A Source-Channel Separation Theorem with Application to the Source Broadcast Problem
A SOURCE-CHANNEL SEPARATION THEOREM WITH APPLICATION TO THE SOURCE BROADCAST PROBLEM

BY
KIA KHEZELI, B.Sc.

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

© Copyright by Kia Khezeli, July 2014
All Rights Reserved
TITLE: A Source-Channel Separation Theorem with Application to the Source Broadcast Problem

AUTHOR: Kia Khezeli
B.Sc., (Electrical Engineering)
Sharif University of Technology, Tehran, Iran

SUPERVISOR: Dr. Jun Chen

NUMBER OF PAGES: x, 73
To my mother and my father
Abstract

A converse method is developed for the source broadcast problem. Specifically, it is shown that the separation architecture is optimal for a variant of the source broadcast problem and the associated source-channel separation theorem can be leveraged, via a reduction argument, to establish a necessary condition for the original problem, which unifies several existing results in the literature. Somewhat surprisingly, this method, albeit based on the source-channel separation theorem, can be used to prove the optimality of non-separation based schemes and determine the performance limits in certain scenarios where the separation architecture is suboptimal.
Acknowledgements

Being a student of Jun Chen has been amazing for the past couple of years. Without his encouragement I would have given up on the problem. He helped me immensely to reach a better understanding of limits of my knowledge and the proper way of expanding it, the skill of problem solving and many other factors which gave me a better understanding of how to be a scholar. I would also like to acknowledge his technical contribution to this work.

I would like to thank the faculty of ECE department especially Steve Hranilovic, Jim Reilly and Tim Davidson. I learnt a lot from these great people. Thanks to Cheryl Gies, the graduate administrative assistant of ECE department.

Thanks also to two distinguished scholars with Sharif University of Technology, Mohammad Reza Aref and Amin Aminzadeh Gohari. I was introduced to information theory by them and they were always helpful and encouraging even after my graduation from SUT.

I consider myself to be lucky to have such wonderful friends. They were great source of energy for me in the long winters of Hamilton. It is hard for me to imagine my life at McMaster without Pedram and Kaveh. I am greatful to my friend Omid who was there for me when not all was going well. Many thanks to my other friends Babak, Reza, Amin, Rita, Farzad, Vanessa and many others.
In the end I want to thank my parents once again. Without their help and support I could not do any of this. They were my first mentors and I owe everything to them.
Notation and abbreviations

Vectors

We use $X^m = (X(1), X(2), \ldots, X(m))$ to show a vector.

Common Functions

The following is the list of the functions which are used frequently. The base of logarithm is 2 unless it is stated otherwise.

- Binary entropy function: $H_b(p) = -p \log p - (1 - p) \log(1 - p)$

- Inverse of binary entropy function: $H_b^{-1}(\cdot)$

- For any $a, b \in [0, 1]$, we define $a * b = a(1 - b) + (1 - a)b$
## Contents

Abstract iv

Acknowledgements v

Notation and abbreviations vii

1 Introduction 1

2 Problem Setup 5

3 Broadcast Channels with Receiver Side Information 11
   3.1 Definitions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
   3.2 Capacity Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
   3.3 Examples . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

4 Optimality of Source-Channel Separation for System $\tilde{\Pi}$ 24

5 A Necessary Condition for the Source Broadcast Problem 33
   5.1 Necessary Condition . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
   5.2 The Binary Uniform Source with the Hamming Distortion Measure . 36
   5.3 BS-BC($p_1, p_2$) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47
5.3.1 BE-BC($\epsilon_1, \epsilon_2$) .................................................. 49
5.3.2 BSC($p$)&BEC($\epsilon$) ...................................................... 50

6 The Quadratic Gaussian Case 53

7 Conclusion 69
List of Figures

2.1 System Π ................................................................. 6
2.2 System ˜Π ................................................................. 9
3.1 Broadcast channel with two private messages ....................... 12
3.2 Broadcast channel with receiver side information ................... 13
3.3 $\mathcal{C}_2(\text{BSC}(p)&\text{BEC}(\epsilon))$ vs. $\mathcal{C}(\text{BEC}(\epsilon)&\text{BSC}(p))$ with $p = 0.3$ and $\epsilon = 0.87$ 22
3.4 $\mathcal{C}_2(\text{BSC}(p)&\text{BEC}(\epsilon))$ vs. $\hat{\mathcal{C}}_2(\text{BSC}(p)&\text{BEC}(\epsilon)) \triangleq \{(R_1, R_2) \in \mathbb{R}_+^2 :$ $R_2 \leq 1 - \epsilon, R_1 + R_2 \leq 1 - H_b(p)\}$ with $p = 0.3$ and $\epsilon = 0.9$ 22
Chapter 1

Introduction

In the source broadcast problem, a source is sent over a broadcast channel through suitable encoding and decoding so that the reconstructions at the receivers satisfy the prescribed constraints. The special case of sending a Gaussian source over a Gaussian broadcast channel has received particular attention. For this special case, it is known that source-channel separation is in general suboptimal [1] and hybrid digital-analog coding schemes can outperform pure digital/analog schemes [2–5]. The extension of the hybrid coding architecture to the non-Gaussian setting can be found in [6].

In contrast, the progress on the converse side is still somewhat limited. To the best of our knowledge, the first non-trivial result in this direction was obtained by Reznic et al. [3] for the scalar version of the aforementioned Gaussian case. The converse argument in [3] involves an auxiliary random variable, which is generated by the source via an additive Gaussian noise channel. This auxiliary random variable is constructed in exactly the same manner as the one in Ozarow’s celebrated work on the Gaussian multiple description problem [7]. However, this resemblance is, in a certain sense, rather superficial. Indeed, on a more technical level, the auxiliary random variable
introduced by Ozarow (as elucidated in [8–11]) plays the role of exploiting an implicit conditional independence structure whereas the role of the auxiliary random variable in [3] is apparently different and still largely elusive. Recent years have seen several new converse results [12–14] for the source broadcast problem. These results are based on arguments similar to the original one by Reznic et al., especially in terms of the way the auxiliary random variables are constructed and exploited. It is worth noting that such arguments can only handle a restricted class of auxiliary random variables (essentially those that can be generated by the source via certain additive noise channels); this restriction typically leads to certain constraints on the set of sources, channels, or distortion measures that can be analyzed.

The present work is, to a certain extent, an outcome of our effort in seeking a conceptual understanding of the converse argument by Reznic et al. in general and the role of the associated auxiliary random variable in particular. We shall show that one can establish a source-channel separation theorem for a variant of the source broadcast problem and leverage it to derive a necessary condition for the original problem. This necessary condition, when specialized to the case of sending a scalar Gaussian source over a Gaussian broadcast channel, recovers the corresponding result by Reznic et al. [3]; moreover, in this way, the converse argument in [3] finds a simple interpretation, and the associated auxiliary random variable acquires an operational meaning. It should be pointed out that, in our approach, the auxiliary random variable can be generated by the source in an arbitrary manner. Therefore, the restriction imposed in the existing arguments [12–14] is in fact unnecessary. On the other hand, the problem of identifying the optimal auxiliary random variable naturally arises due to this additional freedom. It will be seen that the analytical
solutions for this problem can be found in some special cases; interestingly, these solutions indicate that the specific choices of auxiliary random variables in [3,13] are actually optimal in their respective contexts.

Our work is also partly motivated by the problem of sending a bivariate Gaussian source over a Gaussian broadcast channel first studied by Bross et al. [15]. For this problem, it is known that the achievable distortion region of a certain hybrid digital-analog coding scheme [16] matches the outer bound in [15] whereas separate source-channel coding is in general suboptimal [16, 17]. An alternative proof of the outer bound in [15] was recently obtained by Song et al. [18] (see [19] for its conference version). This new proof [18] bears some similarity to the aforementioned converse argument by Reznic et al. [3]. We will clarify their connection by giving a unified proof for the vector Gaussian case, which implies, among other things, that the outer bound in [15] can be deduced from the general necessary condition for the source broadcast problem found in the present report. Therefore, our converse method, albeit based on the source-channel separation theorem, can be used to prove the optimality of non-separation based schemes and determine the performance limits in certain scenarios where the separation architecture is suboptimal.

The rest of this report is organized as follows. We present the problem setup in Chapter 2 and the relevant capacity results for broadcast channels with receiver side information in Chapter 3. We establish a source-channel separation theorem for a variant of the source broadcast problem in Chapter 4. It is shown in Chapter 5 that this separation theorem can be used in conjunction with a simple reduction argument to derive a necessary condition for the original source broadcast problem; moreover, this necessary condition is evaluated for the special case of the binary uniform source
with the Hamming distortion measure. The quadratic Gaussian case is treated in Chapter 6. We conclude it in Chapter 7.
Chapter 2

Problem Setup

The source broadcast system (System II) consists of the following components (see Fig. 2.1):

- an i.i.d. source \( \{S(t)\}_{t=1}^{\infty} \) with marginal distribution \( p_S \) over alphabet \( S \),

- a discrete memoryless broadcast channel \( p_{Y_1,Y_2|X} \) with input alphabet \( X \) and output alphabets \( Y_i, i = 1, 2, \)

- a transmitter, which is equipped with an encoding function \( f^{(m,n)} : S^m \to X^n \) that maps a block of source samples \( S^m \) of length \( m \) to a channel input block \( X^n \) of length \( n \) (the number of channel uses per source sample, i.e., \( \frac{n}{m} \), is referred to as the bandwidth expansion ratio),

- two receivers, where receiver \( i \) is equipped with a decoding function \( g_i^{(n,m)} : Y^n_i \to \hat{S}_i^m \) that maps the channel output block \( Y_i^n \) generated by \( X^n \) to a source reconstruction block \( \hat{S}_i^m, i = 1, 2. \)

Unless stated otherwise, we assume that \( S, \hat{S}_1, \hat{S}_2, X, Y_1, \) and \( Y_2 \) are finite sets.
Let \( \mathcal{P}_{S\times\hat{S}_i}(p_S) \) denote the set of joint distributions over \( S \times \hat{S}_i \) with the marginal distribution on \( S \) fixed to be \( p_S, i = 1, 2 \).

**Definition 1** Let \( \kappa \) be a non-negative number and \( \mathcal{Q}_i \) be a non-empty compact subset of \( \mathcal{P}_{S\times\hat{S}_i}(p_S) \), \( i = 1, 2 \). We say \( (\kappa, \mathcal{Q}_1, \mathcal{Q}_2) \) is achievable for System II if, for every \( \epsilon > 0 \), there exist an encoding function \( f^{(m,n)} : S^m \rightarrow X^n \) and decoding functions \( g_i^{(n,m)} : Y_i^n \rightarrow \hat{S}_i^m \), \( i = 1, 2 \), such that

\[
\frac{n}{m} \leq \kappa + \epsilon, \tag{2.1}
\]

\[
\min_{q_i \in \mathcal{Q}_i} \left\| \frac{1}{m} \sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)} - q_i \right\| \leq \epsilon, \quad i = 1, 2, \tag{2.2}
\]

where \( \| \cdot \| \) is the 1-norm. The set of all achievable \( (\kappa, \mathcal{Q}_1, \mathcal{Q}_2) \) for System II is denoted by \( \Gamma \).

Remark: It is easy to verify that

\[
\frac{1}{m} \sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)} \in \mathcal{P}_{S\times\hat{S}_i}(p_S), \quad i = 1, 2.
\]

Now consider the following more conventional definition.
Definition 2 Let \( w_i : S \times \hat{S}_i \to [0, \infty) \) be two distortion measures. For non-negative numbers \( \kappa, d_1, \) and \( d_2, \) we say \( (\kappa, d_1, d_2) \) is achievable for System \( \Pi \) under distortion measures \( w_1 \) and \( w_2 \) if, for every \( \epsilon > 0, \) there exist an encoding function \( f^{(m,n)} : S^m \to X^n \) and decoding functions \( g_i^{(n,m)} : Y^m_i \to \hat{S}_i^m, \) \( i = 1, 2, \) such that

\[
\begin{align*}
\frac{n}{m} & \leq \kappa + \epsilon, \\
\frac{1}{m} \sum_{t=1}^{m} \mathbb{E}[w_i(S(t), \hat{S}_i(t))] & \leq d_i + \epsilon, \quad i = 1, 2.
\end{align*}
\]

(2.3)

The following results show that Definition 1 is more general than Definition 2.

Proposition 1 \((\kappa, d_1, d_2)\) is achievable for System \( \Pi \) under distortion measures \( w_1 \) and \( w_2 \) if and only if \((\kappa, Q(w_1, d_1), Q(w_2, d_2)) \in \Gamma, \) where \( Q(w_i, d_i) = \{ p_{S_i} \in P_{S \times \hat{S}_i} : \mathbb{E}[w_i(S, \hat{S}_i)] \leq d_i \}, \) \( i = 1, 2. \)

Proof: Let \( T \) be a random variable independent of \((S^m, \hat{S}_1^m, \hat{S}_2^m)\) and uniformly distributed over \( \{1, \cdots, m\}. \) It is easy to verify that (2.2) can be written equivalently as

\[
\min_{q_i \in Q_i} \left\| p_{S(T), \hat{S}_i(T)} - q_i \right\| \leq \epsilon, \quad i = 1, 2,
\]

and (2.3) can be written equivalently as

\[
\mathbb{E}[w_i(S(T), \hat{S}_i(T))] \leq d_i + \epsilon, \quad i = 1, 2.
\]
Note that

\[
E[w_i(S(T), \hat{S}_i(T))] = \sum_{s, \hat{s}_i \in \hat{S}_i} p_{S(T), \hat{S}_i(T)}(s, \hat{s}_i)w_i(s, \hat{s}_i) \\
\leq \sum_{s, \hat{s}_i \in \hat{S}_i} q_i(s, \hat{s}_i)w_i(s, \hat{s}_i) + \sum_{s, \hat{s}_i \in \hat{S}_i} |p_{S(T), \hat{S}_i(T)}(s, \hat{s}_i) - q_i(s, \hat{s}_i)|w_i(s, \hat{s}_i) \\
\leq d_i + \|p_{S(T), \hat{S}_i(T)} - q_i\| \max_{s, \hat{s}_i \in \hat{S}_i} w_i(s, \hat{s}_i)
\]

for any \(q_i \in Q_i(w_i, d_i), i = 1, 2\). Therefore, we have

\[
E[w_i(S(T), \hat{S}_i(T))] \leq d_i + \min_{q_i \in Q_i(w_i, d_i)} \|p_{S(T), \hat{S}_i(T)} - q_i\| \max_{s, \hat{s}_i \in \hat{S}_i} w_i(s, \hat{s}_i), \quad i = 1, 2,
\]

from which the “if” part follows immediately.

Now we proceed to prove the “only if” part. Assume that \((\kappa, d_1, d_2)\) is achievable for System II under distortion measures \(w_1\) and \(w_2\). For every \(\epsilon > 0\), according to Definition 2, we can find encoding function \(f^{(m,n)} : S^m \to X^n\) and decoding functions \(g_i^{(n,m)} : Y_i^n \to S_i^m, i = 1, 2\), satisfying \(\frac{n}{m} \leq \kappa + \epsilon\) and \(E[w_i(S(T), \hat{S}_i(T))] \leq d_i + \epsilon, i = 1, 2\). We shall denote \(S(T)\) simply by \(S\) since the distribution of \(S(T)\) is \(p_S\), and denote \(\hat{S}_1\) and \(\hat{S}_2\) by \(\hat{S}_1^{(\epsilon)}\) and \(\hat{S}_2^{(\epsilon)}\), respectively, to stress their dependence on \(\epsilon\). Note that \(\{p_{S, \hat{S}_1^{(\epsilon)}, \hat{S}_2^{(\epsilon)}} : \epsilon > 0\}\) is contained in a compact set and \(E[w_i(S, \hat{S}_1^{(\epsilon)})] \leq d_i + \epsilon\) for every \(\epsilon > 0, i = 1, 2\). Therefore, one can find a sequence \(\epsilon_1, \epsilon_2, \ldots\) converging to zero such that

\[
\lim_{k \to \infty} p_{S, \hat{S}_1^{(\epsilon_k)}, \hat{S}_2^{(\epsilon_k)}} = p_{S, \hat{S}_1, \hat{S}_2}
\]
for some $p_{s,\tilde{s}_1,\tilde{s}_2}$ with $p_{s,\tilde{s}_i} \in Q_i(w_i,d_i)$, $i = 1,2$. This completes the proof of the “only if” part.

Source-channel separation is known to incur a performance loss for System II in general. However, it turns out that, for the following variant of System II (see Fig. 2.2), separate source-channel coding is in fact optimal. This system (System $\tilde{\Pi}$) is the same as System II except for two differences.

1. The source is an i.i.d. vector process $\{(\tilde{s}_1(t),\tilde{s}_2(t))\}_{t=1}^\infty$ with marginal distribution $p_{\tilde{s}_1,\tilde{s}_2}$ over finite alphabet $\tilde{S}_1 \times \tilde{S}_2$.

2. $\tilde{S}_2^m$ is available at receiver 1 and can be used together with $Y_1^n$ to construct $\hat{S}_1^m$.

Let $\mathcal{P}_{\tilde{s}_1 \times \tilde{s}_2 \times \tilde{s}_1}(p_{\tilde{s}_1,\tilde{s}_2})$ denote the set of joint distributions over $\tilde{S}_1 \times \tilde{S}_2 \times \tilde{S}_1$ with the marginal distribution on $\tilde{S}_1 \times \tilde{S}_2$ fixed to be $p_{\tilde{s}_1,\tilde{s}_2}$. Moreover, let $\mathcal{P}_{\tilde{s}_2 \times \tilde{s}_1}(p_{\tilde{s}_2})$ denote the set of joint distributions over $\tilde{S}_2 \times \tilde{S}_2$ with the marginal distribution on $\tilde{S}_2$ fixed to be $p_{\tilde{s}_2}$.

**Definition 3** Let $\tilde{\kappa}$ be a non-negative number, $\tilde{Q}_1$ be a non-empty compact subset of $\mathcal{P}_{\tilde{s}_1 \times \tilde{s}_2 \times \tilde{s}_1}(p_{\tilde{s}_1,\tilde{s}_2})$, and $\tilde{Q}_2$ be a non-empty compact subset of $\mathcal{P}_{\tilde{s}_2 \times \tilde{s}_1}(p_{\tilde{s}_2})$. We say...
\((\bar{\kappa}, \tilde{Q}_1, \tilde{Q}_2)\) is achievable for System \(\tilde{\Pi}\) if, for every \(\epsilon > 0\), there exist an encoding function \(f^{(m,n)}: \tilde{S}_1^m \times \tilde{S}_2^m \to X^n\) as well as decoding functions \(g_1^{(n,m)}: Y_1^n \times \hat{S}_1^m \to \hat{S}_1^m\) and \(g_2^{(n,m)}: Y_2^n \to \hat{S}_2^m\) such that

\[
\frac{n}{m} \leq \bar{\kappa} + \epsilon, \quad (2.4)
\]

\[
\min_{\tilde{q}_1 \in \tilde{Q}_1} \left\| \frac{1}{m} \sum_{t=1}^{m} p_{\tilde{S}_1(t),\tilde{S}_2(t),\hat{S}_1(t)} - \tilde{q}_1 \right\| \leq \epsilon, \quad (2.5)
\]

\[
\min_{\tilde{q}_2 \in \tilde{Q}_2} \left\| \frac{1}{m} \sum_{t=1}^{m} p_{\tilde{S}_2(t),\hat{S}_2(t)} - \tilde{q}_2 \right\| \leq \epsilon. \quad (2.6)
\]

The set of all achievable \((\bar{\kappa}, \tilde{Q}_1, \tilde{Q}_2)\) for System \(\tilde{\Pi}\) is denoted by \(\tilde{\Gamma}\).

Remark: For the ease of subsequent applications, here we allow \(f^{(m,n)}\), \(g_1^{(n,m)}\), and \(g_2^{(n,m)}\) to be non-deterministic functions as long as the Markov chains \((\tilde{S}_1^m, \tilde{S}_2^m) \leftrightarrow X^n \leftrightarrow (Y_1^n, Y_2^n), \tilde{S}_1^m \leftrightarrow (Y_1^n, \tilde{S}_2^m) \leftrightarrow \hat{S}_1^m, \) and \(\tilde{S}_2^m \leftrightarrow Y_2^n \leftrightarrow \hat{S}_2^m\) are preserved. It will be clear in the proof of the separation theorem that such a relaxation does not affect \(\tilde{\Gamma}\).

To discuss source-channel separation for System \(\tilde{\Pi}\), we need to specify the source coding component and the channel coding component. It will be seen that the source coding part is the conventional lossy source coding scheme. The channel coding part is more involved and is described in the next section.
Chapter 3

Broadcast Channels with Receiver Side Information

3.1 Definitions

Let $p_{Y_1,Y_2|X}$ be a discrete memoryless broadcast channel with input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}_i$, $i = 1, 2$. A length-$n$ coding scheme (see Fig. 3.1) for $p_{Y_1,Y_2|X}$ consists of

- two private messages $M_1$ and $M_2$, where $(M_1, M_2)$ is uniformly distributed over $\mathcal{M}_1 \times \mathcal{M}_2$,
- an encoding function $f^{(n)} : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{X}^n$ that maps $(M_1, M_2)$ to a channel input block $X^n$,
- two decoding functions $g_i^{(n)} : \mathcal{Y}_i^n \rightarrow \hat{M}_i$, $i = 1, 2$, where $g_i^{(n)}$ maps the channel output block at receiver $i$, i.e., $Y_i^n$, to $\hat{M}_i$, $i = 1, 2$. 

11
\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig3.1}
\caption{Broadcast channel with two private messages}
\end{figure}

\textbf{Definition 4} A rate pair \((R_1, R_2) \in \mathbb{R}_+^2\) is said to be achievable for broadcast channel \(p_{Y_1,Y_2|X}\) if there exists a sequence of encoding functions \(f^{(n)} : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n\) with \(\frac{1}{n} \log |\mathcal{M}_i| \geq R_i, \ i = 1, 2,\) and decoding functions \(g_i^{(n)} : \mathcal{Y}_i^n \to \mathcal{M}_i, \ i = 1, 2,\) such that

\[\lim_{n \to \infty} \Pr\{ (\hat{M}_1, \hat{M}_2) \neq (M_1, M_2) \} = 0.\]

The private-message capacity region \(\mathcal{C}(p_{Y_1,Y_2|X})\) is the closure of the set of all achievable \((R_1, R_2)\) for broadcast channel \(p_{Y_1,Y_2|X}\).

A computable characterization of \(\mathcal{C}(p_{Y_1,Y_2|X})\) is still largely unknown. Interestingly, the problem becomes significantly simpler if message \(M_2\) is available at receiver 1 or message \(M_1\) is available at receiver 2; in fact, this is the setting that is most relevant to the present work. Specifically, consider the scenario where two private messages \(M_1\) and \(M_2\) need to be sent over broadcast channel \(p_{Y_1,Y_2|X}\) to receiver 1 and receiver 2, respectively, and \(M_2\) is available at receiver 1. In this case, a length-\(n\) coding scheme (see Fig. 3.2) consists of

- two private messages \(M_i, \ i = 1, 2,\) where \((M_1, M_2)\) is uniformly distributed over \(\mathcal{M}_1 \times \mathcal{M}_2,\)
Figure 3.2: Broadcast channel with receiver side information

- an encoding function \( f^{(n)} : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{X}^n \) that maps \((M_1, M_2)\) to a channel input block \(X^n\),

- two decoding functions \( g_1^{(n)} : Y_1^n \times \mathcal{M}_2 \rightarrow \hat{\mathcal{M}}_1 \) and \( g_2^{(n)} : Y_2^n \rightarrow \hat{\mathcal{M}}_2 \), where \( g_1^{(n)} \) maps \((Y_1^n, M_2)\) to \(\hat{M}_1\) and \( g_2^{(n)} \) maps \(Y_2^n\) to \(\hat{M}_2\).

**Definition 5** A rate pair \((R_1, R_2)\) is said to be achievable for broadcast channel \( p_{Y_1, Y_2|X} \) with message \( M_2 \) available at receiver 1 if there exists a sequence of encoding functions \( f^{(n)} : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{X}^n \) with \( \frac{1}{n} \log |\mathcal{M}_i| \geq R_i, \ i = 1, 2 \), as well as decoding functions \( g_1^{(n)} : Y_1^n \times \mathcal{M}_2 \rightarrow \mathcal{M}_1 \) and \( g_2^{(n)} : Y_2^n \rightarrow \mathcal{M}_2 \) such that

\[
\lim_{n \to \infty} \Pr\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\} = 0.
\]

The capacity region \( \mathcal{C}_1(p_{Y_1, Y_2|X}) \) is the closure of the set of all such achievable \((R_1, R_2)\).

The capacity region \( \mathcal{C}_2(p_{Y_1, Y_2|X}) \) for broadcast channel \( p_{Y_1, Y_2|X} \) with message \( M_1 \) available at receiver 2 can be defined in an analogous manner.
3.2 Capacity Results

It is known [20, Theorem 3] that $C_1(p_{Y_1,Y_2|X})$ is given by the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \leq I(X; Y_1), \quad (3.1)$$
$$R_2 \leq I(V; Y_2), \quad (3.2)$$
$$R_1 + R_2 \leq I(X; Y_1|V) + I(V; Y_2) \quad (3.3)$$

for some $p_{V,X,Y_1,Y_2} = p_{V,X} p_{Y_1,Y_2|X}$; moreover, it suffices to assume that $|\mathcal{V}| \leq |\mathcal{X}| + 1$. By symmetry, $C_2(p_{Y_1,Y_2|X})$ is given by the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \leq I(V; Y_1), \quad (3.4)$$
$$R_2 \leq I(X; Y_2), \quad (3.5)$$
$$R_1 + R_2 \leq I(V; Y_1) + I(X; Y_2|V) \quad (3.6)$$

for some $p_{V,X,Y_1,Y_2} = p_{V,X} p_{Y_1,Y_2|X}$; again, it suffices to assume that $|\mathcal{V}| \leq |\mathcal{X}| + 1$.

A class of distributions $\mathcal{P}$ on the input alphabet $\mathcal{X}$ is said to be a sufficient class of distributions [21, Definition 1] for broadcast channel $p_{Y_1,Y_2|X}$ if, for any $p_{V_i,V_2,X,Y_1,Y_2} = \tilde{p}_{V_1,V_2,X} \tilde{p}_{Y_1,Y_2|X}$, there exists $p_{\tilde{V}_1,\tilde{V}_2,\tilde{X},\tilde{Y}_1,\tilde{Y}_2}$ with $p_{\tilde{X}} \in \mathcal{P}$ and $p_{\tilde{Y}_1,\tilde{Y}_2|\tilde{X}} = p_{Y_1,Y_2|X}$ such
that

$$
I(V_1; Y_1) \leq I(\tilde{V}_1; \tilde{Y}_1),
I(V_2; Y_2) \leq I(\tilde{V}_2; \tilde{Y}_2),
I(V_1; Y_1) + I(X; Y_2|V_1) \leq I(\tilde{V}_1; \tilde{Y}_1) + I(\tilde{X}; \tilde{Y}_2|V_1),
I(X; Y_1|V_2) + I(V_2; Y_2) \leq I(\tilde{X}; \tilde{Y}_1|V_2) + I(\tilde{V}_2; \tilde{Y}_2).
$$

For broadcast channel $p_{Y_1, Y_2|X}$, we say that $p_{Y_1|X}$ is essentially less noisy than $p_{Y_2|X}$ if there exists a sufficient class of distributions $\mathcal{P}$ such that $I(V; Y_1) \geq I(V; Y_2)$ for any $p_{V,X,Y_1,Y_2} = p_{V,X}p_{Y_1,Y_2|X}$ with $p_X \in \mathcal{P}$ [21, Definition 2]; similarly, we say that $p_{Y_1|X}$ is essentially more capable than $p_{Y_2|X}$ if there exists a sufficient class of distributions $\mathcal{P}$ such that $I(X; Y_1|V) \geq I(X; Y_2|V)$ for any $p_{V,X,Y_1,Y_2} = p_{V,X}p_{Y_1,Y_2|X}$ with $p_X \in \mathcal{P}$ [21, Definition 3]. It is known that “less noisy” (“more capable”) implies “essentially less noisy” (“essentially more capable”), and “less noisy” implies “more capable”, but the converses are not true in general.

**Proposition 2** If $p_{Y_1|X}$ is essentially less noisy than $p_{Y_2|X}$, then $C_1(p_{Y_1,Y_2|X}) = C(p_{Y_1,Y_2|X})$.

**Proof:** To compute $C_1(p_{Y_1,Y_2|X})$ defined by (3.1)-(3.3), it suffices to consider those $p_X$ in a sufficient class $\mathcal{P}$. It is easy to see that

$$
I(X; Y_1|V) + I(V; Y_2) \leq I(X; Y_1|V) + I(V; Y_1)
= I(X; Y_1)
$$

for any $p_{V,X,Y_1,Y_2} = p_{V,X}p_{Y_1,Y_2|X}$ with $p_X \in \mathcal{P}$, where (3.7) is due to the fact that $p_{Y_1|X}$ is essentially less noisy than $p_{Y_2|X}$. Therefore, (3.1) is redundant if $p_X$ is restricted to
Proposition 3 If $p_{Y_1|X}$ is essentially more capable than $p_{Y_2|X}$, then $C_2(p_{Y_1,Y_2|X})$ is given by the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

\begin{align*}
R_2 &\leq I(X;Y_2), \\
R_1 + R_2 &\leq I(X;Y_1)
\end{align*}

for some $p_{X,Y_1,Y_2} = p_X p_{Y_1,Y_2|X}$.

Proof: To compute $C_2(p_{Y_1,Y_2|X})$ defined by (3.4)-(3.6), it suffices to consider those $p_X$ in a sufficiently class $\mathcal{P}$. Note that

\begin{equation}
I(V;Y_1) + I(X;Y_2|V) \leq I(V;Y_1) + I(X;Y_1|V)
\end{equation}

\begin{equation}
= I(X;Y_1)
\end{equation}

for any $p_{V,X,Y_1,Y_2} = p_{V,X} p_{Y_1,Y_2|X}$ with $p_X \in \mathcal{X}$, where (3.8) is due to the fact that $p_{Y_1|X}$ is essentially more capable than $p_{Y_2|X}$. Therefore, given $p_X \in \mathcal{P}$, the right-hand side of inequality (3.6) attains its maximum value $I(X;Y_1)$ when $V = X$. Clearly, given $p_X$, the right-hand side of inequality (3.4) also attains its maximum value $I(X;Y_1)$ when $V = X$. As a consequence, $C_2(p_{Y_1,Y_2|X})$ can be expressed as the set of $(R_1, R_2) \in \mathbb{R}_+^2$...
satisfying

\[ R_2 \leq I(X;Y_2), \]
\[ R_1 + R_2 \leq I(X;Y_1) \]

for some \( p_{X,Y_1,Y_2} = p_X p_{Y_1,Y_2|X} \) with \( p_X \in \mathcal{P} \). Removing the redundant constraint \( p_X \in \mathcal{P} \) completes the proof of Proposition 3.

### 3.3 Examples

Consider a broadcast channel \( p_{Y_1,Y_2|X} \) with \( \mathcal{X} = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0,1\} \), where \( p_{Y_i|X} \) is a binary symmetric channel with crossover probability \( p_i \), \( i = 1,2 \); such a channel will be denoted by BS-BC\((p_1,p_2)\). Without loss of generality, we shall assume \( 0 \leq p_1 \leq p_2 \leq \frac{1}{2} \). It is well known that \( C(\text{BS}(p_1,p_2)) \) is given by the set of \( (R_1,R_2) \in \mathbb{R}^2_+ \) satisfying

\[ R_1 \leq H_b(\alpha \ast p_1) - H_b(p_1), \]
\[ R_2 \leq 1 - H_b(\alpha \ast p_2) \]

for some \( \alpha \in [0,\frac{1}{2}] \). Next consider a broadcast channel \( p_{Y_1,Y_2|X} \) with \( \mathcal{X} = \{0,1\} \) and \( \mathcal{Y}_i = \{0,1,e\}, i = 1,2 \), where \( p_{Y_i|X} \) is a binary erasure channel with erasure probability \( \epsilon_i \), \( i = 1,2 \); such a channel will be denoted by BE-BC\((\epsilon_1,\epsilon_2)\). Without loss of generality, we shall assume \( 0 \leq \epsilon_1 \leq \epsilon_2 \leq 1 \). It is well known that \( C(\text{BE-BC}(\epsilon_1,\epsilon_2)) \)
is given by the set of \((R_1, R_2) \in \mathbb{R}^2_+\) satisfying

\[
\begin{align*}
R_1 & \leq \beta (1 - \varepsilon_1), \\
R_2 & \leq (1 - \beta) (1 - \varepsilon_2)
\end{align*}
\]

for some \(\beta \in [0, 1]\).

The following results are simple consequences of Proposition 2 and Proposition 3.

**Proposition 4** For BS-BC\((p_1, p_2)\) with \(0 \leq p_1 \leq p_2 \leq \frac{1}{2}\),

\[
\begin{align*}
C_1(\text{BS-BC}(p_1, p_2)) &= C(\text{BS-BC}(p_1, p_2)), \\
C_2(\text{BS-BC}(p_1, p_2)) &= \{(R_1, R_2) \in \mathbb{R}^2_+ : R_2 \leq 1 - H_b(p_2), R_1 + R_2 \leq 1 - H_b(p_1)\}.
\end{align*}
\]

**Proposition 5** For BE-BC\((\varepsilon_1, \varepsilon_2)\) with \(0 \leq \varepsilon_1 \leq \varepsilon_2 \leq 1\),

\[
\begin{align*}
C_1(\text{BE-BC}(\varepsilon_1, \varepsilon_2)) &= C(\text{BE-BC}(\varepsilon_1, \varepsilon_2)), \\
C_2(\text{BE-BC}(\varepsilon_1, \varepsilon_2)) &= \{(R_1, R_2) \in \mathbb{R}^2_+ : R_2 \leq 1 - \varepsilon_2, R_1 + R_2 \leq 1 - \varepsilon_1\}.
\end{align*}
\]

Now consider a broadcast channel \(p_{Y_1,Y_2|X}\) with \(\mathcal{X} = \mathcal{Y}_1 = \{0, 1\}\) and \(\mathcal{Y}_2 = \{0, 1, e\}\), where \(p_{Y_1|X}\) is a binary symmetric channel with crossover probability \(p\), and \(p_{Y_2|X}\) is a binary erasure channel with erasure probability \(\varepsilon\); such a channel will be denoted by BSC\((p)\)&BEC\((\varepsilon)\). Without loss of generality, we shall assume \(p \in [0, \frac{1}{2}]\) and \(\varepsilon \in [0, 1]\). One can obtain the following explicit characterization of \(C(\text{BSC}(p)\&\text{BEC}(\varepsilon))\) [21, Theorem 4].

1. \(\varepsilon \in [0, 4p(1 - p)]\): \(C(\text{BSC}(p)\&\text{BEC}(\varepsilon))\) is given by the set of \((R_1, R_2) \in \mathbb{R}^2_+\)
satisfying

\[ R_1 \leq 1 - H_b(\alpha * p), \]
\[ R_2 \leq (1 - \epsilon)H_b(\alpha) \]

for some \( \alpha \in [0, \frac{1}{2}] \).

2. \( \epsilon \in (4p(1-p), H_b(p)) \): \( C(BSC(p)\&BEC(\epsilon)) \) is given by the set of \( (R_1, R_2) \in \mathbb{R}_+^2 \) satisfying

\[ R_1 \leq 1 - H_b(\alpha * p), \]
\[ R_2 \leq (1 - \epsilon)H_b(\alpha) \]

for some \( \alpha \in [0, \hat{\alpha}] \), or

\[ R_1 \leq 1 - H_b(\alpha * p), \]
\[ R_2 \leq H_b(\alpha * p) - \epsilon \]

for some \( \alpha \in (\hat{\alpha}, \frac{1}{2}] \), where \( \hat{\alpha} \) is the unique number in \( (0, \frac{1}{2}) \) satisfying

\[ 1 - H_b(\hat{\alpha} * p) + (1 - \epsilon)H_b(\hat{\alpha}) = 1 - \epsilon. \]
3. $\epsilon \in [H_b(p), 1]$: $C(BSC(p)\&BEC(\epsilon))$ is given by the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \leq \beta [1 - H_b(p)],$$
$$R_2 \leq (1 - \beta)(1 - \epsilon)$$

for some $\beta \in [0, 1]$.

**Proposition 6** $C_1(BSC(p)\&BEC(\epsilon))$ has the following explicit characterization.

1. $\epsilon \in [0, H_b(p)]$: $C_1(BSC(p)\&BEC(\epsilon)) = \{(R_1, R_2) \in \mathbb{R}_+^2 : R_1 \leq 1 - H_b(p), R_1 + R_2 \leq 1 - \epsilon\}$.

2. $\epsilon \in (H_b(p), 1]$: $C_1(BSC(p)\&BEC(\epsilon)) = C(BSC(p)\&BEC(\epsilon))$.

**Proof**: According to [21, Theorem 3], BEC($\epsilon$) is more capable than BSC($p$) when $\epsilon \in [0, H_b(p)]$. Therefore, one can readily prove Part 1) by invoking Proposition 3 as well as the fact that $I(X; Y_1)$ and $I(X; Y_2)$ are simultaneously maximized when $p_X(0) = p_X(1) = \frac{1}{2}$. Part 2) follows from Proposition 2 and the fact that BSC($p$) is essentially less noisy than BEC($\epsilon$) when $\epsilon \in (H_b(p), 1]$ [21, Theorem 3].

**Proposition 7** $C_2(BSC(p)\&BEC(\epsilon))$ has the following explicit characterization.

1. $\epsilon \in [0, 4p(1 - p)]$: $C_2(BSC(p)\&BEC(\epsilon)) = C(BSC(p)\&BEC(\epsilon))$.

2. $\epsilon \in (4p(1 - p), 1)$ and $p \neq 0$: $C_2(BSC(p)\&BEC(\epsilon))$ is given by the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \leq 1 - H_b(\alpha \ast p),$$
$$R_2 \leq (1 - \epsilon)H_b(\alpha)$$
for some $\alpha \in [0, \tilde{\alpha}]$, or

$$R_1 \leq 1 - H_b(\tilde{\alpha} \ast p),$$

$$R_2 \leq 1 - \epsilon,$$

$$R_1 + R_2 \leq 1 - H_b(\tilde{\alpha} \ast p) + (1 - \epsilon)H_b(\tilde{\alpha})$$

for some $\alpha \in (\tilde{\alpha}, \frac{1}{2}]$, where $\tilde{\alpha}$ is the unique number in $(0, \frac{1}{2})$ satisfying

$$(1 - 2p) \log \left(\frac{1 - \tilde{\alpha} \ast p}{\tilde{\alpha} \ast p}\right) = (1 - \epsilon) \log \left(\frac{1 - \tilde{\alpha}}{\tilde{\alpha}}\right).$$

3. $\epsilon = 1$ or $p = 0$: $C_2(BSC(p) \& BEC(\epsilon)) = \{(R_1, R_2) \in \mathbb{R}_+^2 : R_2 \leq 1 - \epsilon, R_1 + R_2 \leq 1 - H_b(p)\}$.

**Proof:** Part 1) follows from Proposition 2 and the fact that BEC($\epsilon$) is less noisy than BSC($p$) when $\epsilon \in [0, 4p(1 - p)]$ [21, Theorem 3]. Part 3) is trivial. For Part 2), one can readily show that $C_2(BSC(p) \& BEC(\epsilon))$ is given by the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \leq 1 - H_b(\alpha \ast p),$$

$$R_2 \leq 1 - \epsilon,$$

$$R_1 + R_2 \leq 1 - H_b(\alpha \ast p) + (1 - \epsilon)H_b(\alpha)$$

for some $\alpha \in [0, \frac{1}{2}]$ by following the proof of [21, Claim 2 and Claim 3]. In light of [11, Lemma 6], when $\epsilon \in (4p(1 - p), 1)$ and $p \neq 0$, the following optimization
Figure 3.3: $C_2(\text{BSC}(p)\&\text{BEC}(\epsilon))$ vs. $C(\text{BEC}(\epsilon)\&\text{BSC}(p))$ with $p = 0.3$ and $\epsilon = 0.87$

problem

$$\max_{\alpha\in[0,1]} 1 - H_b(\alpha * p) + (1 - \epsilon)H_b(\alpha)$$

has a unique maximizer at $\alpha = \tilde{\alpha}$. This completes the proof of Proposition 7.

Remark: It might be tempting to conjecture that Proposition 2 continues to hold if “essentially less noisy” is replaced by “essentially more capable”. However, this conjecture turns out to be false. Indeed, for $\text{BSC}(p)\&\text{BEC}(\epsilon)$, it is known [21, Theorem 3] that BEC($\epsilon$) is more capable (but not less noisy) than BSC($p$) when $\epsilon \in (4p(1-p), H_b(p)]$, yet Part 2) of Proposition 7 indicates that in this case $C_2(\text{BSC}(p)\&\text{BEC}(\epsilon))$ is strictly larger than $C(\text{BSC}(p)\&\text{BEC}(\epsilon))$ (see Fig. 3.3). Analogously, Proposition 3 is not true in general if “essentially more capable” is replaced by “essentially less noisy”. For example, according to [21, Theorem 3], BSC($p$) is essentially less noisy than BEC($\epsilon$) when $\epsilon \in [H_b(p), 1)$ and $p \neq 0$, but Part 2) of Proposition 7 shows that in this case $C_2(\text{BSC}(p)\&\text{BEC}(\epsilon))$ is strictly larger than $\{(R_1, R_2) \in \mathbb{R}^2_+ : R_2 \leq 1 - \epsilon, R_1 + R_2 \leq 1 - H_b(p)\}$ (see Fig. 3.4).
Finally consider the case where \( p_{Y_1,Y_2|X} \) is a scalar Gaussian broadcast channel with power constraint \( P \) and noise variances \( N_1 \) and \( N_2 \) (\( 0 < N_1 \leq N_2 \)); such a channel will be denoted by \( G-BC(P, N_1, N_2) \). It is well known that \( \mathcal{C}(G-BC(P, N_1, N_2)) \) is given by the set of \((R_1, R_2) \in \mathbb{R}_+^2\) satisfying

\[
R_1 \leq \frac{1}{2} \log \left( \frac{\beta P + N_1}{N_1} \right),
\]

\[
R_2 \leq \frac{1}{2} \log \left( \frac{P + N_2}{\beta P + N_2} \right)
\]

for some \( \beta \in [0, 1] \). One can readily prove the following result by adapting Proposition 2 and Proposition 3 to this channel model.

**Proposition 8** For \( G-BC(P, N_1, N_2) \) with \( 0 < N_1 \leq N_2 \),

\[
\mathcal{C}_1(G-BC(P, N_1, N_2)) = \mathcal{C}(G-BC(P, N_1, N_2)),
\]

\[
\mathcal{C}_2(G-BC(P, N_1, N_2)) = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 : R_2 \leq \frac{1}{2} \log \left( \frac{P + N_2}{N_2} \right), R_1 + R_2 \leq \frac{1}{2} \log \left( \frac{P + N_1}{N_1} \right) \right\}.
\]
Chapter 4

Optimality of Source-Channel Separation for System $\tilde{\Pi}$

Now we are in a position to state the following source-channel separation theorem, which shows that a separation-based scheme that consists of lossy source coding and broadcast channel coding (see Fig. 3.2 and the associated description) is optimal for System $\tilde{\Pi}$. This result can be viewed as an extension of [17, Lemma 3] from degraded broadcast channels to general broadcast channels.

Theorem 1 $(\tilde{\kappa}, \tilde{Q}_1, \tilde{Q}_2) \in \tilde{\Gamma}$ if and only if $(R_{\tilde{S}_1|\tilde{S}_2}(\tilde{Q}_1), R_{\tilde{S}_2}(\tilde{Q}_2)) \in \tilde{\kappa}C_1(p_{Y_1,Y_2|X})$, where

\[
R_{\tilde{S}_1|\tilde{S}_2}(\tilde{Q}_1) = \min_{p_{\tilde{S}_1,\tilde{S}_2,\hat{S}_1}\in\tilde{Q}_1} I(\hat{S}_1; \tilde{S}_1|\tilde{S}_2),
\]

\[
R_{\tilde{S}_2}(\tilde{Q}_2) = \min_{p_{\tilde{S}_2,\hat{S}_2}\in\tilde{Q}_2} I(\hat{S}_2; \tilde{S}_2).
\]
Proof: The proof of the “if” part hinges on a separation-based scheme. We shall only give a sketch here since the argument only involves standard techniques. Let \( \hat{S}_1 \) be jointly distributed with \((\tilde{S}_1, \tilde{S}_2)\) such that \( p_{\tilde{S}_1, \tilde{S}_2, \hat{S}_1} \in \tilde{Q}_1 \) and \( I(\tilde{S}_1; \hat{S}_1|\tilde{S}_2) = R_{\tilde{S}_1|\tilde{S}_2}(\tilde{Q}_1) \). Let \( \hat{S}_2 \) be jointly distributed with \( \tilde{S}_2 \) such that \( p_{\tilde{S}_2, \hat{S}_2} \in \tilde{Q}_2 \) and \( I(\tilde{S}_2; \hat{S}_2) = R_{\tilde{S}_2}(...\tilde{Q}_2) \). By the functional representation lemma [22, p. 626] (see also [23, Lemma 1]), we can find a random variable \( W \) of cardinality \(|W| \leq |\tilde{S}_2|(|\hat{S}_1| - 1) + 1 \) with the following properties:

- \( W \) is independent of \( \tilde{S}_2 \);
- \( \hat{S}_1 = \psi(\tilde{S}_2, W) \) for some deterministic function \( \psi : \tilde{S}_2 \times W \rightarrow \hat{S}_1 \);
- \( \tilde{S}_1 \leftrightarrow (\tilde{S}_2, \hat{S}_1) \leftrightarrow W \) form a Markov chain.

It is easy to see that

\[
I(\tilde{S}_1; \hat{S}_1|\tilde{S}_2) = I(\tilde{S}_1; W|\tilde{S}_2)
= I(\tilde{S}_1, \tilde{S}_2; W).
\]

For any \( \delta > 0 \), let \( R_1 = (1 + \delta)I(\tilde{S}_1; \hat{S}_1|\tilde{S}_2) \) and \( R_2 = (1 + \delta)I(\tilde{S}_2; \hat{S}_2) \). We independently generate \( 2^{mR_1} \) codewords \( W^m(m_1) \), \( m_1 = 1, \ldots, 2^{mR_1} \), each according to \( \prod_{t=1}^m p_W \), and independently generate \( 2^{mR_2} \) codewords \( \hat{S}_2^m(m_2) \), \( m_2 = 1, \ldots, 2^{mR_2} \), each according to \( \prod_{t=1}^m p_{\hat{S}_2} \). Codebooks \( \{W^m(m_1)\}_{m_1=1}^{2^{mR_1}} \) and \( \{\hat{S}_2^m(m_2)\}_{m_2=1}^{2^{mR_2}} \) are revealed to the transmitter and the receivers. It can be shown that, given \((\tilde{S}_1^m, \tilde{S}_2^m)\), with high probability one can find an index \( M_1 \) such that \((\tilde{S}_1^m, \tilde{S}_2^m, W^m(M_1))\) are jointly typical with respect to \( p_{\tilde{S}_1, \tilde{S}_2, W} \) when \( m \) is large enough (see [22] for the definition of typical sequences and the related properties). Similarly, given \( \tilde{S}_2^m \), with high
probability one can find an index $M_2$ such that $(\tilde{S}_2^m, \hat{S}_2^m(M_2))$ are jointly typical with respect to $p_{\tilde{S}_2, \hat{S}_2}$. If there is more than one such $M_1$ (or $M_2$), we choose the smallest index among them; if no such $M_1$ (or $M_2$) exists, we set $M_1 = 1$ (or $M_2 = 1$). Now a length-$n$ coding scheme is used to send messages $M_1$ and $M_2$ over broadcast channel $p_{Y_1, Y_2|X}$ to receiver 1 and receiver 2, respectively. Given $\tilde{S}_2^m$, receiver 1 can recover $M_2$ and use it together with $Y_1^n$ to produce an estimate $\hat{M}_1$. Receiver 2 can use $Y_2^n$ to produce an estimate $\hat{M}_2$. We assume that this length-$n$ coding scheme is good in the sense that $M_i = \hat{M}_i$, $i = 1, 2$, with high probability. Note that the existence of such a good length-$n$ coding scheme is guaranteed by Definition 5 when $\frac{n}{m} \geq \tilde{\kappa}(1 + 2\delta)$ and $n$ is large enough. Receiver 1 then constructs $\hat{S}_1^m$ with

$$\hat{S}_1(t) = \psi(\tilde{S}_2(t), W(\hat{M}_1, t)), \quad t = 1, \ldots, m,$$

where $W(\hat{M}_1, t)$ is the $t$-th entry of $W_1^m(\hat{M}_1)$. Receiver 2 sets $\hat{S}_2^m = \hat{S}_2^m(\hat{M}_2)$. It is easy to show that $(\hat{S}_1^m, \hat{S}_2^m, \tilde{S}_1^m)$ are jointly typical with respect to $p_{\tilde{S}_1, \tilde{S}_2, \hat{S}_1}$ with high probability, and $(\tilde{S}_2^m, \hat{S}_2^m)$ are jointly typical with respect to $p_{\tilde{S}_2, \hat{S}_2}$ with high probability. This completes the proof of the “if” part.

Now we proceed to prove the “only if” part. Consider an arbitrary tuple $(\tilde{\kappa}, \tilde{Q}_1, \tilde{Q}_2) \in \tilde{\Gamma}$. Given any $\epsilon > 0$, according to Definition 3, we can find encoding function $f_{(m,n)}: \tilde{S}_1^m \times \tilde{S}_2^m \rightarrow \mathcal{X}^m$ as well as decoding functions $g_{1{(n,m)}}: \mathcal{Y}_1^m \times \tilde{S}_2^m \rightarrow \hat{S}_1^m$ and $g_{2{(n,m)}}: \mathcal{Y}_2^m \rightarrow \hat{S}_2^m$ such that (2.4)-(2.6) are satisfied. Let $Q$ be a random variable independent of $(\tilde{S}_1^m, \tilde{S}_2^m, X^n, Y_1^n, Y_2^n)$ and uniformly distributed over $\{1, \ldots, n\}$. Define $X = X(Q), Y_i = Y_i(Q), i = 1, 2$, and $V = (V(Q), Q)$, where $V(t) = (Y_1^{t-1}, Y_2^{n-t+1})$. Note that
for all $t$. It is easy to verify that $V \leftrightarrow X \leftrightarrow (Y_1, Y_2)$ form a Markov chain. Note that

$$I(\tilde{S}_1^m; \tilde{S}_1^m | \tilde{S}_2^m) \leq I(\tilde{S}_1^m; Y_1^n | \tilde{S}_2^m)$$

$$\leq I(\tilde{S}_1^m; \tilde{S}_2^m; Y_1^n)$$

$$\leq I(X^n; Y_1^n)$$

$$= \sum_{t=1}^{n} I(X^n; Y_1(t) | Y_1^{t-1})$$

$$\leq \sum_{t=1}^{n} I(X^n, Y_1^{t-1}; Y_1(t))$$

$$= \sum_{t=1}^{n} I(X(t); Y_1(t))$$

$$= n I(X(Q); Y_1(Q) | Q)$$

$$\leq n(Q, X(Q); Y_1(Q))$$

$$= n I(X(Q); Y_1(Q))$$

$$= n I(X; Y_1) \quad (4.1)$$
and

\[
I(\tilde{S}_2^m; \tilde{S}_2^m) \leq I(\tilde{S}_2^m; Y_2^n)
\]

\[
= \sum_{t=1}^{n} I(\tilde{S}_2^m; Y_2(t)|Y_{2,t+1}^n)
\]

\[
\leq \sum_{t=1}^{n} I(Y_{1,t-1}^n, Y_{2,t+1}^n, \tilde{S}_2^m; Y_2(t))
\]

\[
= \sum_{t=1}^{n} I(V(t); Y_2(t))
\]

\[
= nI(V(Q); Y_2(Q)|Q)
\]

\[
\leq nI(V(Q), Q; Y_2(Q))
\]

\[
= nI(V; Y_2). \tag{4.2}
\]
Moreover,

\[
I(\tilde{S}_1^m; \tilde{S}_1^m | \tilde{S}_2^m) + I(\tilde{S}_2^m; \tilde{S}_2^m) \\
\leq I(\tilde{S}_1^m; Y_1^n | \tilde{S}_2^m) + I(\tilde{S}_2^m; Y_2^n) \\
= \sum_{t=1}^{n} [I(\tilde{S}_1^m; Y_1(t) | Y_1^{t-1}, \tilde{S}_2^m) + I(\tilde{S}_2^m; Y_2(t) | Y_2^{t-1}, \tilde{S}_2^m)] \\
\leq \sum_{t=1}^{n} [I(X(t); Y_1(t) | Y_1^{t-1}, \tilde{S}_2^m) + I(\tilde{S}_2^m; Y_2(t) | Y_2^{t-1}, \tilde{S}_2^m)] \\
\leq \sum_{t=1}^{n} [I(X(t); Y_1(t) | Y_1^{t-1}, \tilde{S}_2^m) + I(Y_2^{t-2}; Y_2(t) | Y_2^{t-1}, \tilde{S}_2^m)] \\
= \sum_{t=1}^{n} [I(X(t); Y_1(t) | Y_1^{t-1}, \tilde{S}_2^m) + I(Y_2^{t-2}; Y_2(t)]) \\
= \sum_{t=1}^{n} [I(X(t); Y_1(t) | Y_1^{t-1}, \tilde{S}_2^m) + I(V(t); Y_2(t))]
\]

(4.3)

\[
= \sum_{t=1}^{n} [I(X(t); Y_1(t) | Y_1^{t-1}, \tilde{S}_2^m) + I(V(t); Y_2(t)]) \\
= n[I(X(Q); Y_1(Q) | V(Q), Q) + I(V(Q); Y_2(Q) | Q)] \\
\leq n[I(X(Q); Y_1(Q) | V(Q), Q) + I(V(Q), Q; Y_2(Q))] \\
= nI(X; Y_1 | V) + nI(V; Y_2),
\]

(4.4)

where (4.3) follows by the Csiszár sum identity [22, p. 25]. Let \( T \) be a random variable independent of \( (\tilde{S}_1^m, \tilde{S}_2^m, \tilde{S}_1^m, \tilde{S}_2^m) \) and uniformly distributed over \( \{1, \cdots, m\} \).
Define $\tilde{S}_i = \hat{S}_i(T)$ and $\hat{S}_i^{(e)} = \tilde{S}_i(T)$, $i = 1, 2$. Note that

$$p_{\tilde{S}_1, \tilde{S}_2, \hat{S}_1^{(e)}, \hat{S}_2^{(e)}} = \frac{1}{m} \sum_{t=1}^{m} p_{\tilde{S}_1(t), \tilde{S}_2(t), \hat{S}_1(t), \hat{S}_2(t)}.$$  

Moreover, we have

$$I(\tilde{S}_1^m; \tilde{S}_1^m | \tilde{S}_2^m) = \sum_{t=1}^{m} I(\tilde{S}_1(t); \tilde{S}_1^m | \tilde{S}_1^{t-1}, \tilde{S}_2^m)$$

$$= \sum_{t=1}^{m} I(\tilde{S}_1(t); \tilde{S}_1^m, \tilde{S}_1^{t-1}, \tilde{S}_2^{t-1}, \tilde{S}_2^{m,t+1} | \tilde{S}_2(t))$$

$$\geq \sum_{t=1}^{m} I(\tilde{S}_1(t); \tilde{S}_1(t) | \tilde{S}_2(t))$$

$$= mI(\tilde{S}_1(T); \tilde{S}_1(T) | \tilde{S}_2(T), T)$$

$$= mI(\tilde{S}_1(T); \tilde{S}_1(T), T | \tilde{S}_2(T))$$

$$\geq mI(\tilde{S}_1(T); \hat{S}_1(T) | \tilde{S}_2(T))$$

$$= mI(\tilde{S}_1; \hat{S}_1^{(e)} | \tilde{S}_2)$$

(4.5)
and

\[
I(\tilde{S}_2^m; \hat{S}_2^m) = \sum_{t=1}^{m} I(\tilde{S}_2(t); \hat{S}_2^m | \tilde{S}_2^{t-1}) \\
= \sum_{t=1}^{m} I(\tilde{S}_2(t); \hat{S}_2^m , \tilde{S}_2^{t-1}) \\
\geq \sum_{t=1}^{m} I(\tilde{S}_2(t); \hat{S}_2(t)) \\
= mI(\tilde{S}_2(T); \hat{S}_2(T)|T) \\
= mI(\tilde{S}_2(T); \hat{S}_2(T), T) \\
\geq mI(\tilde{S}_2(T); \hat{S}_2(T)) \\
= mI(\tilde{S}_2; \hat{S}_2^{(\epsilon)}).
\]  

(4.6)

It follows by (4.1), (4.2), (4.4), (4.5), and (4.6) that

\[
(I(\tilde{S}_1; \hat{S}_1^{(\epsilon)} | \tilde{S}_2), I(\tilde{S}_2; \hat{S}_2^{(\epsilon)})) \in \frac{n}{m} C_1(p_{Y_1,Y_2|X}).
\]

Since \(\{p_{\tilde{S}_1,\tilde{S}_2,\hat{S}_1^{(\epsilon)},\hat{S}_2^{(\epsilon)} : \epsilon > 0}\}\) is contained in a compact set and

\[
\min_{\tilde{q}_1 \in Q_1} \|p_{\tilde{S}_1,\tilde{S}_2,\hat{S}_1^{(\epsilon)} - \tilde{q}_1\| \leq \epsilon, \\
\min_{\tilde{q}_2 \in Q_2} \|p_{\tilde{S}_2,\hat{S}_2^{(\epsilon)} - \tilde{q}_2\| \leq \epsilon
\]

for every \(\epsilon > 0, i = 1, 2\), one can find a sequence \(\epsilon_1, \epsilon_2, \cdots\) converging to zero such that

\[
\lim_{k \to \infty} p_{\tilde{S}_1,\tilde{S}_2,\hat{S}_1^{(\epsilon_k)},\hat{S}_2^{(\epsilon_k)}} = p_{\tilde{S}_1,\tilde{S}_2,\hat{S}_1,\hat{S}_2}
\]
for some \( p_{\hat{S}_1, \hat{S}_2, \tilde{S}_1, \tilde{S}_2} \) with \( p_{\hat{S}_1, \hat{S}_2, \tilde{S}_1} \in \tilde{Q}_1 \) and \( p_{\hat{S}_2, \tilde{S}_2} \in \tilde{Q}_2 \). It is clear that

\[
I(\hat{S}_1; \tilde{S}_1|\tilde{S}_2) \geq R_{\tilde{S}_1|\tilde{S}_2}(\tilde{Q}_1),
\]

\[
I(\hat{S}_2; \tilde{S}_2) \geq R_{\tilde{S}_2}(\tilde{Q}_2).
\]

Now the proof can be completed via a simple limiting argument.
Chapter 5

A Necessary Condition for the Source Broadcast Problem

5.1 Necessary Condition

We shall show that the source-channel separation theorem for System $\check{\Pi}$ (i.e., Theorem 1) can be leveraged to establish a necessary condition for System $\Pi$ via a simple reduction argument. Let $\mathcal{R}_1(p_{S,S_1,S_2})$ denote the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \leq I(S; \hat{S}_1|U),$$
$$R_2 \leq I(U; \hat{S}_2)$$
for some \( p_{U,S,S_1,S_2} = p_{U|S} p_{S,S_1,S_2} \). Similarly, let \( R_2(p_{S,S_1,S_2}) \) denote the set of \( (R_1, R_2) \in \mathbb{R}_+^2 \) satisfying

\[
R_1 \leq I(U; \hat{S}_1), \\
R_2 \leq I(S; \hat{S}_2|U)
\]

for some \( p_{U,S,S_1,S_2} = p_{U|S} p_{S,S_1,S_2} \).

**Theorem 2** For any \((\kappa, Q_1, Q_2) \in \Gamma\), there exists \( p_{S,S_1,S_2} \) with \( p_{S_i} \in Q_i, \ i = 1, 2\), such that

\[
R_i(p_{S,S_1,S_2}) \subseteq \kappa C_i(p_{Y_1,Y_2|X}), \ i = 1, 2.
\] (5.1)

**Proof**: By symmetry, it suffices to prove (5.1) for \( i = 1\). We augment the probability space by introducing a remote source \( \{(\tilde{S}_1(t), \tilde{S}_2(t))\}_{t=1}^{\infty} \) such that \((\tilde{S}_1(t), \tilde{S}_2(t), S(t))\), \( t = 1, 2, \ldots \), are independent and identically distributed over finite alphabet \( \tilde{S}_1 \times \tilde{S}_2 \times S \). Consider an arbitrary tuple \((\kappa, Q_1, Q_2) \in \Gamma\). Given any \( \epsilon > 0 \), according to Definition 1, we can find encoding function \( f^{(m,n)} : S^m \rightarrow X^n \) and decoding functions \( g_i^{(m,n)} : Y_i^n \rightarrow \hat{S}_i^m, \ i = 1, 2 \), satisfying (2.1) and (2.2). Let \( T \) be a random variable independent of \((\tilde{S}_1^m, \tilde{S}_2^m, S^m, \hat{S}_1^m, \hat{S}_2^m)\) and uniformly distributed over \( \{1, \ldots, m\} \). Define \( \tilde{S}_i = \tilde{S}_i(T), \ i = 1, 2, S = S(T) \), and \( \hat{S}_i^{(\epsilon)} = \hat{S}_i(T), \ i = 1, 2 \). It is clear that the distribution of \((\tilde{S}_1, \tilde{S}_2, S)\) is identical with that of \((\tilde{S}_1(t), \tilde{S}_2(t), S(t))\) for every \( t \), and \((\tilde{S}_1, \tilde{S}_2) \leftrightarrow S \leftrightarrow (\hat{S}_1^{(\epsilon)}, \hat{S}_2^{(\epsilon)})\) form a Markov chain. Moreover, we have

\[
\frac{1}{m} \sum_{t=1}^{m} p_{\tilde{S}_1(t),\tilde{S}_2(t),S(t),\hat{S}_1(t),\hat{S}_2(t)} = p_{\tilde{S}_1,\tilde{S}_2,S,\hat{S}_1^{(\epsilon)},\hat{S}_2^{(\epsilon)}}.
\]
Since \( \min_{q \in Q_i} \| p_{S,\hat{S}_i} - q_i \| \leq \epsilon \) for every \( \epsilon > 0 \), \( i = 1, 2 \), one can find a sequence \( \epsilon_1, \epsilon_2, \ldots \) converging to zero such that

\[
\lim_{k \to \infty} p_{\tilde{S}_1, \tilde{S}_2, s_{\hat{S}_1}, s_{\hat{S}_2}} = p_{S_1, S, s_{\hat{S}_1}, s_{\hat{S}_2}}
\]

for some \( p_{\tilde{S}_1, \tilde{S}_2, s_{\hat{S}_1}, s_{\hat{S}_2}} \) with \( p_{S_i, \hat{S}_i} \in Q_i, i = 1, 2 \). Note that (5.2) implies \( (\kappa, \{ p_{\tilde{S}_1, \tilde{S}_2, s_{\hat{S}_1}}, \{ \tilde{p}_{S_2, s_{\hat{S}_2}} \} ) \) \( \in \tilde{\Gamma} \). Therefore, it follows from Theorem 1 that

\[
(I(\tilde{S}_1; \hat{S}_1 | S_2), I(\tilde{S}_2; \hat{S}_2)) \in \kappa C_1(p_{Y_1, Y_2} | X).
\]

Here one can fix \( p_{S, \hat{S}_1, \hat{S}_2} \) and choose \( p_{S_i, \hat{S}_i} | s_i \) arbitrarily. Since \( I(\tilde{S}_1; \hat{S}_1 | S_2) \leq I(S; \hat{S}_1 | S_2) \), there is no loss of generality in setting \( \tilde{S}_1 = S \). Denoting \( \tilde{S}_2 \) by \( U \) completes the proof of Theorem 2.

Remark: Since \( C_1(p_{Y_1, Y_2} | X) \) and \( C_2(p_{Y_1, Y_2} | X) \) are convex sets, it follows that (5.1) holds if and only if \( \kappa C_1(p_{Y_1, Y_2} | X) \) contains all the extreme points of \( R_i(p_{S, \hat{S}_1, \hat{S}_2}, i = 1, 2 \). One can show via a standard application of the support lemma [22, p. 631] that, in contrast with the cardinality bound \( |U| \leq |S| + 1 \) for preserving \( R_i(p_{S, \hat{S}_1, \hat{S}_2}, i = 1, 2 \), it suffices to have \( |U| \leq |S| \) for the purpose of realizing all their extreme points.
5.2 The Binary Uniform Source with the Hamming Distortion Measure

In this section we set $S = \hat{S}_1 = \hat{S}_2 = \{0, 1\}$, $p_S(0) = p_S(1) = \frac{1}{2}$, and $w_1 = w_2 = w_H$, where $w_H$ is the Hamming distortion measure, i.e.,

$$w_H(s, \hat{s}) = \begin{cases} 
0, & s = \hat{s} \\
1, & \text{otherwise}
\end{cases}.$$

The problem is trivial\(^1\) when $d_1 = \frac{1}{2}$ or $d_2 = \frac{1}{2}$. Therefore, we shall focus on the non-degenerate case $d_i \in [0, \frac{1}{2})$, $i = 1, 2$, and assume

$$C(p_{Y_i|X}) \triangleq \max_{p_X} I(X; Y_i) > 0, \quad i = 1, 2,$$

correspondingly.

**Proposition 9** If $\mathbb{E}[w_H(S, \hat{S}_i)] \leq d_i$, $i = 1, 2$, with $d_1 \leq d_2$, then

$$\mathcal{R}_1(p_{S, \hat{S}_1, \hat{S}_2}) \supseteq C(BS-BC(d_1, d_2)), \quad (5.3)$$

$$\mathcal{R}_2(p_{S, \hat{S}_1, \hat{S}_2}) \supseteq \tilde{C}(BS-BC(d_1, d_2)), \quad (5.4)$$

where $\mathcal{C}(BS-BC(d_1, d_2))$ (see Section 3.3 for its definition) is given by the set of

\(^1\)In fact, it reduces to a point-to-point problem.
\[(R_1, R_2) \in \mathbb{R}_2^+ \text{ satisfying} \]

\[R_1 \leq H_b(\alpha \ast d_1) - H_b(d_1),\]
\[R_2 \leq 1 - H_b(\alpha \ast d_2)\]

for some \(\alpha \in [0, \frac{1}{2}]\), and \(\tilde{C}(BS-BC(d_1, d_2))\) is given by the set of \((R_1, R_2) \in \mathbb{R}_2^+\) satisfying

\[R_1 \leq \beta[1 - H_b(d_1)],\]
\[R_2 \leq (1 - \beta)[1 - H_b(d_2)]\]

for some \(\beta \in [0, 1]\). Moreover,

\[R_1(p_{S,\hat{S}_1,\hat{S}_2}) = C(BS-BC(d_1, d_2)), \quad (5.5)\]
\[R_2(p_{S,\hat{S}_1,\hat{S}_2}) = \tilde{C}(BS-BC(d_1, d_2)) \quad (5.6)\]

when \(p_{\hat{S}_1,\hat{S}_2|S}\) is a BS-BC\((d_1, d_2)\) with \(d_1 \leq d_2\).

**Proof:** Let \(p_{U,S,\hat{S}_1,\hat{S}_2} = p_{U|S}p_{S,\hat{S}_1,\hat{S}_2}\), where \(p_{U|S}\) is a BSC(\(\alpha\)) with \(\alpha \in [0, \frac{1}{2}]\). We have

\[\min_{p_{\hat{S}_1|S,\tilde{H}(S,\hat{S}_1)|\leq d_1}} I(S;\hat{S}_1|U)\]
\[= \min_{p_{\hat{S}_1|S,\tilde{H}(S,\hat{S}_1)|\leq d_1}} H(U|\hat{S}_1) - H(S|\hat{S}_1)\]
\[= \min_{d'_1 \in [0, d_1]} H_b(\alpha \ast d'_1) - H_b(d'_1) \quad (5.7)\]
\[= H_b(\alpha \ast d_1) - H_b(d_1), \quad (5.8)\]
where (5.7) follows from [11, Lemma 2], and (5.8) is due to the fact that $H_b(\alpha * d'_1) - H_b(d'_1)$ is a monotonically decreasing function of $d'_1$ for $d'_1 \in [0, \frac{1}{2}]$. Similarly, it can be shown that

$$\min_{p_{\hat{S}_1} | S : E[w_H(S,\hat{S}_2) \leq d_2]} I(U; \hat{S}_2) = 1 - H_b(\alpha * d_2).$$

(5.9)

Combining (5.8) and (5.9) proves (5.3).

It is easy to see that $(I(S; \hat{S}_1), 0)$ and $(0, I(S; \hat{S}_2))$ are contained in $\mathcal{R}_2(p_{S,\hat{S}_1,\hat{S}_2})$. Note that

$$I(S; \hat{S}_i) \geq 1 - H_b(d_i)$$

if $E[w_H(S,\hat{S}_i)] \leq d_i$, $i = 1, 2$. Now one can readily prove (5.4) by invoking the fact that $\mathcal{R}_2(p_{S,\hat{S}_1,\hat{S}_2})$ is a convex set.

Since (5.5) is obviously true, only (5.6) remains to be proved. If $p_{\hat{S}_1,\hat{S}_2|S}$ is a BS-BC($d_1, d_2$) with $d_1 \leq d_2$, then, for any $\lambda \in [0, 1]$,

$$\lambda I(U; \hat{S}_1) + (1 - \lambda) I(S; \hat{S}_2|U)$$

$$= \lambda(1 - H(\hat{S}_1|U)) + (1 - \lambda)[H(\hat{S}_2|U) - H_b(d_2)]$$

$$\leq \max_{u \in U} \lambda(1 - H(\hat{S}_1|U = u)) + (1 - \lambda)[H(\hat{S}_2|U = u) - H_b(d_2)]$$

$$\leq \max_{\alpha \in [0, \frac{1}{2}]} \lambda(1 - H_b(\alpha * d_1)) + (1 - \lambda)[H_b(\alpha * d_2) - H_b(d_2)].$$
Define $v = H_b(\alpha * d_1)$, which is a monotonically increasing function of $\alpha$. Note that

$$\lambda(1 - H_b(\alpha * d_1)) + (1 - \lambda)[H_b(\alpha * d_2) - H_b(d_2)]$$
$$= \lambda(1 - v * d)) + (1 - \lambda)[H_b(H_b^{-1}(v)) - H_b(d_2)],$$

where $d = \frac{d_2 - d_1}{1 - 2d_1}$. It follows by the convexity of $H_b(H_b^{-1}(v) * d)$ in $v$ [24, Lemma 2] that

$$\max_{\alpha \in [0, \frac{1}{2}]} \lambda(1 - H_b(\alpha * d_1)) + (1 - \lambda)[H_b(\alpha * d_2) - H_b(d_2)]$$
$$= \max_{\alpha \in (0, \frac{1}{2})} \lambda(1 - H_b(\alpha * d_1)) + (1 - \lambda)[H_b(\alpha * d_2) - H_b(d_2)]. \quad (5.10)$$

Therefore, we must have $R_2(p_{S_1, S_2}) \subseteq \tilde{C}(BS-BC(d_1, d_2))$, which together with (5.4), proves (5.6).

Remark: The proof of Proposition 9 indicates that, for the binary uniform source with the Hamming distortion measure, there is no loss of optimality (as far as Theorem 2 is concerned) in restricting $p_U|S$ to be a binary symmetric channel, which provides a certain justification for the choice of the auxiliary random variable in [13].

Note that the rate pairs $(C(p_{Y_1|X}), 0)$ and $(0, C(p_{Y_2|X}))$ are contained in both $C_1(p_{Y_1,Y_2|X})$ and $C_2(p_{Y_1,Y_2|X})$. It is easy to see that $C(BS-BC(d_1, d_2)) \subseteq \kappa C_1(p_{Y_1,Y_2|X})$ implies

$$1 - H_b(d_i) \leq \kappa C(p_{Y_i|X}), \quad i = 1, 2,$$

which further implies $\tilde{C}(BS-BC(d_1, d_2)) \subseteq \kappa C_2(p_{Y_1,Y_2|X})$ when $d_1 \leq d_2$. This observation, together with Proposition 9, shows that, for the binary uniform source with the
Hamming distortion measure, Theorem 2 is equivalent to the following more explicit result.

**Theorem 3** For any \((\kappa, Q(w_H, d_1), Q(w_H, d_2)) \in \Gamma\) with \(d_1 \leq d_2\),

\[
C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C_1(p_{Y_1, Y_2|X}).
\]

By symmetry, for any \((\kappa, Q(w_H, d_1), Q(w_H, d_2)) \in \Gamma\) with \(d_1 \geq d_2\),

\[
C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C_2(p_{Y_1, Y_2|X}).
\]

Define \(\kappa^\ast = \min\{\kappa \geq 0 : C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C_1(p_{Y_1, Y_2|X})\}\) if \(d_1 \leq d_2\), and \(\kappa^\ast = \min\{\kappa \geq 0 : C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C_2(p_{Y_1, Y_2|X})\}\) if \(d_1 \geq d_2\). It is obvious that

\[
\kappa^\ast \geq \kappa^\dagger \triangleq \max\left\{\frac{1 - H_b(d_1)}{C(p_{Y_1|X})}, \frac{1 - H_b(d_2)}{C(p_{Y_2|X})}\right\},
\]

(5.11)
i.e., the necessary condition stated in Theorem 3 is at least as strong as the one implied by the source-channel separation theorem for point-to-point communication systems. We shall show that in some cases it is possible to determine whether \(\kappa^\ast\) is equal to or strictly greater than \(\kappa^\dagger\) without an explicit characterization of \(C_i(p_{Y_1, Y_2|X}), i = 1, 2\).

Recall that \(C(\text{BS-BC}(d_1, d_2))\) with \(d_1 \leq d_2\) is given by the the set of \((R_1, R_2) \in \mathbb{R}^2_+\) satisfying

\[
R_1 \leq R_1(\alpha) \triangleq H_b(\alpha \ast d_1) - H_b(d_1),
\]

(5.12)
\[
R_2 \leq R_2(\alpha) \triangleq 1 - H_b(\alpha \ast d_2)
\]

(5.13)
for some $\alpha \in [0, \frac{1}{2}]$. It can be verified that

$$\frac{dR_2(\alpha)}{dR_1(\alpha)} = -\frac{(1 - 2d_2) \log \left(\frac{1 - \alpha + d_2}{\alpha + d_2}\right)}{(1 - 2d_1) \log \left(\frac{1 - \alpha + d_1}{\alpha + d_1}\right)}$$

for $\alpha \in (0, \frac{1}{2})$. Moreover, we have$^2$

$$\left.\frac{dR_2(\alpha)}{dR_1(\alpha)}\right|_{\alpha=0} \triangleq \lim_{\alpha \downarrow 0} \frac{R_2(\alpha) - C(p_{Y_2|X})}{R_1(\alpha)}$$

$$= \lim_{\alpha \downarrow 0} \frac{dR_2(\alpha)}{dR_1(\alpha)}$$

$$= \frac{(1 - 2d_2) \log \left(\frac{1 - d_2}{d_2}\right)}{(1 - 2d_1) \log \left(\frac{1 - d_1}{d_1}\right)}$$

and

$$\left.\frac{dR_2(\alpha)}{dR_1(\alpha)}\right|_{\alpha=\frac{1}{2}} \triangleq \lim_{\alpha \uparrow \frac{1}{2}} \frac{R_2(\alpha)}{R_1(\alpha) - C(p_{Y_1|X})}$$

$$= \lim_{\alpha \uparrow \frac{1}{2}} \frac{dR_2(\alpha)}{dR_1(\alpha)}$$

$$= -\frac{(1 - 2d_2)^2}{(1 - 2d_1)^2}.$$
if one of the following conditions are satisfied:

1. \( 1 - H_b(d_1) \leq \kappa C(p_{Y_1|X}) \) and \( \frac{(1-2d_1)^2}{(1-2d_2)^2} \geq \frac{C(p_{Y_1|X})}{C(p_{Y_2|X})} \),

2. \( 1 - H_b(d_2) \leq \kappa C(p_{Y_2|X}) \) and \( \frac{(1-2d_2)^2}{(1-2d_1)^2} \leq \frac{C(p_{Y_2|X})}{C(p_{Y_1|X})} \).

This observation, together with (5.11) as well as the fact that

\[
\{ (R_1, R_2) \in \mathbb{R}_+^2 : \frac{R_1}{C(p_{Y_1|X})} + \frac{R_2}{C(p_{Y_2|X})} \leq 1 \} \subseteq C_1(p_{Y_1,y_2|X}),
\]

yields the following result.

**Proposition 10** If \( d_1 \leq d_2 \), then

\[
\kappa^* = \kappa^\dagger = \begin{cases} 
1 - H_b(d_1) \quad & \frac{(1-2d_1)^2}{(1-2d_2)^2} \geq \frac{C(p_{Y_1|X})}{C(p_{Y_2|X})} \\
1 - H_b(d_2) \quad & \frac{(1-2d_2)^2}{(1-2d_1)^2} \leq \frac{C(p_{Y_2|X})}{C(p_{Y_1|X})} 
\end{cases}
\]

By symmetry, if \( d_1 \geq d_2 \), then

\[
\kappa^* = \kappa^\dagger = \begin{cases} 
1 - H_b(d_2) \quad & \frac{(1-2d_2)^2}{(1-2d_1)^2} \geq \frac{C(p_{Y_2|X})}{C(p_{Y_1|X})} \\
1 - H_b(d_1) \quad & \frac{(1-2d_1)^2}{(1-2d_2)^2} \leq \frac{C(p_{Y_1|X})}{C(p_{Y_2|X})} 
\end{cases}
\]

Remark: A simple sufficient condition for \( (\kappa, Q(w_H, d_1), Q(w_H, d_2)) \in \Gamma \) is that

\[
\max\{1 - H_b(d_1), 1 - H_b(d_2)\} \leq \kappa C(p_{Y_1|X}, p_{Y_2|X}),
\]
where \( C(p_{Y_1|X}, p_{Y_2|X}) \) is the capacity of the compound channel \( \{p_{Y_1|X}, p_{Y_2|X}\}\). Proposition 10 indicates that this sufficient condition is also necessary if \( C(p_{Y_1|X}, p_{Y_2|X}) = C(p_{Y_1|X}) \) and \( d_1 \leq d_2 \) (or \( C(p_{Y_1|X}, p_{Y_2|X}) = C(p_{Y_2|X}) \) and \( d_1 \geq d_2 \)). For the special case \( d_1 = d_2 = d \), it can be shown that \((\kappa, Q(w_H, d), Q(w_H, d)) \in \Gamma\) if and only if

\[
1 - H_b(d) \leq \kappa C(p_{Y_1|X}, p_{Y_2|X}).
\]

On the other hand, for this special case, Proposition 10 gives

\[
\kappa^* = \kappa^\dagger = \max \left\{ \frac{1 - H_b(d)}{C(p_{Y_1|X})}, \frac{1 - H_b(d)}{C(p_{Y_2|X})} \right\}.
\]

Since \( C(p_{Y_1|X}, p_{Y_2|X}) \) can be strictly smaller than \( \min \{C(p_{Y_1|X}), C(p_{Y_2|X})\} \), the necessary condition stated in Theorem 3 is not sufficient in general.

For every \( R_1 \in [0, C(p_{Y_1|X})] \), we set

\[
\phi(R_1) = \max \{ R_2 : (R_1, R_2) \in C_1(p_{Y_1, Y_2|X}) \}.
\]

Note that \( \phi : [0, C(p_{Y_1|X})] \rightarrow [0, C(p_{Y_2|X})] \) is monotonically decreasing and concave. Define

\[
\phi'_+(0) = \lim_{R_1 \downarrow 0} \frac{C(p_{Y_2|X}) - \phi(R_1)}{R_1},
\]

\[
\phi'_-(C(p_{Y_1|X})) = \lim_{R_1 \uparrow C(p_{Y_1|X})} \frac{\phi(R_1)}{C(p_{Y_1|X}) - R_1}.
\]
Similarly, we set

$$\varphi(R_2) = \max\{R_1 : (R_1, R_2) \in C_2(p_{Y_1,Y_2|X})\}$$

for every $R_2 \in [0, C(p_{Y_2|X})]$, and define

$$\varphi'_+(0) = \lim_{R_2 \downarrow 0} \frac{C(p_{Y_1|X}) - \varphi(R_2)}{R_2},$$

$$\varphi'_-(C(p_{Y_2|X})) = \lim_{R_2 \uparrow C(p_{Y_2|X})} \frac{\varphi(R_2)}{C(p_{Y_2|X}) - R_2}.$$

Now consider the case $d_1 \leq d_2$. It is clear that we must have $1 - H_b(d_1) < \kappa^* C(p_{Y_1|X})$ if

$$\frac{(1 - 2d_2)^2}{(1 - 2d_1)^2} > \varphi'_-(C(p_{Y_1|X})); \quad (5.14)$$

similarly, we must have $1 - H_b(d_2) < \kappa^* C(p_{Y_2|X})$ if

$$\frac{(1 - 2d_2) \log \left(\frac{1-d_2}{d_2}\right)}{(1 - 2d_1) \log \left(\frac{1-d_1}{d_1}\right)} < \varphi'_+(0); \quad (5.15)$$

moreover, since $\varphi'_+(0) \leq \varphi'_-(C(p_{Y_1|X}))$, it follows that (5.14) and (5.15) cannot be satisfied simultaneously when $d_1 = d_2$. The following result is a simple consequence of this observation.
Proposition 11 When \( d_1 < d_2 \), we have \( \kappa^* > \kappa^\dagger \) if

\[
\frac{(1 - 2d_2) \log \left( \frac{1 - d_2}{d_2} \right)}{(1 - 2d_1) \log \left( \frac{1 - d_1}{d_1} \right)} < \varphi'_+(0),
\]

\[
\frac{(1 - 2d_2)^2}{(1 - 2d_1)^2} > \varphi'_-(C(p_{Y_1|X})).
\]

By symmetry, when \( d_1 > d_2 \), we have \( \kappa^* > \kappa^\dagger \) if

\[
\frac{(1 - 2d_1) \log \left( \frac{1 - d_1}{d_1} \right)}{(1 - 2d_2) \log \left( \frac{1 - d_2}{d_2} \right)} < \varphi'_+(0),
\]

\[
\frac{(1 - 2d_1)^2}{(1 - 2d_2)^2} > \varphi'_-(C(p_{Y_2|X})).
\]

A channel \( p_{Y|X} : \mathcal{X} \rightarrow \mathcal{Y} \) with \( \mathcal{X} = \{0, 1, \ldots, M - 1\} \) for some integer \( M \geq 2 \) is said to be circularly symmetric [25, Definition 1] (see also [21, Definition 4]) if there exists a bijective function \( \mu : \mathcal{Y} \rightarrow \mathcal{Y} \) such that \( \mu^M(y) = y \) and \( p_{Y|X}(\mu^k(y)|x) = p_{Y|X}(y|0) \) for all \( (x, y) \in \mathcal{X} \times \mathcal{Y} \), where \( \mu^k \) denotes the \( k \)-times self-composition of \( \mu \) (with \( \mu^0 \) being the identity function). Note that the binary symmetric channel is circularly symmetric with \( \mu : \{0, 1\} \rightarrow \{0, 1\} \) given by

\[
\mu(y) = \begin{cases} 
1, & y = 0 \\
0, & y = 1 
\end{cases}.
\]
the binary erasure channel is also circularly symmetric, and the associated $\mu : \{0, 1, e\} \rightarrow \{0, 1, e\}$ is given by

$$\mu(y) = \begin{cases} 
1, & y = 0 \\
0, & y = 1 \\
e, & y = e 
\end{cases}.$$ 

**Proposition 12** If both $p_{Y_1|X}$ and $p_{Y_2|X}$ are circularly symmetric, then

$$\kappa^* = \min\{\kappa \geq 0 : C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C(p_{Y_1,Y_2|X})\}.$$ 

**Proof**: By symmetry, it suffices to consider the case $d_1 \leq d_2$. Let $C_{\text{SC}}(p_{Y_1,Y_2|X})$ denote the superposition coding inner bound of $C(p_{Y_1,Y_2|X})$, i.e., the set of $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_2 \leq I(V; Y_2),$$

$$R_1 + R_2 \leq I(X; Y_1|V) + I(V; Y_2),$$

$$R_1 + R_2 \leq I(X; Y_1)$$

for some $p_{V,X,Y_1,Y_2} = p_{V,X}p_{Y_1,Y_2|X}$. In light of [21, Lemma 2], the uniform distribution on $\mathcal{X}$ forms a sufficient class of distributions for broadcast channel $p_{Y_1,Y_2|X}$ if both $p_{Y_1|X}$ and $p_{Y_2|X}$ are circularly symmetric. As a consequence, one can readily show that

$$C_{\text{SC}}(p_{Y_1,Y_2|X}) = C_1(p_{Y_1,Y_2|X}) \cap \{(R_1, R_2) : R_1 + R_2 \leq C(p_{Y_1|X})\},$$
Note that, if $C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C_1(p_{Y_1, Y_2|X})$, then we must have

$$1 - H_b(d_1) \leq \kappa C(p_{Y_1|X}),$$

which, together with the fact that $\frac{dR_2(\alpha)}{dH_1(\alpha)} \in [-1, 0]$ for $\alpha \in [0, \frac{1}{2}]$, implies

$$C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa \{(R_1, R_2) : R_1 + R_2 \leq C(p_{Y_1|X})\}.$$ 

Therefore,

$$C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C_1(p_{Y_1, Y_2|X})$$

$$\Rightarrow C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C_{sc}(p_{Y_1, Y_2|X}).$$

Since $C_{sc}(p_{Y_1, Y_2|X}) \subseteq C(p_{Y_1, Y_2|X}) \subseteq C_1(p_{Y_1, Y_2|X})$, the proof is complete.

Now we proceed to consider several concrete examples.

### 5.3 BS-BC$(p_1, p_2)$

First consider the case where $p_{Y_1, Y_2|X}$ is a BS-BC$(p_1, p_2)$ with $0 \leq p_1 \leq p_2 < \frac{1}{2}$.

Without loss of generality, we shall assume $d_1 \leq d_2$. By Theorem 3 and Proposition 4 (or by Theorem 3 and Proposition 12), if $(\kappa, Q(w_H, d_1), Q(w_H, d_2)) \in \Gamma$, then

$$C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C(\text{BS-BC}(p_1, p_2)).$$ (5.16)
On the other hand, the necessary condition implied by the source-channel separation theorem for point-to-point communication systems is

\[ 1 - H_b(d_i) \leq \kappa [1 - H_b(p_i)], \quad i = 1, 2. \]  

(5.17)

For the special case \( \kappa = 1 \), both (5.16) and (5.17) reduce to

\[ d_i \geq p_i, \quad i = 1, 2, \]

which is achievable by the uncoded scheme.

In view of Proposition 4 and the calculation of \( \frac{dr_2(\alpha)}{dr_1(\alpha)} \) for \( \alpha = 0 \) and \( \alpha = \frac{1}{2} \), we have

\[
\phi'_+(0) = \frac{(1 - 2p_2) \log \left( \frac{1-p_2}{p_2} \right)}{(1 - 2p_1) \log \left( \frac{1-p_1}{p_1} \right)},
\]

\[
\phi'_-(C(p_{Y_1}|X)) = \frac{(1 - 2p_2)^2}{(1 - 2p_1)^2}.
\]

Hence, it follows from Proposition 11 that \( \kappa^* > \kappa^\dagger \) if

\[
\frac{(1 - 2d_2) \log \left( \frac{1-d_2}{d_2} \right)}{(1 - 2d_1) \log \left( \frac{1-d_1}{d_1} \right)} < \frac{(1 - 2p_2) \log \left( \frac{1-p_2}{p_2} \right)}{(1 - 2p_1) \log \left( \frac{1-p_1}{p_1} \right)},
\]

(5.18)

\[
\frac{(1 - 2d_2)^2}{(1 - 2d_1)^2} > \frac{(1 - 2p_2)^2}{(1 - 2p_1)^2}.
\]

(5.19)

For example, (5.18) and (5.19) are satisfied when \( d_1 = 0.035 \), \( d_2 = 0.095 \), \( p_1 = 0.15 \), and \( p_2 = 0.2 \).
5.3.1 \textbf{BE-BC}(\epsilon_1, \epsilon_2)

Next consider the case where \( p_{Y_1,Y_2|X} \) is a BE-BC(\(\epsilon_1, \epsilon_2\)) with \(0 \leq \epsilon_1 \leq \epsilon_2 < 1\). Without loss of generality, we shall assume \(d_1 \leq d_2\). By Proposition 5 (or by Proposition 12),

\[
\kappa^* = \min \{ \kappa \geq 0 : C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C(\text{BE-BC}(\epsilon_1, \epsilon_2)) \},
\]

where the expressions of \(C(\text{BS-BC}(d_1, d_2))\) and \(C(\text{BE-BC}(\epsilon_1, \epsilon_2))\) can be found in (5.12)-(5.13) and (3.9)-(3.10), respectively. It is clear that, for any \(\alpha \in [0, \frac{1}{2}]\), there exists \(\beta \in [0, 1]\) such that

\[
\begin{align*}
H_b(\alpha * d_1) - H_b(d_1) &\leq \kappa^* \beta (1 - \epsilon_1), \\
1 - H_b(\alpha * d_2) &\leq \kappa^* (1 - \beta) (1 - \epsilon_2),
\end{align*}
\]

which implies

\[
\kappa^* \geq \frac{H_b(\alpha * d_1) - H_b(d_1)}{1 - \epsilon_1} + \frac{1 - H_b(\alpha * d_2)}{1 - \epsilon_2}
\]

for any \(\alpha \in [0, \frac{1}{2}]\). Moreover, the equalities must hold in (5.20) and (5.21) for some \(\alpha \in [0, \frac{1}{2}]\) and \(\beta \in [0, 1]\); as a consequence, the equality must hold in (5.22) for some \(\alpha \in [0, \frac{1}{2}]\). Therefore, we have

\[
\kappa^* = \max_{\alpha \in [0, \frac{1}{2}]} \frac{H_b(\alpha * d_1) - H_b(d_1)}{1 - \epsilon_1} + \frac{1 - H_b(\alpha * d_2)}{1 - \epsilon_2},
\]

for any \(\alpha \in [0, \frac{1}{2}]\).
from which one can readily recover [13, Theorem 1] by invoking Theorem 3. In light of [11, Lemma 2], for the optimization problem in (5.23), the maximum value is not attained at $\alpha = 0$ or $\alpha = \frac{1}{2}$ if and only if

$$\frac{(1 - 2d_2) \log \left( \frac{1 - d_2}{d_2} \right)}{(1 - 2d_1) \log \left( \frac{1 - d_1}{d_1} \right)} < \frac{1 - \epsilon_2}{1 - \epsilon_1} < \frac{(1 - 2d_2)^2}{(1 - 2d_1)^2},$$

which gives the necessary and sufficient condition for $\kappa^* > \kappa^\dagger$ to hold. The same condition can be obtained through Proposition 10 and Proposition 11.

### 5.3.2 BSC($p$)&BEC($\epsilon$)

Finally consider the case where $p_{Y_1,Y_2|X}$ is a BSC($p$)&BEC($\epsilon$) with $p \in [0, \frac{1}{2})$ and $\epsilon \in [0, 1)$. By Proposition 12,

$$\kappa^* = \min\{\kappa \geq 0 : C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C(\text{BSC}(p)&\text{BEC}(\epsilon))\}. \quad (5.24)$$

Note that

$$\kappa^* \geq \kappa^\dagger = \max \left\{ \frac{1 - H_b(d_1)}{1 - H_b(p)}, \frac{1 - H_b(d_2)}{1 - \epsilon} \right\}.$$  

For the case $d_1 \leq d_2$, in view of the expression of $C(\text{BSC}(p)&\text{BEC}(\epsilon))$ and the fact that $\frac{dR_2(\alpha)}{dR_1(\alpha)} \in [-1, 0]$ for $\alpha \in [0, \frac{1}{2}]$, one can readily verify that

$$C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C(\text{BSC}(p)&\text{BEC}(\epsilon))$$

$$\Leftrightarrow C(\text{BS-BC}(d_1, d_2)) \subseteq \kappa C(\text{BE-BC}(H_b(p), \epsilon));$$
as a consequence, 

\[ \kappa^* = \max_{\alpha \in [0, \frac{1}{2}]} \frac{H_b(\alpha \ast d_1) - H_b(d_1)}{1 - H_b(p)} + \frac{1 - H_b(\alpha \ast d_2)}{1 - \epsilon}, \]

and we have \( \kappa^* > \kappa^\dagger \) if and only if

\[
\frac{(1 - 2d_2) \log \left( \frac{1-d_2}{d_2} \right)}{(1 - 2d_1) \log \left( \frac{1-d_1}{d_1} \right)} < \frac{1 - \epsilon}{1 - H_b(p)} < \left( \frac{1 - 2d_2}{1 - 2d_1} \right)^2.
\]

For the case \( d_1 \geq d_2 \), we shall show that

\[
\mathcal{C}(\text{BS-BC}(d_1, d_2)) \subseteq \kappa \mathcal{C}(\text{BSC}(p) \& \text{BEC}(\epsilon))
\]

\[ \Leftrightarrow \mathcal{C}(\text{BS-BC}(d_1, d_2)) \subseteq \kappa \hat{\mathcal{C}}(\text{BSC}(p) \& \text{BEC}(\epsilon)), \tag{5.25} \]

where \( \hat{\mathcal{C}}(\text{BSC}(p) \& \text{BEC}(\epsilon)) \) is given by the set\(^3\) of \((R_1, R_2) \in \mathbb{R}_{+}^2\) satisfying

\[
R_1 \leq 1 - H_b(\alpha \ast p),
\]

\[
R_2 \leq (1 - \epsilon)H_b(\alpha)
\]

for some \( \epsilon \in [0, \frac{1}{2}] \). It is easy to see that (5.25) is true when \( \epsilon \in [H_b(p), 1) \); moreover,

\[
\mathcal{C}(\text{BSC}(p) \& \text{BEC}(\epsilon)) = \hat{\mathcal{C}}(\text{BSC}(p) \& \text{BEC}(\epsilon)) \cap \{(R_1, R_2) : R_1 + R_2 \leq 1 - \epsilon\}
\]

\(^3\)It follows from [24, Lemma 2] that \( \hat{\mathcal{C}}(\text{BSC}(p) \& \text{BEC}(\epsilon)) \) is a convex set.
when $\epsilon \in [0, H_b(p))$. Combining this observation with the fact that

$$
\mathcal{C}(\text{BS-BC}(d_1, d_2)) \subseteq \kappa \hat{\mathcal{C}}(\text{BSC}(p) \& \text{BEC}(\epsilon))
$$

$$
\Rightarrow 1 - H_b(d_2) \leq \kappa(1 - \epsilon)
$$

$$
d_1 \geq d_2 \mathcal{C}(\text{BS-BC}(d_1, d_2)) \subseteq \kappa\{(R_1, R_2) : R_1 + R_2 \leq 1 - \epsilon\}
$$

proves (5.25). Now we proceed to show that if $\kappa^* = \kappa^\dagger$ if $\kappa^\dagger \geq 1$. In view of (5.24) and (5.25), it suffices to show that, if $\kappa \geq 1$, then

$$
1 - H_b(\alpha \ast d_1) \leq \kappa^\dagger[1 - H_b(\alpha \ast p)], \quad (5.26)
$$

$$
H_b(\alpha \ast d_2) - H_b(d_2) \leq \kappa^\dagger(1 - \epsilon)H_b(\alpha) \quad (5.27)
$$

for any $\alpha \in [0, \frac{1}{2}]$. Note that (5.26) and (5.27) hold when $\alpha = 0$ or $\alpha = \frac{1}{2}$. Moreover, $\kappa^\dagger \geq 1$ implies $p \geq d_1$. Therefore, an argument similar to that for (5.10) can be used here to finish the proof.

\footnote{This result is not implied by Proposition 10.}
Chapter 6

The Quadratic Gaussian Case

Let \( \{S(t)\}_{t=1}^{\infty} \) in System \( \Pi \) be an i.i.d. vector Gaussian process, where each \( S(t) \) is an \( \ell \times 1 \) zero-mean Gaussian random vector with positive definite covariance matrix \( \Sigma_S \). The following definition is the quadratic Gaussian counterpart of Definition 1.

**Definition 6** Let \( \kappa \) be a non-negative number and \( \mathcal{D}_i \) be a non-empty compact set of \( \ell \times \ell \) positive semi-definite matrices, \( i = 1, 2 \). We say \( (\kappa, \mathcal{D}_1, \mathcal{D}_2) \) is achievable for System \( \Pi \) if, for every \( \epsilon > 0 \), there exist encoding function \( f^{(m,n)} : \mathbb{R}^{\ell \times m} \to \mathcal{X}^n \) and decoding functions \( g_i^{(n,m)} : \mathcal{Y}_i^n \to \mathbb{R}^{\ell \times m} \), \( i = 1, 2 \), such that

\[
\frac{n}{m} \leq \kappa + \epsilon,
\]

\[
\min_{\mathcal{D}_i \in \mathcal{D}_i} \left\| \frac{1}{m} \sum_{t=1}^{m} \mathbb{E}[ (S(t) - \hat{S}_i(t))(S(t) - \hat{S}_i(t))^T ] - D_i \right\| \leq \epsilon, \quad i = 1, 2.
\]

The set of all achievable \( (\kappa, \mathcal{D}_1, \mathcal{D}_2) \) for System \( \Pi \) is denoted by \( \Gamma_G \).
Remark: It is clear that \((\kappa, D_1, D_2) \in \Gamma_G\) if and only if \((\bar{\kappa}, D_1, D_2) \in \Gamma_G\), where

\[
\bar{D}_i = \bigcup_{D_i \in \mathcal{D}_i} \{D'_i : 0 \leq D'_i \leq D_i\}, \quad i = 1, 2.
\]

Furthermore, to determine whether or not \((\kappa, \bar{D}_1, \bar{D}_2) \in \Gamma_G\), there is no loss of generality in setting \(\hat{\mathbf{S}}_m \equiv \mathbb{E}[S^m | Y^n], i = 1, 2\), for which we have

\[
\frac{1}{m} \sum_{t=1}^{m} \mathbb{E}[(S(t) - \hat{S}_i(t))(S(t) - \hat{S}_i(t))^T] \preceq \Sigma_s, \quad i = 1, 2.
\]

Therefore, it suffices to consider those \(D_1\) and \(D_2\) with the property that

\[
D_i = \bar{D}_i \cap \{D : 0 \leq D \leq \Sigma_s\}, \quad i = 1, 2. \tag{6.1}
\]

Henceforth we shall implicitly assume that (6.1) is satisfied.

Now we proceed to introduce the corresponding System \(\tilde{\Pi}\) in the quadratic Gaussian setting and establish its associated source-channel separation theorem. Let \(\tilde{S} \equiv (\tilde{S}_1^T, \tilde{S}_2^T)^T\) be an \(\tilde{\ell} \times 1\) zero-mean Gaussian random vector with positive definite covariance matrix \(\Sigma_{\tilde{S}}\), where \(\tilde{S}_i\) is an \(\tilde{\ell}_i \times 1\) random vector, and its covariance matrix is denoted by \(\Sigma_{\tilde{S}_i}, i = 1, 2\). Let \(\{(\tilde{S}(t), \tilde{S}_1(t), \tilde{S}_2(T))\}_{t=1}^{\infty}\) be i.i.d. copies of \((\tilde{S}, \tilde{S}_1, \tilde{S}_2)\).

**Definition 7** Let \(\tilde{\kappa}\) be a non-negative number, \(\tilde{D}_1\) be a non-empty compact subset of \(\{\tilde{D}_1 : 0 \leq \tilde{D}_1 \leq \Sigma_{\tilde{S}_1}\}\), and \(\tilde{D}_2\) be a non-empty compact subset of \(\{\tilde{D}_2 : 0 \leq \tilde{D}_2 \leq \Sigma_{\tilde{S}_2}\}\). We say \((\tilde{\kappa}, \tilde{D}_1, \tilde{D}_2)\) is achievable for System \(\tilde{\Pi}\) if, for every \(\epsilon > 0\), there exist an encoding function \(f^{(m,n)} : \mathbb{R}^{\ell_1 \times m} \times \mathbb{R}^{\ell_2 \times m} \to \mathcal{X}^n\) as well as decoding functions
\[ g_1^{(n,m)} : \mathcal{Y}_1^n \times \mathbb{R}^{\ell_2 \times m} \to \mathbb{R}^{\ell \times m} \text{ and } g_2^{(n,m)} : \mathcal{Y}_2^n \to \mathbb{R}^{\ell_2 \times m} \text{ such that} \]
\[
\frac{n}{m} \leq \bar{\kappa} + \epsilon,
\]
\[
\min_{\tilde{D}_1 \in \tilde{D}_1} \left\| \sum_{t=1}^{m} \mathbb{E}[(\tilde{S}(t) - \hat{S}_1(t))(\tilde{S}(t) - \hat{S}_1(t))^T] - \tilde{D}_1 \right\| \leq \epsilon,
\]
\[
\min_{\tilde{D}_2 \in \tilde{D}_2} \left\| \sum_{t=1}^{m} \mathbb{E}[(\tilde{S}_2(t) - \hat{S}_2(t))(\tilde{S}_2(t) - \hat{S}_2(t))^T] - \tilde{D}_2 \right\| \leq \epsilon.
\]

The set of all achievable \((\bar{\kappa}, \tilde{D}_1, \tilde{D}_2)\) for System \(\tilde{\Pi}\) is denoted by \(\tilde{\Gamma}_G\).

Remark: Similar to Definition 3, here we allow \(f^{(m,n)}\), \(g_1^{(n,m)}\), and \(g_2^{(n,m)}\) to be non-deterministic functions as long as the Markov chains \((\tilde{S}_1^n, \tilde{S}_2^n) \leftrightarrow X^n \leftrightarrow (Y_1^n, Y_2^n)\), \(\tilde{S}_1^n \leftrightarrow (Y_1^n, \tilde{S}_2^n) \leftrightarrow \hat{S}_1^n\), and \(\tilde{S}_2^n \leftrightarrow Y_2^n \leftrightarrow \hat{S}_2^n\) are preserved.

Note that
\[
\Sigma_{\tilde{S}} = \begin{pmatrix}
\Sigma_{\tilde{S}_1} & \Sigma_{\tilde{S}_1, \tilde{S}_2} \\
\Sigma_{\tilde{S}_2, \tilde{S}_1} & \Sigma_{\tilde{S}_2}
\end{pmatrix},
\]

where \(\Sigma_{\tilde{S}_1, \tilde{S}_2} = \mathbb{E}[\tilde{S}_1 \tilde{S}_2^T]\) and \(\Sigma_{\tilde{S}_2, \tilde{S}_1} = \mathbb{E}[\tilde{S}_2 \tilde{S}_1^T]\). Moreover, we write
\[
\tilde{D}_1 = \begin{pmatrix}
\tilde{D}_{1,1} & \tilde{D}_{1,2} \\
\tilde{D}_{2,1} & \tilde{D}_{2,2}
\end{pmatrix}
\]

for any \(\tilde{D}_1 \in \tilde{D}_1\), where \(\tilde{D}_{i,i}\) is an \(\ell_i \times \ell_i\) matrix. The following source-channel separation theorem is a simple translation of Theorem 1 to the quadratic Gaussian setting. Its proof is omitted.

Theorem 4 \((\bar{\kappa}, \tilde{D}_1, \tilde{D}_2) \in \tilde{\Gamma}_G \text{ if and only if } (R_{\tilde{S}_1 | \tilde{S}_2}(\tilde{D}_1), R_{\tilde{S}_2}(\tilde{D}_2)) \in \bar{\kappa}C_1(p_{Y_1, Y_2 | X})\),

55
where

\[ R_{\tilde{S}_1|\tilde{S}_2}(\tilde{D}_1) = \min_{\tilde{D}_1 \in \tilde{D}_1} \frac{1}{2} \log \left( \frac{||\Sigma_{\tilde{S}_1} - \Sigma_{\tilde{S}_1,\tilde{S}_2}\Sigma_{\tilde{S}_2,\tilde{S}_1}^{-1}\Sigma_{\tilde{S}_2}|}{|\tilde{D}_1,1 - K\tilde{D}_2,1|} \right), \]

\[ R_{\tilde{S}_2}(\tilde{D}_2) = \min_{\tilde{D}_2 \in \tilde{D}_2} \frac{1}{2} \log \left( \frac{||\Sigma_{\tilde{S}_2}|}{|\tilde{D}_2|} \right), \]

with \( K \) being any solution\(^1\) of \( K\tilde{D}_2,2 = \tilde{D}_1,2 \).

Remark: It can be verified that

\[ R_{\tilde{S}_1|\tilde{S}_2}(\tilde{D}_1) = \min_{\tilde{S}_1|\tilde{S}_2}:E[(\tilde{S}_1 - \tilde{S}_1)(\tilde{S}_1 - \tilde{S}_1)^T] \in \tilde{D}_1 I(\tilde{S}_1; \tilde{S}_1|\tilde{S}_2), \]

\[ R_{\tilde{S}_2}(\tilde{D}_2) = \min_{\tilde{S}_2|\tilde{S}_2}:E[(\tilde{S}_2 - \tilde{S}_2)(\tilde{S}_2 - \tilde{S}_2)^T] \in \tilde{D}_2 I(\tilde{S}_2; \tilde{S}_2), \]

which highlights the similarity between Theorem 1 and Theorem 4.

Again, in the quadratic Gaussian setting, the source-channel separation theorem for System \( \Pi \) can be leveraged to derive a necessary condition for System \( \Pi \). For any \( D_i \in \mathcal{D}_i, \ i = 1, 2 \), let \( R_1(\Sigma_S, D_1, D_2) \) denote the convex closure of the set of \((R_1, R_2) \in \mathbb{R}_+^2\) satisfying

\[ R_1 \leq \frac{1}{2} \log \left( \frac{||\Sigma_S||D_1 + \Sigma_Z||}{|D_1||\Sigma_S + \Sigma_Z|} \right), \]

\[ R_2 \leq \frac{1}{2} \log \left( \frac{||\Sigma_S + \Sigma_Z||}{D_2 + \Sigma_Z} \right). \]

---

\(^1\)If \( \tilde{D}_2,2 \) is invertible, then \( K = \tilde{D}_1,2\tilde{D}_2,2^{-1}. \)
for some \( \Sigma_Z \succ 0 \), and let \( \mathcal{R}_2(\Sigma_S, D_1, D_2) \) denote the convex closure of the set of \((R_1, R_2) \in \mathbb{R}_+^2\) satisfying

\[
R_1 \leq \frac{1}{2} \log \left( \frac{|\Sigma_S + \Sigma_Z|}{|D_1 + \Sigma_Z|} \right), \\
R_2 \leq \frac{1}{2} \log \left( \frac{|\Sigma_S||D_2 + \Sigma_Z|}{|D_2||\Sigma_S + \Sigma_Z|} \right)
\]

for some \( \Sigma_Z \succ 0 \). By setting \( \Sigma_U = \Sigma_S(\Sigma_S + \Sigma_Z)^{-1}\Sigma_S \), we can write \( \mathcal{R}_1(\Sigma_S, D_1, D_2) \) equivalently as the convex hull of the set of \((R_1, R_2) \in \mathbb{R}_+^2\) such that

\[
R_1 \leq \frac{1}{2} \log \left( \frac{|\Sigma_U \Sigma_S^{-1}D_1 + \Sigma_S - \Sigma_U|}{|D_1|} \right), \\
R_2 \leq \frac{1}{2} \log \left( \frac{|\Sigma_U|}{|\Sigma_U \Sigma_S^{-1}D_2 + \Sigma_S - \Sigma_U|} \right)
\]

for some \( \Sigma_U \) satisfying \( 0 \preceq \Sigma_U \preceq \Sigma_S \); similarly, \( \mathcal{R}_1(\Sigma_S, D_1, D_2) \) can be written equivalently as the convex hull of the set of \((R_1, R_2) \in \mathbb{R}_+^2\) such that

\[
R_1 \leq \frac{1}{2} \log \left( \frac{|\Sigma_S|}{|\Sigma_U \Sigma_S^{-1}D_1 + \Sigma_S - \Sigma_U|} \right), \\
R_2 \leq \frac{1}{2} \log \left( \frac{|\Sigma_U \Sigma_S^{-1}D_2 + \Sigma_S - \Sigma_U|}{|D_2|} \right)
\]

for some \( \Sigma_U \) satisfying \( 0 \preceq \Sigma_U \preceq \Sigma_S \).

Let \( S \) be an \( \ell \times 1 \) zero-mean Gaussian random vector with positive definite covariance matrix \( \Sigma_S \). Recall the definition of \( \mathcal{R}_i(p_S, \hat{S}_1, \hat{S}_2), i = 1, 2 \) in Section 5. The following result provides a connection between \( \mathcal{R}_i(\Sigma_S, D_1, D_2) \) and \( \mathcal{R}_i(p_S, \hat{S}_1, \hat{S}_2), i = 1, 2 \).
Proposition 13  If $\mathbb{E}[(S - \hat{S}_i)(S - \hat{S}_i)^T] = D_i \in \mathcal{D}_i$, $i = 1, 2$, then

$$\mathcal{R}_i(p_{S,\hat{S}_1,\hat{S}_2}) \supseteq \mathcal{R}_i(\Sigma_S, D_1, D_2), \quad i = 1, 2. \quad (6.2)$$

Moreover, if $S - \hat{S}_i$ and $\hat{S}_i$ are independent zero-mean Gaussian random vectors with covariance matrices $D_i$ and $\Sigma_S - D_i$, respectively, $i = 1, 2$, where $0 \preceq D_1 \preceq D_2 \preceq \Sigma_S$, then

$$\mathcal{R}_1(p_{S,\hat{S}_1,\hat{S}_2}) = \mathcal{R}_1(\Sigma_S, D_1, D_2), \quad (6.3)$$
$$\mathcal{R}_2(p_{S,\hat{S}_1,\hat{S}_2}) \subseteq \{ (R_1, R_2) \in \mathbb{R}_+^2 : R_2 \leq \frac{1}{2} \log \left( \frac{|\Sigma_S|}{|D_2|} \right), R_1 + R_2 \leq \frac{1}{2} \log \left( \frac{|\Sigma_S|}{|D_1|} \right) \} \quad (6.4)$$

Proof: By symmetry, it suffices to prove (6.2) for $i = 1$. Given any $\Sigma_U$ satisfying $0 \preceq \Sigma_U \preceq \Sigma_S$, we can find $U$ jointly distributed with $S$ such that $U$ and $S - U$ are independent zero-mean Gaussian random vectors with covariance matrices $\Sigma_U$ and $\Sigma_S - \Sigma_U$, respectively. Note that for any $(\hat{S}_1, \hat{S}_2)$ jointly distributed with such $(U, S)$ subject to the constraints that $\mathbb{E}[(S - \hat{S}_i)(S - \hat{S}_i)^T] = D_i \in \mathcal{D}_i$, $i = 1, 2$, and that $U \leftrightarrow S \leftrightarrow (\hat{S}_1, \hat{S}_2)$ form a Markov chain, we have

$$I(S; \hat{S}_1|U) \geq \frac{1}{2} \log \left( \frac{|\Sigma_U \Sigma_S^{-1} D_1 + \Sigma_S - \Sigma_U|}{|D_1|} \right), \quad (6.5)$$
$$I(U; \hat{S}_2) \geq \frac{1}{2} \log \left( \frac{|\Sigma_S|}{|\Sigma_U \Sigma_S^{-1} D_2 + \Sigma_S - \Sigma_U|} \right), \quad (6.6)$$

where the equalities in (6.5) and (6.6) hold when $S - \hat{S}_i$ and $\hat{S}_i$ are independent zero-mean Gaussian random vectors with covariance matrices $D_i$ and $\Sigma_S - D_i$, respectively, $i = 1, 2$. Now the desired result follows by the convexity of $\mathcal{R}_1(p_{S,\hat{S}_1,\hat{S}_2})$. 

58
To prove (6.3), it suffices to consider the non-degenerate case $0 \prec D_1 \preceq D_2 \prec \Sigma_S$; the general case $0 \preceq D_1 \preceq D_2 \preceq \Sigma_S$ can be proved via a simple limiting argument. Let $O_i$ be a zero-mean Gaussian random vector, independent of $(U, S)$, with covariance matrix $\Sigma_{O_i} = (D_i^{-1} - \Sigma_S^{-1})^{-1}$, $i = 1, 2$. It is clear that

$$I(S; \hat{S}_1|U) = I(S; S + O_1|U),$$

$$I(U; \hat{S}_2) = I(U; S + O_2).$$

For any $\lambda \in [0, 1],$

$$\max_{(R_1, R_2) \in \mathcal{R}_1(p_S, S_1, S_2)} \lambda R_1 + (1 - \lambda) R_2$$

$$= \max_{pu|S} \lambda I(S; \hat{S}_1|U) + (1 - \lambda) I(U; \hat{S}_2)$$

$$= \max_{pu|S} \lambda I(S; S + O_1|U) + (1 - \lambda) I(U; S + O_2)$$

$$= \max_{0 \preceq \Sigma_U \preceq \Sigma_S} \frac{\lambda}{2} \log \left( \frac{\|\Sigma_S - \Sigma_U + \Sigma_{O_1}\|}{\|\Sigma_{O_1}\|} \right) + \frac{1 - \lambda}{2} \log \left( \frac{\|\Sigma_S + \Sigma_{O_2}\|}{\|\Sigma_S - \Sigma_U + \Sigma_{O_2}\|} \right)$$

$$= \max_{0 \preceq \Sigma_U \preceq \Sigma_S} \frac{\lambda}{2} \log \left( \frac{\|\Sigma_U \Sigma_S^{-1} D_1 + \Sigma_S - \Sigma_U\|}{|D_1|} \right) + \frac{1 - \lambda}{2} \log \left( \frac{\|\Sigma_S \Sigma_U^{-1} D_2 + \Sigma_S - \Sigma_U\|}{|\Sigma_U \Sigma_S^{-1} D_2 + \Sigma_S - \Sigma_U|} \right)$$

$$= \max_{(R_1, R_2) \in \mathcal{R}_1(\Sigma_S, D_1, D_2)} \lambda R_1 + (1 - \lambda) R_2,$$

where (6.7) is due to the conditional version of [26, Corollary 4]. This together with the convexity of $\mathcal{R}_1(p_S, S_1, S_2)$ and $\mathcal{R}_1(\Sigma_S, D_1, D_2)$ proves (6.3). It can be verified that

$$I(S; \hat{S}_2|U) \leq I(S; \hat{S}_2)$$

$$\leq \frac{1}{2} \log \left( \frac{\|\Sigma_S\|}{|D_2|} \right)$$
and

\[
I(U; \hat{S}_1) + I(S; \hat{S}_2|U) \leq I(U; \hat{S}_1) + I(S; \hat{S}_1|U) \\
= I(S; \hat{S}_1) \\
= \frac{1}{2} \log \left( \frac{|\Sigma_S|}{|D_1|} \right),
\]

from which (6.3) follows immediately.

**Theorem 5** For any \((\kappa, D_1, D_2) \in \Gamma_G\), there exist \(D_i \in D_i\), \(i = 1, 2\), such that

\[
\mathcal{R}_i(\Sigma_S, D_1, D_2) \subseteq \kappa \mathcal{C}_i(p_{Y_1,Y_2|X}), \quad i = 1, 2. \tag{6.8}
\]

**Proof:** By symmetry, it suffices to prove (6.8) for \(i = 1\). Let \(\{Z(t)\}_{t=1}^{\infty}\) be an i.i.d. vector Gaussian process, independent of \(\{S(t)\}_{t=1}^{\infty}\), where each \(Z(t)\) is an \(\ell \times 1\) zero-mean Gaussian random vector with positive definite covariance matrix \(\Sigma_Z\). Define \(\hat{S}_1(t) = S(t)\) and \(\hat{S}_2(t) = S(t) + Z(t)\) for \(t = 1, 2, \ldots\). Now consider an arbitrary tuple \((\kappa, D_1, D_2) \in \Gamma_G\). Given any \(\epsilon > 0\), according to Definition 6, there exist encoding function \(f^{(n,m)} : \mathbb{R}^{\ell \times m} \rightarrow X^n\) and decoding functions \(g_i^{(n,m)} : Y_i^n \rightarrow \mathbb{R}^{\ell \times m}, \quad i = 1, 2\), satisfying\(^2\)

\[
\frac{n}{m} \leq \kappa + \epsilon, \\
\min_{D_i \in D_i} \left\| \frac{1}{m} \sum_{t=1}^{m} \mathbb{E}[(S(t) - \hat{S}_i^{(\epsilon)}(t))(S(t) - \hat{S}_i^{(\epsilon)}(t))^T] - D_i \right\| \leq \epsilon, \quad i = 1, 2.
\]

\(^2\)We have denoted \(\hat{S}_i(t)\) by \(\hat{S}_i^{(\epsilon)}(t)\) to stress its dependence on \(\epsilon\).
Therefore, one can find a sequence $\epsilon_1, \epsilon_2, \cdots$ converging to zero such that

$$\lim_{k \to \infty} \frac{1}{m} \sum_{t=1}^{m} \mathbb{E}[(S(t) - \hat{S}_i^{(\epsilon_k)}(t))(S(t) - \hat{S}_i^{(\epsilon_k)}(t))^T] = D_i$$

for some $D_i \in D_i$, $i = 1, 2$. Note that

$$\lim_{k \to \infty} \frac{1}{m} \sum_{t=1}^{m} \left( \mathbb{E}[(\tilde{S}_1(t) - \hat{S}_1^{(\epsilon_k)}(t))(\tilde{S}_1(t) - \hat{S}_1^{(\epsilon_k)}(t))^T] \quad \mathbb{E}[(\tilde{S}_1(t) - \hat{S}_1^{(\epsilon_k)}(t))(\tilde{S}_2(t) - \hat{S}_1^{(\epsilon_k)}(t))^T] \right)$$

$$= \tilde{D}_1 \triangleq \begin{pmatrix} D_1 & D_1 \\ D_1 & D_1 + \Sigma_Z \end{pmatrix},$$

$$\lim_{k \to \infty} \frac{1}{m} \sum_{t=1}^{m} \mathbb{E}[(\tilde{S}_2(t) - \hat{S}_2^{(\epsilon_k)}(t))(\tilde{S}_2(t) - \hat{S}_2^{(\epsilon_k)}(t))^T]$$

$$= \tilde{D}_2 \triangleq D_2 + \Sigma_Z.$$

As a consequence, we must have $(\kappa, \tilde{D}_1, \tilde{D}_2) \in \tilde{\Gamma}_G$. It then follows from Theorem 1 that

$$\left( \frac{1}{2} \log \left( \frac{\Sigma_S - \Sigma_S(\Sigma_S + \Sigma_Z)^{-1}\Sigma_S}{\left| D_1 - D_1(D_1 + \Sigma_Z)^{-1}D_1 \right|} \right), \frac{1}{2} \log \left( \frac{|\Sigma_S + \Sigma_Z|}{|D_2 + \Sigma_Z|} \right) \right) \in \kappa C_1(p_{Y_1,Y_2|X}).$$

Here one can fix $(D_1, D_2)$ and choose the positive definite covariance matrix $\Sigma_Z$ arbitrarily; moreover, it can be verified that

$$\frac{|\Sigma_S - \Sigma_S(\Sigma_S + \Sigma_Z)^{-1}\Sigma_S|}{|D_1 - D_1(D_1 + \Sigma_Z)^{-1}D_1|} = \frac{|D_1^{-1} + \Sigma_Z^{-1}|}{|\Sigma_S^{-1} + \Sigma_Z^{-1}|} = \frac{|\Sigma_S||D_1 + \Sigma_Z|}{|D_1||\Sigma_S + \Sigma_Z|}.$$
This completes the proof of Theorem 5.

Note that \( R_1(\Sigma, D_1, D_2) \) coincides with the capacity region of vector Gaussian broadcast channel with covariance power constraint \( \Sigma \) and noise covariances \( \Delta_i \triangleq (D_i^{-1} - \Sigma^{-1})^{-1}, i = 1, 2 \), when \( 0 < D_1 \preceq D_2 \preceq \Sigma \). For this reason, we shall denote \( R_1(\Sigma, D_1, D_2) \) alternatively by \( C(\text{G-BC}(\Sigma, \Delta_1, \Delta_2)) \) (even when \( \Delta_1 \) and \( \Delta_2 \) are not well-defined). One can obtain the following refined necessary condition for the case where \( p_{Y_1,Y_2|X} \) is a scalar Gaussian broadcast channel.

**Theorem 6** If \( p_{Y_1,Y_2|X} \) is a G-BC(\( P, N_1, N_2 \)) with \( 0 < N_1 \leq N_2 \), then, for any \((\kappa, D_1, D_2) \in \Gamma_G\), there exist \( D_i \in \mathcal{D}_i, i = 1, 2 \), with \( D_1 \preceq D_2 \) such that

\[
C(\text{G-BC}(\Sigma, \Delta_2)) \subseteq \kappa C(\text{G-BC}(P, N_1, N_2)).
\]

**Proof:** According to the remark after Definition 6, there is no loss of generality in setting \( \hat{S}_i^m = \mathbb{E}[S_i^m|Y_i^n], i = 1, 2 \). As a consequence, in (6.9) we must have \( D_1 \preceq D_2 \) if \( p_{Y_2|X} \) is degraded with respect to \( p_{Y_1|X} \). Now one can readily adapt the proof of Theorem 5 to the current setting to show that, for any \((\kappa, D_1, D_2) \in \Gamma_G\), there exist \( D_i \in \mathcal{D}_i, i = 1, 2 \), with \( D_1 \preceq D_2 \), such that

\[
\mathcal{R}_i(\Sigma, D_1, D_2) \subseteq \kappa \mathcal{C}_i(\text{G-BC}(P, N_1, N_2)), \quad i = 1, 2.
\]
It follows from Proposition 8 that \( C_1(G-BC(P, N_1, N_2)) = C(G-BC(P, N_1, N_2)) \), and \( C_2(G-BC(P, N_1, N_2)) \) is given by the set of \((R_1, R_2) \in \mathbb{R}_+^2\) satisfying
\[
R_2 \leq \frac{1}{2} \log \left( \frac{P + N_2}{N_2} \right),
\]
\[
R_1 + R_2 \leq \frac{1}{2} \log \left( \frac{P + N_1}{N_1} \right).
\]

Note that \( R_1(\Sigma_S, D_1, D_2) \subseteq \kappa C_1(G-BC(P, N_1, N_2)) \) implies
\[
\frac{1}{2} \log \left( \frac{\lvert \Sigma_S \rvert}{\lvert D_i \rvert} \right) \leq \frac{\kappa}{2} \log \left( \frac{P + N_i}{N_i} \right), \quad i = 1, 2.
\]

Moreover, in view of (6.2) and (6.4) in Proposition 13, we have
\[
\mathcal{R}_2(\Sigma_S, D_1, D_2) \subseteq \left\{(R_1, R_2) \in \mathbb{R}_+^2 : R_2 \leq \frac{1}{2} \log \left( \frac{\lvert \Sigma_S \rvert}{\lvert D_2 \rvert} \right), R_1 + R_2 \leq \frac{1}{2} \log \left( \frac{\lvert \Sigma_S \rvert}{\lvert D_1 \rvert} \right) \right\}.
\]

Therefore,
\[
\mathcal{R}_1(\Sigma_S, D_1, D_2) \subseteq \kappa C_1(G-BC(P, N_1, N_2))
\]
\[
\Rightarrow \mathcal{R}_2(\Sigma_S, D_1, D_2) \subseteq \kappa C_2(G-BC(P, N_1, N_2))
\]
when \( 0 \preceq D_1 \preceq D_2 \preceq \Sigma_S \). This completes the proof of Theorem 6.

For the case \( 0 \preceq D_1 \preceq D_2 \preceq \Sigma_S \), one can show by leveraging Proposition 13 that (6.10) is equivalent to the existence of \((\hat{S}_1, \hat{S}_2)\) with \( \mathbb{E}[(S - \hat{S}_i)(S - \hat{S}_i)^T] = D_i \in \mathcal{D}_i, \) \( i = 1, 2 \), such that
\[
\mathcal{R}_i(p_{S, \hat{S}_1, \hat{S}_2}) \subseteq \kappa C_i(G-BC(P, N_1, N_2)), \quad i = 1, 2;
\]
in fact, there is no loss of generality in assuming that $S - \hat{S}_i$ and $\hat{S}_i$ are independent zero-mean Gaussian random vectors with covariance matrices $D_i$ and $\Sigma_S - D_i$, respectively, $i = 1, 2$. Note that $U$ is not restricted to the form $U = S + Z$ (or equivalently $U = \mathbb{E}[S|S + Z]$) in the definition of $\mathcal{R}_i(p_{S,S_1,S_2})$, $i = 1, 2$, where $Z$ is a zero-mean Gaussian random vector independent of $S$. Therefore, removing this restriction does not lead to a stronger necessary condition. This provides a certain justification for the choice of the auxiliary random variable in [3].

With no essential loss of generality, henceforth we focus on the non-degenerate case $\kappa > 0$. Define

$$P^* = \min\{P \geq 0 : C(G-BC(\Sigma_S, \Delta_1, \Delta_2)) \subseteq \kappa C(G-BC(P, N_1, N_2))\}.$$ 

It is clear that, for any $\Sigma_Z \succ 0$, there exists $\beta \in [0, 1]$ such that

$$\frac{1}{2} \log \left( \frac{|\Sigma_S| |D_1 + \Sigma Z|}{|D_1||\Sigma_S + \Sigma Z|} \right) \leq \frac{\kappa}{2} \log \left( \frac{\beta P^* + N_1}{N_1} \right),$$

$$\frac{1}{2} \log \left( \frac{|\Sigma_S + \Sigma Z|}{|D_2 + \Sigma Z|} \right) \leq \frac{\kappa}{2} \log \left( \frac{P^* + N_2}{\beta P^* + N_2} \right),$$

which can be rewritten as

$$\beta P^* \geq N_1 \left( \frac{|\Sigma_S| |D_1 + \Sigma Z|}{|D_1||\Sigma_S + \Sigma Z|} \right)^{\frac{1}{\kappa}} - N_1,$$

$$\beta P^* \leq (P^* + N_2) \left( \frac{|D_2 + \Sigma Z|}{|\Sigma_S + \Sigma Z|} \right)^{\frac{1}{\kappa}} - N_2.$$
Hence, for any $\Sigma_Z \succ 0$, we have

\[
(P^\star + N_2)\left(\frac{|D_2 + \Sigma_Z|}{|\Sigma_S + \Sigma_Z|}\right)^{\frac{1}{\kappa}} - N_2 \geq N_1\left(\frac{|\Sigma_S||D_1 + \Sigma_Z|}{|D_1||\Sigma_S + \Sigma_Z|}\right)^{\frac{1}{\kappa}} - N_1,
\]

i.e.,

\[
P^\star \geq N_1\left(\frac{|\Sigma_S||D_1 + \Sigma_Z|}{|D_1||D_2 + \Sigma_Z|}\right)^{\frac{1}{\kappa}} + (N_2 - N_1)\left(\frac{|\Sigma_S + \Sigma_Z|}{|D_2 + \Sigma_Z|}\right)^{\frac{1}{\kappa}} - N_2. \tag{6.11}
\]

Moreover, there must exist some $\beta \in [0, 1]$ and a sequence of positive definite matrices $\Sigma_Z^{(k)}$, $k = 1, 2, \cdots$, such that

\[
\lim_{k \to \infty} \frac{1}{2} \log\left(\frac{|\Sigma_S||D_1 + \Sigma_Z^{(k)}|}{|D_1||\Sigma_S + \Sigma_Z^{(k)}|}\right) = \frac{\kappa}{2} \log\left(\frac{\beta P^\star + N_1}{N_1}\right),
\]

\[
\lim_{k \to \infty} \frac{1}{2} \log\left(\frac{|\Sigma_S + \Sigma_Z^{(k)}|}{|D_2 + \Sigma_Z^{(k)}|}\right) = \frac{\kappa}{2} \log\left(\frac{P^\star + N_2}{\beta P^\star + N_2}\right),
\]

which implies

\[
P^\star = \lim_{k \to \infty} N_1\left(\frac{|\Sigma_S||D_1 + \Sigma_Z^{(k)}|}{|D_1||D_2 + \Sigma_Z^{(k)}|}\right)^{\frac{1}{\kappa}} + (N_2 - N_1)\left(\frac{|\Sigma_S + \Sigma_Z^{(k)}|}{|D_2 + \Sigma_Z^{(k)}|}\right)^{\frac{1}{\kappa}} - N_2. \tag{6.12}
\]

Combining (6.11) and (6.12) gives

\[
P^\star = \sup_{\Sigma_Z \succ 0} N_1\left(\frac{|\Sigma_S||D_1 + \Sigma_Z|}{|D_1||D_2 + \Sigma_Z|}\right)^{\frac{1}{\kappa}} + (N_2 - N_1)\left(\frac{|\Sigma_S + \Sigma_Z|}{|D_2 + \Sigma_Z|}\right)^{\frac{1}{\kappa}} - N_2. \tag{6.13}
\]

Therefore, by Theorem 6, if $(\kappa, D_1, D_2) \in \Gamma_G$, then

\[
P \geq \inf_{D_1, D_2 \Sigma_Z \succ 0} \sup N_1\left(\frac{|\Sigma_S||D_1 + \Sigma_Z|}{|D_1||D_2 + \Sigma_Z|}\right)^{\frac{1}{\kappa}} + (N_2 - N_1)\left(\frac{|\Sigma_S + \Sigma_Z|}{|D_2 + \Sigma_Z|}\right)^{\frac{1}{\kappa}} - N_2, \tag{6.14}
\]

65
where the infimum is over $D_1$ and $D_2$ subject to the constraints $D_i \in D_i$, $i = 1, 2$, and $D_1 \preceq D_2$. For the case where $D_i = \{D_i : 0 \preceq D_i \preceq \Theta_i\}$, $i = 1, 2$, for some $\Theta_1$ and $\Theta_2$ satisfying $0 \prec \Theta_1 \preceq \Theta_2 \preceq \Sigma_S$, we can simplify (6.14) to

$$ P \geq \sup_{\Sigma_Z \succ 0} N_1 \left( \frac{||\Sigma_S||\Theta_1 + \Sigma_Z||}{||\Theta_1||\Theta_2 + \Sigma_Z||} \right)^{\frac{1}{2}} + (N_2 - N_1) \left( \frac{||\Sigma_S + \Sigma_Z||}{||\Theta_2 + \Sigma_Z||} \right)^{\frac{1}{2}} - N_2, $$

from which one can readily recover [3, Theorem 1] by setting $\ell = 1$.

Now partition $S(t)$ to the form $S(t) = (S_i^T(t), S_j^T(t))^T$, $t = 1, 2, \cdots$, where each $S_i(t)$ is an $\ell_i \times 1$ zero-mean Gaussian random vector with positive definite covariance matrix $\Sigma_{S_i}$, $i = 1, 2$. We require that $\{S_i(t)\}_{t=1}^\infty$ be reconstructed at receiver $i$ subject to positive definite covariance distortion constraint $\Lambda_i$, $i = 1, 2$. This corresponds to the case where $D_i = D_i(\Lambda_i) \triangleq \{D_i : 0 \preceq D_i \preceq \Sigma_S, D_{i,i} \preceq \Lambda_i\}$ with $D_i$ partitioned to the form

$$ D_i = \begin{pmatrix} D_{i,1} & \# \\ \# & D_{i,2} \end{pmatrix}, \quad i = 1, 2. $$

Therefore, the lower bound in (6.14) is also applicable here. By restricting $\Sigma_Z$ to a special block diagonal form

$$ \Sigma_Z = \begin{pmatrix} \Lambda I & 0 \\ 0 & \Sigma_{Z_2} \end{pmatrix}, $$

\footnote{Here $I$ is an $\ell_1 \times \ell_1$ identity matrix}
one can deduce from (6.14)

\[ P \geq \inf_{D_1,D_2} \sup_{\Sigma Z_2 > 0} \lim_{\lambda \to \infty} N_1 \left( \frac{\|\Sigma S\| D_1 + \Sigma Z}{\|D_1\| D_2 + \Sigma Z} \right)^{\frac{1}{\kappa}} + (N_2 - N_1) \left( \frac{\|\Sigma S + \Sigma Z\|}{\|D_2 + \Sigma Z\|} \right)^{\frac{1}{\kappa}} - N_2 \]

where the infimum is over \( D_1 \) and \( D_2 \) subject to the constraints \( D_i \in \mathcal{D}_i(\Lambda_i), i = 1, 2, \) and \( D_1 \preceq D_2 \). This potentially weakened lower bound, when specialized to the case \( \kappa = 1 \), is at least as tight as [18, Theorem 1]. Note that, for any \( D_i \in \mathcal{D}_i(\Lambda_i), i = 1, 2, \) and any positive definite matrix \( \Sigma Z \) partitioned to the form

\[ \Sigma Z = \begin{pmatrix} \Sigma Z_1 & \# \\ \# & \Sigma Z_2 \end{pmatrix} \]

we have

\[ \frac{\|\Sigma S\| D_{1,2} + \Sigma Z_2}{\|D_1\| D_{2,2} + \Sigma Z_2} \geq \frac{\|\Sigma S + \Sigma Z\| D_{1,2} + \Sigma Z_2}{\|D_1 + \Sigma Z\| D_{2,2} + \Sigma Z_2} \geq \frac{\|\Sigma S + \Sigma Z\|}{\|D_{1,1} + \Sigma Z_1\| D_{2,2} + \Sigma Z_2} \]  

and

\[ \frac{\|\Sigma S_2 + \Sigma Z_2\|}{\|D_{2,2} + \Sigma Z_2\|} \geq \frac{\|\Sigma S_2 + \Sigma Z_2\|}{\|\Lambda_2 + \Sigma Z_2\|}. \]
Substituting (6.17) and (6.18) into (6.15) gives

\[ P \geq \sup_{\Sigma_Z > 0} N_1 \left( \frac{|\Sigma_S + \Sigma_Z|}{|\Lambda_1 + \Sigma_Z_1||\Lambda_2 + \Sigma_Z_2|} \right)^{\frac{1}{\kappa}} + (N_2 - N_1) \left( \frac{|\Sigma_S_2 + \Sigma_Z_2|}{|\Lambda_2 + \Sigma_Z_2|} \right)^{\frac{1}{\kappa}} - N_2, \quad (6.19) \]

where \( \Sigma_Z \) is partitioned to the form in (6.16). Setting \( \kappa = 1 \) in (6.19) recovers [18, Corollary 1]. An equivalent form of the lower bound in (6.19) was first obtained by Bross et al. [15] via a different approach for the special case \( \kappa = \ell_1 = \ell_2 = 1 \). It is worth mentioning that source-channel separation is known to be suboptimal in general for this problem [16, 17]. Somewhat surprisingly, the lower bound in (6.19), derived with the aid of a source-channel separation theorem (i.e., Theorem 4), turns out to be tight when \( \kappa = \ell_2 = 1 \) [18, Theorem 2] and is achievable by a class of hybrid digital-analog coding schemes\(^4\) [18, Section IV.B]. Therefore, the application of source-channel separation theorems is not restricted to the relatively limited scenarios where the separation architecture is optimal; they can also be used to prove the optimality of non-separation based schemes and determine the performance limits in certain scenarios where the separation architecture is suboptimal.

\(^4\)The hybrid scheme in [16] can be viewed as an extremal case of this class of schemes.
Chapter 7

Conclusion

We have established a source-channel separation theorem, which is further leveraged to derive a general necessary condition for the source broadcast problem. It is intriguing to note that, in certain cases (see, e.g., Theorem 3 and Theorem 6), this necessary condition takes the form of comparison of two capacity regions. This is by no means a coincidence. In fact, it suggests a new direction that can be explored to establish stronger converse results for the source broadcast problem [27].
Bibliography


