_{t-1}} \) through (4.2.17), such that

\[
\int \text{d}x_t p(x_t \mid x_{t-1}) \lambda_{x_t \mid x_{t-1}} e^{sp(x_t, \hat{x}_t)} = 1
\]  
(4.2.24)

for almost all \( \hat{x}_t \) for which \( Q(\hat{x}_t \mid x_{t-1}) > 0 \).

Theorem 4.2.5

The slope of \( R(D) \) is continuous in the interval \( 0 < D < D_{\text{max}} \) and tends to \(-\infty\) as \( D \to 0 \). A discontinuity of \( dR(D) \) can occur only at \( D = D_{\text{max}} \).

Only Theorem 3.3.6 concerning the continuity of the rate distortion function \( R(D) \) at \( D = 0 \) does not carry over to the continuous case. For sources of continuous amplitude in most cases \( R(D) \to \infty \) as \( D \to 0 \).

In order to show that the lower bound provided by the solution of (4.2.1)-(4.2.3) is attained, we must either prove a real time coding Theorem.
or present an example of a system which attains the bound. Before we present a system which attains the bound of (4.2.1)-(4.2.3), we will study and interesting class of problems that arise when $\rho(\cdot)$ is a difference distortion measure.

4.2.3 Difference distortion measures - Shannon-type lower bounds

The most commonly studied distortion measures for sources with continuous amplitude are the difference distortion measures, i.e., measures that depend on the difference between $x_t$ and $\hat{x}_t$ for all $t$. In this case theorem 4.2.4 can be used to obtain a lower bound to the solution of (4.2.1)-(4.2.3). The procedure to obtain the bound is the same as that followed in [3], (section 4.3.1), and we briefly describe it here for our formulation.

According to Theorem 4.2.4 we have

$$R(D_t) = \sup_{s \leq 0, \lambda_{x_t|x_{t-1}} \in \Lambda_s} \left[ sD_t + \int dx_{t-1} \ p(x_{t-1}) \int dx_t \ p(x_t|x_{t-1}) \right]$$

$$\log \lambda_{x_t|x_{t-1}}].$$

(4.2.25)

Setting

$$\delta_{x_t|x_{t-1}} = \lambda_{x_t|x_{t-1}} \ p(x_t|x_{t-1})$$

(4.2.26)

we obtain

$$R(D_t) = \sup_{s \leq 0, \lambda_{x_t|x_{t-1}} \in \Lambda_s} \left[ sD_t + h((x_t|x_{t-1})) + \int dx_{t-1} \ p(x_{t-1}) \int dx_t \ p(x_t|x_{t-1}) \log \delta_{x_t|x_{t-1}} \right]$$

(4.2.27)

Take $s \leq 0$, set

$$\lambda_{x_t|x_{t-1}} = \frac{C}{p(x_t|x_{t-1})}$$

(4.2.28)
and assume that \( \rho \) is such that

\[
\int dx_t e^{s\rho(x_t - \hat{x}_t)} < \infty, \forall s < 0
\]  

(4.2.29)

Because of (4.2.28), (4.2.22) gives

\[
c \int dx_t e^{s\rho(x_t - \hat{x}_t)} = c \int dw_t e^{s\rho(w_t)} < 1
\]  

(4.2.30)

and because of (4.2.29) \( c \) can be chosen so that equality holds in (4.2.30).

Then, Theorem 4.2.4 and (4.2.27) imply that for all \( s < 0 \)

\[
R(D_t) \geq h((x_{-}\mid x_{t-1})) + sD_t - \log \int dw_t e^{s\rho(w_t)} \overset{\Delta}{=} R_L(D_t, s)
\]

(4.2.31)

We now have to maximize \( R_L(D_t, s) \) with respect to \( s \) in order to get the lower bound for \( R(D_t) \). \( R_L(D_t, s) \) is concave as a function of \( s \) because

\[
\frac{d^2 R_L(D_t, s)}{ds^2} = - \int dx_t \rho^2(x_t) \Pi_s(x_t) +
\]

\[
+ (\int dx_t \rho(x_t) \Pi_s(x_t))^2 =
\]

\[
= - \text{variance} \left( \rho_s(x_t) \right) \leq 0
\]  

(4.2.32)

where

\[
\Pi_s(x_t) = \frac{e^{s\rho(x_t)}}{\int dw_t e^{s\rho(w_t)}}
\]  

(4.2.33)

Therefore, for fixed \( D_t \), \( R_L(D_t, s) \) has a unique maximum which is attained for the value of \( s \) for which

\[
\frac{dR_L(D_t, s)}{ds} = D_t - \int dx_t \rho(x_t) \Pi_s(x_t) = 0
\]  

(4.2.34)

The maximizing value of \( s \) is given by the solution of the equation

\[
\int dx_t \Pi_s(x_t) \rho(x_t) = D_t
\]  

(4.2.35)

72
Using (4.2.35) and (4.2.33) in (4.2.31) we obtain

\[
R_L(D_t, s) = h(x_t | x_{t-1}) + s \int dx_t \rho(x_t) \Pi_s(x_t) - \\
- \int \Pi_s(x_t) dx_t \log \int dw_t e^{s\rho(w_t)} - \\
= h(x_t | x_{t-1}) + \int dx_t \Pi_s(x_t) [s\rho(x_t) - \log \int dw_t e^{s\rho(w_t)}] \\
= h(x_t | x_{t-1}) + \int dx_t \Pi_s(x_t) \log \Pi_s(x_t) = \\
= h(x_t | x_{t-1}) - h(\Pi_s(x_t))
\]

(4.2.36)

Hence

\[
R(D_t) \geq h((x_t | x_{t-1})) - h(\Pi_s(x_t)) \triangleq R_L(D_t).
\]

(4.2.37)

(4.2.37) provides a Shannon-type lower bound on the rate distortion function resulting from our formulation.

In the case of magnitude error criterion, (i.e., \(\rho(x_t, \hat{x}_t) = (x_t - \hat{x}_t)\)), (4.2.37) gives ([3], section 4.3.2),

\[
R_L(D_t) = h((x_t | x_{t-1})) - \ln(2eD_t)
\]

(4.2.38)

In the case of the squared error criterion (4.2.37) gives ([3], section 4.3.2),

\[
R_L(D_t) = h(x_t | x_{t-1}) - \frac{1}{2} \ln(2\pi eD_t)
\]

(4.2.39)

If \(p(x_t | x_{t-1})\) is Gaussian with arbitrary mean and variance \(V_t^2\) then (4.2.39) gives

\[
R_L(D_t) = \begin{cases} 
\frac{1}{2} \ln \frac{V_t^2}{D_t}, & 0 \leq D_t \leq V_t^2 \\
0, & D_t > V_t^2
\end{cases}
\]

(4.2.40)

and moreover

\[
R_L(D_t) = R(D_t)
\]

(4.2.41)
over the entire range $0 \leq D_t \leq V_t^2$ ([3], section 4.3.3).

Based on the results of rate distortion problems with difference distortion measures, we will present a feedforward system which attains the bound of (4.2.1)-(4.2.3). This system uses unavailable information. Hence the bound should be loose for feedforward systems that satisfy the information constraints.

4.2.4 Example 4.2.1

Consider the system of figure 3.3 (p.39). Assume that the source is Gaussian described by the equation

$$x_{t+1} = x_t + \eta_t$$

and that

(i) $x_1 \sim N(0,\sigma_1^2)$

(ii) $\eta_t \sim N(0,\sigma_t^2)$

(iii) $\sigma_1^2 = \sigma_t^2$ \forall t

(iv) the outputs of the source are transmitted through a memoryless additive Gaussian noise channel of capacity C; the variance of the additive noise in the channel is

$$V = \sigma_1^2 e^{-2c} (1-e^{-2c})$$

We consider the three shot problems, (i.e., $T=3$). According to (4.2.1)-(4.2.3) we have to solve the following problem

$$\text{Minimize } J_3 \triangleq \frac{1}{3} E \left\{ (x_1 - \hat{x}_1)^2 + (x_2 - \hat{x}_2)^2 + (x_3 - \hat{x}_3)^2 \right\}$$

$$\{q(\hat{x}_1|x_1), q(\hat{x}_2|x_1x_2), q(\hat{x}_3|x_1x_2x_3)\}$$
Subject to

\[ I(\hat{x}_1; x_1) \leq C \]  
\[ I(\hat{x}_2; x_2 | x_1) \leq C \]  
\[ I(\hat{x}_3; x_3 | x_2 x_1) \leq C \]  
\[ \int q(\hat{x}_1 | x_1) d\hat{x}_1 = 1 \]  
\[ \int q(\hat{x}_2 | x_1 x_2) d\hat{x}_2 = 1 \]  
\[ \int q(\hat{x}_3 | x_1 x_2 x_3) d\hat{x}_3 = 1 \]  

The details of the solution are given in appendix D. Here we give the results.

**Time t=1**

\[
C = R(D_1) = \begin{cases} 
\frac{1}{2} \ln \frac{\sigma_1^2}{D_1}, & 0 \leq D_1 \leq \frac{\sigma_1^2}{1} \\
0, & D_1 \geq \frac{\sigma_1^2}{1} 
\end{cases}
\]

or

\[
D_1 = \sigma_1^2 e^{-2C} \]

\[
q^*(\hat{x}_1 | x_1) = \frac{1}{\sqrt{2\pi D_1}} \exp \left\{ -\frac{(\hat{x}_1 - B_1 x_1)^2}{2B_1 D_1} \right\}
\]

where

\[
B_1 = 1 - \frac{D_1}{\sigma_1^2}
\]

**Time t=2**

\[
C = R(D_2) = \begin{cases} 
\frac{1}{2} \ln \frac{N_1^2}{D_2}, & 0 \leq D_2 \leq \frac{N_1^2}{1} \\
0, & D_2 \geq \frac{N_1^2}{1} 
\end{cases}
\]

or

\[
D_2 = \frac{N_1^2}{1} e^{-2C}
\]
\[ q^* (\hat{x}_2 | x_1, x_2) = \frac{1}{\sqrt{2\pi B_2 D_2}} \exp \left\{ - \frac{[\hat{x}_2 - B_2 (x_2 - x_1)]^2}{2B_2 D_2} \right\} \]  \hspace{1cm} (4.2.60)

where

\[ B_2 = 1 - \frac{D_2}{N_1^2} \]  \hspace{1cm} (4.2.61)

**Time t=3**

\[ C = R(D_3) = \begin{cases} \frac{1}{2} \ln \frac{N_2^2}{D_3}, & 0 \leq D_3 \leq N_2^2 \\ 0, & D_3 > N_2^2 \end{cases} \]  \hspace{1cm} (4.2.62)

or

\[ D_3 = N_2^2 e^{-2C} \]  \hspace{1cm} (4.2.63)

\[ q^* (\hat{x}_3 | x_1, x_2, x_3) = \frac{1}{\sqrt{2\pi B_3 D_3}} \exp \left\{ - \frac{[\hat{x}_3 - B_3 (x_3 - x_2)]^2}{2B_3 D_3} \right\} \]  \hspace{1cm} (4.2.64)

where

\[ B_3 = 1 - \frac{D_3}{N_2^2} \]  \hspace{1cm} (4.2.65)

Notice that since \( \sigma_1^2 = N_2^2 \) for all \( t \),

\[ D_1 = D_2 = D_3 \]  \hspace{1cm} (4.2.66)

and

\[ B_1 = B_2 = B_3 \]  \hspace{1cm} (4.2.67)

Notice also that at time \( t=1 \) the probability density of \( x_1 \) scaled by \( B_1 \) matches the channel.

Each source output is available to the decoder with one unit of delay.

Consider the following encoding-decoding strategies:
At time \( t=1 \) the output \( x_1 \) is scaled by \( B_1 = 1 - e^{-2C} \) and transmitted through the channel. The output of the channel is taken to be the estimate \( \hat{x}_1 \) of \( x_1 \). At all subsequent times \( t \), the innovations of the source, expressed by \( n_{t-1} \), are scaled by \( B_t = 1 - e^{-2C} \) and transmitted through the channel. The output \( \hat{n}_{t-1} \) of the channel is combined with \( x_{t-1} \) to give
\[
\hat{x}_t = x_{t-1} + \hat{n}_{t-1}.
\] (4.2.68)

We show that under the above encoding-decoding strategies the system of figure 3.3 achieves the lower bound of (4.2.1)-(4.2.3).

Since at time \( t=1 \) the input distribution scaled by \( B_1 \) matches the channel, we will get
\[
I(\hat{x}_1; x_1) = C
\] (4.2.69)
and
\[
E(x_1 - \hat{x}_1)^2 = D_1
\] (4.2.70)
where \( D_1 \) is given by (4.2.55). At time \( t=2 \) the mutual information \( I(\hat{x}_2; x_2 | x_1) \) is
\[
I(\hat{x}_2; x_2 | x_1) = H(x_1 + \hat{n}_1 | x_1) - H(x_1 + \hat{n}_1 | x_1 + n_1, x_1) = H(\hat{n}_1) - H(\hat{n}_1 | n_1) =
\]
\[
= I(\hat{n}_1; n_1)
\] (4.2.71)

The average error is
\[
E(x_2 - \hat{x}_2)^2 = E(x_1 + n_1 - x_1 - \hat{n}_1)^2 = E(n_1 - \hat{n}_1)^2
\] (4.2.72)
Since the probability density of \( n_1 \) scaled by \( B_2 \) matches the channel, (4.2.71) and (4.2.72) give
\[
I(\hat{x}_2; x_2 | x_1) = I(\hat{n}_1; n_1) = C
\] (4.2.73)
(hence the encoding-decoding strategy at t=2 is admissible), and
\[ E(x_2-x_2)^2 = E(n_1-n_1)^2 = D_2 \quad (4.2.74) \]
where \( D_2 \) is given by (4.2.59). At time t=3 the mutual information
\[ I(\hat{x}_3;x_3|x_1x_2) \]
is
\[ I(\hat{x}_3;x_3|x_1x_2) = H(\hat{x}_3|x_1x_2) - H(\hat{x}_3|x_1x_2x_3) = \]
\[ = H(x_2+n_2|x_1x_2) - H(x_2+n_2|x_1x_2x_2+n_2) \]
\[ = H(n_2) - H(n_2|x_2) = I(n_2;n_2) \quad (4.2.75) \]
The average error is
\[ E(x_3-x_3)^2 = E(x_2+n_2-x_2-n_2)^2 = E(n_2-n_2)^2 \quad (4.2.76) \]
Since the probability density of \( n_2 \) scaled by \( \beta_3 \) matches the channel, (4.2.75) and (4.2.76) give
\[ I(\hat{x}_3;x_3|x_1x_2) = I(n_2;n_2) = C \quad (4.2.77) \]
(therefore the encoding-decoding strategy at t=3 is admissible), and
\[ E(x_3-x_3)^2 = E(n_2-n_2)^2 = D_3 \quad (4.2.78) \]
where \( D_3 \) is given by (4.2.63). Thus, the system of figure 3.3 with source channel and encoding-decoding strategies described above, achieve the lower bound of (4.2.1)-(4.2.3).

As discussed in Chapter 3, the system of figure 3.3 has only theoretical value. We cannot implement such a system. It remains to investigate whether we can find a practical system whose performance approaches the bound of (4.2.1)-(4.2.3). This will be the topic of the next section.

4.2.5 Investigation of the tightness of the lower bound of (4.2.1)-(4.2.3).

As in the case of finite state sources, we will present communication problems where it is possible that the decoder sees a probability
distribution of the source that matches the channel for all instants of
time t. For these problems we will formulate a real time distortion rate
problem, we will realize its solution by a real time communication system
and compare the performance of that system to the lower bound of (4.2.1)-
(4.2.3).

The two special communication problems we consider are the same as
problems (R) and (R') of the previous chapter. The real time distortion
rate problem for (R), (R') and continuous amplitude sources can be stated
as follows:

\[
\begin{align*}
\min_{q(\hat{x}_T|\hat{x}_{T-1})} \quad & \frac{1}{T} \quad \mathbb{E} \left\{ \rho(x_T, \hat{x}_T) + \min_{q(\hat{x}_{T-1}|x_{T-1}:\hat{x}_{T-2})} \left\{ \mathbb{E} \rho(x_{T-1}, \hat{x}_{T-1}) \right\} \right. \\
& \quad + \cdots + \left. \min_{q(\hat{x}_1|x_1)} \mathbb{E} \rho(x_1, \hat{x}_1) \right\} \\
\text{Subject to} \\
& I(\hat{x}_T; x_T | \hat{x}_{T-1}) \leq C \quad \forall t \\
& \int q(\hat{x}_T|x_T:\hat{x}_{T-1}) \, d\hat{x}_T = 1 \quad \forall t
\end{align*}
\] (4.2.79)

(4.2.79)-(4.2.81) is the continuous amplitude analog of (3.4.9)-(3.4.12).
It provides a lower bound to problems (R) and (R') and has the same prop-
erties and characteristics as (3.4.9)-(3.4.12). Following the procedure of
section 4.2.3 we can easily show that a Shannong-type lower bound to the
solution of (4.2.79)-(4.2.81) is

\[
R_L(D_t') = h(x_T|\hat{x}_{T-1}) - h(\pi_s(x_T))
\] (4.2.82)

where \( \pi_s(x_T) \) is given by (4.2.33).
In the case of the magnitude error criterion and the squared error criterion (4.2.82) gives

\[ R_L(D'_t) = h(x_t | \hat{x}_{t-1}) - \ln(2\pi D'_t) \] (4.2.83)

and

\[ R_L(D'_t) = h(x_t | \hat{x}_{t-1}) - \frac{1}{2} \ln(2\pi D'_t) \] (4.2.84)

respectively. When \( p(x_t | \hat{x}_{t-1}) \) is Gaussian with arbitrary mean and variance \( \Sigma^2_t \), (4.2.84) gives

\[
R_L(D'_t) = \begin{cases} 
\frac{1}{2} \ln \frac{\Sigma^2_t}{D'_t}, & 0 \leq D'_t \leq \Sigma^2_t \\
0, & D'_t > \Sigma^2_t 
\end{cases} 
\] (4.2.85)

and moreover

\[ R_L(D'_t) = R(D'_t) \] (4.2.86)

over the entire range \( 0 \leq D'_t \leq \Sigma^2_t \). We will use (4.2.85)-(4.2.86) to present an example that achieves the solution of (4.2.79)-(4.2.81) and we will compare the performance of that system to the lower bound of (4.2.1)-(4.2.3).

**Example 4.2.2**

Consider a Gaussian source described by the equation

\[ x_{t+1} = x_t + v_t \] (4.2.87)

Assume that

\begin{align*}
(i) & \quad x_1 \sim N(0, \sigma^2) \quad (4.2.88) \\
(ii) & \quad v_t \sim N(0, v^2) \quad t \geq 1 \\
& \quad v^2 = \sigma^2 (1-e^{-2\epsilon}) \quad (4.2.89)
\end{align*}

\begin{align*}
(iii) & \quad \text{The outputs of the source are transmitted through a memoryless additive Gaussian noise channel of capacity } C. \text{ The variance of the additive Gaussian noise in the channel is}
\end{align*}
\[ \theta^2 = \frac{v^2}{\sigma^2} (\sigma^2 - v^2) = \sigma^2 e^{-2c} (1 - e^{-2c}) \] (4.2.91)

(iv) The distortion measure is the squared error criterion.

We consider the three shot problem. According to (4.2.79)-(4.2.81) we have to solve the problem

\[
\min \frac{1}{3} \mathbb{E} \{ (x_3 - \hat{x}_3)^2 + \min_{q(\hat{x}_2|x_2)} \mathbb{E} \{ (x_2 - \hat{x}_2)^2 + \min_{q(\hat{x}_1|x_1)} \mathbb{E} (x_1 - \hat{x}_1)^2 \} \} \tag{4.2.92}
\]

Subject to

\[ I(\hat{x}_1; x_1) \leq C \] (4.2.93)
\[ I(\hat{x}_2; x_2 | \hat{x}_1) \leq C \] (4.2.94)
\[ l(\hat{x}_3; x_3 | \hat{x}_1 \hat{x}_2) \leq C \] (4.2.95)
\[ \int q(\hat{x}_1|x_1) \, d\hat{x}_1 = 1 \] (4.2.96)
\[ \int q(\hat{x}_2|x_2 \hat{x}_1) \, d\hat{x}_2 = 1 \] (4.2.97)
\[ \int q(\hat{x}_3|x_3 \hat{x}_1 \hat{x}_2) \, d\hat{x}_3 = 1 \] (4.2.98)

The details of the solution are given in Appendix E. Here we give the results.

Time \( t = 1 \)

The minimum is given by

\[ C = R(D_1') = \begin{cases} \frac{1}{2} \ln \frac{\sigma^2}{D_1'}, & 0 \leq D_1' \leq \sigma^2 \\ 0, & D_1' > \sigma^2 \end{cases} \] (4.2.99)

or

81
\[ D_1' = \sigma^2 e^{-2c} \]  \hspace{1cm} (4.2.100)

The test channel is
\[ q(\hat{x}_1|x_1) = \frac{1}{\sqrt{2\pi B_1 D_1'}} \exp \left\{ -\frac{(\hat{x}_1-B_1 x_1)^2}{2B_1 D_1'} \right\} \]  \hspace{1cm} (4.2.101)

where
\[ B_1 = 1 - \frac{D_1'}{\sigma^2} = 1-e^{-2c} \]  \hspace{1cm} (4.2.102)

**Time t=2**

The probability density of the source, as seen by the decoder is
\[ p(x_2|\hat{x}_1) = \frac{1}{\sqrt{2\pi \sigma}} \exp \left\{ -\frac{(\hat{x}_1-x_1)^2}{2\sigma^2} \right\} \]  \hspace{1cm} (4.2.103)

The minimum average distortion \( D_2' \) is given by
\[ C = R(D_2') = \begin{cases} \frac{1}{2} \ln \frac{\sigma^2}{D_2'}, & 0 \leq D_2' \leq \sigma^2 \\ 0, & D_2' > \sigma^2 \end{cases} \]  \hspace{1cm} (4.2.104)

or
\[ D_2' = \sigma^2 e^{-2c} = D_1'. \]  \hspace{1cm} (4.2.105)

The test channel is
\[ q^*(\hat{x}_2|x_2 \hat{x}_1) = \frac{1}{\sqrt{2\pi B_2 D_2'}} \exp \left\{ -\frac{[\hat{x}_2-\beta(x_2-\hat{x}_1)]^2}{2B_2 D_2'} \right\} \]  \hspace{1cm} (4.2.106)

where
\[ B_2 = 1 - \frac{D_2'}{\sigma^2} = 1-e^{2c} = B_1 \]  \hspace{1cm} (4.2.107)

**Time t=3**

The probability density of the source, as seen by the decoder as a result of his information, is
\[ p(x_3 | \hat{x}_1 \hat{x}_2) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ \frac{[x_3 - (\hat{x}_1 + \hat{x}_2)]^2}{2\sigma^2} \right\} \quad (4.2.108) \]

The minimum average distortion \( D'_3 \) is given by
\[
C = R(D'_3) = \begin{cases} 
\frac{1}{2} \ln \frac{D'_3}{\sigma^2}, & 0 \leq D'_3 \leq \sigma^2 \\
0, & D'_3 > \sigma^2 
\end{cases} \quad (4.2.109)
\]

or
\[ D'_3 = \sigma^2 e^{-2c} = D'_2 = D'_1. \quad (4.2.110) \]

The test channel is
\[ q(\hat{x}_3 | x_3 \hat{x}_1 \hat{x}_2) = \frac{1}{\sqrt{2\pi B_3 D'_3}} \exp \left\{ \frac{[\hat{x}_3 - x_3 - \hat{x}_1 - \hat{x}_2]^2}{2B_3 D'_3} \right\} \quad (4.2.111) \]

where
\[ B_3 = 1 - \frac{D'_3}{\sigma^2} = 1 - e^{-2c} = B_2 = B_1 \quad (4.2.112) \]

Hence
\[ D'_1(3) = \frac{1}{3} (D'_1 + D'_2 + D'_3) = \sigma^2 e^{-2c}. \quad (4.2.113) \]

The characteristics of the above solution are:

(i) Originally the probability density of the source scaled by \( B_1 \) matches the channel (i.e., capacity is achieved for an input probability density equal to the source's initial probability density scaled by \( B_1 \)).

(ii) For subsequent instants of time \( t \), the decoder sees, as a result of his information, a probability density of the source which, scaled by \( B_t \), matches the channel.

We will present a real time communication system which achieves the solution of \( (4.2.92)-(4.2.98) \).
Consider the system of figure 4.1 (page 85). At each instant of time t the output of the channel is fed back into the encoder. The encoder forms the sum \( \sum_{i=1}^{t-1} \hat{x}_i \), subtracts it from \( x_t \) and scales the result by a factor \( B = 1 - e^{-2c} \). The decoder output is equal to the output of the channel.

In Appendix F we show that the system of figure 4.1 with the encoding-decoding strategies described above achieves the solution of (4.2.92)-(4.2.98). The feedforward system of figure 4.2 (page 85), gives

\[
\hat{x}_{t_{\text{feedforward}}} = \hat{x}_{t_{\text{feedback}}}
\]

for all t, but it requires a continuously growing channel capacity. The capacity of an additive noise Gaussian channel is given by ([4], page 337).

\[
C = \frac{1}{2} \log \left( 1 + \frac{\text{E}(Bx_t)^2}{\sigma^2} \right)
\]

(4.2.114)

where \( Bx_t \) is the input to the channel and \( \sigma^2 \) is the variance of the additive noise in the channel. Since in our example \( \text{E}(x_t^2) \) grows with time and \( \text{E}(x_t^2) \to \infty \) as \( t \to \infty \), we would need a continuously growing channel capacity for the system of figure 4.2 in order to realize the solution.

We compare now the performance of the system of figure 4.1 to the lower bound of (4.2.1)-(4.2.3). The solution of (4.2.1)-(4.2.3) for the source of this example and the three shot problem gives: (according to example 4.2.1)

Time \( t = 1 \)

\[
R(D_1) = \begin{cases} 
\frac{1}{2} \log \frac{\sigma^2}{D_1}, & 0 \leq D_1 \leq \sigma^2 \\
0, & D_1 > \sigma^2
\end{cases}
\]

(4.2.115)

or

\[
D_1 = \sigma^2 e^{-2c}
\]

(4.2.116)
Figure 4.1

Figure 4.2
Time $t=2$

\[
R(D_2) = \begin{cases} 
\frac{1}{2} \ln \frac{V^2}{D_2}, & 0 \leq D_2 \leq V^2 \\
0, & D_2 > V^2
\end{cases}
\]  

(4.2.117)

or

\[
D_2 = V^2 e^{-2c}
\]  

(4.2.118)

Time $t=3$

\[
R(D_3) = \begin{cases} 
\frac{1}{2} \ln \frac{V^2}{D_3}, & 0 \leq D_3 \leq V^2 \\
0, & D_3 > V^2
\end{cases}
\]  

(4.2.119)

or

\[
D_3 = V^2 e^{-2c}
\]  

(4.2.120)

Hence

\[
D^{(3)} = \frac{1}{3} (D_1 + D_2 + D_3) = \frac{1}{3} \sigma^2 + \frac{2}{3} V^2 e^{-2c}
\]  

(4.2.121)

(4.2.113) and (4.2.121) give

\[
\frac{D^{(3)}}{D'_{(3)}} = \frac{\sigma^2 + 2V^2}{3\sigma^2}
\]  

(4.2.122)

If $V^2$ is not much smaller than $\sigma^2$ then the performance of the system of figure 4.1 with the encoding-decoding strategy described above approaches the lower bound of (4.2.1)-(4.2.3). If for example

$\sigma^2 = 1$, $V^2 = 0.8$

then

\[
\frac{D^{(3)}}{D'_{(3)}} = \frac{2.6}{3} = 86.6%.
\]

Example 4.2.2 shows that with a noiseless feedback communication system we can achieve an average distortion very close to the lower bound provided by the solution of (4.2.1)-(4.2.3).
4.3 Summary

In this chapter we extended the results of Chapter 3 to continuous amplitude sources. We formulated a real time distortion rate problem, namely (4.2.1)-(4.2.3), which provides a lower bound to the real time communication problems (P) and (P'). (4.2.1)-(4.2.3) is the continuous amplitude analog of (3.3.1)-(3.3.4) and has properties similar to those of (3.3.1)-(3.3.4). We developed Shannon-type lower bounds for (4.2.1)-(4.2.3) with a difference distortion measure and using the Shannon-type lower bounds we presented a real time noiseless feedback communication system whose performance approaches the lower bound of (4.2.1)-(4.2.3). In the next chapter we will discuss a special case of the real time communication problem that arises when we fix the encoder. This is the filtering problem.
CHAPTER 5

The Filtering Problem

In this chapter our goal will be to formulate decentralized filtering problems and provide lower bounds to these problems. In order to do this, first we will rederive Galdos' lower bound to the classical nonlinear filtering problem using an information theoretic approach different than Galdos'. Then we will formulate a decentralized filtering problem (DN), and using our approach we will present a real time distortion rate problem which provides a lower bound to (DN). In the case of linear Gaussian sources, linear observations and a squared error criterion, this bound lies above the solution of the corresponding centralized filtering problem. Finally, we will present a decentralized filtering problem (L) involving delays and we will derive a lower bound to (L) by solving a linear filtering problem.

5.1 The Nonlinear Filtering Problem

Consider the system of Figure 5.1. Assume that the source has known initial probability density \( p(x(0)) \) and known transition probability \( p(x_t | x_{t-1}) \) for all \( t, t=1,2,...,T \). The sensor is known and fixed and is described by the observation equation; the observation equation can be, for example

\[
y_t = g_t(x_t) + v_t
\]

(5.1.1)

where \( v_t \) is noise whose statistics are known. The nonlinear filtering problem can be stated as follows:
Figure 5.1
Minimize \[
\frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \rho(x_t, \hat{x}_t)
\]
subject to
\[
\hat{x}_t = f_t(y_t)
\]
\[
I(y_t; x_t | y_{t-1}) = C_t.
\]

In (5.1.2) \(\rho(\cdot)\) is a distortion measure. \(C_t\) is determined by the observation equation. We will use (5.1.4) instead of the observation equation because we will present a rate distortion theoretic approach to the nonlinear filtering problem. (Note that there may be many observation equations of the form (5.1.1) that could satisfy (5.1.3).)

We will formulate a real time distortion rate problem for (N) and discuss its characteristics.

5.2 Formulation of a Real Time Distortion Rate Problem for (N)

5.2.1 Statement of the problem

The real time distortion rate problem for (N) can be stated as follows:

Minimize \[
J_T = \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \rho(x_t, \hat{x}_t) = 
\{q(\hat{x}_t | x_t, y_{t-1})\}_{t=1}^{T}
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} \int p(x_t, y_{t-1}) q(\hat{x}_t | x_t, y_{t-1}) \rho(x_t, \hat{x}_t) \, dx_t \, d\hat{x}_t \, dy_t
\]
subject to
\[
I(\hat{x}_t; x_t | y_{t-1}) = \int dx_t d\hat{x}_t dy_{t-1} p(x_t, y_{t-1}) \, q(\hat{x}_t | x_t, y_{t-1})
\]
\[
\log \frac{q(\hat{x}_t | x_t, y_{t-1})}{Q(\hat{x}_t | y_{t-1})} \leq C_t \forall t
\]
\[ \int q(\hat{x}_t \mid x_t, y_{t-1}) \, d\hat{x}_t = 1 \quad \forall t \quad (5.2.3) \]

\[ Q(\hat{x}_t \mid y_{t-1}) = \int q(\hat{x}_t \mid x_t, y_{t-1}) \, p(x_t \mid y_{t-1}) \, dx_t \quad \forall t \quad (5.2.4) \]

In (5.2.2) \( C_t \) is the capacity of the sensor at time \( t \); it is given by

\[ C_t = I(y_t; x_t \mid y_{t-1}) \quad (5.2.5) \]

In the sequel we will discuss the properties of (5.2.1)-(5.2.5). We will show that the solution of (5.2.1)-(5.2.5) provides a lower bound to the nonlinear filtering problem (N), and will compare our bound to the existing ones. We will also compare (5.2.1)-(5.2.5) to the real time distortion rate problem presented in Chapter 4.

5.2.2 Properties of the formulation

First we show that (5.2.1)-(5.2.5) provides a lower bound to the nonlinear filtering problem (N).

**Theorem 5.2.1**

The minimum of (5.2.1)-(5.2.5) is a lower bound to the nonlinear filtering problem (N).

**Proof**

Consider \( I(\hat{x}_t; x_t \mid y_{t-1}) \) and use (5.2.2) and the data processing theorem. Then we obtain

\[ I(\hat{x}_t; x_t \mid y_{t-1}) \leq I(y_t; x_t \mid y_{t-1}) = I(y_t; x_t \mid y_{t-1}) \quad (5.2.6) \]

Because of (5.2.6) the set of admissible solutions of (N) is a subset of the set of admissible solution of (5.2.1)-(5.2.5). Hence, the minimum of (5.2.1)-(5.2.5) is a lower bound to (N).
Since the decision taken at a specific instance of time $t$, does not affect the decisions taken in the future, (5.2.1)-(5.2.5) decomposes into a series of static optimization problems, each one of which is of the form

$$\text{Minimize} \quad E \rho(x_t, \hat{x}_t) \quad (5.2.7)$$
subject to

$$I(\hat{x}_t;x_t | y_{t-1}) \leq C_t \quad (5.2.8)$$
$$\int q(\hat{x}_t | x_t y_{t-1}) \, d\hat{x}_t = 1 \quad (5.2.9)$$

The problem symmetric to (5.2.7)-(5.2.9) is

$$\text{Minimize} \quad I(\hat{x}_t;x_t | y_{t-1}) \quad (5.2.10)$$
subject to

$$E \rho(x_t, \hat{x}_t) \leq D_t (C_t) \quad (5.2.11)$$
$$\int q(\hat{x}_t | x_t y_{t-1}) \, d\hat{x}_t = 1 \quad (5.2.12)$$

(5.2.10)-(5.2.12) is similar to a classical rate distortion problem and the curves resulting from (5.2.7)-(5.2.9) and (5.2.10)-(5.2.12) by varying the sensor, (i.e., $C_t$), are identical. The rate distortion function defined by (5.2.10)-(5.2.12) has the same properties as the classical rate distortion function. The equations and theorems derived by the variational problem corresponding to (5.2.10)-(5.2.12) are similar to those of the classical rate distortion problem. Theorems similar to those corresponding to the variational problem (5.2.10)-(5.2.12) were presented in Section 4.2.2. The quantity $I(\hat{x}_t;x_t | y_{t-1})$ gives the additional information about the source output $x_t$ that the filter is able to extract from the observation at time $t$.

We will present an example where the bound of (5.2.1)-(5.2.3) is attained.
5.2.3 Example

Consider the source

\[ x_{t+1} = x_t + n_t \]  \hspace{1cm} (5.2.13)

with observations

\[ y_t = x_t + v_t \]  \hspace{1cm} (5.2.14)

Assume that

(i) \( x_1 \sim N(0, \Sigma^2) \)  \hspace{1cm} (5.2.15)

(ii) \( n_t \sim N(0, N_t^2) \)  \hspace{1cm} (5.2.16)

(iii) \( v_t \sim N(0, V_t^2) \)  \hspace{1cm} (5.2.17)

(iv) \( E(n_t n_s) = N_t^2 \delta(t-s) \)  \hspace{1cm} (5.2.18)

(v) \( E(v_t v_s) = V_t^2 \delta(t-s) \)  \hspace{1cm} (5.2.19)

(vi) \( E(n_t v_s) = 0 \)  \hspace{1cm} \forall t, s \)  \hspace{1cm} (5.2.20)

(vii) \( E(n_t x_s) = 0 \)  \hspace{1cm} \forall t, \geq s \)  \hspace{1cm} (5.2.21)

(viii) The distortion measure is the squared error criterion.

According to (5.2.1)-(5.2.5) we have to solve the problem

\[
\text{Minimize} \quad \frac{1}{2} E \{ (x_1 - \hat{x}_1)^2 + (x_2 - \hat{x}_2)^2 \} \\
\{ q(\hat{x}_1 | x_1) \} \quad q(\hat{x}_2 | x_2, y_1) \]  \hspace{1cm} (5.2.22)

subject to

\[
I(\hat{x}_1; x_1) \leq I(y_1; x_1) \]  \hspace{1cm} (5.2.23)

\[
I(\hat{x}_2; x_2 | y_1) \leq I(y_2; x_2 | y_1) \]  \hspace{1cm} (5.2.24)

\[
\int q(\hat{x}_1 | x_1) d\hat{x}_1 = 1 \]  \hspace{1cm} (5.2.25)

\[
\int q(\hat{x}_2 | x_2, y_1) d\hat{x}_2 = 1 \]  \hspace{1cm} (5.2.26)

The details of the solution of (5.2.22)-(5.2.26) are given in Appendix G.

Here we give the results:
Time \( t=1 \)

\[
\text{Min } E(x_1 - \hat{x}_1)^2 \triangleq D_1 = \Sigma^2 \frac{\frac{1}{V_2}}{\Sigma^2 + \frac{1}{V_2}}
\]  
(5.2.27)

Time \( t=2 \)

\[
\text{Min } E(x_2 - \hat{x}_2)^2 = D_2 = \frac{\frac{V_1^2 \Sigma^2}{V_2^2} + \Sigma^2 \frac{V_2^2}{V_1} + \frac{N_1^2 V_1^2}{V_2}}{\frac{V_1^2 \Sigma^2}{V_2^2} + \Sigma^2 \frac{N_1^2 V_1^2}{V_2} + \Sigma^2 \frac{V_2^2}{V_1} + \Sigma^2 \frac{V_2^2}{V_1}}
\]  
(5.2.28)

The solution of the least squares estimation problem for (5.2.13)-(5.2.21) is given by the Kalman filter algorithm. The filter Riccati equation gives:

For \( t=1 \)

\[
\Sigma(1|0) = \Sigma^2
\]  
(5.2.29)

\[
\Sigma(1|1) = \left( \frac{1}{\Sigma^2} + \frac{1}{V_1^2} \right)^{-1} = \Sigma^2 \frac{V_1^2}{\Sigma^2 + \frac{1}{V_1^2}}
\]  
(5.2.30)

For \( t=2 \)

\[
\Sigma(2/2) = [\Sigma(2|1)^{-1} + (V_2^2)^{-1}]^{-1}
\]  
(5.2.31)

But

\[
\Sigma(2|1) = \Sigma(1|1) + N_1^2 = \Sigma^2 \frac{V_1^2}{\Sigma^2 + \frac{1}{V_1^2}} + N_1^2 = \frac{\Sigma^2 V_1^2 + \Sigma^2 N_1^2 V_1^2 + N_1^2 V_1^2}{\Sigma^2 + V_1^2}
\]  
(5.2.32)

Hence

\[
\Sigma(2|2) = \left( \frac{\Sigma^2 + V_1^2}{\Sigma^2 N_1^2 + N_1^2 V_1^2 + \Sigma^2 V_1^2} + \frac{1}{V_2^2} \right)^{-1} = \frac{\Sigma^2 + V_2^2}{\Sigma^2 N_1^2 + N_1^2 V_1^2 + \Sigma^2 V_1^2 + \Sigma^2 V_2^2 + V_1^2 V_1^2}
\]  
(5.2.33)

(5.2.27), (5.2.28), (5.2.30) and (5.2.33) give
\[ D_1 = \Sigma(1|1) \quad (5.2.34) \]
\[ D_2 = \Sigma(2|2) \quad (5.2.35) \]

The above results can be easily extended to the multishot problem and show that the lower bound of (5.2.1)-(5.2.5) is attained in the case of linear Gaussian sources with linear observations and a squared error criterion.

This result has to be intuitively expected for the following reason: The derivation of the lower bound of (5.2.1)-(5.2.5) is based on the fact that there is no loss of information in the filter. On the other hand, the Kalman filter uses all the information that is available from the observations.

As we have pointed out in Section 5.1, we have taken a rate distortion theoretic approach to the nonlinear filtering problem. Other approaches that use information theoretic ideas to obtain lower bounds to the nonlinear filtering problem are those of Zakai and Ziv [17] and Galdos [18]. We will compare our approach to the above mentioned approaches.

5.2.4 Comparison with the results of Zakai-Ziv and Galdos

The derivation of the lower bound of Zakai and Ziv is based upon three important results:

(i) The data processing theorem of information theory ([4], pg. 80).

(ii) The Shannon lower bound on the rate distortion function

(iii) The mutual information of the white Gaussian channel ([24], [25], [26]).

Zakai and Ziv consider the system
\[ dx_1(t) = x_2(t) \, dt \]
\[ dx_2(t) = x_3(t) \, dt \]
\[ \vdots \]
\[ dx_{n-1}(t) = x_n(t) \, dt \]
\[ dx_n(t) = m(x_1(t), x_2(t), \ldots, x_n(t), t) \, dt + dw(t) \] (5.2.36)

with observations described by
\[ dy(t) = g(x_k(t), t) \, dt + \sqrt{N_0} \, dv(t) \] (5.2.37)

and want to minimize the mean square error
\[ \epsilon^2(t) = E \{ (x_j(t) - \hat{x}_j(t))^2 \} \] (5.2.38)

where
\[ \hat{x}_j(t) = f_t(y^t_0) \]

At first
\[ \epsilon^2(t) = E(d^2(t)) = E \{ E \{ (x_j(t) - \hat{x}_j(t))^2 \mid \bar{x}(0) \} \} \] (5.2.39)

where
\[ \bar{x}(0) = (x_1(0), x_2(0), \ldots, x_n(0))' \] (5.2.40)

Using the data processing theorem, the Shannon lower bound and the mutual information of the white Gaussian channel they obtain the following series of inequalities:

Shannon-lower bound \[ \leq h(x_j(t) \mid \bar{x}(0)) - \frac{1}{2} \ln(2\pi e d^2(t)) \]

\[ \leq I(d^2(t)) \leq I(\hat{x}_j(t); x_j(t) \mid \bar{x}(0)) \leq I(x_j(t); y^t_0 \mid \bar{x}(0)) \leq I(x_j(t_0); y^t_0 \mid \bar{x}(0)) \]

\[ = I(x_k(t_0); y^t_0 \mid \bar{x}(0)) = \]

\[ = \frac{1}{2} E_{\bar{x}(0)} \left\{ \int_0^t (g(x_k(\tau), \tau) - \hat{g}(x_k(\tau), \tau))^2 d\tau \right\} \] (5.2.41)

where
\[ \hat{g}(x_k(\tau), \tau) = E \{ g(x_k(\tau), \tau) \mid y^t_0, \bar{x}(0) \} , \]
\( x_j(0) \) denotes the path \( x_j(\tau), 0 \leq \tau \leq t, \) \( y_0^t \) denotes the observations up to time \( t, \) and \( E_{\mathbf{x}(0)} \) denotes the conditional expectation with respect to \( \mathbf{x}(0). \) The last equality in (5.2.41) gives the mutual information of the white Gaussian channel. From (5.2.41) it follows that

\[
d^2(t) > \frac{1}{2\pi e} \exp \{2h(x_j(t)|\mathbf{x}(0)) - \\
- \frac{1}{N_0} \int_0^t (g(x_k(\tau), \tau) - \hat{g}(x_k(\tau), \tau))^2 \ d\tau\} \tag{5.2.42}
\]

(5.2.39), (5.2.42) and the convexity of the exponential function combined together give

\[
\varepsilon^2(t) > \frac{1}{2\pi e} \left[ \exp \{2Eh(x_j(t)|\mathbf{x}(0)) - \right. \\
\left. - \frac{1}{N_0} \int_0^t E \{(g(x_k(\tau), \tau) - \hat{g}(x_k(\tau), \tau))^2 \} \ d\right] \tag{5.2.43}
\]

(5.2.43) gives the Zakai-Ziv lower bound. From (5.2.43) we conclude that the Zakai Ziv formulation allows computational of the bound only for the class of nonlinear filtering problems with linear observations. If the observations are nonlinear, then computing the bound is equivalent to solving the nonlinear filtering problem. Moreover the result of Zakai and Ziv holds only for the continuous time filtering problem. This is because the last equality in (5.2.41) is based upon the fact that the innovations are Gaussian. The innovations are actually Gaussian in the continuous time filtering problem, but they are not necessarily Gaussian in the discrete time nonlinear filtering problem.

Galdos' approach, [18], is in principle the same as our approach. Galdos' derivation of the lower bound assumes that the filter uses all the information that is available from the observations. We will show that our approach gives an alternative derivation of Galdos' bound. Considering
difference distortion measures and following a procedure similar to that of Section 4.2.3, we find that the Shannon-type lower bound to (5.2.1)-(5.2.5) is

\[ h(x_t \mid y_{t-1}) - h(\pi_s(x_t)) \leq I(y_t ; x_t \mid y_{t-1}) \]  

(5.2.44)

where

\[ \pi_s(x_t) = \frac{e^{\rho(x_t)}}{\int_{dz_t} e^{\rho(z_t)}}. \]  

(5.2.45)

(5.2.44) gives

\[ h(x_t \mid y_{t-1}) - h(\pi_s(x_t)) + h(x_t) - h(x_t) \leq h(x_t \mid y_{t-1}) - h(x_t \mid y_t) \]

or

\[ h(x_t) - h(\pi_s(x_t)) \leq h(x_t) - h(x_t \mid y_t) \]

\[ \leq h(x_t \mid y_{t-1}) - h(x_t \mid y_t) \]

(5.2.46)

Consider (5.2.46) applied to the system

\[ x_{t+1} = a(x_t, t) + b(x_k, t)w_t \]  

(5.2.47)

\[ y_t = g_t(x_t, t) + v_t \]  

(5.2.48)

\[ E(v_t, v_s') = R \delta(t-s) \]  

(5.2.49)

\[ E(w_t, w_s') = W \delta(t-s) \]  

(5.2.50)

\[ E(v_t, x_s') = 0 \quad \forall (t, s) \]  

(5.2.51)

\[ E(w_t, x_s') = 0 \quad t \geq s \]  

(5.2.52)

with

\[ \rho(x_t, \hat{x}_t) = (x_t - \hat{x}_t)^2 \]  

(5.2.53)

Using the representation theorem of nonlinear filtering, [31], Jensen's inequality, and the smoothing property of the conditional expectation we get for \( I(x_t ; y_t) \),
\[
I(x_t; y_t) = \mathbb{E}_{x_t \cdot y_t} \log \left( \frac{\mathbb{E}_{x_t} \sigma(x_t)}{\mathbb{E}_{x_t} (\exp J_t)} \right) = \\
= \mathbb{E}_{x_t \cdot y_t} \log \mathbb{E}_{x_t} \sigma(x_t) (\exp J_t) - \mathbb{E}_{x_t \cdot y_t} \log \mathbb{E}_{x_t} (\exp J_t) \\
\leq \log \mathbb{E}_{x_t \cdot y_t} \mathbb{E}_{x_t} \sigma(x_t) (\exp J_t) - \mathbb{E}_{x_t \cdot y_t} \log \mathbb{E}_{x_t} (\exp J_t) \\
= \log \mathbb{E}_{x_t \cdot y_t} (\exp J_t) - \mathbb{E}_{x_t \cdot y_t} \log \mathbb{E}_{x_t} (\exp J_t) \quad (5.2.54)
\]

where

\[
J_t = \frac{1}{2} \sum_{\tau=0}^{t} g(x_t, \tau) R^{-1} g'(x_t, \tau) + \sum_{\tau=0}^{t} g(x_t, \tau) R^{-1} v_{\tau} \quad (5.2.55)
\]

and \(\mathbb{E}_{x_t \cdot y_t}, \mathbb{E}_{x_t}\) denote the expectation with respect to the densities \(p(x_t, y_t)\) and \(p(x_t)\) respectively.

Because of (5.2.53),

\[
h(\pi S(x_t)) = \frac{1}{2} \ln(2\pi\varepsilon^2(t)) \quad (5.2.56)
\]

(5.2.54) and (5.2.56) substituted into (5.2.46) give

\[
\varepsilon^2(t) \geq \frac{1}{2\pi\varepsilon} \exp \{2h(x_t)\} \left\{ \frac{\exp \mathbb{E}_{x_t \cdot y_t} \log \mathbb{E}_{x_t} (\exp J_t)}{\mathbb{E}_{x_t \cdot y_t} (\exp J_t)} \right\} \quad (5.2.57)
\]

(5.2.57) is the discrete time version of relation (5.17) of [18], which is Galdos' lower bound. Since our approach gives an alternative derivation of Galdos' lower bound, it has the following advantages over the Zakai-Ziv approach:

(i) It is applicable to any nonlinear filtering problem and not only to systems of the form (5.2.36)-(5.2.37).

(ii) In certain cases the resulting bound is better than the Zakai and Ziv bound when the signal to noise ratio is low, as shown experimentally in [18], Chapter 6.
A disadvantage of our approach is that the quantity $I(y_t; x_t | y_{t-1})$ which represents the sensor's capacity at time $t$, cannot be computed exactly in most filtering problems. Galdos' bounding techniques and results can be used in this case but they are very complicated. The main advantage of our approach, however, is that it can be used to provide lower bounds to decentralized filtering problems. These bounds are in some cases better than the ones given by the solution of the corresponding centralized filtering problems. The formulation of decentralized filtering problems will be the topic of later sections. Next we will discuss the relation of the nonlinear filtering problem to the real time communication problem.

5.2.5 Relation to the Real Time Communication Problem

The nonlinear filtering problem can be viewed as a special case of the real time communication problem. In the real time communication problem we know the source, we are given a channel of certain capacity and we want to determine the encoder and decoder so that the source outputs are transmitted in real time through the communication system in the most reliable way. In the filtering problem we are given the encoder and the channel which combined together give the sensor, and we want to choose the decoder, (filter), to achieve the most reliable transmission of the source outputs. By fixing the encoder we eliminate the signaling phenomena which occur in the real time communication problem. Thus, in the case of filtering we are able to decompose and solve (5.2.1)-(5.2.5) whereas in the real time communication problem we are unable to decompose and solve (3.2.3)-(3.2.6). It is the prespecified encoder that makes the filtering problem conceptually much simpler than the real time communication problem (although still very hard computationally).
By fixing the encoder we automatically determine the capacity of the sensor, i.e., the amount of information that becomes available to the decoder after each transmission. The capacity of the sensor is usually time varying moreover, for all instants of time it is less than or equal to the capacity of the channel. Therefore, in the nonlinear filtering problem we do not use the channel in its full capacity at each transmission. On the contrary, in the real time communication problem we want to use the channel in its full capacity at each transmission. This is an important difference between the filtering problem and the real time communication problem. A consequence of this difference is that the real time distortion rate problem

$$\text{Minimize}_{\{q(\hat{x}_t | x_t)\}} \sum_{t=1}^{T} \frac{1}{T} E \sum_{t=1}^{T} \rho(x_t, \hat{x}_t)$$

subject to

$$I(\hat{x}_t : x_t | x_{t-1}) \leq C_t, \quad \forall t$$

$$\int q(\hat{x}_t | x_t) d\hat{x}_t = 1, \quad \forall t$$

(where $C_t$ is time varying and the time variation is such that the encoding-decoding problem is simplified), can not be used to provide a lower bound to problems (P) and (P') of Chapter 3. This happens because the capacity of the implemented channel is at least as big as $\max_t C_t$ and the system is not using the maximum available capacity at each instant of time.

The above discussion reveals some of the connections and differences between the real time communication problem and the filtering problem, and shows the limitations and difficulties that exist in coming up with a formulation of a real time distortion rate problem that provides a good lower bound to the real time communication problem.
In the rest of Chapter 5 we will formulate and discuss decentralized filtering problems.

5.3 Formulation of a Decentralized Filtering Problem

Consider the system of Figure 5.3.1 (page 103). The system consists of two subsystems which are correlated. Assume that

(i) Subsystem 1 (SS1) is described by the state $x_1$ and subsystem 2 (SS2) is described by the state $x_2$,

(ii) The dynamics of the system are known as well as the observation equations for $y_1^t$ and $y_2^t$.

(iii) For all instants of time, observation $y_1^t(y_2^t)$ is directly available to filter 1 (2), whereas observation $y_2(y_1)$ is transmitted in real time to filter 1 (2).

(iv) The capacity of the communication system 2→1 (1→2) is $C_{21}(C_{12})$ and the channels used are memoryless.

(v) Filter 1(2) forms its estimate $\hat{x}_1^t(x_2^t)$ based on the observations $y_1^t = (y_1^1, y_1^2, \ldots, y_1^t)$, $(y_2^t)$ and the information $\hat{x}_2^t (\hat{x}_1^t)$ transmitted through the communication system 2→1 (1→2), i.e.,

$$\hat{x}_1^t = f_1^t(y_1^t, \hat{x}_1^t)$$

$$\hat{x}_2^t = f_2^t(y_2^t, \hat{x}_1^t)$$

Under the above assumptions the decentralized filtering problem for filter 1 can be stated as follows:

$$\begin{align*}
\text{Minimize} & \quad \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \rho(x_1^t, \hat{x}_1^t) \\
\text{subject to} & \quad f_1^t, g_2^t
\end{align*}$$

(5.3.3)
Figure 5.3.1
subject to
\[ \hat{x}_1^t = f_1^t (y_1^t, \hat{x}_2^t) \forall t \]  
(5.3.4)
\[ z_2^t = g_2^t (y_2^t) \forall t \]  
(5.3.5)
\[ I(y_2^t; z_2^t) \leq \min (C_{21}, I(y_2^t; x_1^t | y_1^t y_2^t)) \forall t \]  
(DN)  
(5.3.6)
\[ I(y_1^t; x_1^t | y_1^{t-1} y_2^{t-1}) = C_1^t \forall t \]  
(5.3.7)

The problem for filter 2 is the same as (DN). In (5.3.3) \( \rho(\cdot) \) is a distortion measure. \( C_{1}^t \) is determined by the observation equations for \( y_1^t, y_2^t \). (5.3.6) and (5.3.7) present an alternative way of writing the observation equations for \( y_1^t, y_2^t \) and the constraint on the information transfer between the sensors and the filters. We will use (5.3.6) and (5.3.7) instead of the observation equations because we will take a rate distortion theoretic approach to the decentralized filtering problem.

From the above formulation it is clear that the decentralized filtering problem (DN) is the combination of a real time communication problem and a nonlinear filtering problem. The observations \( y_1^t \) are always directly available to filter 1 and the problem of forming \( \hat{x}_1^t \) based only on \( y_1^t \) is the same as the nonlinear filtering problem. On the other hand only a communicated version \( \hat{y}_2^t \) of the observations \( y_2^t \) is available to filter 1 and the problem of forming the estimate \( \hat{x}_1^t \) based solely on \( \hat{y}_2^t \) is the same as the real time communication problem. The existence of communication links between the sensors and the filters changes significantly the nature of the filtering problem. Even the simplest linear filtering problems become nonlinear when there is a limitation on information transfer between the sensor and the filter. Consider for example the system of Figure 5.3.1 and assume that
(i) it is linear Gaussian,

(ii) $C_{21} < I(y_2^t; y_1^t | y_1^{t-1} y_2^{t-1})$,

(iii) the observations are linear.

Under the above conditions the optimal filter is infinite dimensional, as opposed to the centralized filter which is linear and finite dimensional (Kalman filter). The reason we get a nonlinear filter is that we have to use a nonlinear coding technique in order to transmit the maximum amount of information through the communication system; hence $\hat{y}_2^t$ is not necessarily Gaussian for all $t$, and the optimal filter will in general be infinite dimensional.

Finding the solution of the decentralized filtering problem (DN) is equivalent to solving the signaling problem. Since the solution of the signaling problem remains unknown, we will formulate a real time distortion rate problem whose solution will provide a lower bound to (DN) and we will discuss the characteristics of the formulation.

5.4 Formulation of a Real Time Distortion Rate Problem for (DN)

5.4.1 Statement of the problem

The real time distortion rate problem for (DN) can be stated as follows:

$$\text{Minimize} \quad \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \rho(x_1^t, \hat{x}_1^t) =$$

$$= \frac{1}{T} \sum_{t=1}^{T} \int p(x_1^t, y_1^t, y_2^t) \rho(x_1^t, \hat{x}_1^t) dx_1^t dy_1^t dy_2^t$$

subject to

$$\rho(x_1^t, \hat{x}_1^t) \text{ d}x_1^t \text{ d}y_1^t \text{ d}y_2^t \text{ d}x_1^t \quad (5.4.1)$$
\[ I(\hat{x}_1^t; x_1^t y_1^{t-1} y_2) \triangleq \int p(x_1^t y_1^{t-1} y_2) q(\hat{x}_1^t | x_1^t y_1^{t-1} y_2) \]
\[ \log \frac{q(\hat{x}_1^t | x_1^t y_1^{t-1} y_2)}{Q(\hat{x}_1^t | y_1^{t-1} y_2)} \, dx_1^t \, dx_1 \, dy_1 \, dy_2 \leq \]
\[ I(y_1^t; x_1^t y_1^{t-1} y_2) + \min (C_{21}, I(y_2^t; x_1^t y_1^{t-1} y_2)) \]  
(5.4.2)

\[ \int q(\hat{x}_1^t | x_1^t y_1^{t-1} y_2) \, d\hat{x}_1 = 1 \]  
(5.4.3)

\[ Q(\hat{x}_1^t | y_1^{t-1} y_2) = \int q(\hat{x}_1^t | x_1^t y_1^{t-1} y_2) p(x_1^t | y_1^{t-1} y_2) \, dx_1^t \]  
(5.4.4)

In the next section we will discuss the properties of (5.4.1)-(5.4.4), we will show that it provides a lower bound to problem (DN) and compare the bound to the centralized solution.

5.4.2 Properties of (5.4.1)-(5.4.4)

Since the decision taken at a specific time does not affect the decisions taken in the future, (5.4.1)-(5.4.4) decomposes into a series of static optimization problems each one of which has the following form:

\[ \text{Minimize} \quad \mathbb{E} \rho(x_1^t \hat{x}_1^t) \]  
(5.4.5)

subject to

\[ I(\hat{x}_1^t; x_1^t y_1^{t-1} y_2) \leq I(y_1^t; x_1^t y_1^{t-1} y_2) + \]
\[ + \min (C_{21}, I(y_2^t; x_1^t y_1^{t-1} y_2)) \]  
(5.4.6)

\[ \int q(\hat{x}_1^t | x_1^t y_1^{t-1} y_2) \, d\hat{x}_1 = 1 \]  
(5.4.7)

The properties of (5.4.5)-(5.4.7) are the same as those of problem (4.2.1)-(4.2.3) discussed in Chapter 4. We will show that the solution of (5.4.1)-(5.4.4) provides a lower bound to problem (DN).
Theorem 5.4.1

The solution of (5.4.1)-(5.4.4) provides a lower bound to problem (DN).

Proof

Consider the information constraint (5.4.2); applying the data processing theorem we obtain

\[
I(\hat{x}_1^t|x_1^t, y_1^{t-1}, y_2^{t-1}) \leq I(\hat{y}_1^t|y_1^{t-1}, y_2^{t-1})
\]

\[
= I(\hat{y}_1^t; x_1^t | y_1^{t-1}, y_2^{t-1}) + I(\hat{y}_1^t; x_1^t | y_1^{t-1}, y_2^{t-1})
\]

or

\[
I(\hat{x}_1^t|x_1^t, y_1^{t-1}, y_2^{t-1}) \leq I(\hat{y}_1^t; x_1^t | y_1^{t-1}, y_2^{t-1}) +
\]

\[
+ I(\hat{y}_1^t; x_1^t | y_1^{t-1}, y_2^{t-1})
\]

(5.4.8)

Consider each of the terms of the right-hand side of (5.4.8) separately.

\[
I(\hat{y}_1^t; x_1^t | y_1^{t-1}, y_2^{t-1}) \triangleq \frac{c_1^t}{\Delta} \text{ from (5.3.7)}
\]

\[
I(\hat{y}_1^t; x_1^t | y_1^{t-1}, y_2^{t-1}) \leq I(\hat{y}_1^t; x_1^t | y_1^{t-1}, y_2^{t-1}) \triangleq
\]

\[
\triangledown \int dx_2^t dx_1^t dy_1^t dy_2^t \ p(x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1}) x
\]

\[
\times \log \frac{p(\hat{y}_1^t | x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1}) p(\hat{y}_1^t | x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1})}{p(\hat{y}_1^t | y_1^{t-1}, y_2^{t-1}) p(\hat{y}_1^t | y_1^{t-1}, y_2^{t-1})}
\]

(4.4.9)

But

\[
p(\hat{y}_1^t | x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1}) = \frac{p(x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1})}{p(x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1})}
\]

\[
= \frac{p(x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1}) p(\hat{y}_1^t | y_1^{t-1}, y_2^{t-1}) p(y_1^{t-1})}{p(x_1^t, x_2^t, y_1^{t-1}, y_2^{t-1}) p(y_1^{t-1})}
\]

\[
= p(\hat{y}_1^t | y_1^{t-1}, y_2^{t-1})
\]

(5.4.10)
Moreover
\[ p(\hat{y}_2^t | x_1^t x_2^t y_1^{t-1} y_2^{t-1}) = p(\hat{y}_2^t | x_1^t x_2^t y_1^{t-1} y_2^{t-1}) \] (5.4.11)

Because of (5.4.10), (5.4.11) and the assumption that the channel is memoryless, (5.4.9) gives
\[
I(\hat{y}_2^t; x_1^t y_1^{t-1}) \leq I(\hat{y}_2^t; x_1^t x_2^t | y_1^{t-1} y_2^{t-1}) = \\
= I(\hat{y}_2^t; x_1^t x_2^t | y_1^{t-1} y_2^{t-1}) = \\
= H(\hat{y}_2^t | y_1^{t-1} y_2^{t-1}) - H(\hat{y}_2^t | x_1^t x_2^t y_1^{t-1} y_2^{t-1}) \leq \\
\leq H(\hat{y}_2^t) - H(\hat{y}_2^t | z_2^t) = I(\hat{y}_2^t; z_2^t) \] (5.4.12)

(5.4.8), (5.3.7) and (5.4.12) combined give
\[
I(\hat{y}_2^t; x_1^t y_1^{t-1} y_2^{t-1}) \leq I(y_1^t; x_1^t | y_1^{t-1} y_2^{t-1}) + \\
+ I(\hat{y}_2^t; z_2^t) \] (5.4.13)

Because of (5.4.13), the set of admissible solutions of problem (DN) is a subset of the set of admissible solution of problem (5.4.1)-(5.4.4). Therefore, the solution of (5.4.1)-(5.4.4) provides a lower bound to problem (DN) QED.

The quantities \(I(y_1^t; x_1^t | y_1^{t-1} y_2^{t-1})\) and \(I(y_2^t; x_1^t | y_1^{t-1} y_2^{t-1})\) which appear in the information constraint (5.4.2) can not always be exactly computed, but they can be upperbounded using the techniques and results of Galdos [18].

Notice that when for all \(t\)
\[
\min (c_{21}, I(y_2^t; x_1^t | y_1^{t-1} y_2^{t-1})) = I(y_2^t; x_1^t | y_1^{t-1} y_2^{t-1}) \] (5.4.14)
i.e., when there is no limitation on the information transfer between sensor 2 and filter 1, then the information constraint (5.4.2) becomes

\[ I(\hat{x}_1^t; x_1^t | y_1^t \bigcup y_2^t) \leq I(y_1^t y_2^t; x_1^t | y_1^t \bigcup y_2^t) \tag{5.4.15} \]

and (5.4.1)-(5.4.4) is the same as (5.2.1)-(5.2.5) which is the real time distortion rate problem corresponding to the centralized filtering problem (N).

In the sequel we will study the decentralized filtering problem (DN) for linear Gaussian sources and linear observations using (5.4.1)-(5.4.4); we will find a lower bound to problem (DN) and will compare it to the centralized solution.

### 5.4.3 Example 5.4.1

Assume that the system of figure (5.3.1) is linear Gaussian described by

\[
\begin{pmatrix}
x_1(t+1) \\
x_2(t+1)
\end{pmatrix} = \begin{pmatrix}
-1 & 0.5 \\
0.5 & -1
\end{pmatrix} \begin{pmatrix}
x_1(t) \\
x_2(t)
\end{pmatrix} + \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \begin{pmatrix}
\xi_1(t) \\
\xi_2(t)
\end{pmatrix} \tag{5.4.16}
\]

and that the observations are

\[
y_1(t) = x_1(t) + 0.5 x_2(t) + v_1(t) \tag{5.4.17}
\]
\[
y_2(t) = 0.5 x_1(t) + x_2(t) + v_2(t) \tag{5.4.18}
\]

\[
\mathbb{E} \begin{pmatrix}
\xi_1(t) \\
\xi_2(t)
\end{pmatrix} \begin{pmatrix}
\xi_1(s) \\
\xi_2(s)
\end{pmatrix} = \begin{pmatrix}
Q_1 \delta(t-s) & 0 \\
0 & Q_2 \delta(t-s)
\end{pmatrix} \tag{5.4.19}
\]

\[
\mathbb{E} \begin{pmatrix}
v_1(t) \\
v_2(t)
\end{pmatrix} \begin{pmatrix}
v_1(s) \\
v_2(s)
\end{pmatrix} = \begin{pmatrix}
R_1 \delta(t-s) & 0 \\
0 & R_2 \delta(t-s)
\end{pmatrix} \tag{5.4.20}
\]

\[
\mathbb{E} \{x(s) \xi'(t)\} = 0 \quad \forall t \geq s \tag{5.4.21}
\]
\[ E \{ \xi(t) \nu'(s) \} = 0 \quad \forall (t,s) \]  \hspace{1cm} (5.4.22)

The communication system \(2 \rightarrow 1 \) (\(1 \rightarrow 2\)) has capacity \(C_{21}(C_{12})\) and the channel is memoryless. At any time \(t\), the estimate of filter \(1(2)\) is a function of the available data, i.e.

\[ \hat{x}^t_1 = f^t_1(y^t_1 \hat{y}^t_2) \quad \forall t \]  \hspace{1cm} (5.4.23)

\[ \hat{x}^t_2 = f^t_2(y^t_2 \hat{y}^t_1) \quad \forall t \]  \hspace{1cm} (5.4.24)

We consider the least squares filtering problem (i.e., \(\rho(x^*_t, \hat{x}^*_t) = (x^*_t - \hat{x}^*_t)^2\) in (5.3.3) for the system described above, and use (5.4.1)-(5.4.4) in order to find a lower bound to this problem.

According to the results of section 4.2.3, at time \(t\), the Shannon-type lower bound to the solution of (5.4.1)-(5.4.4) is for our example

\[ h(x^t_1, y^t_1, y^t_2, y^t_2) \leq \frac{1}{2} \ln(2\pi e D^t_1) \leq \]  \hspace{1cm} (5.4.25)

\[ \leq I(y^t_1; x^t_1|y^t_2, y^t_2) + \min \{I(y^t_2; x^t_1|y^t_1, y^t_2), C_{21}\} \]

where

\[ D^t_1 = E(x^t_1 - \hat{x}^t_1)^2. \]  \hspace{1cm} (5.4.26)

Because of (5.4.16)-(5.4.22) and the result of section (5.2.3) we get

\[ h(x^t_1, y^t_1, y^t_2, y^t_2) = \frac{1}{2} \ln [2\pi e (\Sigma(t|t-1))]_{11} \]  \hspace{1cm} (5.4.27)

\[ I(y^t_1; x^t_1|y^t_1, y^t_2) = \frac{1}{2} \ln \frac{(\Sigma(t|t-1))_{11}}{((\Sigma^{-1}(t|t-1) + H^t_1 R^{-1}_1 H^t_1)^{-1})_{11}} \]  \hspace{1cm} (5.4.28)

where \((\Sigma(t|t-1))_{11}\) is the \((1,1)\) entry of the solution of the Riccati equation for (5.4.16)-(5.4.18) and \(H_1\) is the observation matrix corresponding to (5.4.17) i.e.,

\[ H_1 = (1 \hspace{0.5cm} 0.5) \]  \hspace{1cm} (5.4.29)
Assume that for all $t$

$$C_{21} < I(y_2^t; x_1^t | y_1^t y_2^t) \quad (5.4.30)$$

Then (5.4.25)-(5.4.30) give

$$\frac{1}{2} \ln \frac{[\Sigma(t|t-1)]_{11}}{D_1^t} \leq \frac{1}{2} \ln \frac{[\Sigma(t|t-1)]_{11}}{\{[\Sigma^{-1}(t|t-1) + H_1' R_1^{-1} H_1]^{-1}\} + c_{21}} \quad (5.4.31)$$

or

$$D_1^t \geq \{[\Sigma^{-1}(t|t-1) + H_1' R_1^{-1} H_1]^{-1}\}_{11} \exp\{-2c_{21}\}. \quad (5.4.32)$$

(5.4.32) provides a lower bound to the least squares decentralized filtering problem (DN) for the system (5.4.16)-(5.4.22). This bound lies above the lower bound of the centralized filtering problem (N) for the system (5.4.16)-(5.4.22) because

$$C_{21} < I(y_2^t; x_1^t | y_1^t y_2^t) \quad \forall t \quad (5.4.33)$$

As we have shown in Section 5.2.3, the lower bound to the least squares centralized filtering problem for linear Gaussian sources and linear observations is equal to the solution of the corresponding Riccati equation. Consequently, the lower bound of (5.4.32) lies above the centralized solution.

When $C_{21} \geq I(y_2^t; x_1^t | y_1^t y_2^t)$ for all $t$, then for this example the lower bound to problem (DN), provided by the solution of (5.4.1)-(5.4.4) is equal to the solution of the centralized least squares filtering problem.

In Sections 5.3 and 5.4 we formulated a decentralized filtering problem (DN) and presented a real time distortion rate problem whose solution provided a lower bound to (DN). In the next section we will present another formulation of a decentralized filtering problem and its corresponding real time distortion rate problem.

111
5.5 Another Formulation of the Decentralized Filtering Problem

5.5.1 Statement of the problem

Consider the system of Figure 5.3,1. Under the assumptions (i)-(v) of page the centralized filtering problem for filter 1 can be stated as follows:

\[
\begin{align*}
\text{Minimize} & \quad \frac{1}{T} \ E \sum_{t=1}^{T} \rho(x_1^t, \hat{x}_1^t) \\
\{f_1^t, g_2^t\} & \text{subject to} \\
\hat{x}_1^t & = f_1^t (y_1^t, \hat{y}_2^t) \quad \forall t \\
\hat{y}_2^t & = g_2^t (y_2^t) \quad \forall t \\
I(y_1^t; x_1^t) & = C_{1}^t \\
I(y_2^t; x_1^t) & \leq \min (tC_{21}, I(y_2^t; x_1^t | y_1^t))
\end{align*}
\] (5.5.1)

The problem for filter 2 is the same. The decentralized filtering problem (DNG) has the same characteristics as problem (DN), i.e., finding the solution of (DNG) is equivalent to solving the signaling problem. Since we cannot solve the problem itself, we will find a lower bound by formulating a real time distortion rate problem for (DNG).

5.5.2 Formulation of a real time distortion rate problem for (DNG)

The real time distortion rate problem for (DNG) can be stated as follows:

\[
\begin{align*}
\text{Minimize} & \quad \frac{1}{T} \ E \sum_{t=1}^{T} \rho(x_1^t, \hat{x}_1^t) \\
\{q(x_1^t | \hat{x}_1^t)\} & \text{subject to} \\
I(x_1^t; x_1^t) & \leq C_{1}^t \quad \forall t \\
\int q(x_1^t | \hat{x}_1^t) \, dx_1^t & = 1 \quad \forall t
\end{align*}
\] (5.5.6)

\] (5.5.7)

\] (5.5.8)
\[ C^*_t = I(y^t_1; x^t_1) + \min \left( tC_{21}, I(y^t_2; x^t_1 | y^t_1) \right) \]  \hspace{1cm} (5.5.9)

(5.5.6)-(5.5.9) is a straightforward extension of Galdos' approach applied to the decentralized filtering problem.

It is easy to show that the solution of (5.5.6)-(5.5.9) gives a lower bound to problem (DNG).

**Theorem 5.5.1**

The solution of (5.5.6)-(5.5.9) is a lower bound to (DNG).

**Proof**

Consider \( I(\hat{x}^t_1; x^t_1) \), use (5.5.2) and apply the data processing theorem.

Then,

\[ I(\hat{x}^t_1; x^t_1) \leq I(y^t_1 y^t_2; x^t_1) = \]

\[ = I(y^t_1; x^t_1) + I(y^t_2; x^t_1 | y^t_1) \]  \hspace{1cm} (5.5.10)

Consider each of the terms of the right-hand side of (5.5.10) separately.

\[ I(y^t_1; x^t_1) \triangleq C^*_1 \text{ from (5.5.4).} \]  \hspace{1cm} (5.5.11)

\[ I(y^t_2; x^t_1 | y^t_1) \leq h(y^t_2) - h(y^t_2 | x^t_1 y^t_1 y^t_2) = \]

\[ = h(y^t_2) - h(y^t_2 | y^t_2) = I(\hat{y}^t_2; y^t_2) \]  \hspace{1cm} (5.5.12)

\[ \leq I(y^t_2; z^t_2) \]

(5.5.10)-(5.5.12) combined give

\[ I(\hat{x}^t_1; x^t_1) \leq I(y^t_1; x^t_1) + I(y^t_2; z^t_2) \]  \hspace{1cm} (5.5.13)

Because of (5.5.13), the set of admissible solutions of problem (DNG) is a subset of the set of admissible solutions of problem (5.5.6)-(5.5.9).
Therefore, the solution of (5.5.6)-(5.5.9) provides a lower bound to problem (DNG) QED.

Let us investigate the solution of (5.5.6)-(5.5.9) for very large \( t \). For very large \( t \) we will have

\[
\min(tC_{21}, I(y_2^t; x_1^t | y_1^t)) = I(y_2^t; x_1^t | y_1^t)
\]

(even if \( C_{21} < I(y_2^t; x_1^t | y_1^t y_2^{t-1}) \) for all \( t \)). Consequently, (5.5.9) will give

\[
C_t^* = I(y_1^t y_2^t; x_1^t).
\]

(5.5.6)-(5.5.8) and (5.5.14) is Galdos' information theoretic formulation of the classical filtering problem. Therefore, for very large \( t \) the bound provided by the solution of (5.5.6)-(5.5.9) is the same as Galdos' bound for the classical nonlinear filtering problem. Hence, when there is a limitation on information transfer between the sensor and the filter, our original approach provides a better bound to the decentralized filtering problem than (5.5.6)-(5.5.9) which is a straightforward extension of Galdos' approach.

In general, the solution of filtering problems where part of the data is directly available to the filter, and another part is communicated through a link with limited capacity, is not known. In order to find the solution we must be able to solve a real time communication problem and a nonlinear filtering problem. Because of the difficulties of these problems lower bounds are valuable since they help us evaluate linear suboptimal solutions.

In the next section a linear filtering problem will be solved to provide a lower bound to a decentralized filtering problem involving delays.
5.6 Formulation of a Decentralized Filtering Problem (L) Involving Delays

5.6.1 Statement of the problem

Consider the system of Figure 5.6.1. Assume that

(i) The system is described by the equation

\[ x(t+1) = ax(t) + w(t) \]  \hspace{1cm} (5.6.1)
\[ x(0) \sim N(m, \sigma^2) \]  \hspace{1cm} (5.6.2)
\[ w(t) \sim N(0, Q) \quad \forall t \geq 1 \]  \hspace{1cm} (5.6.3)
\[ E \{ w(t)x(s) \} = 0 \quad \forall t > s \]  \hspace{1cm} (5.6.4)

(ii) Linear observations \( y_1(t) \) and \( y_2(t) \) are taken at two locations, 1 and 2 respectively. Observation \( y_1(t) \) (\( y_2(t) \)) is prefILTERED as shown in the figure, and the estimate \( \hat{x}_{1e}(t|t) \) (\( \hat{x}_{2e}(t|t) \)) is transmitted through the communication system 1→2 (2→1) to location 2(1). The data available from \( y_2(t) \) (\( y_1(t) \)) and the communicated version of \( \hat{x}_{1e}(t|t) \) (\( \hat{x}_{2e}(t|t) \)) are used to estimate the state of the system at location 2(1), as shown in figure 5.6.1.

(iii) Transmission through the communication system 1→2 (2→1) occurs at a rate that is twice as fast as the rate of observations \( y_1(t) \) (\( y_2(t) \)). Thus, consider that observations are taken every two seconds and every estimate \( \hat{x}_{1e}(t|t) \) (\( \hat{x}_{2e}(t|t) \)) can be transmitted twice before \( y_1(t+1) \) (\( y_2(t+1) \)) is available to the input of the communication system 1→2 (2→1).

(iv) The observation equations are

\[ y_1(2t+1) = c_1 x(2t+1) + v_1(2t+1) \]  \hspace{1cm} (5.6.5)
\[ y_2(2t+1) = c_2 x(2t+1) + v_2(2t+1) \]  \hspace{1cm} (5.6.6)

\[
E \begin{bmatrix}
    v_1(2t+1) \\
    v_2(2t+1)
\end{bmatrix}
= \begin{bmatrix}
    R_1 \delta(t-s) & 0 \\
    0 & R_2 \delta(t-s)
\end{bmatrix}
\]  \hspace{1cm} (5.6.7)
Figure 5.6.1
and
\[ E \{v_i(t)w(s)\} = 0 \quad (t,s) \quad i=1,2 \quad (5.6.8) \]

(v) For all instants of time
\[ C_{12} < I(\hat{x}_{2e}(t); x_1(t) | \hat{x}_{1e}(t-1) \leq 2C_{12} \quad (5.6.9) \]

(and
\[ C_{21} < I(\hat{x}_{2e}(t); x_1(t) | \hat{x}_{2e}(t-1) \leq 2C_{21}) \quad (5.6.10) \]

Thus, ideally, at time \( t=2\tau+2 \) (\( \tau=0,1,2, \)) we can have a perfect reproduction of \( \hat{x}_{1e}(2\tau+1|2\tau+1) \) (\( \hat{x}_{2e}(2\tau+1|2\tau+1) \)) at the output of the communication system 1→2 (2→1).

(vi) The communicated version of \( y_1(2\tau+1) \) (\( y_2(2\tau+1) \)) is used for estimation at location 2(1) after the transmission of \( y_1(2\tau+1) \) (\( y_2(2\tau+1) \)) through the communication system 1→2 (2→1) is completed. Consequently, the communicated version of \( y_1(2\tau+1) \) (\( y_2(2\tau+1) \)) is used for estimation at location 2 with one time unit of delay.

Under the assumptions (i)-(vi) above, the least squares filtering problem at location 2 can be stated as follows:

\[
\text{Minimize} \quad \frac{1}{T} E \sum_{t=1}^{T} (x_t - \hat{x}_t)^2 \quad (5.6.11)
\]

subject to
\[ I(x_t; \hat{x}_t) \leq C_{12} \quad \forall t \quad (5.6.12) \]

and the constraints (i)-(vi) stated above.

The least squares filtering problem at location 2 is exactly the same.
5.6.2 A lower bound to (L)

Because of the information constraint (5.6.9), ideally, at time \( t=2\tau+2 \) we have a perfect reproduction of \( \hat{x}_{le}(2\tau+1|2\tau+1) \) at location 2. Consequently, if we solve the least squares filtering problem at location 2, assuming that \( \hat{x}_{le}(2\tau+1|2\tau+1) \) is available to that location with one unit of delay, we will get a lower bound to (L). We can use a Kalman filter as a prefilter because at each instant of time the Kalman filter estimate \( \hat{x}_{le}(t|t) \) contains all the information about \( x(t) \) provided by the observation up to that time [18]. After each measurement we transmit the innovation

\[
\nu_{le}(2\tau+1) \triangleq y_{1}(2\tau+1) - C_{1} \hat{x}_{le}(2\tau+1|2\tau+1)
\]

through the communication system 1→2. The innovation \( \nu_{le}(2\tau+1) \) gives the new information about the system provided to the prefilter by the observation \( y_{1}(2\tau+1) \). Since in the ideal situation we can perfectly reconstruct the innovations \( \nu_{le}(2\tau+1) \) at the output of the communication system at time \( t=2\tau+2 \), we can perfectly reconstruct \( \hat{x}_{le}(2\tau+1|2\tau+1) \) and \( y_{1}(2\tau+1) \) at the output of the data processor #1 at time \( t=2\tau+2 (\tau=0,1,...) \). Given that \( \hat{x}_{le}(2\tau+1|2\tau+1) \) is available to location 2 at time \( t=2\tau+2 \), for all \( \tau(\tau=0,1,...) \), the optimum processing of the data will be the following:

**Time \( t=1 \)**

At time \( t=1 \) observations \( y_{1}(1) \) and \( y_{2}(1) \) are taken at locations 1 and 2 respectively. Observation \( y_{2}(1) \) is directly available to location 2 but observation \( y_{1}(1) \) is not available to this location until time \( t=2 \). Thus, the best estimate at time \( t=1 \) is equal to the estimate of data processor #2, i.e.,

\[
\hat{x}_{2}(1|1) = \hat{x}_{2e}(1|1) = \hat{x}_{2}(1) + k_{2}(1) [y_{2}(1) - C_{2}\hat{x}_{2}(1)] =
\]

\[
= m + k_{2}(1) [y_{2}(1) - C_{2}m]
\]

(5.6.13)
where

\[ k_2(1) = \sigma^2 c_2 \left[ R_2 + c_2 \sigma^2 c_2 \right]^{-1} \]. \quad (5.6.14)

The resulting mean square error is

\[ \Sigma_2 (1|1) = [c_2 R_2^{-1} c_2 + (\sigma^2)^{-1}]^{-1} \]. \quad (5.6.15)

**Time t=2**

At time t=2 no observations are taken but \( \hat{x}_{1e}(1|1) \) becomes available to location 2. The best estimate at time t=2 is then,

\[
\hat{x}_2(2|1) = a[\hat{x}(1) + (k(1))_{11} \{(k_1(1))^{-1} [\hat{x}_{1e}(1|1) -
- (1-k_1(1)c_1) \hat{x}(1)] - c_1 \hat{x}(1)\} + (k(1))_{12} \{(k_2(1))^{-1}
[\hat{x}_{2e}(1|1) - (1-k_2(1)c_2) \hat{x}(1)] - c_2 \hat{x}(1)\}] \]  

or

\[
\hat{x}_2(2|1) = a[\hat{x}(1) + (k(1))_{11} [y_1(1) - c_1 \hat{x}(1)] +
+ (k(1))_{12} [y_2(1) - c_2 \hat{x}(1)]] = a \hat{x}(1|1) \]  

where

\[
k(1) = \sigma^2 (c_1 c_2) \left[ R_1 \begin{pmatrix} 0 & 0 \\ 0 & R_2 \end{pmatrix} + (c_1 c_2) \right]. \]

\[
\sigma^2 (c_1 c_2) \]

\[
k_1(1) = \sigma^2 c_1 \left[ R_1 + c_1 \sigma^2 c_1 \right]^{-1}. \quad (5.6.19)
\]

\( k_2(1) \) is given by (5.6.15) and \( \hat{x}(1|1) \) is the minimum mean square error estimate of the centralized filter at time t=1. The resulting mean square error is

\[ \Sigma_2(2|1) = a^2 \Sigma(1|1) + Q \]. \quad (5.6.20)
where
\[
\Sigma(1|1) = \begin{bmatrix}
\sigma^2^{-1} + (C_1 C_2)^{-1} R_1^{-1} 0 \\
0 R_2^{-1}
\end{bmatrix}
\begin{bmatrix}
C_1 \\
C_2
\end{bmatrix}
\]  \hspace{1cm} (5.6.21)

Thus, at time \(t=2\) we can recover the solution of the centralized filtering problem. In order to achieve the estimate of the centralized filter the data processor 1 and 2 have to transmit their local estimates \(\hat{x}_{1e}(1|1)\) and \(\hat{x}_{2e}(1|1)\) respectively to the final data processor, and that processor gives \(\hat{x}_2(2|1)\) according to (5.6.16). Note that the procedure for deriving \(\hat{x}_2(2|1)\) is essentially the same as that described in [28].

For any time \(t\) the processing of the data will be the following:

**Time \(t=2t+1\)**

At this time the estimate \(\hat{x}(2t+1|2t-1)\) which is the same as the estimate of the centralized filtering problem, is available to location 2, and the observations \(y_1(2t+1)\) and \(y_2(2t+1)\) are taken. The observation \(y_2(2t+1)\) is directly available to location 2, whereas \(y_1(2t+1)\) is not available to that location until time \(t=2t+2\). Hence at time \(t=2t+1\) the best estimate at location 2 is

\[
\hat{x}_2(2t+1|2t+1) = \hat{x}(2t+1|2t-1) + k_2(2t+1) [y_2(2t+1) - C_2 \hat{x}(2t+1|2t-1)]
\]  \hspace{1cm} (5.6.22)

where
\[
k_2(2t+1) = \Sigma(2t+1|2t-1) C_2 [R_2 + C_2 \Sigma(2t+1|2t-1) C_2]
\]  \hspace{1cm} (5.6.23)

and \(\Sigma(2t+1|2t-1)\) is the error covariance of \(\hat{x}(2t+1|2t-1)\).

The resulting mean square error is
\[
\Sigma_2(2t+1|2t+1) = [C_2 R_2^{-1} C_2 + (\Sigma(2t+1|2t-1))^{-1}]^{-1}
\]  \hspace{1cm} (5.6.24)
At time $t=2\tau+2$ no observation is taken at either location but \( \hat{x}_{1e}(2\tau+1|2\tau+1) \) becomes available to location 2. The best estimate at time $t=2\tau+2$ is then

\[
\hat{x}_2(2\tau+2|2\tau+1) = a \left[ \hat{x}(2\tau+1|2\tau-1) + (k(2\tau+1))_{11} \right.
\]

\[
\{ (k_{1e}(2\tau+1))^{-1} \left[ \hat{x}_{1e}(2\tau+1|2\tau+1) - (1-k_{1e}(2\tau+1))C_1 \hat{x}_{1e}(2\tau+1|2\tau-1) \right. \\
- C_1 \hat{x}(2\tau+1|2\tau-1) \} + (k(2\tau+1))_{12} \left( (k_{2e}(2\tau+1))^{-1} \right)
\]

\[
\left. \left[ \hat{x}_{2e}(2\tau+1|2\tau+1) - (1-k_{2e}(2\tau+1))C_2 \hat{x}_{2e}(2\tau+1|2\tau-1) \right] \\
- C_2 \hat{x}(2\tau+1|2\tau-1) \right]
\]  \hspace{1cm} (5.6.25)

or

\[
\hat{x}_2(2\tau+2|2\tau+1) = a \left( \hat{x}(2\tau+1|2\tau-1) + (k(2\tau+1))_{11} \right) \cdot \\
[y_1(2\tau+1) - C_1 \hat{x}(2\tau+1|2\tau-1)] + (k(2\tau+1))_{12} \cdot \\
[y_2(2\tau+1) - C_2 \hat{x}(2\tau+1|2\tau-1)] = a \hat{x}(2\tau+1|2\tau+1) \hspace{1cm} (5.6.26)
\]

where

\[
k(2\tau+1) \overset{\Delta}{=} (k(2\tau+1))_{11} (k(2\tau+1))_{12} = \\
= \Sigma(2\tau+1|2\tau-1) \left( C_1 C_2 \right) \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} + \\
\left( \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \right) \Sigma(2\tau+1|2\tau-1) \left( C_1 C_2 \right)^{-1} \hspace{1cm} (5.6.27)
\]

\[
k_{1e}(2\tau+1) = \Sigma_{1e}(2\tau+1|2\tau-1) C_1 \left[ R_1 + C_1 \Sigma_{1e}(2\tau+1|2\tau-1) C_1 \right] \hspace{1cm} (5.6.28)
\]

\[
k_{2e}(2\tau+1) = \Sigma_{2e}(2\tau+1|2\tau-1) C_2 \left[ R_2 + C_2 \Sigma_{2e}(2\tau+1|2\tau-1) C_2 \right]. \hspace{1cm} (5.6.29)
\]
\( \Sigma_{ie}(2\tau+1|2\tau-1) \) are the error covariances of the estimates \( x_{ie}(2\tau+1|2\tau-1) \) (\( i=1,2 \)) and \( \hat{x}(2\tau+1|2\tau-1) \) is the minimum mean squares error estimate of the centralized filter at \( t = 2\tau+1 \).

The resulting mean square error is

\[
\Sigma_2(2\tau+2|2\tau+1) = a^2 \Sigma(2\tau+1|2\tau+1) + Q
\]

where

\[
\Sigma(2\tau+1|2\tau+1) = \left[ \Sigma^{-1}(2\tau+1|2\tau-1) + (C_1 \ C_2) \begin{pmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{pmatrix} (C_1 \ C_2)^T \right]^{-1}
\]

(5.6.31)

Thus, at time \( t=2\tau+2 \) we recover the solution of the centralized filtering problem. In order to achieve the estimate of the centralized filter, the data processors 1 and 2 have to transmit their local estimates \( \hat{x}_{1e}(2\tau+1|2\tau+1) \) and \( \hat{x}_{2e}(2\tau+1|2\tau+1) \) respectively, to the final data processor, and that processor combines them with \( \hat{x}_{1e}(2\tau+1|2\tau-1) \), and \( \hat{x}_{2e}(2\tau+1|2\tau-1) \) and \( \hat{x}(2\tau+1|2\tau-1) \) according to (5.6.25).

Note that the solution of the centralized filtering problem is achieved only at times \( t=2\tau+2 \) (\( \tau=0,1,... \)); at times \( t=2\tau+1 \), (\( \tau=0,1,2,... \)), \( \hat{x}_2(2\tau+1|2\tau+1) \) is worse than the estimate of the centralized filter because only \( y_2(2\tau+1) \) is available to location 2 at time \( t=2\tau+1 \).

The mean square errors \( \Sigma_2(2\tau+1|2\tau+1) \), given by (5.6.24), and \( \Sigma_2(2\tau+2|2\tau+1) \) given by (5.6.30), provide a lower bound to the filtering problem (L) for all instants of time \( \tau \). (\( \tau=0,1,2,... \)). The optimal filter will in general be nonlinear for the reasons we discussed in Section 5.4.1.

122
5.7 **Summary**

In this chapter we presented a rate distortion theoretic approach to filtering problems.

At the beginning we formulated a real time distortion rate problem, namely, (5.2.1)-(5.2.5), which provides a lower bound to the centralized filtering problem. We compared our approach to Zakai's-Ziv's approach and showed that it leads to the same lower bound as Galdos' approach. We also showed that the filtering problem is a special case of the real time communication problem, resulting when we fix the encoder.

Finally, we formulated the decentralized filtering problems (DN) and (L) which are combinations of a real time communication problem and a filtering problem. We provided a lower bound to problem (DN) by solving a real time distortion rate problem, namely (5.4.1)-(5.4.4), and showed that for linear Gaussian systems and the squared error criterion this lower bound lies above the solution of the centralized filtering problem corresponding to (DN). We also provided a lower bound to problem (L) by solving a linear filtering problem.
CHAPTER 6

Summary, Conclusions and Suggestions for Further Research

This chapter summarizes the results of the thesis. A brief discussion of the conclusions that can be drawn from the research, as well as topics for further investigation are also included.

6.1 Summary

The thesis consisted of three parts. In the first part we formulated a real time feedforward communication problem (P) and a real time noiseless feedback communication problem (P'). We presented two real time distortion rate problems whose solutions provided lower bounds to problems (P) and (P'). We showed that the bounds resulting from the real time reformulations of the distortion rate problem lie above the lower bound provided by the classical distortion rate problem. We presented a noiseless feedback communication system whose performance approaches the lower bounds.

In the second part we presented an information theoretic approach to the centralized filtering problem (N). We formulated a real time distortion rate problem whose solution provided a lower bound to problem (N). We compared the resulting lower bound to the existing ones and showed that it is achieved in the case of the least squares filtering problem for linear Gaussian sources and linear observations. We also discussed the connections and differences between the centralized filtering problem and the real time feedforward communication problem.

In the third part we formulated a decentralized filtering problem (DN). We presented a real time distortion rate problem whose solution is a lower
bound to (DN), and showed that for linear Gaussian sources, linear observations and a squared error criterion our bound lies above the solution of the corresponding centralized filtering problem. We also formulated a decentralized filtering problem (L) involving delays, and provided a lower bound to problem (L) via the solution of a linear filtering problem.

6.2 Conclusions, Suggestions for Further Research

Using a real time reformulation of the distortion rate problem, namely (3.3.1)-(3.3.4), we found a lower bound to the real time communication problem. This bound is better than the lower bound provided by the classical rate distortion problem. We were able to solve (3.3.1)-(3.3.4) because we considered an input-output description of the communication system and did not worry about the detailed way the data are processed within the system. We showed that further improvement of the bound requires knowledge of the processing of the data within the system. It requires the solution of problem (3.2.3)-(3.2.6). Problem (3.2.3)-(3.2.6) is not well posed until the encoder is defined. Determining the encoder however is equivalent to solving the signaling problem of nonclassical stochastic control theory. If the encoder is fixed then we have to solve a classical filtering problem. The centralized filtering problem is conceptually much easier than the real time communication problem. It is possible to solve problem (3.2.3)-(3.2.6) in the case of filtering because the encoder is fixed and signaling phenomena are not present. Signaling phenomena are present however in the decentralized filtering problem presented in this thesis. That is why we again have to consider an input-output description of the system, namely (5.4.1)-(5.4.4), in order to be able to derive a lower bound that is computable. In some cases, (e.g., the LQG decentralized filtering problem), the lower bound
resulting from the solution of (5.4.1)-(5.4.4) lies above the solution of the corresponding centralized filtering problem. This happens because the constraints on information transfer between the sensors and the filters are taken into account in (5.4.1)-(5.4.4).

It is difficult to find a feedforward real time communication system that approaches the solution of (3.3.1)-(3.3.4). The difficulty is due to the fact that the two decision makers, the encoder and the decoder, have, at every instant of time, different information about the source, and no one knows what the information of the other is. Thus, it is very difficult for the two decision makers to cooperate. Cooperation is easier when one of the decision makers, (the encoder, for example) knows the information of the other. Then, he can adjust his strategy to help the other decision maker so that the overall performance is improved. This is exactly the situation in the noiseless feedback communication system. The encoder always knows the information of the decoder and adjusts his strategy so that the performance of the system can approach (as shown in examples 3.4.1 and 4.2.2), the lower bound provided by the solution of (3.3.1)-(3.3.4).

The results of this thesis indicate the following:

(i) If we want to improve the lower bounds on the real time communication problem and the decentralized filtering problem we have to be able to solve problems of the form of (3.2.3)-(3.2.6); solving these problems is equivalent to solving the signaling problem.

(ii) If we want to find real time feedforward communication systems whose performance approaches the lower bound found in this thesis, we have to determine the strategies that achieve the desired performance. Determining good strategies requires a better understanding of the signaling problem.
Consequently any progress in the problems considered in this thesis depends upon the progress in understanding the signaling phenomenon.

That is why the study of the signaling problem has to be the main topic of the future research related to the problems presented in this thesis.
APPENDIX A
Solution of (3.3.53)-(3.3.59)

According to our formulation the problem (3.3.53)-(3.3.59) decomposes into two static optimization problems.

At time $t = 1$ we have to

\[
\text{Minimize } E \rho(x_1, \hat{x}_1), \quad q(\hat{x}_1 | x_1) \tag{A.1}
\]

subject to

\[
I(\hat{x}_1; x_1) \leq C, \tag{A.2}
\]

\[
qu(\hat{x}_1 | x_1) \geq 0 \quad \forall (x_1, \hat{x}_1), \tag{A.3}
\]

\[
\sum_{\hat{x}_1} q(\hat{x}_1 | x_1) = 1 \quad \forall x_1. \tag{A.4}
\]

At time $t = 2$ we have to

\[
\text{Minimize } E \rho(x_2, \hat{x}_2), \quad q(\hat{x}_2 | x_1 x_2) \tag{A.5}
\]

subject to

\[
I(\hat{x}_2; x_2 | x_1) \leq C, \tag{A.6}
\]

\[
qu(\hat{x}_2 | x_1 x_2) \geq 0 \quad \forall (x_1, x_2, \hat{x}_2), \tag{A.7}
\]

\[
\sum_{\hat{x}_2} q(\hat{x}_2 | x_1 x_2) = 1 \quad \forall (x_1, x_2). \tag{A.8}
\]

First shot; time $t = 1$

The problem symmetric to (A.1)-(A.4) is
\[
\begin{align*}
\text{Minimize} & \quad I(\hat{x}_1; x_1), \\
q(\hat{x}_1 | x_1) & \quad (A.9) \\
\text{subject to} & \\
E \quad \rho(x_1, \hat{x}_1) & \leq D(C), \quad (A.10) \\
q(\hat{x}_1 | x_1) & \geq 0 \quad \forall (x_1, \hat{x}_1), \quad (A.11) \\
\sum_{\hat{x}_1} q(\hat{x}_1 | x_1) & = 1 \quad \forall x_1. \quad (A.12)
\end{align*}
\]

(A.9)-(A.12), considered for the source described in Example 3.3.1, is the same as Example 2.7.1 of [3]. Its solution is derived by forming the augmented function \( \mathcal{F}(q) \) (defined in Section 3.3), determining its stationary point from the requirements

\[
\frac{d \mathcal{F}(q(\hat{x}_1 | x_1))}{dq(\hat{x}_1 | x_1)} = 0,
\]

and then following the procedure of Section 2.6 (case 1) of [3]. We get

\[
Q(\hat{x}_1 = 0) = \frac{1}{1 - e^s} [\pi_0 - e^s(1 - \pi_0)], \quad (A.13)
\]

\[
Q(\hat{x}_1 = 1) = \frac{1}{1 - e^s} [(1 - \pi_0) - \pi_0 e^s], \quad (A.14)
\]

\[
D_1 = \frac{e^s}{1 + e^s}, \quad (A.15)
\]

\[
R(D_1) \triangleq \min_{q(\hat{x}_1 | x_1)} I(\hat{x}_1; x_1) = H_b(\pi_0) - H_b(D_1), \quad (A.16)
\]

(\text{where } H_b(\cdot) \text{ denotes binary entropy}),
\[ u_o = u_1 = \frac{1}{1 + e^s}, \quad (A.17) \]

\[ \lambda_o = \frac{u_o}{\pi_o} = \frac{1}{\pi_o (1 + e^s)}, \quad (A.18) \]

\[ \lambda_1 = \frac{u_1}{1 - \pi_o} = \frac{1}{(1 - \pi_o)(1 + e^s)}, \quad (A.19) \]

\[ q(\hat{x}_1 = 0|\hat{x}_1 = 0) = \lambda_o \quad Q(\hat{x}_1 = 0) \quad e^o = \frac{\pi_o - e^s (1 - \pi_o)}{\pi_o (1 - e^{2s})}, \quad (A.20) \]

\[ q(\hat{x}_1 = 1|\hat{x}_1 = 0) = \lambda_o \quad Q(\hat{x}_1 = 1) \quad e^s = \frac{(1 - \pi_o) - e^s \pi_o}{\pi_o (1 - e^{2s})} e^s, \quad (A.21) \]

\[ q(\hat{x}_1 = 1|\hat{x}_1 = 1) = \lambda_1 \quad Q(\hat{x}_1 = 1) e^o = \frac{(1 - \pi_o) - e^s \pi_o}{(1 - \pi_o)(1 - e^{2s})}, \quad (A.22) \]

\[ q(\hat{x}_1 = 0|\hat{x}_1 = 1) = \lambda_1 \quad Q(\hat{x}_1 = 1) e^s = \frac{e^s (\pi_o - e^s (1 - \pi_o))}{(1 - \pi_o)(1 - e^{2s})}. \quad (A.23) \]

Since (A.9)-(A.12) and (A.1)-(A.4) have the same solution, we can use
(A.16) to compute \( D_1 \) when the channel capacity \( C = R(D_1) \) is given.
Then from (A.15) we can get the value of \( e^s \) and use it in (A.13), (A.14),
(A.20)-(A.23) to obtain \( Q(\hat{x}_1 = i) \) \((i = 0, 1)\) and \( q(\hat{x}_1 = k|\hat{x}_1 = j) \) \((k, j = 0, 1)\).

**Second shot: time \( t = 2 \)**

At time \( t = 2 \) we have again two inputs to the communication system,
\( x_2 = 0, 1, \) and two outputs, \( \hat{x}_2 = 0, 1. \) However, the probabilities of the
inputs \( x_2 \) depend on the value of \( x_1. \) Therefore, we have to solve two
rate distortion problems corresponding to the values of the source output
at time $t=1$. Each one of these problems is the same as the problem at $t=1$.

For $x_1 = 0$ we will have inputs $x_2 = 0, 1$ to the communication system with probabilities

$$p(x_2 = 0|x_1 = 0) = 1 - p_1,$$  \hspace{1cm} \text{(A.24)}

$$p(x_2 = 1|x_1 = 0) = p_1.$$  \hspace{1cm} \text{(A.25)}

Using the same procedure as in the first shot and the results of Example 2.7.1 of [3] we get

$$Q(x_2 = 0|x_1 = 0) = \frac{(1-p_1) - e^{s'}}{1-e^{s'}},$$  \hspace{1cm} \text{(A.26)}

$$Q(x_2 = 1|x_1 = 0) = \frac{p_1 - e^{s'}(1-p_1)}{1-e^{s'}}.$$  \hspace{1cm} \text{(A.27)}

$$\min_{q(x_2|x_2x_1 = 0)} I(x_2;x_2|x_1 = 0) \triangleq R(D_{20}) = \mathcal{H}_b(p_1) - \mathcal{H}_b(D_{20})$$  \hspace{1cm} \text{(A.28)}

$$D_2 = \frac{e^{s'}}{1+e^{s'}},$$  \hspace{1cm} \text{(A.29)}

$$u_{0/0} = u_{1/0} = \frac{1}{1+e^{s'}}.$$  \hspace{1cm} \text{(A.30)}

$$\lambda_{0/0} = \frac{u_{0/0}}{1-p_1} = \frac{1}{(1-p_1)(1+e^{s'})},$$  \hspace{1cm} \text{(A.31)}

$$\lambda_{1/0} = \frac{u_{1/0}}{p_1} = \frac{1}{p_1(1+e^{s'})},$$  \hspace{1cm} \text{(A.32)}
and
\[ q(\hat{x}_2=0|x_2=0, x_1=0) = \lambda_0/Q(\hat{x}_2=0|x_1=0)e^0 = \frac{(1-p_1)e^{s'}}{(1-p_1)(1-e^{2s'})}, \quad (A.33) \]

\[ q(\hat{x}_2=1|x_2=0, x_1=0) = \lambda_0/Q(\hat{x}_2=1|x_1=0)e^{s'} = \frac{[p_1 - (1-p_1)e^{s']e^{s'}}}{(1-p_1)(1-e^{2s'})}, \quad (A.34) \]

\[ q(\hat{x}_2=1|x_2=1, x_1=0) = \lambda_1/Q(\hat{x}_2=1|x_1=0) = \frac{p_1 - e^{s'}(1-p_1)}{(1-e^{2s'})p_1}, \quad (A.35) \]

\[ q(\hat{x}_2=0|x_2=1, x_1=0) = \lambda_1/Q(\hat{x}_2=0|x_1=0)e^{s'} = \frac{[(1-p_1) - e^{s'}p_1]e^{s'}}{p_1(1-e^{2s'})}. \quad (A.36) \]

Similarly for \( x=1 \) we will have two inputs \( x_2=0,1 \) to the communication system with probabilities

\[ p(x_2=0|x_1=1) = p_1, \quad (A.37) \]
\[ p(x_2=1|x_1=1) = 1-p_1. \quad (A.38) \]

Then
\[ Q(\hat{x}_2=1|x_2=1) = Q(\hat{x}_2=0|x_1=0) = \frac{(1-p_1) - e^u p_1}{1 - e^u}, \quad (A.39) \]

\[ Q(\hat{x}_2=0|x_1=1) = Q(\hat{x}_2=1|x_1=0) = \frac{p_1 - e^u(1-p_1)}{1 - e^u}, \quad (A.40) \]
\[
\begin{align*}
\text{Min } & I(\hat{x}_2; x_2 | x_1 = 1) = R(D_{21}) = \mathcal{H}_b(p_1) - \mathcal{H}_b(D_{21}) \quad \text{(A.41)} \\
q(\hat{x}_2 | x_2, x_1 = 1) & = R(D_{21}) = \mathcal{H}_b(p_1) - \mathcal{H}_b(D_{21}) \quad \text{(A.42)} \\
D_{21} & = \frac{e^\mu}{1 + e^\mu}, \\
\lambda_{0/1} & = \mu_0/1 = \frac{1}{1 + e^\mu}, \\
\lambda_{1/1} & = \mu_0/0 = \frac{1}{p_1 (1 + e^\mu)}, \\
\lambda_{0/1} & = \mu_0/0 = \frac{1}{p_1 (1 + e^\mu)}, \\
q(\hat{x}_2 = 1 | x_2 = 1, x_1 = 1) & = \frac{(1 - p_1) - e^\mu p_1}{(1 - p_1)(1 - e^{2\mu})}, \\
q(\hat{x}_2 = 0 | x_2 = 1, x_1 = 1) & = \frac{p_1 - (1 - p_1)e^\mu}{(1 - p_1)(1 - e^{2\mu})}, \\
q(\hat{x}_2 = 1 | x_2 = 0, x_1 = 1) & = \frac{e^\mu [(1 - p_1) - e^\mu p_1]}{p_1 (1 - e^{2\mu})}, \\
q(\hat{x}_2 = 0 | x_2 = 0, x_1 = 1) & = \frac{p_1 - e^\mu (1 - p_1)}{p_1 (1 - e^{2\mu})}. \\
\end{align*}
\]

Thus

\[
R(D_2) = \text{Min } I(\hat{x}_2; x_2 | x_1) = \mathcal{H}_b(p_1) - \mathcal{H}_b(D_2), \\
q(\hat{x}_2 | x_2, x_1) = \mathcal{H}_b(p_1) - \mathcal{H}_b(D_2), \\
D_2 = D_{20} = D_{21}. \\
\]
Similarly to the first shot, when the capacity of the channel \( C = R(D_2) \) is given, we can determine \( D_2 \) from (A.50) and we can find \( e^{s'} \) and \( e^{\mu} \) from (A.29) and (A.42), respectively \( (e^{s'} = e^{\mu}) \). We can then use these values to determine \( Q(\hat{x}_2^1 = i|x_1^j = j), (i, j = 0, 1) \) and \( q(\hat{x}_2^2 = \lambda|x_1^k x_2^m = m), (\lambda, k, m = 0, 1) \), from (A.26)-(A.27), (A.39)-(A.40) and (A.33)-(A.36), (A.46)-(A.49), respectively.

The same procedure can be used to solve the multishot problem and also the problem where the transitions of the source are not symmetric \( (i.e., \ p(x_t^1|x_{t-1}^0) \neq p(x_t^0|x_{t-1}^1)) \).
APPENDIX B

Solution of (3.4.19)-(3.4.28)

First Shot; Time \( t=1 \)

At time \( t=1 \) we have to solve the problem

\[
\text{Minimize } E \log p(x_1, \hat{x}_1, x_1), \quad q(\hat{x}_1|x_1) \tag{B.1}
\]

subject to

\[
I(\hat{x}_1; x_1) \leq C, \quad \tag{B.2}
\]

\[
q(\hat{x}_1|x_1) \geq 0 \quad \forall (x_1, \hat{x}_1), \quad \tag{B.3}
\]

\[
\sum_{\hat{x}_1} q(\hat{x}_1|x_1) = 1 \quad \forall x_1. \tag{B.4}
\]

The solution of (B.1)-(B.4) is identical to the solution of problem (A.1)-(A.4) (of Appendix A) and is given by (A.13)-(A.23).

Using the values \( \pi_1 = 0.3, \ C = 0.6 \) bits we obtain

\[
C = \mathcal{H}_b(0.3) - \mathcal{H}_b(D_1), \tag{B.5}
\]

\[
\mathcal{H}_b(D_1) = 0.2813, \tag{B.6}
\]

\[
D_1 = 0.0488, \tag{B.6}
\]

\[
0.0488 = \frac{e^s}{1+e^s} \Rightarrow e^s = 0.00514, \tag{B.7}
\]

\[
q(\hat{x}_1=0|x_1=0) = 0.895, \quad \tag{B.8}
\]

\[
q(\hat{x}_1=1|x_1=0) = 0.105, \quad \tag{B.9}
\]
\begin{align*}
q(\hat{x}_1=0|x_1=1) &= 0.0487, \\
nq(\hat{x}_1=1|x_1=1) &= 0.9513, \\
Q(\hat{x}_1=0) &= 0.302, \\
Q(\hat{x}_1=1) &= 0.698.
\end{align*}

Second Shot; Time $t=2$

At time $t=2$ we have to solve the problem

\begin{align*}
\text{Minimize} & \quad E \rho(x_2,\hat{x}_2), \\
nq(\hat{x}_2|x_2,\hat{x}_1)
\end{align*}

subject to

\begin{align*}
I(\hat{x}_2;x_2|\hat{x}_1) &\leq C, \\
q(\hat{x}_2|x_2,\hat{x}_1) &\geq 0 \quad \forall (\hat{x}_1,\hat{x}_2, x_2), \\
\sum_{\hat{x}_2} q(\hat{x}_2|x_2,\hat{x}_1) &= 1 \quad \forall (\hat{x}_1, x_2).
\end{align*}

The problem symmetric to (B.14)-(B.17) is

\begin{align*}
\text{Minimize} & \quad I(\hat{x}_2;x_2|\hat{x}_1), \\
nq(\hat{x}_2|x_2,\hat{x}_1)
\end{align*}

subject to

\begin{align*}
E \rho(x_2,\hat{x}_2) &\leq D(C), \\
q(\hat{x}_2|x_2,\hat{x}_1) &\geq 0 \quad \forall (x_2,\hat{x}_1,\hat{x}_2), \\
\sum_{\hat{x}_2} q(\hat{x}_2|x_1, x_2) &= 1 \quad \forall (\hat{x}_1, x_2).
\end{align*}

The solutions (B.14)-(B.17) and (B.18)-(B.21) are the same. (B.18)-(B.21) is the same as Example 2.7.1 of [3].
The minimum is equal to

\[ R(D_2) = C = C_b(x_2|\hat{x}_1) - C_b(D_2) \] \hspace{1cm} (B.22)

To compute \( C_b(x_2|\hat{x}_1) \) we need to know \( p(x_2|\hat{x}_1) \). We compute \( p(x_2|\hat{x}_1) \)

\[ p(x_2=0|\hat{x}_1=0) = p(x_2=0|x_1=0) \, p(x_1=0|\hat{x}_1=0) + 
\]

\[ + \, p(x_2=0|x_1=1) \, p(x_1=1|\hat{x}_1=0) = 
\]

\[ = p(x_2=0|x_1=0) \, \frac{q(\hat{x}_1=0|x_1=0) \, p(x_1=0)}{Q(\hat{x}_1=0)} + 
\]

\[ + \, p(x_2=0|x_1=1) \, \frac{q(\hat{x}_1=0|x_1=1) \, p(x_1=1)}{Q(\hat{x}_1=0)}. \] \hspace{1cm} (B.23)

Similarly we find

\[ p(x_2=1|\hat{x}_1=0) = p(x_2=1|x_1=1) \, \frac{q(\hat{x}_1=0|x_1=1) \, p(x_1=1)}{Q(\hat{x}_1=0)} + 
\]

\[ + \, p(x_2=1|x_1=0) \, \frac{q(\hat{x}_1=0|x_1=0) \, p(x_1=0)}{Q(\hat{x}_1=0)}. \] \hspace{1cm} (B.24)

\[ p(x_2=1|\hat{x}_1=1) = p(x_2=1|x_1=1) \, \frac{q(\hat{x}_1=1|x_1=1) \, p(x_1=1)}{Q(\hat{x}_1=1)} + 
\]

\[ + \, p(x_2=1|x_1=0) \, \frac{q(\hat{x}_1=1|x_1=0) \, p(x_1=0)}{Q(\hat{x}_1=1)}. \] \hspace{1cm} (B.25)

\[ p(x_2=0|\hat{x}_1=1) = p(x_2=0|x_1=0) \, \frac{q(\hat{x}_1=1|x_1=0) \, p(x_1=0)}{Q(\hat{x}_1=1)} + 
\]

\[ + \, p(x_2=0|x_1=1) \, \frac{q(\hat{x}_1=1|x_1=1) \, p(x_1=1)}{Q(\hat{x}_1=1)}. \] \hspace{1cm} (B.26)
Substituting the initial condition, the transition probabilities and
the solution of the first step in (B.23)-(B.26) we obtain

\[ p(x_2=0|\hat{x}_1=0) = \frac{1-p}{1+e^s} + \frac{pe^s}{1+e^s}, \]  
(B.27)

\[ p(x_2=1|\hat{x}_1=0) = \frac{(1-p)e^s}{1+e^s} + \frac{p}{1+e^s}, \]  
(B.28)

\[ p(x_2=0|\hat{x}_1=1) = \frac{(1-p)e^s}{1+e^s} + \frac{p}{1+e^s}, \]  
(B.29)

\[ p(x_2=1|\hat{x}_1=1) = \frac{1-p}{1+e^s} + \frac{pe^s}{1+e^s}. \]  
(B.30)

Using \( e^s = 0.00514 \) and \( p = 0.278 \) we get

\[ p(x_2=1|\hat{x}_1=1) = 0.7, \]  
(B.31)

\[ p(x_2=0|\hat{x}_1=1) = 0.3, \]  
(B.32)

\[ p(x_2=0|\hat{x}_1=0) = 0.7, \]  
(B.33)

\[ p(x_2=1|\hat{x}_1=0) = 0.3. \]  
(B.34)

The input probabilities are again 0.7 and 0.3 (as in the first shot);
consequently, the transition probabilities \( q(\hat{x}_2|x_2,\hat{x}_1) \) will be given by

\[ q(\hat{x}_2=0|x_2=0,\hat{x}_1=0) = q(\hat{x}_2=1|x_2=1,\hat{x}_1=1) = q(\hat{x}_1=2|x_1=1) = 0.9513, \]  
(B.35)

\[ q(\hat{x}_2=0|x_2=1,\hat{x}_1=1) = q(\hat{x}_2=1|x_2=0,\hat{x}_1=0) = q(\hat{x}_1=0|x_1=1) = 0.0487, \]  
(B.36)

\[ q(\hat{x}_2=1|x_2=0,\hat{x}_1=1) = q(\hat{x}_2=0|x_2=1,\hat{x}_1=0) = q(\hat{x}_1=1|x_1=0) = 0.105. \]  
(B.37)
\[ q(\hat{x}_2 = 0 | x_2 = 0, \hat{x}_1 = 1) = q(\hat{x}_2 = 1 | x_2 = 1, \hat{x}_1 = 0) = q(\hat{x}_1 = 0 | x_1 = 0) = 0.895 , \]

and

\[ Q(\hat{x}_2 = 0 | \hat{x}_1 = 0) = Q(\hat{x}_2 = 1 | \hat{x}_1 = 1) = 0.698 , \quad \text{(B.39)} \]

\[ Q(\hat{x}_2 = 1 | \hat{x}_1 = 0) = Q(\hat{x}_2 = 0 | \hat{x}_1 = 1) = 0.302 . \quad \text{(B.40)} \]

The minimum average distortion is

\[ D_2 = 0.0488 . \]

Third Shot; Time \( t = 3 \)

At time \( t = 3 \) we have to solve the problem

\[
\begin{align*}
\text{Minimize} & \quad E \rho(x_3, \hat{x}_3) , \\
& q(\hat{x}_3 | x_3, \hat{x}_2, \hat{x}_1) \\
\text{subject to} & \quad I(\hat{x}_3; x_3 | \hat{x}_2, \hat{x}_1) \leq C , \quad \text{(B.42)} \\
& q(\hat{x}_3 | x_3, \hat{x}_2, \hat{x}_1) \geq 0 \quad \forall (\hat{x}_1, \hat{x}_2, x_3) , \quad \text{(B.43)} \\
& \sum_{\hat{x}_3} q(\hat{x}_3 | x_3, \hat{x}_2, \hat{x}_1) = 1 \quad \forall (\hat{x}_1, \hat{x}_2, x_3) . \quad \text{(B.44)}
\end{align*}
\]

The problem symmetric to (B.41)-(B.44) is

\[
\begin{align*}
\text{Minimize} & \quad I(\hat{x}_3; x_3 | \hat{x}_1, \hat{x}_2) , \\
& q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) \\
\text{subject to} & \quad E \rho(x_3, \hat{x}_3) \leq D(C) , \quad \text{(B.46)} \\
& q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) \geq 0 \quad \forall (x_3, \hat{x}_3, \hat{x}_2, \hat{x}_1) , \quad \text{(B.47)} \\
& \sum_{\hat{x}_3} q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) = 1 \quad \forall (x_3, \hat{x}_1, \hat{x}_2) . \quad \text{(B.48)}
\end{align*}
\]
(B.41)-(B.44) and (B.45)-(B.48) have the same solution. (B.45)-(B.48)
is similar to Example 2.7.1 of [3]. Its minimum is equal to

\[
R(D_3) = C = H_b(x_3|\hat{x}_1,\hat{x}_2) - H_b(D_3). \tag{B.49}
\]

To compute \(H_b(x_3|\hat{x}_1,\hat{x}_2)\) we need to know \(p(x_3|\hat{x}_1,\hat{x}_2)\). The input probabilities are

\[
p(x_3=0|\hat{x}_1=0, \hat{x}_2=0) = p(x_3=0|x_2=0) \; p(x_2=0|\hat{x}_1=0, \hat{x}_2=0) \\
+ p(x_3=0|x_2=1) \; p(x_2=1|\hat{x}_2=0, \hat{x}_1=0) = \\
= p(x_3=0|x_2=0) \frac{q(\hat{x}_2=0|x_2=0, \hat{x}_1=0) \; p(x_2=0, \hat{x}_1=0)}{Q(\hat{x}_2=0|\hat{x}_1=0)} + \\
+ p(x_3=0 \; x_2=1) \frac{q(\hat{x}_2=0|x_2=1, \hat{x}_1=0) \; p(x_2=1, \hat{x}_1=0)}{Q(\hat{x}_2=0|\hat{x}_1=0)}, \tag{B.50}
\]

Similarly

\[
p(x_3=1|\hat{x}_1=0, \hat{x}_2=0) = p(x_3=1|x_2=0) \frac{q(\hat{x}_2=0|x_2=0, \hat{x}_1=0) \; p(x_2=0, \hat{x}_1=0)}{Q(\hat{x}_2=0|\hat{x}_1=0)} + \\
+ p(x_3=1|x_2=1) \frac{q(\hat{x}_2=0|x_2=1, \hat{x}_1=0) \; p(x_2=1, \hat{x}_1=0)}{Q(\hat{x}_2=0|\hat{x}_1=0)}, \tag{B.51}
\]

\[
p(x_3=0|\hat{x}_2=1, \hat{x}_1=0) = p(x_3=0|x_2=0) \frac{q(\hat{x}_2=1|x_2=0, \hat{x}_1=0) \; p(x_2=0, \hat{x}_1=0)}{Q(\hat{x}_2=1|\hat{x}_1=0)} + \\
+ p(x_3=0|x_2=1) \frac{q(\hat{x}_2=1|x_2=1, \hat{x}_1=0) \; p(x_2=1, \hat{x}_1=0)}{Q(\hat{x}_2=1|\hat{x}_1=0)}, \tag{B.52}
\]
\[ p(x_3 = 1 | \hat{x}_2 = 1 \hat{x}_1 = 0) = p(x_3 = 1 | x_2 = 1) \frac{q(\hat{x}_2 = 1 | x_2 = 1 \hat{x}_1 = 0) p(x_2 = 1 | \hat{x}_1 = 0)}{Q(\hat{x}_2 = 1 | \hat{x}_1 = 0)} + \\
+ p(x_3 = 1 | x_2 = 0) \frac{q(\hat{x}_2 = 1 | x_2 = 0 \hat{x}_1 = 0) p(x_2 = 0 | \hat{x}_1 = 0)}{Q(\hat{x}_2 = 1 | \hat{x}_1 = 0)} \\
(B.53) \]

\[ p(x_3 = 0 | \hat{x}_2 = 1 \hat{x}_1 = 1) = p(x_3 = 0 | x_2 = 0) \frac{q(\hat{x}_2 = 1 | x_2 = 0 \hat{x}_1 = 1) p(x_2 = 0 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 1 | \hat{x}_1 = 1)} + \\
+ p(x_3 = 0 | x_2 = 1) \frac{q(\hat{x}_2 = 1 | x_2 = 1 \hat{x}_1 = 1) p(x_2 = 1 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 1 | \hat{x}_1 = 1)} \\
(B.54) \]

\[ p(x_3 = 1 | \hat{x}_2 = 1 \hat{x}_1 = 1) = p(x_3 = 1 | x_2 = 0) \frac{q(\hat{x}_2 = 1 | x_2 = 0 \hat{x}_1 = 1) p(x_2 = 0 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 1 | \hat{x}_1 = 1)} + \\
+ p(x_3 = 1 | x_2 = 1) \frac{q(\hat{x}_2 = 1 | x_2 = 1 \hat{x}_1 = 1) p(x_2 = 1 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 1 | \hat{x}_1 = 1)} \\
(B.55) \]

\[ p(x_3 = 0 | \hat{x}_2 = 0 \hat{x}_1 = 1) = p(x_3 = 0 | x_2 = 0) \frac{q(\hat{x}_2 = 0 | x_2 = 0 \hat{x}_1 = 1) p(x_2 = 0 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 0 | \hat{x}_1 = 1)} + \\
+ p(x_3 = 0 | x_2 = 1) \frac{q(\hat{x}_2 = 0 | x_2 = 1 \hat{x}_1 = 1) p(x_2 = 1 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 0 | \hat{x}_1 = 1)} \\
(B.56) \]

\[ p(x_3 = 1 | \hat{x}_2 = 0 \hat{x}_1 = 1) = p(x_3 = 1 | x_2 = 0) \frac{q(\hat{x}_2 = 0 | x_2 = 0 \hat{x}_1 = 1) p(x_2 = 0 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 0 | \hat{x}_1 = 1)} + \\
+ p(x_3 = 1 | x_2 = 1) \frac{q(\hat{x}_2 = 0 | x_2 = 1 \hat{x}_1 = 1) p(x_2 = 1 | \hat{x}_1 = 1)}{Q(\hat{x}_2 = 0 | \hat{x}_1 = 1)} \\
(B.57) \]
Substituting the transition probabilities and the solution of the second step in (B.50)-(B.57) we get

\[
p(x_3=0|\hat{x}_2=0 \hat{x}_1) = p(x_3=1|\hat{x}_2=1 \hat{x}_1) =
\]

\[
= p(x_2=0|\hat{x}_1=0) = p(x_2=1|\hat{x}_1=1) = p(x_1=1) = 0.7 ,
\]

(B.58)

\[
p(x_3=0|\hat{x}_2=1 \hat{x}_1) = p(x_3=1|\hat{x}_2=0 \hat{x}_1) =
\]

\[
= p(x_2=0|\hat{x}_1=1) = p(x_2=1|\hat{x}_1=0) = p(x_1=0) = 0.3 .
\]

(B.59)

Again the input probabilities are 0.7 and 0.3; hence, the transition probabilities \( q(\hat{x}_3|x_3\hat{x}_2\hat{x}_1) \) will be given by

\[
q(\hat{x}_3=0|x_3=0 \hat{x}_2=0 \hat{x}_1) = q(\hat{x}_3=1|x_3=1 \hat{x}_2=1 \hat{x}_1) = 0.9513 ,
\]

(B.60)

\[
q(\hat{x}_3=1|x_3=0 \hat{x}_2=0 \hat{x}_1) = q(\hat{x}_3=0|x_3=1 \hat{x}_2=1 \hat{x}_1) = 0.0487 ,
\]

(B.61)

\[
q(\hat{x}_3=0|x_3=0 \hat{x}_2=1 \hat{x}_1) = q(\hat{x}_3=1|x_3=1 \hat{x}_2=0 \hat{x}_1) = 0.895 ,
\]

(B.62)

\[
q(\hat{x}_3=1|x_3=0 \hat{x}_2=1 \hat{x}_1) = q(\hat{x}_3=0|x_3=1 \hat{x}_2=0 \hat{x}_1) = 0.105 ,
\]

(B.63)

and

\[
Q(\hat{x}_3=0|\hat{x}_2=0 \hat{x}_1) = Q(\hat{x}_3=1|\hat{x}_2=1 \hat{x}_1) = 0.698 ,
\]

(B.64)

\[
Q(\hat{x}_3=0|\hat{x}_2=1 \hat{x}_1) = Q(\hat{x}_3=1|\hat{x}_2=0 \hat{x}_1) = 0.302 .
\]

(B.65)

The minimum average distortion is

\[
D_3 = 0.0488 .
\]

(B.66)

It can be easily shown that the input probabilities \( p(x_t|\hat{x}_{t-1}) \) for the multishot problem are always 0.7 and 0.3. Consequently, the test channel resulting from the solution of (3.4.9)-(3.4.12) for the source of Example 3.4.1 is time invariant.
APPENDIX C

Solution of (3.4.19)-(3.4.28)

We show that the system of figure 3.5 with the encoding and decoding strategies described in Example 3.4.1 achieves the solution of (3.4.19)-(3.4.28).

First we check the information constraint

\[ I(\hat{x}_t; x_t | \hat{x}_{t-1}) = \sum_{\hat{x}_{t-1}} p(\hat{x}_{t-1}) \sum_{x_t, \hat{x}_t} p(x_t | \hat{x}_{t-1}) \cdot \]

\[ \cdot q(\hat{x}_t | x_t \hat{x}_{t-1}) \log \frac{q(\hat{x}_t | x_t \hat{x}_{t-1})}{q(\hat{x}_t | \hat{x}_{t-1})}. \quad (C.1) \]

For fixed \( \hat{x}_{t-1} \),

\[ \sum_{x_t, \hat{x}_t} p(x_t | \hat{x}_{t-1}) q(\hat{x}_t | x_t \hat{x}_{t-1}) \log \frac{q(\hat{x}_t | x_t \hat{x}_{t-1})}{q(\hat{x}_t | \hat{x}_{t-1})} = \]

\[ = \sum_{x_1, \hat{x}_1} p(x_1) q(\hat{x}_1 | x_1) \log \frac{q(\hat{x}_1 | x_1)}{q(\hat{x}_1)} = c, \quad (C.2) \]

because

\[ p(x_t = 0 | \hat{x}_{t-1} = 0, x_{t-2}) = p(x_t = 1 | \hat{x}_{t-1} = 1, x_{t-2}) = p(x_1 = 1) = 0.7, \quad (C.3) \]

\[ p(x_t = 0 | \hat{x}_{t-1} = 1, \hat{x}_{t-2}) = p(x_t = 1 | \hat{x}_{t-1} = 0, \hat{x}_{t-2}) = p(x_1 = 0) = 0.3, \quad (C.4) \]

\[ q(\hat{x}_t = 1 | x_t = 1, \hat{x}_{t-1} = 1, \hat{x}_{t-2}) = q(\hat{x}_t = 0 | x_t = 0, \hat{x}_{t-1} = 0, \hat{x}_{t-2}) = \]

\[ = q(\hat{x}_1 = 1 | x_1 = 1), \quad (C.5) \]

\[ q(\hat{x}_t = 1 | x_t = 1, \hat{x}_{t-1} = 0, \hat{x}_{t-2}) = q(\hat{x}_t = 0 | x_t = 0, \hat{x}_{t-1} = 1, \hat{x}_{t-2}) = \]

\[ = q(\hat{x}_1 = 0 | x_1 = 0), \quad (C.6) \]
and $p(x_1)$ matches the channel. Hence,

Hence

$$I(\hat{x}_t; x_t | \hat{x}_{t-1}) = \sum_{\hat{x}_{t-1}} p(\hat{x}_{t-1}) C = C, \ \forall t, \ t=1,2,3,$$

and the information constraint is satisfied. The error is

$$D'_t = \sum_{\hat{x}_{t-1}} p(\hat{x}_{t-1}) \sum_{x_t, \hat{x}_t} q(\hat{x}_t | x_t \hat{x}_{t-1}) p(x_t | \hat{x}_{t-1}) \rho(x_t, \hat{x}_t). \ (C.7)$$

For fixed $\hat{x}_{t-1}$,

$$\sum_{x_t, \hat{x}_t} p(x_t | \hat{x}_{t-1}) q(\hat{x}_t | x_t \hat{x}_{t-1}) \rho(x_t, \hat{x}_t) =$$

$$= \sum_{x_1, \hat{x}_1} p(x_1) q(\hat{x}_1 | x_1) \rho(x_1, \hat{x}_1) = D = 0.0488,$$

because of (C.3)-(C.6).

Hence

$$D'_t = \sum_{\hat{x}_{t-1}} p(\hat{x}_{t-1}) D = D = 0.0488 \ \forall t \ \text{QED}.$$
APPENDIX D

Solution of (4.2.47)-(4.2.53)

According to our formulation the problem (4.2.47)-(4.2.53) decomposes into three static optimization problems.

At time $t=1$ we have to

$$\text{Minimize } E(x_1 - \hat{x}_1)^2, \quad q(\hat{x}_1|\mathbf{x}_1)$$

subject to

$$I(\hat{x}_1; \mathbf{x}_1) \leq C,$$

$$\mathbf{f}_q(\hat{x}_1|\mathbf{x}_1)\mathbf{d}\hat{x}_1 = 1.$$

At time $t=2$ we have to

$$\text{Minimize } E(x_2 - \hat{x}_2)^2, \quad q(\hat{x}_2|\mathbf{x}_1\mathbf{x}_2)$$

subject to

$$I(\hat{x}_2; \mathbf{x}_2|\mathbf{x}_1) \leq C,$$

$$\mathbf{f}_q(\hat{x}_2|\mathbf{x}_1\mathbf{x}_2)\mathbf{d}\hat{x}_2 = 1.$$

At time $t=3$ we have to

$$\text{Minimize } E(x_3 - \hat{x}_3)^2, \quad q(\hat{x}_3|\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3)$$

subject to

$$I(\hat{x}_3; \mathbf{x}_3|\mathbf{x}_1\mathbf{x}_2) \leq C,$$

$$\mathbf{f}_q(\hat{x}_3|\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3)\mathbf{d}\hat{x}_3 = 1.$$
Time $t = 1$

The problem symmetric to (D.1)-(D.3) is

$$\text{Minimize } I(\hat{x}_1; x_1),$$
$$q(\hat{x}_1 | x_1)$$

subject to

$$E(\hat{x}_1 - x_1)^2 \leq D_1(C),$$

$$\int q(\hat{x}_1 | x_1) d\hat{x}_1 = 1.$$ \hspace{1cm} (D.10, D.11, D.12)

The curves resulting from (D.1)-(D.3) and (D.10)-(D.12) by varying $C$ are the same. (D.10)-(D.12), considered for the source described in Example 4.2.1, is the same as the example of Section 4.3.3 of [3]. Its solution is given by

$$C = R(D_1) = \begin{cases} \frac{1}{2} \ln \frac{\sigma_1^2}{D_1}, & 0 \leq D_1 \leq \sigma_1^2, \\ 0, & D_1 > \sigma_1^2, \end{cases}$$ \hspace{1cm} (D.13)

and the test channel is

$$q^*(\hat{x}_1 | x_1) = \frac{1}{\sqrt{2\pi B_1 D_1}} \exp \left\{ -\frac{(\hat{x}_1 - B_1 x_1)^2}{2B_1 D_1} \right\},$$ \hspace{1cm} (D.14)

where

$$B_1 = 1 - \frac{D_1}{\sigma_1^2}.$$ \hspace{1cm} (D.15)

Since (D.1)-(D.3) and (D.10)-(D.12) have the same solution (for fixed $C$), we can use (D.13) to compute $D_1$ when the channel capacity $C$ is given.
\[ D_1 = \sigma_1^2 e^{-2C}. \]  

(D.16)

\( D_1 \) can be substituted into (D.14) and (D.15) to give the test channel.

**Time \( t = 2 \)**

The problem symmetric to (D.4)-(D.6) is

\[
\text{Minimize } I(\delta_2; x_1 | x) \quad \text{subject to} \quad \mathbb{E}(x_2 - \delta_2)^2 \leq D_2(C), \quad \mathbb{E}(\delta_2 | x_1 x_2) \quad (D.17)
\]

subject to

\[ E(x_2 - \delta_2)^2 \leq D_2(C), \]  

(D.18)

\[ \int q(\delta_2 | x_1 x_2) d\delta_2 = 1. \]  

(D.19)

The curves resulting from (D.4)-(D.6) and (D.17)-(D.19) by varying \( C \) are the same. (D.17)-(D.19), considered for the source of Example 4.2.1, is the same as the example of Section 4.3.3 of [3]. Its solution is given by

\[
C = R(D_2) = \begin{cases} 
\frac{1}{2} \ln \frac{N_1^2}{D_2}, & 0 \leq D_2 \leq N_1^2, \\
0, & D_2 > N_1^2,
\end{cases} \quad (D.20)
\]

and the test channel is

\[
q(\delta_2 | x_1 x_2) = \frac{1}{\sqrt{2\pi B_2 D_2}} \exp \left\{ - \frac{[\delta_2 - B_2(x_2 - x_1)]^2}{2B_2 D_2} \right\}, \quad (D.21)
\]

where

\[
B_2 = 1 - \frac{D_2}{N_1^2}. \quad (D.22)
\]
Since (D.4)-(D.6) and (D.17)-(D.19) have the same solution (for fixed C), we can use (D.20) to compute $D_2$ when the channel capacity $C$ is given.

$$D_2 = N_1^2 e^{-2C}.$$  \hspace{1cm} (D.23)

$D_2$ can be substituted into (D.21),(D.22) to give the test channel.

**Time $t = 3$**

The solution of the problem at time $t = 3$ is similar to the solutions at $t = 1, 2$. The problem symmetric to (D.7)-(D.9) is

$$\text{Minimize } I(\hat{x}_3; x_3 | x_1 x_2),$$

subject to

$$E(x_3 - \hat{x}_3)^2 \leq D_3(C).$$  \hspace{1cm} (D.25)

$$\int q(\hat{x}_3 | x_1 x_2 x_3) d\hat{x}_3 = 1.$$  \hspace{1cm} (D.26)

The curves, resulting from (D.7)-(D.9) and (D.24)-(D.26) by varying $C$, are the same. (D.24)-(D.26), considered for the source of Example 4.2.1, is the same as the example of Section 4.3.3 of [3]. The minimum is given by

$$C = R(D_3) = \begin{cases} \frac{1}{2} \ln \frac{N_2^2}{D_3}, & 0 \leq D_3 \leq N_2^2, \\ 0, & D_3 > N_2^2, \end{cases}$$  \hspace{1cm} (D.27)

and the test channel is

$$q^*(\hat{x}_3 | x_1 x_2 x_3) = \frac{1}{\sqrt{2\pi B_3 D_3}} \exp \left\{ - \frac{[\hat{x}_3 - B_3(x_3 - x_2)]^2}{2B_3 D_3} \right\}.$$  \hspace{1cm} (D.28)
where
\[ B_3 = 1 - \frac{D_3}{N_2^2}. \]  \hfill (D.29)

Since (D.7)-(D.9) and (D.24)-(D.26) have the same solution, we can use (D.27) to compute \( D_3 \) when the channel capacity \( C \) is given.

\[ D_3 = N_2^2 e^{-2C}. \]  \hfill (D.30)

\( D_3 \) can be substituted into (D.28) and (D.29) to give the test channel.
APPENDIX E

Solution of (4.2.92)-(4.2.98)

First Shot; Time \( t = 1 \)

At time \( t = 1 \) we have to solve the problem

\[
\text{Minimize } \quad E(x_1 - \hat{x}_1)^2, \quad \text{subject to}
\]

\[
I(\hat{x}_1; x_1) \leq C ,
\]

\[
\int q(\hat{x}_1 | x_1) d\hat{x}_1 = 1.
\]

(E.1)-(E.3) is the same as (D.1)-(D.3). Its solution is given by (D.13)-(D.16). The backward channel is given by ([3], Section 4.3.3)

\[
p(x_1 | \hat{x}_1) = \frac{1}{\sqrt{2\pi D_1}} \exp \left\{ -\frac{(x_1 - \hat{x}_1)^2}{2D_1} \right\}.
\]

Second Shot; Time \( t = 2 \)

At \( t = 2 \) we have to

\[
\text{Minimize } \quad E(x_2 - \hat{x}_2)^2, \quad \text{subject to}
\]

\[
I(\hat{x}_2; x_2 | \hat{x}_1) \leq C ,
\]

\[
\int q(\hat{x}_2 | x_2 \hat{x}_1) d\hat{x}_2 = 1.
\]

The problem symmetric to (E.5)-(E.7) is

\[
\text{Minimize } \quad I(\hat{x}_2; x_2 | \hat{x}_1),
\]

150
subject to
\[ E(x_2 - \hat{x}_2)^2 \leq D_2(C), \quad (E.9) \]
\[ \int q(\hat{x}_2 | x_2 \hat{x}_1) d\hat{x}_2 = 1. \quad (E.10) \]

The curves resulting from (E.5)-(E.7) and (E.8)-(E.10) by varying \( C \) are the same. In order to compute the solution of (E.8)-(E.10) we have to know \( p(x_2 | \hat{x}_1) \)

\[ p(x_2 | \hat{x}_1) = \int p(x_2 | x_1) p(x_1 | \hat{x}_1) dx_1, \quad (E.11) \]

\[ p(x_2 | x_1) = \frac{1}{\sqrt{2\pi N_1}} \exp \left\{ -\frac{(x_2 - x_1)^2}{2N_1^2} \right\}, \quad (E.12) \]

and \( p(x_2 | \hat{x}_1) \) is given by (E.4). Because of (E.12) and (E.4), (E.11) gives

\[ p(x_2 | \hat{x}_1) = \frac{1}{\sqrt{2\pi D_1}} \frac{1}{\sqrt{2\pi N_1}} \int \exp \left\{ -\frac{(x_2 - x_1)^2}{2N_1^2} \right\} \exp \left\{ -\frac{(x_1 - \hat{x}_1)^2}{2D_1} \right\} dx_1 = \]

\[ = \frac{1}{\sqrt{2\pi N_1}} \frac{1}{\sqrt{2\pi D_1}} \exp \left\{ -\frac{x_1^2}{2N_1^2} \right\} \exp \left\{ -\frac{\hat{x}_1^2}{2D_1} \right\}. \]

\[ \cdot \int \exp \left\{ -x_1^2 \left( \frac{1}{2N_1^2} + \frac{1}{2D_1} \right) + 2x_1 \left( \frac{\hat{x}_1}{2D_1} + \frac{x_2}{2N_1^2} \right) \right\} dx_1 = \]

\[ = \frac{1}{\sqrt{2\pi N_1}} \frac{1}{\sqrt{2\pi D_1}} \exp \left\{ -\frac{x_2^2}{2N_1^2} \right\} \exp \left\{ -\frac{\hat{x}_1^2}{2D_1} \right\} \exp \left\{ \frac{x_2}{2N_1^2} + \frac{\hat{x}_1}{2D_1} \right\} \left( \frac{2N_1^2 D_1}{2N_1^2 + 2D_1} \right). \]

\[ \cdot \int dx_1 \exp \left\{ -x_1^2 \left( \frac{1}{2N_1^2} + \frac{1}{2D_1} \right) + 2x_1 \left( \frac{x_2}{2N_1^2} + \frac{\hat{x}_1}{2D_1} \right) - \left( \frac{x_2}{2N_1^2} + \frac{\hat{x}_1}{2D_1} \right)^2 \frac{2N_1^2 D_1}{2N_1^2 + 2D_1} \right\} = \]
\[
= \frac{1}{\sqrt{2\pi N_1}} \frac{1}{\sqrt{2\pi D_1}} \exp \left\{ -\frac{x_2^2}{2N_1} \right\} \exp \left\{ -\frac{\hat{x}_1^2}{2D_1} \right\}.
\]

\[
\cdot \exp \left\{ \left( \frac{x_2}{2N_1} + \frac{\hat{x}_1}{2D_1} \right)^2 \frac{2N_1^2 D_1}{N_1^2 + D_1} \right\} \Rightarrow
\]

\[
p(x_2 | \hat{x}_1) = \frac{1}{\sqrt{2\pi (N_1^2 + D_1)}} \exp \left\{ -\frac{(x_2 - \hat{x}_1)^2}{2(N_1^2 + D_1)} \right\}.
\]

(E.13)

\( p(x_2 | \hat{x}_1) \) is Gaussian with

\[
\text{mean} = \hat{x}_1 ,
\]

(E.14)

and

\[
\text{variance} = D_1 + N_1^2 .
\]

(E.15)

Since \( p(x_2 | \hat{x}_1) \) is Gaussian, (E.8)-(E.10) is the same as the example of Section 4.3.3 of [3]. The minimum is given by

\[
C = R(D'_2) = \begin{cases} 
\frac{1}{2} \ln \frac{N_1^2 + D_1}{D'_2} , & 0 \leq D'_2 \leq N_1^2 + D_1 , \\
0 , & \text{otherwise} .
\end{cases}
\]

(E.16)

The forward and backward test channels are given by

\[
q(\hat{x}_2 | \hat{x}_1 x_2) = \frac{1}{\sqrt{2\pi B_2 D'_2}} \exp \left\{ -\frac{[\hat{x}_2 - B_2(x_2 - \hat{x}_1)^2]}{2B_2 D'_2} \right\} ,
\]

(E.17)

and

\[
p(x_2 | \hat{x}_1 \hat{x}_2) = \frac{1}{\sqrt{2\pi D'_2}} \exp \left\{ -\frac{[x_2 - (\hat{x}_1 + \hat{x}_2)]^2}{2D'_2} \right\} ,
\]

(E.18)

respectively, where

\[
B_2 = 1 - \frac{D'_2}{D_1 + N_1^2} .
\]

(E.19)
Since from the solution of the first step we must have ([3], Section 4.3.3),

Variance of noise in the channel = \( B_1 D_1 \),

and moreover by assumption

\[
\text{Variance of noise in the channel} = \frac{N_1^2}{\sigma_1^2} (\sigma_1^2 - N_1^2),
\]

we get

\[
B_1 D_1 = \left(1 - \frac{D_1}{\sigma_1^2}\right) D_1 = \frac{N_1^2}{\sigma_1^2} (\sigma_1^2 - N_1^2) \quad \text{or}
\]

\[
D_1 + N_1^2 = \sigma_1^2.
\]  \( \text{(E.20)} \)

Because of (E.20), (E.13), (E.16) and (E.19) give

\[
p(x_2 | \hat{x}_1) = \frac{1}{\sqrt{2\pi} \sigma_1} \exp \left\{ - \frac{(x_2 - \hat{x}_1)^2}{2 \sigma_1^2} \right\},
\]  \( \text{(E.21)} \)

\[
C = R(D_2^1) = \begin{cases} \frac{1}{2} \ln \frac{\sigma_1^2}{D_2^1}, & 0 \leq D_2^1 \leq \sigma_1^2, \\ 0, & D_2^1 > \sigma_1^2, \end{cases}
\]  \( \text{(E.22)} \)

\[
B_2 = 1 - \frac{D_2^1}{\sigma_1^2}.
\]  \( \text{(E.23)} \)

(E.22) gives

\[
D_2^1 = \sigma_1^2 e^{-2C}.
\]  \( \text{(E.24)} \)

\( D_2^1 \) can be substituted into (E.17), (E.18) to give the forward and backward test channel for \( t = 2 \).
Third Shot; Time \( t = 3 \)

At \( t = 3 \) we have to

\[
\text{Minimize } \mathbb{E}(x_3 - \hat{x}_3)^2, \quad q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) \tag{E.25}
\]

subject to

\[
I(\hat{x}_3; x_3 | \hat{x}_1, \hat{x}_2) \leq C, \tag{E.26}
\]

\[
\int q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) \, d\hat{x}_3 = 1. \tag{E.27}
\]

The problem symmetric to (E.25)-(E.27) is

\[
\text{Minimize } I(\hat{x}_3; x_3 | \hat{x}_1, \hat{x}_2), \quad q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) \tag{E.28}
\]

subject to

\[
\mathbb{E}(x_3 - \hat{x}_3)^2 \leq D_3(C), \tag{E.29}
\]

\[
\int q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) \, d\hat{x}_3 = 1. \tag{E.30}
\]

The curves resulting from (E.25)-(E.27) and (E.28)-(E.30) by varying \( C \) are the same.

In order to compute the solution of (E.28)-(E.30) we have to know

\[
p(x_3 | \hat{x}_1, \hat{x}_2) = \int p(x_3 | x_2) \, p(x_2 | \hat{x}_1, \hat{x}_2) \, dx_2, \tag{E.31}
\]

\[
p(x_3 | x_2) = \frac{1}{\sqrt{2\pi N_2}} \exp \left\{ - \frac{(x_3 - x_2)^2}{2N_2^2} \right\}, \tag{E.32}
\]

and \( p(x_2 | \hat{x}_1, \hat{x}_2) \) is given by (E.18). Because of (E.18) and (E.32), (E.30) gives
Third Shot; Time $t = 3$

At $t = 3$ we have to

$$\begin{align*}
\text{Minimize} & \quad \mathbb{E}(\hat{x}_3 - \hat{x}_3)^2, \\
& \quad q(\hat{x}_3 | x_3 \hat{x}_1 \hat{x}_2) \\
\text{subject to} & \quad I(\hat{x}_3 ; x_3 | \hat{x}_1 \hat{x}_2) \leq C, \\
& \quad \int q(\hat{x}_3 | x_3 \hat{x}_1 \hat{x}_2) d\hat{x}_3 = 1.
\end{align*}$$

(E.25) (E.26) (E.27)

The problem symmetric to (E.25)-(E.27) is

$$\begin{align*}
\text{Minimize} & \quad I(\hat{x}_3 ; x_3 | \hat{x}_1 \hat{x}_2), \\
& \quad q(\hat{x}_3 | x_3 \hat{x}_1 \hat{x}_2) \\
\text{subject to} & \quad \mathbb{E}(x_3 - \hat{x}_3)^2 \leq D_3(C), \\
& \quad \int q(\hat{x}_3 | x_3 \hat{x}_1 \hat{x}_2) d\hat{x}_3 = 1.
\end{align*}$$

(E.28) (E.29) (E.30)

The curves resulting from (E.25)-(E.27) and (E.28)-(E.30) by varying $C$ are the same.

In order to compute the solution of (E.28)-(E.30) we have to know

$$p(x_3 | \hat{x}_1 \hat{x}_2) = \int p(x_3 | x_2) \ p(x_2 | \hat{x}_1 \hat{x}_2) dx_2,$$

(E.31)

and $p(x_2 | \hat{x}_1 \hat{x}_2)$ is given by (E.18). Because of (E.18) and (E.32), (E.30) gives
\[ p(x_3 | \hat{x}_1, \hat{x}_2) = \int dx_2 \frac{1}{\sqrt{2\pi N_2}} \exp \left\{ -\frac{(x_3 - x_2)^2}{2N_2^2} \right\} \frac{1}{\sqrt{2\pi D_2^I}} \exp \left\{ -\frac{(x_2 - (\hat{x}_1 + \hat{x}_2))^2}{2D_2^I} \right\} = \] 

\[ = \frac{1}{\sqrt{2\pi N_2}} \frac{1}{\sqrt{2\pi D_2^I}} \exp \left\{ -\frac{x_3^2}{2N_2^2} \right\} \exp \left\{ -\frac{(\hat{x}_1 + \hat{x}_2)^2}{2D_2^I} \right\} \cdot \exp \left\{ \frac{(x_3 - x_2)^2}{2N_2^2} \frac{2N_2^2D_2^I}{N_2^2 + D_2^I} \right\} \cdot \int \exp \left\{ \left( \frac{1}{2N_2^2} + \frac{1}{2D_2^I} \right)x_2^2 + 2x_2 \left( \frac{x_3}{2N_2^2} + \frac{(\hat{x}_1 + \hat{x}_2)}{2D_2^I} \right) - \left( \frac{x_3}{2N_2^2} + \frac{\hat{x}_1 + \hat{x}_2}{2D_2^I} \right) \frac{2N_2^2D_2^I}{N_2^2 + D_2^I} \right\} dx_2 = \] 

\[ = \frac{1}{\sqrt{2\pi N_2}} \frac{1}{\sqrt{2\pi D_2^I}} \exp \left\{ -\frac{x_3^2}{2N_2^2} \left( 1 - \frac{D_2^I}{N_2^2 + D_2^I} \right) \right\} \exp \left\{ \frac{(\hat{x}_1 + \hat{x}_2)^2}{2D_2^I} \left( 1 - \frac{N_2^2}{D_2^I + N_2^2} \right) + \frac{2(\hat{x}_1 + \hat{x}_2)x_3}{N_2^2 + D_2^I} \right\} \] 

or

\[ p(x_3 | \hat{x}_1, \hat{x}_2) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{N_2^2 + D_2^I}{2}}} \exp \left\{ -\frac{(x_3 - (\hat{x}_1 + \hat{x}_2))^2}{2(N_2^2 + D_2^I)} \right\}. \quad (E.33) \]

\[ p(x_3 | \hat{x}_1, \hat{x}_2) \] is Gaussian with

\[ \text{mean} = \hat{x}_1 + \hat{x}_2, \quad (E.34) \]

and

\[ \text{variance} = \frac{N_2^2 + D_2^I}{2}. \quad (E.35) \]
Since \( p(x_3 | x_1 x_2) \) is Gaussian (E.28)-(E.30) is the same as the example of Section 4.3.3 of [3]. The minimum is given by

\[
C = R(D'_3) = \begin{cases} 
\frac{1}{2} \ln \frac{N_2^2 + D'_2}{D'_3}, & 0 \leq D'_3 \leq N_2^2 + D'_2, \\
0, & D'_3 > N_2^2 + D'_2.
\end{cases} \tag{E.36}
\]

The forward and backward channels are given by

\[
q(\hat{x}_3 | x_3, \hat{x}_1, \hat{x}_2) = \frac{1}{\sqrt{2\pi B_3 D'_3}} \exp \left\{ -\frac{\{\hat{x}_3 - B_3 [x_3 - (\hat{x}_1 + \hat{x}_2)]\}^2}{2B_3 D'_3} \right\}, \tag{E.37}
\]

and

\[
p(x_3 | \hat{x}_1, \hat{x}_2, \hat{x}_3) = \frac{1}{\sqrt{2\pi D'_3}} \exp \left\{ -\frac{[x_3 - (\hat{x}_1 + \hat{x}_2 + \hat{x}_3)]^2}{2D'_3} \right\}, \tag{E.38}
\]

respectively, where

\[
B_3 = 1 - \frac{D'_3}{D'_2 + N_2^2}. \tag{E.39}
\]

Since from the solution of the problem at \( t = 2 \) we must have ([3], Section 4.3.3),

\[
\text{Variance of noise in the channel} = B_2 D'_2,
\]

and moreover by assumption

\[
\text{Variance of noise in the channel} = \frac{N_2^2}{\sigma_1^2} (\sigma_1^2 - N_2^2),
\]

we get

\[
B_2 D'_2 = \left(1 - \frac{D'_2}{\sigma_1^2}\right) D'_2 = \frac{N_2^2}{\sigma_1^2} (\sigma_1^2 - N_2^2) \quad \text{or}
\]

\[
D'_2 + N_2^2 = \sigma_1^2. \tag{E.40}
\]
Because of (E.40), (E.33), (E.36), and (E.39) give

\[ p(x_3 | \hat{x}_1, \hat{x}_2) = \frac{1}{\sqrt{2\pi\sigma_1}} \exp \left\{ -\frac{[x_3 - (\hat{x}_1 + \hat{x}_2)]^2}{2\sigma_1^2} \right\} , \quad (E.41) \]

\[ C = R(D'_3) = \begin{cases} 
\frac{1}{2} \ln \frac{\sigma_1^2}{D'_3} , & 0 \leq D'_3 \leq \sigma_1^2 , \\
0 , & D'_3 > \sigma_1^2 , 
\end{cases} \quad (E.42) \]

\[ B_3 = 1 - D'_3 \frac{3}{\sigma_1^2} . \quad (E.43) \]

From (E.43) we get

\[ D'_3 = \sigma_1^2 e^{-2C} . \quad (E.44) \]

\( D'_3 \) can be substituted into (E.37), (E.38) to give the forward and backward test channel for \( t = 3 \).
APPENDIX F

We show that the system of Figure 4.1 with the encoding and decoding strategies described in Example 4.2.2 achieves the solution of (4.2.92)-(4.2.98).

At first we check the information constraint. For $t=1$ we have

$$I(\hat{x}_1;x_1) = \int p(x_1) q(\hat{x}_1|x_1) \log \frac{q(\hat{x}_1|x_1)}{Q(\hat{x}_1)} \, dx_1 d\hat{x}_1 = C \quad (F.1)$$

because the probability density of $x_1$ scaled by $\beta_1$ matches the channel.

For $t=2,3$ we have

$$I(\hat{x}_t;x_t|\hat{x}_{t-1}) = \int dx_{t-1} p(\hat{x}_{t-1}) \int dx_t d\hat{x}_t p(x_t|\hat{x}_{t-1})$$

$$q(\hat{x}_t|x_t) \log \frac{q(\hat{x}_t|x_t\hat{x}_{t-1})}{Q(\hat{x}_t)} \quad (F.2)$$

But

$$\int p(x_t|\hat{x}_{t-1}) q(\hat{x}_t|x_t\hat{x}_{t-1}) \log \frac{q(\hat{x}_t|x_t\hat{x}_{t-1})}{Q(\hat{x}_t)} \, dx_t d\hat{x}_t =$$

$$= C \text{ for } t=2,3 \quad (F.3)$$

because the conditional density $p(x_t|\hat{x}_{t-1})$ of $x_t$ scaled by $\beta_t$ ($t=2,3$) matches the channel.

Consequently

$$I(\hat{x}_t;x_t|\hat{x}_{t-1}) = \int p(\hat{x}_{t-1}) d\hat{x}_{t-1} C = C \quad (F.4)$$

Thus the information constraint is satisfied for all $t$, ($t=1,2,3$), hence the encoding-decoding strategies described in Example 4.2.2 are admissible.

The average error, resulting when these strategies are used, is:

At time $t=1$

$$D''_1 = \int p(x_1) q(\hat{x}_1|x_1)(x_1-\hat{x}_1)^2 \, dx_1 d\hat{x}_1 = D'_1 \quad (F.5)$$
(where \( D^i_1 \) is given by (4.2.99)). The last equality in (F.5) holds because \( q(\hat{x}_t | x_t) \) is equal to the test channel for \( t=1 \).

At time \( t=2,3 \)

\[
D''_t = \int d\hat{x}_{t-1} \int p(\hat{x}_t | \hat{x}_{t-1}) \int p(x_t | \hat{x}_{t-1}) q(\hat{x}_t | x_t \hat{x}_{t-1}) (x_t - \hat{x}_t)^2
\]

\[
dx_t d\hat{x}_t = D^i_t \tag{F.6}
\]

where \( D^i_t \) is given by (4.2.104) and (4.2.109). The last equality in (F.6) holds because \( q(\hat{x}_t | x_t \hat{x}_{t-1}) \) is equal to the test channel for \( t=2,3 \).
APPENDIX G

Solution of (5.2.22)-(5.2.26)

We have to solve two static optimization problems.

First Shot; Time \( t = 1 \)

At \( t = 1 \) we have to

\[
\text{Minimize } E(x_1 - \hat{x}_1)^2, \quad q(\hat{x}_1 | x_1) \quad (G.1)
\]

subject to

\[
I(\hat{x}_1; x_1) \leq I(y_1; x_1), \quad (G.2)
\]

\[
\int q(\hat{x}_1 | x_1) d\hat{x}_1 = 1. \quad (G.3)
\]

From the dynamics of the source and the observations we know that

\[
y_1 \sim N(0, \Sigma^2 + v_1^2). \quad (G.4)
\]

Consequently,

\[
C_1 \triangleq I(y_1; x_1) = H(y_1) - H(y_1 | x_1) = H(y_1) - H(v_1) =
\]

\[
= \frac{1}{2} \ln \frac{\Sigma^2 + v_1^2}{v_1^2}. \quad (G.5)
\]

We also know, [3], that for the Gaussian source

\[
C_1 = R(D_1) = \frac{1}{2} \ln \frac{\Sigma^2}{D_1}, \quad (G.5)
\]

where \( D_1 = \min E(x_1 - \hat{x}_1)^2. \)
Equating (G.5) and (G.6) we get

\[
\frac{1}{2} \ln \frac{\Sigma^2}{D_1} = \frac{1}{2} \ln \frac{\Sigma^2 + v_1^2}{v_1^2}, \quad \text{or}
\]

\[
D_1 = \Sigma^2 \frac{v_1^2}{v_1^2 + \Sigma^2}.
\]  \hspace{1cm} (G.7)

(G.7) provides a lower bound on the filtering problem at time \( t = 1 \).

**Second Shot; Time \( t = 2 \)**

At time \( t = 2 \) we have to

\[
\begin{align*}
\text{Minimize} & \quad E(x_2 - \hat{x}_2)^2, \\
q(\hat{x}_2 | x_2 y_1) & \quad (G.8)
\end{align*}
\]

subject to

\[
I(\hat{x}_2; x_2 | y_1) \leq I(y_2; x_2 | y_1) \triangleq C_2, \quad (G.9)
\]

\[
\int q(\hat{x}_2 | x_2 y_1) d\hat{x}_2 = 1. \quad (G.10)
\]

We have to compute the quantity

\[
I(y_2; x_2 | y_1) = H(y_2 | y_1) - H(y_2 | x_2 y_1) = H(y_2 | y_1) - H(y_2 | x_2). \quad (G.11)
\]

Because of (5.2.14) and (5.2.17)

\[
H(y_2 | x_2) = H(V_2) = \frac{1}{2} \ln 2\pi \sigma V_2^2. \quad (G.12)
\]

We have to find an expression for \( p(y_2 | y_1) \) in order to compute \( H(y_2 | y_1) \)

\[
p(y_2 | y_1) = \int p(y_2 | x_2) p(x_2 | y_1) dx_2 = \int p(y_2 - x_2) p(x_2 | y_1) dx_2, \quad (G.13)
\]
\[ p(y_2 \mid x_2) = p(v_2) = \frac{1}{\sqrt{2\pi v_2}} \exp \left\{ -\frac{u_2}{2v_2} \right\}. \] (G.14)

Thus, we only have to compute \( p(x_2 \mid y_1) \) in order to find \( p(y_2 \mid y_1) \)

\[ p(x_2 \mid y_1) = \int p(x_2 \mid x_1) p(x_1 \mid y_1) dx_1 = \int p(x_2 - x_1) p(x_1 \mid y_1) dx_1. \] (G.15)

We know that

\[ p(x_2 - x_1) = \frac{1}{\sqrt{2\pi N_1}} \exp \left\{ -\frac{(x_2 - x_1)^2}{2N_1^2} \right\}, \] (G.16)

\[ p(x_1 \mid y_1) = \frac{p(y_1 \mid x_1) p(x_1)}{p(y_1)} = \\
= \frac{1}{\sqrt{2\pi \nu_1}} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(y_1 - x_1)^2}{2\nu_1^2} \right\} \exp \left\{ -\frac{x_1^2}{2\Sigma^2} \right\} \cdot \frac{\sqrt{\nu_1^2 + \Sigma^2}}{\nu_1} \exp \left\{ -\frac{y_1^2}{2(\Sigma^2 + \nu_1^2)} \right\}. \] (G.17)

or

\[ p(x_1 \mid y_1) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\nu_1}} \frac{1}{\sqrt{\Sigma^2 + \nu_1^2}} \exp \left\{ -\frac{1}{2\nu_1^2}\left( x_1 - \frac{\Sigma^2 y_1}{\Sigma^2 + \nu_1^2} \right)^2 \right\}. \] (G.18)

Using (G.16) and (G.18) in (G.15) we get

\[ p(x_2 \mid y_1) = \frac{1}{\sqrt{2\pi N_1}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\nu_1}} \frac{1}{\sqrt{\Sigma^2 + \nu_1^2}} \int \exp \left\{ -\frac{x_2^2}{2N_1^2} -\frac{x_1^2}{2N_1^2} + \frac{2x_2 x_1}{N_1} \right\}. \]
\[
\cdot \exp \left\{ \frac{x_1^2}{\nu_1^2} - \frac{y_1^2}{\nu_1^2(\nu_1^2 + \Sigma^2)} + \frac{2x_1y_1}{2\nu_1^2} \right\} \, dx_1. \tag{G.19}
\]

Completing the square in (G.19) we obtain

\[
p(x_2|y_1) = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\Sigma^2 + \nu_1^2}}{\sqrt{N_1 \nu_1^2 + \Sigma^2 N_1 + \Sigma^2 \nu_1^2}} \cdot 
\left( \frac{\sqrt{\Sigma^2 + \nu_1^2}}{2(\nu_1^2 N_1 + \Sigma^2 N_1 + \Sigma^2 \nu_1^2)} \left( x_2 - y_1 \frac{\Sigma^2}{\nu_1^2 + \Sigma^2} \right)^2 \right). \tag{G.20}
\]

(G.20) shows that \( p(x_2|y_1) \) is Gaussian with

\[
\text{mean} = \frac{\Sigma^2}{\Sigma^2 + \nu_1^2}, \tag{G.21}
\]

and

\[
\text{variance} = \frac{N_1 \nu_1^2 + \Sigma^2 \nu_1^2 + \Sigma^2 N_1}{\Sigma^2 + \nu_1^2}. \tag{G.22}
\]

Using (G.20) and (G.15) and (G.16), we can compute \( p(y_2|y_1) \). We have

\[
p(y_2|y_1) = \int p(y_2|x_2) p(x_2|y_1) \, dx_2 =
\]

\[
= \frac{1}{2\pi} \frac{\sqrt{\Sigma^2 + \nu_1^2}}{2\nu_2 \sqrt{N_1 \nu_1^2 + \Sigma^2 \nu_1^2 + N_1 \nu_1^2}} \int \exp \left\{ -\frac{y_2^2}{2\nu_2^2} - \frac{x_2^2}{2\nu_2^2} + \frac{2y_2x_2}{2\nu_2^2} \right\} \exp \left\{ -\frac{x_2^2}{2\nu_2^2} \right\} \frac{\Sigma^2 + \nu_1^2}{\nu_2^2 \Sigma^2 + \Sigma^2 N_1 + N_1 \nu_1^2}.
\]

163
\[
- \frac{y_1^2}{2} \frac{\Sigma^4}{(V_1^2 + \Sigma^2)(\Sigma^2 N_1^2 + \Sigma^2 V_1^2 + N_1^2 V_1^2)} + \frac{2x_2 y_1 \Sigma^2}{V_1^2 \Sigma^2 + \Sigma^2 N_1^2 + N_1^2 V_1^2} = \\
= \frac{1}{2\pi} \frac{\sqrt{\Sigma^2 + V_1^2}}{V_2 \sqrt{\Sigma^2 N_1^2 + N_1^2 V_1^2 + \Sigma^2 V_1^2}} \exp \left\{ - \frac{y_1^2 \Sigma^4}{(V_1^2 + \Sigma^2)(\Sigma^2 N_1^2 + V_1^2 \Sigma^2 + N_1^2 V_1^2)} \right\} \\
- \frac{y_2^2}{2V_2^2} \left\{ \exp \left\{ - \frac{x_2^2}{2V_2^2} + \frac{2y_2 x_2}{2V_2^2} - \frac{x_2^2}{2} \frac{\Sigma^2 + V_1^2}{\Sigma^2 V_1^2 + \Sigma^2 N_1^2 + N_1^2 V_1^2} + \frac{2x_2 y_1 \Sigma^2}{V_1^2 \Sigma^2 + \Sigma^2 N_1^2 + N_1^2 V_1^2} \right\} \right\} dx_2 .
\]

(G.23)

Completing the square in (G.23) we obtain

\[
p(y_2 | y_1) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{\Sigma^2 + V_1^2}{V_1^2 \Sigma^2 + \Sigma^2 N_1^2 + N_1^2 V_1^2 + \Sigma^2 V_1^2 + V_1^2 V_2^2}}} \\
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \frac{
\frac{V_1^2 \Sigma^2}{\Sigma^2 + V_1^2} + \frac{\Sigma^2 N_1^2}{\Sigma^2 + V_1^2} + \frac{\Sigma^2 V_1^2}{\Sigma^2 + V_1^2} + \frac{V_1^2 V_2^2}{\Sigma^2 + V_1^2}}{\Sigma^2 + V_1^2}.
\]

(G.25)

Hence \( p(y_2 | y_1) \) is Gaussian with

\[
\begin{align*}
\text{mean} &= y_1 \frac{\Sigma^2}{\Sigma^2 + V_1^2} , \\
\text{variance} &= \frac{V_1^2 \Sigma^2 + \Sigma^2 N_1^2 + \Sigma^2 V_1^2 + V_1^2 V_2^2}{\Sigma^2 + V_1^2} .
\end{align*}
\]

(G.26)
Consequently,

\[
H(y_2 | y_1) = \frac{1}{2} \ln 2\pi e \frac{\nu_1^2 \Sigma^2 + \Sigma^2 N_1^2 + N_1^2 V_1^2 + V_2^2 \nu_2^2 + v_2^2 v_2^2}{\Sigma^2 + v_1^2}.
\]  (G.27)

Using (G.27) and (G.12) we obtain

\[
C_2 = I(y_2; x_2 | y_1) = \frac{1}{2} \ln \frac{\nu_1^2 \Sigma^2 + \Sigma^2 N_1^2 + N_1^2 V_1^2 + V_2^2 \nu_2^2 + v_2^2 v_2^2}{(\Sigma^2 + v_1^2) v_2^2}.
\]  (G.28)

Moreover, we know that, [3],

\[
C_2 = R(D_2) = \frac{1}{2} \ln \frac{\text{var}[p(x_2 | y_1)]}{D_2},
\]

or because of (G.22)

\[
C_2 = R(D_2) = \frac{1}{2} \ln \frac{N_1^2 \Sigma^2 + \Sigma^2 V_1^2 + N_1^2 V_1^2}{D_2 (\Sigma^2 + v_1^2)},
\]  (G.29)

where \( D_2 = \min E(x_2 - \hat{x}_2)^2 \).

(G.28) and (G.29) give

\[
D_2 = \frac{\nu_1^2 (N_1^2 \Sigma^2 + \Sigma^2 V_1^2 + N_1^2 V_1^2)}{\nu_1^2 \Sigma^2 + \Sigma^2 N_1^2 + N_1^2 V_1^2 + V_1^2 V_2^2 + \Sigma^2 V_2^2}.
\]  (G.30)

(G.7) and (G.30) provide the lower bounds to the least squares filtering problem for the system (5.2.13)-(5.2.21) at times \( t = 1 \) and \( t = 2 \), respectively.
References


