DUALITY BETWEEN SOURCE CODING AND CHANNEL CODING WITH SIDE INFORMATION

by

S. Sandeep Pradhan, Jim Chou and Kannan Ramchandran

Memorandum No. UCB/ERL M01/34

5 December 2001

DUALITY BETWEEN SOURCE CODING AND CHANNEL CODING WITH SIDE INFORMATION

by

S. Sandeep Pradhan, Jim Chou and Kannan Ramchandran

Memorandum No. UCB/ERL M01/34

5 December 2001

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Duality between source coding and channel coding with side information †

S. Sandeep Pradhan, Jim Chou and Kannan Ramchandran
Electrical Engineering and Computer Science Department
University of California, Berkeley, CA-94720
email: {pradhan5,jimchou,kannanr}@eecs.berkeley.edu.

December 5, 2001

Abstract

We explore the information-theoretic duality between source coding with side information at the decoder and channel coding with side information at the encoder. We begin with a mathematical characterization of the functional duality between classical source and channel coding, formulating the precise conditions under which the optimal encoder for one problem is functionally identical to the optimal decoder for the other problem. We then extend this functional duality to the case of coding with side information. By applying this duality, we are able to generalize the result of Wyner-Ziv [1] relating to no rate-loss for source coding with side information from Gaussian to more arbitrary distributions. We consider several examples corresponding to both discrete-valued and continuous-valued cases to illustrate our formulation. Our treatment inspires the construction and dual use of practical coset codes for a large class of emerging applications for coding with side information, such as distributed sensor networks, watermarking and information-hiding communication systems.

Keywords

Source coding with side information, Channel coding with side information, Duality, Rate-distortion function, Capacity-cost function

1 Introduction

Classical source coding under a distortion constraint and channel coding under a channel cost constraint have long been considered as information-theoretic duals[1] of each other starting from Shannon's landmark paper in 1959 [3]. In the source coding (rate-distortion) problem, the goal is to find, for a given source distribution, the conditional distribution between the input source alphabet and the output reconstruction alphabet that minimizes the mutual information [4] between the input source and the output reconstruction subject to a target distortion constraint corresponding to a given distortion measure. In the channel coding (capacity-cost) problem, the goal is to find, for a given channel input-output conditional distribution, the channel input distribution that maximizes the mutual information between input and output subject to a target channel cost (or power) constraint corresponding to a channel cost measure.

†This work was presented in part at the 33rd Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, 1999, titled "On the duality between distributed source coding and data hiding".

[1]This should not be confused for Lagrangian duality in optimization theory [2].

Cover and Thomas [4] point out that these two problems are information-theoretic duals of each other. This duality has also been exploited in the formulation of numerical optimization algorithms for the channel capacity and rate-distortion problems using the Blahut-Arimoto algorithm [5]. In [4] (see Sec 13.5 in [4]), a packing versus covering interpretation of duality in Gaussian source and channel coding has been formulated to illustrate this duality. Indeed, some of the solutions to practical source coding (respectively channel coding) have been inspired by solutions to channel coding (source coding): e.g., trellis-coded quantization [6, 7] in source coding has been inspired by trellis-coded modulation [8, 7] in channel coding, and shaping of constellation points [9] in channel coding has been inspired by constraining of entropy of quantization points [10] in source coding. In [7] a duality between quantization and modulation has been analyzed.

In this work, we address the problem of duality between source [1] and channel [11] coding in the presence of side-information at the decoder and encoder, respectively. As a first step towards this goal, we first address the more classical duality between source and channel coding with the objective of providing a mathematical characterization of their *functional* duality. To be specific, given an optimal source (respectively channel) coding scheme (given by the encoder and decoder), we detail the conditions under which such a scheme is a functional dual to a channel (source) coding scheme (given by the encoder and decoder), in the sense that the optimal encoder mapping for one problem is functionally identical to the optimal decoder mapping for the other problem. We show in the sequel that this functional duality is enabled through appropriate choices of measures and constraints for the dual source (or channel) coding problem. Our inspiration for this formulation is derived from recent work by Gastpar et al. on the optimality conditions for uncoded communications [12, 13].

We then extend this duality to the case of coding with side information, where we study the duality between source coding with side information at the decoder, and channel coding with side information at the encoder. These problems have attracted considerable attention in recent times due to a large class of relevant applications related to both problems. A partial list of applications for source coding with side information includes distributed sensor networks [14], sensor arrays, digital upgrade of analog television signals [15], and communication in ad-hoc networks. The application set for the problem of channel coding with side information includes data hiding, watermarking [16], broadcast[17], ISI precoding, multi-antenna communication systems and steganography. In this paper, we explore the information-theoretic duality between these two problems, and provide a mathematical characterization of the conditions under which they become functional duals, i.e. the encoder for one problem is exchangeable for the decoder of the other.

As an example, we will show that the problem of encoding a memoryless Gaussian source with correlated

Gaussian side-information available at the decoder studied by Wyner and Ziv in [1] is a functional dual to the Gaussian watermarking problem studied in [16], related to the problem of "writing on dirty paper" considered by Costa in [18]. Recently, Costa's result [18] has been generalized from the case of Gaussian side information to more arbitrary distributions by Cohen and Lapidoth in [19]. By invoking our duality concepts, we are able to find a similar generalization of the result of Wyner-Ziv [1] relating to no rate-loss for source coding with side information from Gaussian to more arbitrary distributions. Rate-loss refers to the performance loss due to the presence of side information at only one end (decoder only in source coding and encoder only in channel coding) rather than at both ends. A goal of this paper is to inspire the practical dual use of functional blocks for the two problems, building on our preliminary work in [20]. It is illuminating to note that both problems have recently inspired promising constructive frameworks based on coset codes that are underpinned by the relevant information-theoretic concepts, e.g. [21, 22, 20, 23, 24] for the channel coding problem, and [25, 26] for the source coding problem.

To summarize, the main contributions of this paper are 1) a mathematical characterization of the functional duality between conventional source and channel coding, and 2) extension of this to the case of source and channel coding with side information at the decoder and encoder respectively. The paper is organized as follows. In the next section we review the problems of source coding with side information and channel coding with side information. Sections 3 formulates the precise notion of duality between general source and channel coding problems. Section 4 concludes the paper.

2 Coding with side information

In this section we give a brief review of source coding and channel coding with side information.

2.1 Source coding with side information at decoder (SCSI)

Consider the problem [1] of rate-distortion optimal lossy encoding of a source X with the side information S available (losslessly) at the decoder as shown in Fig. 1. X and S are correlated random variables with

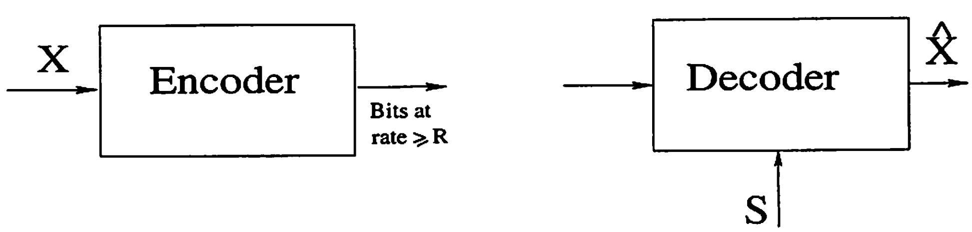


Figure 1: Source encoding with side information at the decoder. The encoder transmits at a rate greater than or equal to R bits/ source sample.

joint distribution $\bar{p}(x,s)$, such that the sequence pair $\{X_k, S_k\}_{k=1}^{\infty}$ denotes independent realizations of the given random variables. Let the alphabets of X and S be $\mathcal{X}, \mathcal{S}$ respectively. The encoding and decoding are done in blocks of length L. The encoder is a mapping: $\mathcal{X}^L \longrightarrow \{1, 2, \ldots, 2^{LR}\}$, and the decoder is a mapping $\{1, 2, \ldots, 2^{LR}\} \times \mathcal{S}^L \longrightarrow \hat{\mathcal{X}}^L$ where $\hat{\mathcal{X}}$ is the reconstruction alphabet and the distortion criterion is given by $E\left[\frac{1}{L}\sum_{k=1}^{L} d(X_k, \hat{X}_k, S_k)\right]$, where $d : \mathcal{X} \times \hat{\mathcal{X}} \times \mathcal{S} \longrightarrow \mathbb{R}^+$ is the per-letter additive distortion measure[2] and $E(.)$ is the expectation operator and R is the rate of transmission.

Fact 1 [1]: The rate-distortion function for this set-up, $R^*(D)$, is given by

$$R^*(D) = \min_{p(u|x)} \min_{f : \mathcal{U} \times \mathcal{S} \rightarrow \hat{\mathcal{X}}} \left[I(U;X) - I(U;S)\right], \tag{1}$$

such that $S \rightarrow X \rightarrow U$ form a Markov chain and $E\left[d(X, f(U,S), S)\right] \leq D$, where U is an auxiliary random variable with alphabet $\mathcal{U}$. This is a natural Markov chain associated with the definition of the problem, which characterizes the fact that only the decoder has access to the side information. In other words, the dependency between the side information S and the auxiliary variable U is completely captured through the source X. Using this Markov chain it can be seen that $I(U;X) - I(U;S) = I(U;X|S)$. From this identity, the above optimization problem can be rewritten as follows:

$$R^*(D) = \min_{p(u|x), p(\hat{x}|u,s)} \left[I(X;U|S)\right], \tag{2}$$

such that $S \rightarrow X \rightarrow U$ and $E\left[d(X, \hat{X}, S)\right] \leq D$. Further, since the decoder does not have access to the source X, in (2) we have the constraint that $X \rightarrow (U,S) \rightarrow \hat{X}$ which implies that U and S completely determine the reconstruction $\hat{X}$.

Now the rate-distortion function $R_{x|s}(D)$ when both the encoder and the decoder have access to the side information is given by:

$$R_{x|s}(D) = \min_{p(\hat{x}|x,s)} I(X;\hat{X}|S), \tag{3}$$

such that $E[d(X, \hat{X}, S)] \leq D$. As a way of measuring the performance loss due to the presence of side information only at the decoder (see Fig. 1), we note that the SCSI problem (2) differs from (3) in two regards: the inclusion of auxiliary random variable U, and the presence of the Markov chain. In order to address this, it is insightful to recast the rate-distortion problem in (3) as in Fig. 2 by introducing an intermediate processing stage between X and $\hat{X}$ that includes U, and imposing the Markov constraint $S \rightarrow X \rightarrow U$ as follows:

$$R_{x|s}(D) = \min_{p(u|x), p(\hat{x}|u,s)} I(X;\hat{X}|S), \tag{4}$$

[2]Although $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$ is typically considered in the literature, the extension to a more general d which includes S is straightforward. It can be interpreted as the distortion between x and $\hat{x}$ when the outcome of the side information is s. This is necessary for a precise formulation of duality between source and channel coding with side information. A similar generalization of cost measure is considered in Section 2.2 for the case of channel coding with side information.

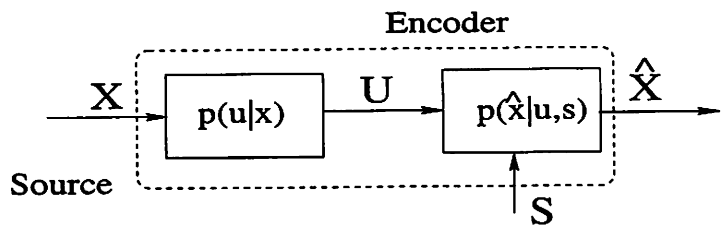


Figure 2: Source coding with side information present at both encoder and decoder: an intermediate stage is introduced with U.

such that $S \rightarrow X \rightarrow U$ and $E[d(X,\hat{X},S)] \leq D$. As in (2), we have the additional Markov constraint $X \rightarrow (U,S) \rightarrow \hat{X}$ that emphasizes the fact that the reconstruction $\hat{X}$ depends only on U and S. Note that this formulation subsumes that of (3) when $U = X$, i.e. the first processing stage is an identity operation. It can be shown by invoking the data processing inequality that this expanded formulation does not perform any better than the original one of (3), i.e. the two formulations are identical. The only reason for this expanded version is to exactly match the constraints of the SCSI problem of (2) for a fair comparison.

In general, under the given set of constraints $I(X;\hat{X}|S) \leq I(X;U|S)$. Thus $R^*(D) = R_{x|s}(D)$ if and only if at optimality in (4), $I(X;\hat{X}|S) = I(X;U|S)$, which means that conditioned on the side information S, the information conveyed about the source X by observing U is the same as that by observing $\hat{X}$. This implies the Markov constraint: $X \rightarrow (\hat{X},S) \rightarrow U$. Thus, given the side information S, the auxiliary variable U and the reconstruction $\hat{X}$ are information equivalent. So the encoder can as well generate U (which depends only on the source X) instead of $\hat{X}$ (which depends on both U and side information Y), thus obviating the need for the side information Y at the encoder.

More formally, there is no rate loss, i.e., $R^*(D) = R_{x|s}(D)$ if and only if [1] the distribution achieving the minimization in (3) (say $p^*(x,s,\hat{x})$) can be represented in the form having the auxiliary random variable U: $p^*(x,s,\hat{x},u)$ such that

(a) $\sum_u p^*(x,s,\hat{x},u) = p^*(x,s,\hat{x})$,

(b) the reconstruction $\hat{X}$ can be represented as $\hat{x} = f^*(u,s)$, which is obtained by the SCSI problem (1), and

(c) the following two Markov chains are satisfied: $S \rightarrow X \rightarrow U$ and $X \rightarrow (\hat{X},S) \rightarrow U$.

It is important to note that these constraints are associated with the optimizing distribution for the problem of source coding with side information at both encoder and decoder as given by (3,4). It is not in general possible to guarantee no rate loss if we start the other way around, i.e. with the SCSI problem of (1,2) and impose these constraints.

2.2 Channel coding with side information at encoder (CCSI)

Channel coding with side information [11, 21, 18] about the channel at the encoder is shown schematically in Fig. 3. The encoder wishes to communicate over a channel given by the conditional distribution $\bar{p}(y|x, s)$, where

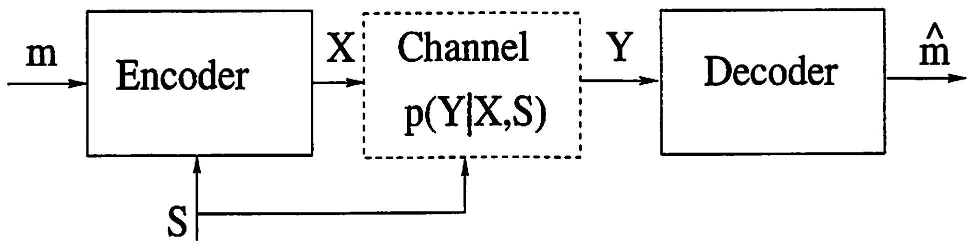


Figure 3: Channel coding with side information: the encoder has access to side information S about the channel; m represents the message, and X and Y are the input and the output of the channel respectively.

Y is the channel output, X is the channel input which the encoder can control and S, the side information with a given distribution $\bar{p}(s)$, is the second channel input which the encoder cannot control but can only observe, and has alphabet $\mathcal{S}$ and there is a cost measure $\omega : \mathcal{X} \times \mathcal{S} \rightarrow \mathbb{R}^+$. This can be interpreted as the cost of transmitting x when the side information outcome is s. The encoder is a mapping:$\{1, 2, \ldots, 2^{LC}\} \times \mathcal{S}^L \longrightarrow \mathcal{X}^L$, and the decoder is a mapping: $\mathcal{Y}^L \longrightarrow \{1, 2, \ldots, 2^{LC}\}$, where C is the rate of transmission.

Fact 2 [11]: The capacity $C^*(W)$, for a cost constraint W, of this channel coding set-up is given by

$$C^*(W) = \max_{p(u|s)} \max_{f : \mathcal{U} \times \mathcal{S} \rightarrow \mathcal{X}} [I(U;Y) - I(U;S)], \tag{5}$$

such that $Y \rightarrow (X, S) \rightarrow U$ form a Markov chain[3] and $E[\omega(X, S)] \leq W$, where U is an auxiliary random variable with alphabet $\mathcal{U}$. This is a natural Markov chain associated with the definition of the problem which means that the dependency between the auxiliary variable U and the channel output Y is completely captured through the channel input X and the side information S. The encoder can influence the channel output Y only through the input X. The above optimization can be rewritten as

$$C^*(W) = \max_{p(u|s), p(x|u,s)} [I(U;Y) - I(U;S)], \tag{6}$$

such that $Y \rightarrow (X, S) \rightarrow U$ and $E[\omega(X, S)] \leq W$. Further, since the encoder does not have access to the channel output Y, in (6) we have the constraint that $Y \rightarrow (U, S) \rightarrow X$, which implies that U and S completely determine the channel input X.

Now the capacity of the system $C_{y|s}(W)$, when both the encoder and the decoder have access to the side

[3]See equation (2.4) in [11].

information is given by:

$$C_{y|s}(W) = \max_{p(x|s)} I(X;Y|S), \tag{7}$$

such that $E[\omega(X,S)] \leq W$. As was done in SCSI, we note that the CCSI problem (6) differs from (7) in two

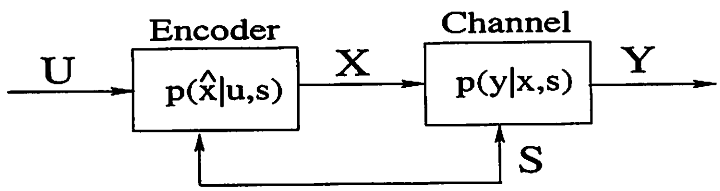


Figure 4: Channel coding with side information present at both encoder and decoder: an intermediate stage is introduced with U.

regards: the inclusion of auxiliary random variable U, and the presence of the Markov chain. To address this, it is insightful to recast the capacity-cost problem in (7) as in Fig. 4 by introducing an intermediate processing stage that includes U, and imposing the Markov chain $Y \to (X,S) \to U$ as follows:

$$C_{y|s}(W) = \max_{p(u|s),p(x|u,s)} I(U;Y|S), \tag{8}$$

such that $Y \to (X,S) \to U$ and $E[\omega(X,S)] \leq W$. This is because, a) in (8) we have the additional constraint that $Y \to (U,S) \to X$, which implies that U and S completely determine the channel input X, and b) using the Markov chain $Y \to (X,S) \to U$, we have $I(X;Y|S) = I(U;Y|S)$. Note that this formulation subsumes that of (7) when $U = X$, i.e., the first processing stage is an identity operation. It can be shown by invoking the data processing inequality that this expanded formulation does not perform any better than the original one of (7), i.e. the two formulations are identical.

In general, under the given set of constraints $I(U;Y|S) \geq I(U;Y) - I(U;S)$. Thus $C^*(W) = C_{y|s}(W)$ if and only if at optimality in (8), $I(U;Y|S) = I(U;Y) - I(U;S)$, which means that the information conveyed about the auxiliary variable U by observing (Y,S) is the same as that by observing Y. This implies that $S \to Y \to U$. Thus given the channel output Y, the side information S does not give any more information about U (which contains the message sent by the encoder) thus obviating the need for the side information at the decoder.

More formally, there is no rate-loss, i.e., $C^*(W) = C_{y|s}(W)$ if and only if the distribution achieving the maximization in (7) (say $p^*(x,s,y)$) can be represented in the form having the auxiliary variable U: $p^*(x,s,y,u)$ such that

(a) $\sum_u p^*(x,s,y,u) = p^*(x,s,y)$,

(b) the channel input X can be represented as $x = f^*(u,s)$, which is obtained by the CCSI problem (5), and

(c) the following two Markov chains are satisfied: $Y \to (X,S) \to U$ and $S \to Y \to U$.

It is important to note that these constraints are associated with the optimizing distribution for the problem of channel coding with side information at both encoder and decoder as given by (7,8). It is not in general possible to guarantee no rate loss if we start the other way around, i.e. with the CCSI problem of (5,6) and impose these constraints.

3 Duality

Traditionally, the conventional source and channel coding [4] problems have been considered as information-theoretic duals of each other. Shannon in his 1959 paper [3] on rate-distortion theory stated (italicized for emphasis) "*...duality between the properties of a source with a distortion measure and those of a channel. This duality is enhanced if we consider channels in which there is a cost associated with the different input letters..*". In this section we obtain a precise characterization of functional duality between conventional source and channel coding and then extend this notion for source and channel coding with side information. In other words, for a given source coding problem, we obtain a channel coding problem, where the roles of encoder and decoder are functionally exchangeable and the input/output joint distribution is the same with some renaming of variables. For a given source, the rate-distortion function [4] denoted by $R(D)$ is the minimum rate of information required to represent it with a distortion constraint D. Similarly, for a given channel, the capacity-cost function [4] denoted by $C(W)$ is the maximum rate of information that can be reliably transmitted with a cost constraint W.

3.1 Duality in conventional source and channel coding

We would like to emphasize that the notion of duality pervades the literature starting from Shannon in 1959. To study these concepts in one framework, let us state a correspondence between the variables involved in the two coding problems.

$$\text{source coding}\left\{\begin{array}{ccccc}\text{source input} & X & \longleftrightarrow & Y & \text{channel output}\\ \text{source reconstruction} & \hat{X} & \longleftrightarrow & X & \text{channel input}\end{array}\right\}\text{channel coding} \tag{9}$$

To avoid confusion, we stick to the conventional notation of source coding and use the same for channel coding. Let us now recall two facts which have been recently studied in [12, 13] in a totally different context of optimality of uncoded transmission. It was shown in [12, 13] that for a given source $\bar{p}(x)$ and a channel $\bar{p}(\hat{x}|x)$, there exist distortion and cost measures such that uncoded transmission is optimal. The essential concepts are presented in the following form for completeness. We denote the given distribution by $\bar{p}(.)$ and the distribution which optimizes the given objective function by $p^*(.)$.

Fact 3 [12]: For a conventional channel coding problem with a channel $\bar{p}(x|\hat{x})$, input and output alphabets $\hat{\mathcal{X}}$ and $\mathcal{X}$ respectively, given a channel input distribution $p^*(\hat{x})$, $\exists$ a cost measure $\omega : \hat{\mathcal{X}} \to \mathbb{R}^+$ and a cost constraint W such that

$$p^*(\hat{x}) = \underset{p(\hat{x}) : (X|\hat{X}) \sim \bar{p}(x|\hat{x}), E\omega(\hat{X}) \leq W}{\text{argmax}} I(X; \hat{X}), \tag{10}$$

where the cost measure and cost constraint are given respectively by

$$\omega(\hat{x}) = c_1 D\left(\bar{p}(x|\hat{x}) \| p^*(x)\right) + \theta \text{ and } W = E_{p^*(\hat{x})}\left[\omega(\hat{X})\right], \tag{11}$$

$D(.\|.)$ is information divergence [4], $c_1 > 0$, θ are arbitrary constants, $p^*(x) = \sum_{\hat{x}} \bar{p}(x|\hat{x}) p^*(\hat{x})$.

Fact 4 [12]: For a conventional source coding problem, with a source $\bar{p}(x)$, input and reconstruction alphabets $\mathcal{X}$, $\hat{\mathcal{X}}$ respectively, given a conditional distribution $p^*(\hat{x}|x)$, $\exists$ a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$, and a distortion constraint D such that

$$p^*(\hat{x}|x) = \underset{p(\hat{x}|x) : X \sim \bar{p}(x), Ed(X, \hat{X}) \leq D}{\text{argmin}} I(X; \hat{X}), \tag{12}$$

where the distortion measure and distortion constraint are given respectively by

$$d(x, \hat{x}) = -c_2 \log p^*(x|\hat{x}) + d_0(x) \text{ and } D = E_{\bar{p}(x) p^*(\hat{x}|x)}\left[d(X, \hat{X})\right], \tag{13}$$

$c_2 > 0$ and $d_0(x)$ are arbitrary, $p^*(x|\hat{x}) = \frac{\bar{p}(x) p^*(\hat{x}|x)}{\sum_x \bar{p}(x) p^*(\hat{x}|x)}$, $p^*(\hat{x}) = \sum_x \bar{p}(x) p^*(\hat{x}|x)$.

Remark 1: In Fact 3, for a range of cost measures ω, the channel input distribution $p^*(\hat{x})$ maximizes $I(X; \hat{X})$. This cost measure has a nice interpretation in some interesting cases such as Gaussian and binary symmetric channels. It can be shown [12] that for an additive memoryless Gaussian noise channel, and Gaussian channel input distribution, this cost measure is given by $\hat{x}^2$, for an appropriate choice of c_1 and θ. In general, this cost measure penalizes those input distributions which does not result in the output distribution given by $p^*(x)$. Similarly, in Fact 4, the conditional distribution $p^*(\hat{x}|x)$ minimizes $I(X; \hat{X})$ for a range of distortion measures. As in Fact 3, it has nice closed-form solutions for the Gaussian and binary memoryless sources. For example, for the Gaussian memoryless source, and additive Gaussian conditional distribution, this distortion measure is given by $(x - \hat{x})^2$ for an appropriate choice of c_2 and $d_0(x)$. In general, this distortion measure somewhat corresponds to the conditional length of description of the source value x, given a reconstruction value $\hat{x}$. See [12] for details and other examples.

Using this, we have the following theorem which connects the source and channel coding problems.

Theorem 1a: For a source coding problem, with a given source $\bar{p}(x)$, input and reconstruction alphabets $\mathcal{X}$

and $\hat{\mathcal{X}}$ respectively, a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$, and a distortion constraint D, let the conditional distribution achieving rate-distortion optimality $R(D)$ be given by:

$$p^*(\hat{x}|x) \triangleq \underset{p(\hat{x}|x) : X \sim \bar{p}(x), Ed(X,\hat{X}) \leq D}{\operatorname{argmin}} I(X;\hat{X}), \tag{14}$$

inducing the following distributions:

$$p^*(x|\hat{x}) \triangleq \frac{p^*(\hat{x}|x)\bar{p}(x)}{\sum_x \bar{p}(x)p^*(\hat{x}|x)}, \quad p^*(\hat{x}) \triangleq \sum_x \bar{p}(x)p^*(\hat{x}|x). \tag{15}$$

Then $\exists$ a *dual* channel coding problem for the channel $p^*(x|\hat{x})$, having input and output alphabets $\hat{\mathcal{X}}$ and $\mathcal{X}$ respectively, a cost measure $\omega : \hat{\mathcal{X}} \to \mathbb{R}^+$, and a cost constraint W, such that:

- the rate-distortion bound $R(D)$ is equal to the capacity-cost bound $C(W)$, i.e.,

$$\min_{p(\hat{x}|x) : X \sim \bar{p}(x), Ed(X,\hat{X}) \leq D} I(X;\hat{X}) = \max_{p(\hat{x}) : (X|\hat{X}) \sim p^*(x|\hat{x}), E\omega(\hat{X}) \leq W} I(X;\hat{X}); \tag{16}$$

- the channel input distribution given by $p^*(\hat{x})$ achieves capacity-cost optimality, i.e.,

$$p^*(\hat{x}) = \underset{p(\hat{x}) : X|\hat{X} \sim p^*(x|\hat{x}), E\omega(\hat{X}) \leq W}{\operatorname{argmax}} I(X;\hat{X}), \quad . \tag{17}$$

where the cost measure and cost constraint are given respectively by:

$$\omega(\hat{x}) \triangleq c_1 D\left(p^*(x|\hat{x}) \| \bar{p}(x)\right) + \theta, \quad W \triangleq E_{p^*(\hat{x})}(\omega(\hat{X})), \tag{18}$$

for arbitrary $c_1 > 0$ and θ.

Proof: Omitted

Theorem 1b: For a channel coding problem, with a given channel $\bar{p}(x|\hat{x})$, input and output alphabets $\hat{\mathcal{X}}$ and $\mathcal{X}$ respectively, a cost measure $\omega : \hat{\mathcal{X}} \to \mathbb{R}^+$, and a cost constraint W, let the channel input distribution achieving the capacity-cost optimality $C(W)$ be given by:

$$p^*(\hat{x}) \triangleq \underset{p(\hat{x}) : (X|\hat{X}) \sim \bar{p}(x|\hat{x}), E\omega(\hat{X}) \leq W}{\operatorname{argmax}} I(X;\hat{X}), \tag{19}$$

inducing the following distributions:

$$p^*(x) \triangleq \sum_{\hat{x}} p^*(\hat{x})\bar{p}(x|\hat{x}), \quad p^*(\hat{x}|x) \triangleq \frac{\bar{p}(x|\hat{x})p^*(\hat{x})}{p^*(x)}. \tag{20}$$

Then $\exists$ a *dual* source coding problem for the source $p^*(x)$, having input and reconstruction alphabets $\mathcal{X}$ and $\hat{\mathcal{X}}$ respectively, a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$, and a distortion constraint D, such that:

- the capacity-cost bound $C(W)$ is equal to the rate-distortion bound $R(D)$, i.e.,

$$\max_{p(\hat{x}) : (X|\hat{X}) \sim \bar{p}(x|\hat{x}), E\omega(\hat{X}) \leq W} I(X;\hat{X}) = \min_{p(\hat{x}|x) : X \sim p^*(x), Ed(X,\hat{X}) \leq D} I(X;\hat{X}); \tag{21}$$

- the conditional distribution given by $p^*(\hat{x}|x)$ achieves rate-distortion optimality, i.e.,

$$p^*(\hat{x}|x) = \operatorname*{argmin}_{p(\hat{x}|x) : X \sim p^*(x), Ed(X,\hat{X}) \leq D} I(X;\hat{X}), \tag{22}$$

where the distortion measure and distortion constraint are given respectively by:

$$d(x,\hat{x}) \triangleq -c_2 \log \bar{p}(x|\hat{x}) + d_0(x), \quad D \triangleq E_{p^*(\hat{x})\bar{p}(x|\hat{x})}(d(X,\hat{X})), \tag{23}$$

for arbitrary $c_2 > 0$ and $d_0(x)$.

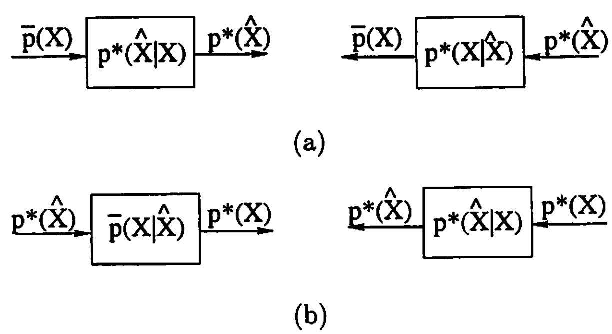


Figure 5: (a) For a given source coding problem with source $\bar{p}(x)$, the dual channel coding problem is defined in a reverse order. (b) For a given channel coding problem with channel $\bar{p}(x|\hat{x})$, the dual source coding problem is defined in a reverse order.

Remark 2: The first property of the dual problem as given by (16) implies that the minimum value of $I(X;\hat{X})$ with respect to $p(\hat{x}|x)$ under one set of constraints is exactly equal to the maximum value of $I(X;\hat{X})$ with respect to $p(\hat{x})$ under another set of constraints. The second property as given by (17) implies that the solution to the dual channel coding problem is exactly the distribution $p^*(\hat{x})$ induced from the solution to the given source coding problem, provided we are careful about our choice of cost measure as given by (18). Similar interpretations can be given to the properties given by (21), (22) and (23). In other words, for a given source coding problem, we can associate a dual channel coding problem with the same input-output joint distribution, and the rate-distortion bound is equal to the capacity-cost bound, $R(D) = C(W)$. Note that for a given source coding problem, although the conditional distribution is $p^*(\hat{x}|x)$ with $\bar{p}(x)$ as the source input, the dual channel coding problem is defined for the channel $p^*(x|\hat{x})$ with input distribution $p^*(\hat{x})$ in the reverse order as shown in Fig. 5. For a given source coding problem, the test channel [4] characterized by $p^*(x|\hat{x})$ becomes the channel for the dual channel coding problem and vice versa.

Using the forward part of Shannon source and channel coding theorems, for a pair of dual problems, the random codebooks are constructed using the same distribution, $p^*(\hat{x})$ and the jointly typical encoding and decoding (done respectively in source and channel coding) are the same in the sense of finding a codeword from the codebook which is jointly typical (in the sense of $p^*(x,\hat{x}) \triangleq \bar{p}(x)p^*(\hat{x}|x)$ or $p^*(x,\hat{x}) \triangleq p^*(\hat{x})\bar{p}(x|\hat{x})$) with the observed L-length sequence of source input and channel output respectively with the same distribution $\bar{p}(x)$. Thus the distribution $p^*(\hat{x}|x)$ characterizing the function of the encoder of source coding also characterizes the function of the decoder in the dual channel coding. Thus the encoder of the one is functionally identical to the decoder of the other and vice versa. Summarizing, we have:

- Source encoder and channel decoder are characterized by a mapping which has the same domain and range: $\mathcal{X}^L \rightarrow \{1,2,\ldots,2^{LR}\}$, and similarly the source decoder and channel encoder are characterized by a mapping which has the same domain and range: $\{1,2,\ldots,2^{LR}\} \rightarrow \hat{\mathcal{X}}^L$, where R is the rate of transmission. Thus the roles of encoder and decoder are reversed in these two problems.
- Using the forward part [4] of Shannon[4] source and channel coding theorems, in the limit of large block-length, a rate-distortion optimal encoder of the former is functionally identical to a capacity-cost optimal decoder of the latter and vice versa in the sense of Theorem 1a and 1b.

Let us consider the following examples as illustrations of duality.

Example 1: *Source coding*[5] (see Fig. 6(a)): the source is Gaussian with mean 0 and variance σ^2, i.e., $\bar{p}(x) = \mathcal{N}(0,\sigma^2)$, with a quadratic (squared error) distortion measure, $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$ which is given by $d(x,\hat{x}) = (x-\hat{x})^2$, and a distortion constraint D. The optimal test channel conditional distribution [4] satisfying (14) is of the form $\hat{X} = \mu X + q$, where $p(q) = \mathcal{N}(0, \frac{D(\sigma^2-D)}{\sigma^2})$ and $\mu = \frac{\sigma^2-D}{\sigma^2}$, i.e., $p^*(\hat{x}|x) = \mathcal{N}\left(\frac{\sigma^2-D}{\sigma^2}x, \frac{D(\sigma^2-D)}{\sigma^2}\right)$. The induced distributions are given by $X = \hat{X} + Z$, where $p(Z) = \mathcal{N}(0,D)$, i.e., $p^*(x|\hat{x}) = \mathcal{N}(\hat{x}, D)$ and $p^*(\hat{x}) = \mathcal{N}(0,\sigma^2 - D)$. Note that $D < \sigma^2$.

Channel coding dual (see Fig. 6(b)): given the additive memoryless Gaussian noise channel as given above, $X = \hat{X} + Z$, i.e., $p^*(x|\hat{x}) = \mathcal{N}(\hat{x}, D)$, and a cost measure $\omega : \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$, a cost constraint W; the channel input distribution obtained from the above source coding problem, given by $p^*(\hat{x}) = \mathcal{N}(0,\sigma^2 - D)$ maximizes $I(X;\hat{X})$ over all $p(\hat{x})$ such that $E(\omega(\hat{X})) \leq W$, where using Theorem 1a,

$$\omega(\hat{x}) = [c_1 D(p^*(x|\hat{x}) \| \bar{p}(x)) + \theta] = \hat{x}^2 \tag{24}$$

[4] Recall that in the limit of large block-length, in source coding, the rate-distortion bound $R(D)$ is approached from above, while in channel coding the capacity-cost bound $C(W)$ is approached from below.

[5] With a slight abuse of notation, we denote a Gaussian random variable X with mean μ and variance σ^2 as either $X \sim \mathcal{N}(\mu,\sigma^2)$ or $p(x) = \mathcal{N}(\mu,\sigma^2)$.

for an appropriate choice of parameters: $c_1 = 2\sigma^2$, $\theta = c_1\left[\frac{1}{2} - \log\sqrt{\frac{\sigma^2}{D} - \frac{D}{2\sigma^2}}\right]$, and the cost constraint is $W = \sigma^2 - D$. This cost measure, $\omega(\hat{x})$ corresponds to the traditional average power constraint on the channel input. Note from Fig. 6 that the roles of the encoder and the decoder of these dual problems are exactly reversed. The test channel in the source coding problem becomes the channel in the dual channel coding problem. It is also interesting to note that we could have started with this channel coding problem and obtained the above source coding problem as its dual by noting that $d(x,\hat{x})$ can be put in the form (given in (13)), $-c_2\log p^*(x|\hat{x}) + d_0(x)$, where $c_2 = 2D$, $d_0(x) = -D\log(2\pi D)$.

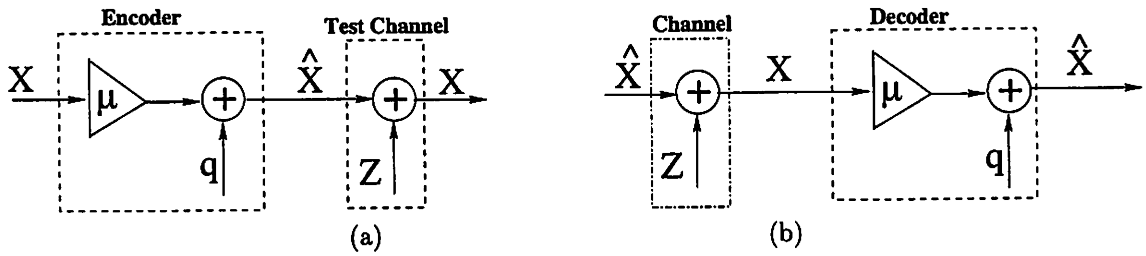


Figure 6: (a) Gaussian Source coding: $\bar{p}(x) = \mathcal{N}(0,\sigma^2)$, $\mu = \frac{\sigma^2 - D}{\sigma^2}$, $p(q) = \mathcal{N}(0, \frac{D(\sigma^2-D)}{\sigma^2})$, $p(Z) = \mathcal{N}(0,D)$. The effect of encoding on X is characterized as $\mu X + q$. Test channel gives the relation between $\hat{X}$ and X. (b) Channel coding dual with the same joint distribution. The decoding is depicted as one of recovering $\hat{X}$ from X.

Example 2: *Channel coding*: consider a binary symmetric channel with transition probability p, characterized as: $X = \hat{X} \oplus Z$, where $X, \hat{X}, Z \in \{0,1\}$, $p(Z=1) = p$ and $\oplus$ denotes binary modulo-2 addition. The cost measure is given by $\omega(i) = i$ for $i = 0,1$, which corresponds to a constraint on the duty cycle of the channel input. The capacity of this channel with expected cost $E\omega(\hat{X}) \leq W$, for some $0 \leq W \leq 0.5$ can be shown [4] to be[6] $C(W) = h(p * W) - h(p)$, where $p * W = p(1-W) + (1-p)W$. The optimal channel input distribution is $p^*(\hat{X} = 1) = W$. Note that $p^*(X=1) = p * W$.

Source coding dual: given the binary source with distribution from the above, i.e., $p^*(X=1) = p * W$, with a distortion measure $d(.,.)$, and a distortion constraint D; the test channel conditional distribution $p^*(\hat{x}|x)$ (induced from the above channel coding problem) minimizes $I(X;\hat{X})$ over all $p(\hat{x}|x)$ such that $Ed(X,\hat{X}) \leq D$, where using Theorem 1b,

$$d(x,\hat{x}) = [-c_2\log\bar{p}(x|\hat{x}) + d_0(x)] = x \oplus \hat{x}, \tag{25}$$

is the Hamming distortion measure if we choose $c_2 = \frac{1}{[\log(1-p) - \log(p)]}$, $d_0(x) = \frac{\log(1-p)}{[\log(1-p)-\log(p)]}$, and the distortion constraint is $D = p$. Thus in this binary example, the duty cycle cost measure for the channel coding problem corresponds to the Hamming distortion measure for the dual source coding problem.

[6] Note that when $W = 0.5$, we get the classical capacity of the binary symmetric channel, $1 - h(p)$.

3.2 Duality between source and channel coding with common side information

Let us consider a similar formulation of duality for source and channel coding with common side information present at both the encoder and the decoder (denoted by SCSIB and CCSIB respectively, where the appended "B" stands for "both"). This is a straightforward extension of the treatment of the previous section to the case where we condition on the common side information. The motivation for addressing this extension is that this is a necessary step (as will be evident in Section 3.3) that will serve as a bridge to establishing the duality between SCSI and CCSI. Another motivation is to establish the need for the generalized notion of cost and distortion measures introduced in Section 2. The following lemmas extend the ideas given in Facts 3 and 4 to this case in a straightforward way and are presented here for completeness.

Lemma 1a: For CCSIB, with a given channel $\bar{p}(x|\hat{x}, s)$, side information $\bar{p}(s)$, input, side information and output alphabets $\hat{\mathcal{X}}$, $\mathcal{S}$ and $\mathcal{X}$ respectively, given a conditional channel input distribution $p^*(\hat{x}|s)$, $\exists$ a cost measure $\omega : \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$ and a cost constraint W such that

$$p^*(\hat{x}|s) = \underset{p(\hat{x}|s)}{\text{argmax}} \; I(X; \hat{X}|S) \tag{26}$$

under the constraints $(X|\hat{X}, S) \sim \bar{p}(x|\hat{x}, s)$, $S \sim \bar{p}(s)$ and $E\omega(\hat{X}, S) \leq W$, where

$$\omega(\hat{x}, s) = c_1 D\left(\bar{p}(x|\hat{x}, s) \| p^*(x|s)\right) + \theta(s), \text{ and } W = E_{\bar{p}(s)p^*(\hat{x}|s)}(\omega(\hat{X}, S)), \tag{27}$$

with $c_1 > 0$ and $\theta(s)$ being arbitrary and $p^*(x|s) = \sum_{\hat{x}} p^*(\hat{x}|s)\bar{p}(x|\hat{x}, s)$ from standard Bayes' rule.

Proof: Omitted

Lemma 1b: For SCSIB, with a given source $\bar{p}(x|s)$, side information $\bar{p}(s)$, input, side information and reconstruction alphabets $\mathcal{X}$, $\mathcal{S}$, and $\hat{\mathcal{X}}$ respectively, given a conditional distribution, $p^*(\hat{x}|x, s)$, $\exists$ a distortion measure, $d : \mathcal{X} \times \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$ and a distortion constraint D such that

$$p^*(\hat{x}|x, s) = \underset{p(\hat{x}|x, s)}{\text{argmin}} \; I(X; \hat{X}|S) \tag{28}$$

under the constraints $(X|S) \sim \bar{p}(x|s)$, $S \sim \bar{p}(s)$, and $Ed(X, \hat{X}, S) \leq D$, where

$$d(x, \hat{x}, s) = -c_2 \log p^*(x|\hat{x}, s) + d_0(x, s), \text{ and } D = E_{\bar{p}(x|s)\bar{p}(s)p^*(\hat{x}|x,s)}(d(X, \hat{X}, S)), \tag{29}$$

with $c_2 > 0$ and $d_0(x, s)$ being arbitrary, and $p^*(x|\hat{x}, s) = \frac{\bar{p}(x|s)p^*(\hat{x}|x,s)}{\sum_x \bar{p}(x|s)p^*(\hat{x}|x,s)}$, $p^*(\hat{x}|s) = \sum_x \bar{p}(x|s)p^*(\hat{x}|x, s)$, from standard Bayes' rule.

As noted earlier, with the more general distortion and cost measures, we are able to extend the concepts of Facts 3 and 4 to the case when a common side information is present at both the encoder and the decoder. We will give an example of a Gaussian source to illustrate the behavior of these measures. We now connect these two coding problems using the following theorem (which is an extension of Theorem 1a), which says that for a given SCSIB, we can associate a dual CCSIB having the same joint distribution.

Theorem 2a: For a SCSIB, with a given source $\bar{p}(x|s)$, side information $\bar{p}(s)$, input, side information and reconstruction alphabets $\mathcal{X}$, $\mathcal{S}$ and $\hat{\mathcal{X}}$ respectively, a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$, and a distortion constraint D, let the conditional distribution achieving the rate-distortion optimality $R_{x|s}(D)$ be given by:

$$p^*(\hat{x}|x,s) \triangleq \underset{p(\hat{x}|x,s)\,:\,(X|S) \sim \bar{p}(x|s),\, S \sim \bar{p}(s),\, Ed(X,\hat{X},S) \leq D}{\text{argmin}} I(X;\hat{X}|S), \tag{30}$$

inducing the following distributions:

$$p^*(x|\hat{x},s) \triangleq \frac{p^*(\hat{x}|x,s)\bar{p}(x|s)}{\sum_x \bar{p}(x|s)p^*(\hat{x}|x,s)}, \quad p^*(\hat{x}|s) \triangleq \sum_x \bar{p}(x|s)p^*(\hat{x}|x,s). \tag{31}$$

Then $\exists$ a *dual* CCSIB for the channel $p^*(x|\hat{x},s)$, having side information $\bar{p}(s)$, input, side information and output alphabets $\hat{\mathcal{X}}$, $\mathcal{S}$ and $\mathcal{X}$ respectively, a cost measure $\omega : \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$ and a cost constraint W, such that:

- the rate-distortion bound with side information $R_{x|s}(D)$ is equal to the capacity-cost bound[7] with side information $C_{x|s}(W)$, i.e.,

$$\min_{p(\hat{x}|x,s)\,:\,\left\{\begin{array}{c}(X|S) \sim \bar{p}(x|s)\\ S \sim \bar{p}(s)\\ Ed(X,\hat{X},S) \leq D\end{array}\right\}} I(X;\hat{X}|S) = \max_{p(\hat{x}|s)\,:\,\left\{\begin{array}{c}(X|\hat{X},S) \sim p^*(x|\hat{x},s)\\ S \sim \bar{p}(s)\\ E\omega(\hat{X},S) \leq W\end{array}\right\}} I(X;\hat{X}|S); \tag{32}$$

- the channel input distribution given by $p^*(\hat{x}|s)$ achieves capacity-cost optimality, i.e.,

$$p^*(\hat{x}|s) = \underset{p(\hat{x}|s)\,:\,(X|\hat{X},S) \sim p^*(x|\hat{x},s),\, S \sim \bar{p}(s),\, E\omega(\hat{X},S) \leq W}{\text{argmax}} I(X;\hat{X}|S), \tag{33}$$

where the cost measure and cost constraint are given respectively by

$$\omega(\hat{x},s) \triangleq c_1 D\left(p^*(x|\hat{x},s) \| \bar{p}(x|s)\right) + \theta(s), \quad W \triangleq E_{\bar{p}(s)p^*(\hat{x}|s)}(\omega(\hat{X},S)), \tag{34}$$

for arbitrary $c_1 > 0$ and $\theta(s)$.

Proof: Omitted

Remark 3: The first property of the dual problem as given by (32) implies that the minimum of $I(X;\hat{X}|S)$ with respect to $p(\hat{x}|x,s)$ under one set of source coding constraints is equal to the maximum of the same objective function $I(X;\hat{X}|S)$ with respect to $p(\hat{x}|s)$ under another set of channel coding constraints. The second property

[7] $C_{x|s}(W)$ denotes the capacity of CCSIB using (7) and (9).

as given by (33) implies that the solution to the dual CCSIB problem is exactly the distribution $p^*(\hat{x}|s)$ induced from the solution to the given SCSIB, provided we are careful about our choice of cost measure as in (34). We emphasize that for a given SCSIB, we can associate a dual CCSIB with the *same joint distribution*, and the rate-distortion bound is equal to the capacity-cost bound, $R_{x|s}(D) = C_{x|s}(W)$.

A similar result holds true for the case of CCSIB which is an extension of Theorem 1b. It essentially says that for a given CCSIB problem, there exists a dual SCSIB problem with an optimal distortion measure. We are not presenting this here. In other words, for a given SCSIB, there exists a dual CCSIB such that in the limit of large blocklength, a rate-distortion optimal encoder of the former is functionally identical to a capacity-cost optimal decoder of the latter and vice versa. Further, using the forward part of the coding theorems [27], we see that the random conditional codebooks are constructed with the conditional distribution $p^*(\hat{x}|s)$ corresponding to the given outcome of the side information. The jointly typical encoding and decoding in SCSIB and CCSIB respectively are identical. Let us consider an example to illustrate the dual problems.

Example 3: *SCSIB* (see Fig. 7(a)): consider a Gaussian source X, given by $X = S + V$, with $p(V) = \mathcal{N}(0, N)$, i.e., $\bar{p}(x|s) = \mathcal{N}(s, N)$ and S is the Gaussian side information given by $\bar{p}(s) = \mathcal{N}(0, Q)$ for some positive real N and Q, the distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$ is quadratic, $(x - \hat{x} - s)^2$ and the distortion constraint D. This corresponds to the case of reconstruction of V, the difference between X and S (instead of X) using a mean squared error as distortion measure. Let $\alpha = \frac{N-D}{N}$. The optimal conditional distribution[8] [27] satisfying (30) is given by $\hat{X} = \alpha V + q$, with $p(q) = \mathcal{N}(0, \frac{D(N-D)}{N})$, i.e., $p^*(\hat{x}|x, s) = \mathcal{N}\left(\alpha(x - s), \frac{D(N-D)}{N}\right)$. The induced conditional test channel distribution is given by $X = \hat{X} + S + Z$, where $p(Z) = \mathcal{N}(0, D)$, i.e., $p^*(x|\hat{x}, s) = \mathcal{N}(\hat{x} + s, D)$, and $\hat{X}$ is independent of S, i.e., $p^*(\hat{x}|s) = \mathcal{N}(0, N - D)$. Note that $D < N$. Further, α is equal to the coefficient in the Minimum Mean Squared Error (MMSE) estimate of $\hat{X}$ from $\hat{X} + Z$.

Dual CCSIB (see Fig. 7(b)): given the additive memoryless Gaussian channel obtained from the above SCSIB, characterized by $X = \hat{X} + S + Z$, i.e., $p^*(x|\hat{x}, s) = \mathcal{N}(\hat{x} + s, D)$, with a cost measure $\omega : \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$, and a cost constraint W; the channel input distribution induced from the above SCSIB, $p^*(\hat{x}|s) = \mathcal{N}(0, N - D)$ maximizes $I(X; \hat{X}|S)$ over all $p(\hat{x}|s)$ such that $E\omega(\hat{X}, S) \leq W$, where using Theorem 2a, we have the cost measure

$$\omega(\hat{x}, s) = c_1 D\left(p^*(x|\hat{x}, s) \| \bar{p}(x|s)\right) + \theta(s) = \hat{x}^2 \tag{35}$$

if we choose $c_1 = 2N$, $\theta(s) = c_1 \left[\frac{1}{2}\log\frac{eD}{N} - \frac{D}{2N}\right]$, and the cost constraint is $W = N - D$. This corresponds to a simple average power constraint on the channel input, independent of the side information S. Thus the encoder completely ignores the presence of the side information, and the decoder subtracts the side information from

[8] given by the rate-distortion bound on $V = X - S$

the channel output before further processing. Note that the functional roles of the encoder and decoder of these dual problems are exactly reversed: the test channel of SCSIB becomes the channel of the dual CCSIB, and the source encoder for SCSIB becomes the channel decoder for CCSIB. We could just as easily have started with the CCSIB problem and found the dual SCSIB problem.

Remark 4: If in this example, for SCSIB with the same source X, and side information S, if we choose the

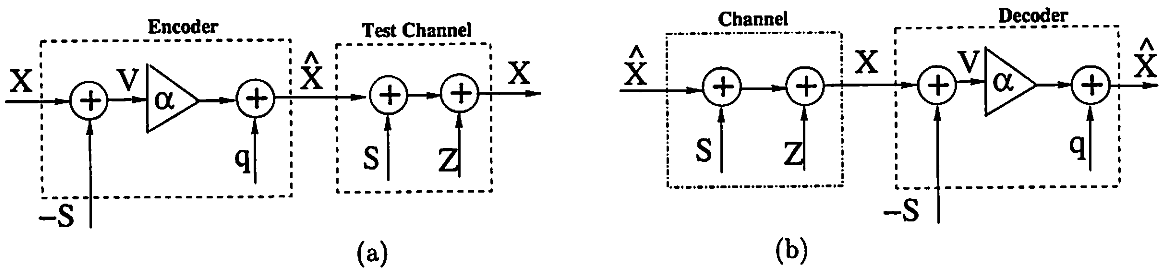


Figure 7: Duality: $\alpha = (N-D)/N$, $p(q) = \mathcal{N}(0, \frac{D(N-D)}{N})$, $p(Z) = \mathcal{N}(0, D)$. (a) SCSIB as given in Example 3, with distortion $d(x, \hat{x}, s) = (x - \hat{x} - s)^2$. (b) CCSIB dual of (a) with cost $\omega(\hat{x}, s) = \hat{x}^2$. Note the roles of encoder and decoder are exchanged in (a) and (b).

distortion measure to be $d(x, \hat{x}, s) = (x - \hat{x})^2$ (i.e., if we are interested in the conventional MSE reconstruction of the source X rather than that of the difference between the source and the side information, $(X - S)$), then the optimal conditional distribution is given by $\hat{X} = \alpha X + (1 - \alpha)S + q$, with $p(q) = \mathcal{N}(0, \frac{D(N-D)}{N})$, i.e., $p^*(\hat{x}|x, s) = \mathcal{N}\left(\alpha(x - s) + s, \frac{D(N-D)}{N}\right)$. This induces the test channel given by $X = \hat{X} + Z$, i.e., $p^*(x|\hat{x}, s) = \mathcal{N}(\hat{x}, D)$, and the correlation between $\hat{X}$ and S is given by $p^*(\hat{x}|s) = \mathcal{N}(s, N - D)$. The dual CCSIB will have the channel $p^*(x|\hat{x}, s) = \mathcal{N}(\hat{x}, D)$, under the channel cost measure $\omega(\hat{x}, s) = (\hat{x} - s)^2$. This corresponds to a constraint on the average power of the difference between the channel input $\hat{X}$ and side information S (In data hiding and watermarking applications [16, 22] such a cost measure is popular as it represents the amount of distortion induced on the original host signal.). Summarizing, we see that the distortion measure $(x - \hat{x})^2$ in SCSIB corresponds to the channel cost measure $(\hat{x} - s)^2$ in the dual CCSIB, where we saw earlier that the distortion measure $(x - \hat{x} - s)^2$ in SCSIB corresponds to the cost measure $\hat{x}^2$ in the dual CCSIB.

3.3 Duality of SCSI and CCSI

This section contains the main result of this paper. First let us state a correspondence between the variables used in the definition of SCSI and CCSI.

$$SCSI \left\{ \begin{array}{cccccc} \text{source input} & X & \longleftrightarrow & Y & \text{channel output} \\ \text{side information} & S & \longleftrightarrow & S & \text{side information} \\ \text{auxiliary variable} & U & \longleftrightarrow & U & \text{auxiliary variable} \\ \text{source reconstruction} & \hat{X} & \longleftrightarrow & X & \text{channel input} \end{array} \right\} CCSI \qquad (36)$$

We follow the notation of SCSI and use the same for CCSI. The duality between SCSI and CCSI has been studied in [20, 28]. In this section, we formulate the notion of functional duality for the SCSI and CCSI similar to the one developed in the previous sections.

3.3.1 Duality formulation

Recall that in the Wyner-Ziv formulation of SCSI (Fig. 1), the problem is to rate-distortion optimally compress X with side information S present only at the decoder. The objective function to be minimized is $[I(U;X) - I(U;S)]$ with respect to $p(u|x)$ and $\hat{x} = f(u,s)$, where U is an auxiliary random variable. It is important to note that there is a natural Markov chain associated with the definition of the problem: $S \to X \to U$ which essentially captures the fact that side information S is available at only the decoder. Suppose the objective function $[I(U;X) - I(U;S)]$ is optimized by $\{p^*(u|x), f^*(u,s)\}$. This[9], with the Markov chain $S \to X \to U$, completely determines the joint distribution $p^*(x,s,\hat{x},u) = \bar{p}(s)\bar{p}(x|s)p^*(u|x)p^*(\hat{x}|u,s)$, thus fixing all conditional distributions including $p^*(u|s)$. Similarly, in the Gelfand-Pinsker formulation of CCSI (Fig. 2), the problem is to reliably transmit maximal amount of information across a channel in the presence of side information only at the encoder. The objective function to be maximized is $[I(U;X) - I(U;S)]$ with respect to $p(u|s)$ and $\hat{x} = f(u,s)$, where U is an auxiliary random variable. A natural Markov chain associated with the definition of this problem is given by $X \to (\hat{X},S) \to U$, which essentially captures the fact that the channel output is governed only by the channel input $\hat{X}$ and side information S. Suppose the objective function is optimized by $\{p^*(u|s), f^*(u,s)\}$. This, with the Markov chain $X \to (\hat{X},S) \to U$, completely determines the joint distribution $p^*(x,s,\hat{x},u) = \bar{p}(s)p^*(u|s)p^*(\hat{x}|u,s)\bar{p}(x|\hat{x},s)$, thus fixing all the conditional distributions including $p^*(u|x)$.

In order for general duality between the CCSI and SCSI problems, we have to reconcile the fact that these two problems come with completely different Markov chain conditions in their very definition. Thus for a general SCSI problem, the distribution minimizing $[I(U;X) - I(U;S)]$ cannot be used for any CCSI problem unless $X \to (\hat{X},S) \to U$ is satisfied, which is a necessary initial condition for the latter. Therefore, unless the optimizing joint distribution $p^*(x,s,\hat{x},u)$ satisfies *both* Markov chains, there can be no duality. However, fortuitously, there exists a rich subset of these problems (including the important Gaussian case), where these Markov chains are satisfied, and we do have a functional duality. We have the following theorems characterizing this.

Theorem 3a: For a SCSI, with a given source $\bar{p}(x|s)$, side information $\bar{p}(s)$, input, side information and

[9] Recall the convention relating to $\bar{p}(.)$ and $p^*(.)$ given in Section 3.1.

reconstruction alphabets $\mathcal{X}, \mathcal{S}$ and $\hat{\mathcal{X}}$ respectively, a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$, and a distortion constraint D, let the conditional distribution achieving the rate-distortion optimality $R^*(D)$ be given by:

$$\left\{ \begin{array}{c} p^*(u|x), \\ \hat{x} = f^*(u,s) \end{array} \right\} \triangleq \underset{\substack{p(u|x), \\ \hat{x} = f(u,s)}}{\operatorname{argmin}} : \left\{ \begin{array}{c} (X|S) \sim \bar{p}(x|s), S \sim \bar{p}(s) \\ Ed(X,\hat{X},S) \le D, (S \to X \to U) \end{array} \right\} I(U;X) - I(U;S), \tag{37}$$

inducing the following distributions: $p^*(x,s,\hat{x},u) \triangleq \bar{p}(s)\bar{p}(x|s)p^*(u|x)p^*(\hat{x}|u,s)$,

$$p^*(x|\hat{x},s) \triangleq \frac{\sum_u p^*(x,s,\hat{x},u)}{\sum_{u,x} p^*(x,s,\hat{x},u)}, \quad p^*(u|s) \triangleq \frac{\sum_{x,\hat{x}} p^*(x,s,\hat{x},u)}{\bar{p}(s)}, \quad p^*(\hat{x}|s) \triangleq \frac{\sum_{u,x} p^*(x,s,\hat{x},u)}{\bar{p}(s)}, \tag{38}$$

where $p^*(\hat{x}|u,s)$ is given by $\hat{x} = f^*(u,s)$. If $p^*(x,s,\hat{x},u)$ is such that $X \to (\hat{X},S) \to U$, then $\exists$ a dual CCSI, for the channel $p^*(x|\hat{x},s)$, having side information $\bar{p}(s)$, input, side information and output alphabets $\hat{\mathcal{X}}$, $\mathcal{S}$ and $\mathcal{X}$ respectively, a cost measure $\omega : \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$ and a cost constraint W, such that:

- the rate-distortion bound with side information $R^*(D)$ is equal to the capacity-cost bound with side information $C^*(W)$, i.e.,

$$\underset{\substack{p(u|x), \\ \hat{x} = f(u,s)}}{\min} : \left\{ \begin{array}{c} (X|S) \sim \bar{p}(x|s) \\ S \sim \bar{p}(s) \\ Ed(X,\hat{X},S) \le D \\ S \to X \to U \end{array} \right\} I(U;X) - I(U;S) = \underset{\substack{p(u|s), \\ \hat{x} = g(u,s)}}{\max} : \left\{ \begin{array}{c} (X|\hat{X},S) \sim p^*(x|\hat{x},s) \\ S \sim \bar{p}(s) \\ E\omega(\hat{X},S) \le W \\ X \to (\hat{X},S) \to U \end{array} \right\} I(U;X) - I(U;S); \tag{39}$$

- the conditional distribution $p^*(u|s)$ and the function $\hat{x} = f^*(u,s)$ achieve capacity-cost optimality, i.e.,

$$\left\{ \begin{array}{c} p^*(u|s), \\ \hat{x} = f^*(u,s) \end{array} \right\} = \underset{\substack{p(u|s) \\ \hat{x} = g(u,s)}}{\operatorname{argmax}} : \left\{ \begin{array}{c} (X|\hat{X},S) \sim p^*(x|\hat{x},s), S \sim \bar{p}(s) \\ E\omega(\hat{X},S) \le W, (X \to (\hat{X},S) \to U) \end{array} \right\} I(U;X) - I(U;S), \tag{40}$$

where the cost measure and cost constraint are given respectively by

$$\omega(\hat{x},s) \triangleq c_1 D\left(p^*(x|\hat{x},s) \| \bar{p}(x|s)\right) + \theta(s), \quad W \triangleq E_{\bar{p}(s)p^*(\hat{x}|s)}(\omega(\hat{X},S)), \tag{41}$$

for arbitrary $c_1 > 0$ and $\theta(s)$.

Proof: See Appendix

Theorem 3b: For a CCSI, with a given channel $\bar{p}(x|\hat{x},s)$, side information $\bar{p}(s)$, input, side information and output alphabets $\hat{\mathcal{X}}$, $\mathcal{S}$, and $\mathcal{X}$ respectively, a cost measure $\omega : \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$, and a cost constraint W, let the conditional distribution achieving the capacity-cost optimality $C^*(W)$ be given by:

$$\left\{ \begin{array}{c} p^*(u|s) \\ \hat{x} = f^*(u,s) \end{array} \right\} \triangleq \underset{\substack{p(u|s) \\ \hat{x} = f(u,s)}}{\operatorname{argmax}} : \left\{ \begin{array}{c} (X|\hat{X},S) \sim \bar{p}(x|\hat{x},s), S \sim \bar{p}(s) \\ E\omega(\hat{X},S) \le W, (X \to (\hat{X},S) \to U) \end{array} \right\} I(U;X) - I(U;S), \tag{42}$$

inducing the following distributions: $p^*(x,s,\hat{x},u) \triangleq \bar{p}(s)p^*(u|s)p^*(\hat{x}|u,s)\bar{p}(x|\hat{x},s)$,

$$p^*(x|s) \triangleq \frac{\sum_{u,\hat{x}} p^*(x,s,\hat{x},u)}{\bar{p}(s)}, \quad p^*(u|x) \triangleq \frac{\sum_{s,\hat{x}} p^*(x,s,\hat{x},u)}{\sum_{u,s,\hat{x}} p^*(x,s,\hat{x},u)}, \quad p^*(\hat{x}|s) \triangleq \frac{\sum_{u,x} p^*(x,s,\hat{x},u)}{\bar{p}(s)}, \tag{43}$$

where $p^*(\hat{x}|u,s)$ is given by $\hat{x} = f^*(u,s)$. If $p^*(x,s,\hat{x},u)$ is such that $S \to X \to U$, then $\exists$ a dual SCSI for the source $p^*(x|s)$, having side information $\bar{p}(s)$, input, side information and reconstruction alphabets $\mathcal{X}, \mathcal{S}$, and $\hat{\mathcal{X}}$ respectively, a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$ and a distortion constraint D, such that:

- the capacity-cost bound with side information $C^*(W)$ is equal to the rate-distortion bound with side information $R^*(D)$, i.e.,

$$\max_{\substack{p(u|s) \\ \hat{x}=f(u,s)}} : \left\{ \begin{array}{c} (X|\hat{X},S) \sim \bar{p}(x|\hat{x},s) \\ S \sim \bar{p}(s) \\ E\omega(\hat{X},S) \le W \\ X \to (\hat{X},S) \to U \end{array} \right\} I(U;X) - I(U;S) = \min_{\substack{p(u|x) \\ \hat{x}=g(u,s)}} : \left\{ \begin{array}{c} (X|S) \sim p^*(x|s) \\ S \sim \bar{p}(s) \\ Ed(X,\hat{X},S) \le D \\ S \to X \to U \end{array} \right\} I(U;X) - I(U;S); \tag{44}$$

- the conditional distribution $p^*(u|x)$ and the function $\hat{x} = f^*(u,s)$ achieve rate-distortion optimality, i.e.,

$$\left\{ \begin{array}{c} p^*(u|x), \\ \hat{x} = f^*(u,s) \end{array} \right\} = \operatorname*{argmin}_{\substack{p(u|x) \\ \hat{x}=g(u,s)}} : \left\{ \begin{array}{c} (X|S) \sim p^*(x|s), S \sim \bar{p}(s) \\ Ed(X,\hat{X},S) \le D, (S \to X \to U) \end{array} \right\} I(U;X) - I(U;S), \tag{45}$$

where the distortion measure and distortion constraint are given respectively by

$$d(x,\hat{x},s) \triangleq -c_2 \log \bar{p}(x|\hat{x},s) + d_0(x,s), \quad D \triangleq E_{\bar{p}(s)p^*(\hat{x}|s)\bar{p}(x|\hat{x},s)}(d(X,\hat{X},S)), \tag{46}$$

for arbitrary $c_2 > 0$ and $d_0(x,s)$.

Remark 5: Before we interpret these theorems in terms of functional duality, we want to point out (see an illustration of the Gaussian case in Fig. 8) that using the forward part of the coding theorems [1, 11], in SCSI and CCSI, a codebook with blocklength L, consisting of roughly $2^{LI(U;X)}$ codewords is partitioned (so-called random binning as given in [4]) into 2^{LR} cosets containing roughly $2^{LI(U;S)}$ codewords (see the correspondence in (36)). In SCSI, this is a partition of a source codebook for quantizing X into cosets of "channel codebooks" for the fictitious channel between U and S. In CCSI, this is a partition of a channel codebook for the fictitious channel between U and X into cosets of "source codebooks" for quantizing S. This concept of coset codes also ties these two problems together.

Remark 6: In Theorem 3a, the property of the dual CCSI, given by (39), implies that the minimum of $[I(U;X) - I(U;S)]$ with respect to $p(u|x)$ and $\hat{x} = f(u,s)$ under one set of constraints is equal to the maximum of the same objective function $[I(U;X) - I(U;S)]$ with respect to $p(u|s)$ and $\hat{x} = f(u,s)$ under another set of

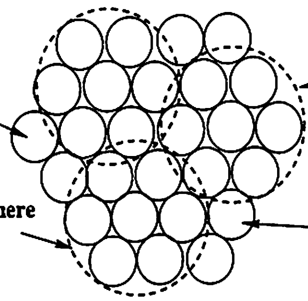


Figure 8: Coset code structure for SCSI and CCSI for the Gaussian case: In SCSI (respectively CCSI) the set of source (channel) codewords denoted by small spheres are partitioned into cosets of channel (source) codewords denoted by big spheres.

constraints. Further, the property as given by (40) implies that the solution to the dual CCSI is exactly the distribution $p^*(u|s)$ and $\hat{x} = f^*(u, s)$ induced from the solution to the given SCSI, provided we are careful about our choice of channel cost measure as given in (41). A similar interpretation can be given to the properties given by (44) and (45) in Theorem 3b. In other words, Theorem 3(a) states that for a given SCSI, we can associate a dual CCSI such that the joint distribution is identical, with the appropriate choice of the channel cost measure, and the rate-distortion bound is equal to the capacity cost bound, $R^*(D) = C^*(W)$.

Using the forward part [1, 11], the random codebooks are constructed with distribution $p^*(u)$. These codebooks are randomly partitioned into approximately $2^{LR^*(D)} = 2^{LC^*(W)}$ cosets (see Remark 5). The test channel characterized by $p^*(x|\hat{x}, s)$ of SCSI becomes the channel of the dual CCSI and vice versa. The jointly typical encoding operation in SCSI is identical to the jointly typical decoding operation in CCSI in the following sense: find a codeword U from the composite codebook which is jointly typical (in the sense of $p^*(u, x)$) with the observed (source input in SCSI and channel output in CCSI) L-sequence coming from the same distribution $p^*(x)$. Then the encoder of SCSI sends the index of the coset containing this typical codeword as a message to the decoder, while the decoder of CCSI declares the index of the coset containing this typical codeword as the message sent from the encoder. The corresponding decoder and encoder of SCSI and CCSI respectively have access to a message. The jointly typical decoding operation in SCSI is identical to the jointly typical encoding operation in CCSI, and is done as follows: find a codeword U from the coset whose index is given by the message that is jointly typical (in the sense of $p^*(u, s)$) with the observed side information. Thus, $p^*(u|x)$ ($p^*(u|s)$ respectively) characterizing the function of the encoder (decoder) of a SCSI also characterizes the function of the decoder (encoder) of the dual CCSI. The reconstruction and the channel input are the same, taken as a function of this typical codeword and the observed side information. *Thus the encoder (decoder respectively) of SCSI and the decoder (encoder) of CCSI are functionally identical.* We will illustrate this shortly using concrete examples.

Remark 7: Note that for a general SCSI (say $SCSI_1$), with a given source $\bar{p}(x|s)$, side information $\bar{p}(s)$, and distortion measure d, even if the joint distribution (say $p^*(x,s,\hat{x},u)$) induced by the optimizer $p^*(u|x)$ and $\hat{x} = f^*(u,s)$ has the Markov property $X \rightarrow (\hat{X}, S) \rightarrow U$ (needed to define the dual CCSI problem), there might be rate-loss (see end of Section 2.1 for conditions for rate-loss) compared to the corresponding SCSIB (say $SCSIB_1$). We want to point out that its dual CCSI (say $CCSI_1$), with the cost measure given in Theorem 3a, which uses the same joint distribution $p^*(x,s,\hat{x},u)$ has no rate-loss (see end of Section 2.2) compared to its corresponding CCSIB (say $CCSIB_1$), as can be seen from the proof of Theorem 3a. Further, if we had started with $CCSI_1$ as the given problem, then its dual say $SCSI_2$ with the associated distortion measure (say d') as given by Theorem 3b, will have no rate-loss compared to its corresponding $SCSIB_2$. Clearly both $SCSI_1$ and $SCSI_2$ use the same joint distribution $p^*(x,s,\hat{x},u)$, but different distortion measures. Using the forward part of Wyner-Ziv theorem, the encoder and the decoder of $SCSI_1$ are respectively identical to that of $SCSI_2$. This is illustrated as given below.

$$\begin{array}{ccccc} SCSIB_1 & & CCSIB_1 & \overset{\text{dual}}{\longleftrightarrow} & SCSIB_2 \\ \text{rate-loss} \Downarrow & & \text{NO rate-loss} \Downarrow & & \text{NO rate-loss} \Downarrow \\ SCSI_1 & \overset{\text{dual}}{\longleftrightarrow} & CCSI_1 & \overset{\text{dual}}{\longleftrightarrow} & SCSI_2 \end{array}$$

A similar sequence of arguments holds true for a general CCSI with the joint distribution satisfying $S \rightarrow X \rightarrow U$. Thus in SCSI and CCSI, we can not guarantee uniqueness of functionally dual problems. This is summarized by the following corollary.

Corollary 1: For every SCSI (respectively CCSI) with a given source $\bar{p}(x|s)$ (channel $\bar{p}(x|\hat{x},s)$), side information $\bar{p}(s)$, a distortion measure d (cost measure ω) and a distortion constraint D (cost constraint W), if the optimal joint distribution $p^*(x,s,\hat{x},u)$ satisfies $X \rightarrow (\hat{X},S) \rightarrow U$ ($S \rightarrow X \rightarrow U$), then $\exists$ a distortion measure d' (cost measure ω') and a distortion constraint D' (cost constraint W') such that the SCSI (CCSI), using this distortion (cost) measure has $p^*(x,s,\hat{x},u)$ as the optimal joint distribution, and there is no rate-loss compared to its corresponding SCSIB (CCSIB).

The well-studied Gaussian cases for SCSI [1] and CCSI [18] tempt one to associate duality with no rate-loss. *But in general, functional duality does not equate to no rate-loss in the given SCSI and similarly in the given CCSI.* In other words, even if a given SCSI has rate-loss compared to its corresponding SCSIB, we can find a dual CCSI if the optimizer in the given SCSI satisfies the two Markov chains as mentioned in Theorem 3a. This is illustrated using an example to be given in Section 3.3.4. A similar observation can be made for a given CCSI with the joint distribution induced by the optimizer $p^*(u|s)$ and $\hat{x} = f^*(u,s)$, satisfying $S \rightarrow X \rightarrow U$. In the proof of Theorem 3a and 3b, the corresponding CCSIB and SCSIB problems serve as a bridge in obtaining the dual problems for SCSI and CCSI. Summarizing, we note the following:

- The objective function $[I(U;X)-I(U;S)]$ is convex in one variable $p(u|x)$, given $p(x|s)$, $p(s)$, $S \to X \to U$ and $\hat{x} = f(u,s)$ and concave in the other variable $p(u|s)$, given $p(x|\hat{x},s)$, $p(s)$, $X \to (\hat{X},S) \to U$ and $\hat{x} = f(u,s)$.

- For SCSI, $p(x|s)$ and $p(s)$ are fixed; the optimization involves minimizing $[I(U;X)-I(U;S)]$ over $p(u|x)$.

- For CCSI, $p(x|\hat{x},s)$ and $p(s)$ are fixed; the optimization involves maximizing $[I(U;X)-I(U;S)]$ over $p(u|s)$.

- The encoder of SCSI and the decoder of CCSI are characterized by a mapping having the same domain and range: $\mathcal{X}^L \to \{1,2,\ldots,2^{RL}\}$. Similarly the decoder of SCSI and the encoder of CCSI are characterized by a mapping having the same domain and range: $\{1,2,\ldots,2^{RL}\} \times \mathcal{S}^L \to \hat{\mathcal{X}}^L$. Thus the mappings of the encoder and decoder are reversed in these two problems.

- *Using the forward*[10] *part of the Wyner-Ziv [1] and Gelfand-Pinsker [11] theorems, in the limit of large block-length, a rate-distortion optimal encoder of a SCSI is functionally identical to a capacity-cost optimal decoder of the dual CCSI and vice versa in the sense of Theorem 3a and 3b.*

In the following, we consider an important generalization of the result of Wyner-Ziv [1] relating to no rate-loss for SCSI from Gaussian to more arbitrary distributions. We have obtained this generalization as a dual to the recent generalization of Costa's result to arbitrary side information, obtained in [19]. Then we consider an example [1, 18] with the noise and the side information both being Gaussian, illustrating the duality. Finally, we consider a discrete example [1] of SCSI having a rate-loss compared to its corresponding SCSIB and obtain its dual CCSI.

3.3.2 Generalization of Gaussian case of Wyner and Ziv with no rate-loss

Cohen and Lapidoth [19] have recently generalized the "dirty paper" result of Costa [18] to hold even if the channel side information is not Gaussian (the channel noise has to be Gaussian). That is, for the CCSI problem with $X = S + \hat{X} + Z$ (see correspondence in (36)), where Z is i.i.d. Gaussian, there is no capacity loss over the corresponding CCSIB problem where S is known at both ends, with S being arbitrary. By the above duality,

[10] Recall that in the limit of large block-length, in SCSI, the rate-distortion bound $R^*(D)$ is approached from above, while in CCSI, the capacity-cost bound $C^*(W)$ is approached from below.

we are now in a position to correspondingly generalize the Wyner-Ziv result (with no rate-loss) for the Gaussian source characterized by $X = S + V$ (where X, S, V are all Gaussian) to the more general case where V is Gaussian but X and S are otherwise arbitrary (but related by $X = S + V$): this follows directly from the above duality, and is a direct example of how generalization of a result in one problem leads to a corresponding impact in the dual problem. This important generalization is presented as a corollary in the following. While the SCSI claim of this corollary can be obtained "from first principles" similar to [19], it is far more insightful and elegant to deduce this using the duality concepts developed in Theorem 3.

CCSI [19] (see Fig. 9(a)): Let the channel be additive white Gaussian, given by $X = \hat{X} + S + Z$, with $p(Z) = \mathcal{N}(0, D)$, i.e., $\bar{p}(x|\hat{x}, s) = \mathcal{N}(\hat{x} + s, D)$ for some $D > 0$ and let the side information, given by $\bar{p}(s)$ be *arbitrary*, and let the cost measure be quadratic: $\omega(\hat{x}, s) = \hat{x}^2$ and let $W = N - D$ for some[11] $N > D$. Let $\alpha = \frac{N-D}{N}$ as before. It was shown in [19] that $C^*(N - D) = \frac{1}{2}\log\frac{N}{D}$ (no rate-loss) and that the optimizer given by the distribution, $U = \hat{X} + \alpha S$, i.e., $p^*(u|s) = \mathcal{N}(\alpha s, (N - D))$ and the function characterizing the channel input, $\hat{X} = f^*(U, S) = U - \alpha S$ induce the relation between U and X as given by $U = \alpha X + q$, where $p(q) = \mathcal{N}(0, \frac{D(N-D)}{N})$, i.e., $p^*(u|x) = \mathcal{N}\left(\alpha x, \frac{D(N-D)}{N}\right)$.

Note (as in Section 3.2) that α is equal to the coefficient in the MMSE estimate of $\hat{X}$ from $\hat{X} + Z$. We now provide an interpretation for the joint distribution depicted in Fig. 9(a) by connecting it to that depicted in the CCSIB problem in Fig. 7(b). It is observed that the former (CCSI) can be obtained from the latter (CCSIB) by moving the side-information adder (arrow labeled "-S" in Fig. 7(b)) to the right past the addition by "q" and wrapping around to the encoder with input as $U = \alpha X + q$, resulting in $\hat{X} = U - \alpha S$. This is akin to moving of the Decision Feedback Equalizer (DFE) from the receiver to the transmitter as a precoding unit in inter-symbol interference channels [29]. Note that the channel input $\hat{X}$ is statistically independent of side information S, i.e., $p^*(\hat{x}|s) = \mathcal{N}(0, N - D)$. Further, it can be verified (see the Appendix for a note regarding this) that the joint distribution satisfies the Markov chain $S \to X \to U$ (necessary condition for finding the dual SCSI). The channel output X is related to the side information S in the following way: $X = S + V$, where $p(V) = \mathcal{N}(0, N)$, i.e., $p^*(x|s) = \mathcal{N}(s, N)$.

Corollary 2: For the above CCSI, we can associate a dual SCSI (see Fig. 9(b)) given as follows: given a conditional source X characterized by $X = S + V$, where $p(V) = \mathcal{N}(0, N)$, i.e., $p^*(x|s) = \mathcal{N}(s, N)$ and side information $\bar{p}(s)$ as above (chosen arbitrarily in CCSI); the conditional distribution given by the above CCSI, $U = \alpha X + q$, i.e., $p^*(u|x) = \mathcal{N}(\alpha x, \frac{D(N-D)}{N})$ and the function $\hat{X} = f^*(U, S) = U - \alpha S$ minimize

[11] Using the correspondence as given in (36) and renaming $P = N - D$, $N = D$, we get the notation for CCSI conventionally used in the literature.

$[I(U;X) - I(U;S)]$, such that $S \to X \to U$ and $E(d(X, \hat{X}, S)) \le D$, implying $R^*(D) = C^*(N - D)$ and the distortion measure is given by

$$d(x, \hat{x}, s) = -c_2 \log \bar{p}(x|\hat{x}, s) + d_0(x, s) = (x - \hat{x} - s)^2 \tag{47}$$

by choosing $c_2 = 2D$ and $d_0(x, s) = -D\log(2\pi D)$ with distortion constraint D.

Proof: Follows from Theorem 3b and the two Markov chains.

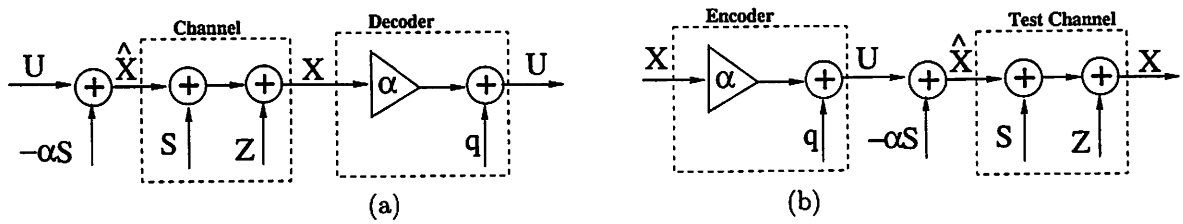


Figure 9: Duality: α, q and Z are the same as in Fig. 7. The side information S has arbitrary but fixed distribution. (a) CCSI: the cost measure is $\omega(\hat{x}, s) = \hat{x}^2$. The joint distribution is represented in this form. The decoding operation is depicted as one of recovering U from X. (b) SCSI: the distortion measure is $d(x, \hat{x}, s) = (x - \hat{x} - s)^2$. The joint distribution is represented in this form. The effect of encoding is illustrated as $\alpha X + q$.

The reconstruction $\hat{X} = U - \alpha S$ is the MMSE estimate of $(X - S)$ from U and S. It is observed that the joint distribution of SCSI as depicted in Fig. 9(b) can be obtained from that of SCSIB as depicted in Fig. 7(a), by moving the side-information adder (arrow labeled "-S") in the encoder to the right past the addition by "q", and wrapping around to the decoder with $U = \alpha X + q$, resulting in $\hat{X} = U - \alpha S$. Note that in both the problems there is no rate loss compared to the corresponding CCSIB and SCSIB even though $\bar{p}(s)$ is chosen arbitrarily. This CCSI with the given quadratic cost measure is the problem of transmission over a Gaussian channel with known interference similar to the one considered by Costa [18]. Thus, its dual is the SCSI with the decoder wishing to reconstruct $(X - S)$ under a mean squared error constraint.

As discussed in Section 3.2, if we consider the channel (see Fig. 10(a)) to be additive white Gaussian, $X = \hat{X} + Z$, i.e., $\bar{p}(x|\hat{x}, s) = \mathcal{N}(\hat{x}, D)$, with the side information being arbitrary but known and $\omega(\hat{x}, s) = (\hat{x} - s)^2$ as the cost measure with the cost constraint given by $W = N - D$, then the optimal distributions $p^*(u|s)$, $p^*(x|s)$ and $p^*(u|x)$ will remain the same as before and the channel input becomes $\hat{X} = U + (1 - \alpha)S$. The dual SCSI (see Fig. 10(b)) will have $(x - \hat{x})^2$ as the distortion measure, similar to the observation made in Remark 4, with $p^*(x|s) = \mathcal{N}(s, N)$ as the conditional source and all the dual properties as given by Theorem 3b. With this cost measure, this CCSI corresponds to the watermarking channel [16, 22], i.e., a power constraint on $(\hat{X} - S)$, and its dual is the SCSI with the decoder wishing to reconstruct X under a mean squared error constraint. In both the above pairs of dual problems, the channel of CCSI becomes the test channel of SCSI, and the encoder of

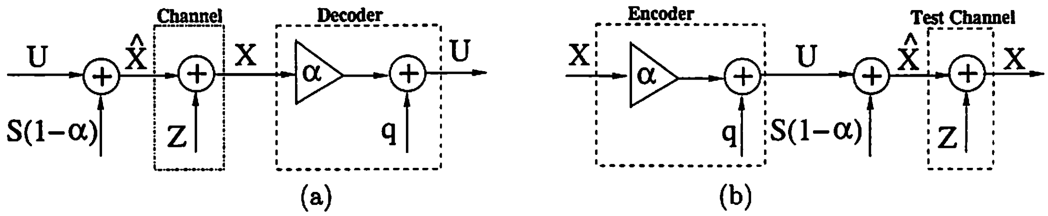


Figure 10: Duality: α, q and Z are the same as in Fig. 9. The side information S has arbitrary but fixed distribution. (a) CCSI: the cost measure is $\omega(\hat{x}, s) = (\hat{x} - s)^2$. (b) SCSI: the distortion measure is $d(x, \hat{x}, s) = (x - \hat{x})^2$.

one of the problems is functionally identical to the decoder of the dual problem. This is further illustrated in the following example, which is a special case where S is Gaussian. This is presented for completeness due to the special significance of Gaussian distribution. To break up monotony, we illustrate the duality starting from SCSI.

3.3.3 Example of Gaussian SCSI and CCSI

SCSI: given a Gaussian source X, Gaussian side information S given by: $X = S + V$, with $p(V) = \mathcal{N}(0, N)$, i.e., $\bar{p}(x|s) = \mathcal{N}(s, N)$, Gaussian side information $\bar{p}(s) = \mathcal{N}(0, Q)$ for some $N, Q > 0$, and a quadratic distortion measure $d(x, \hat{x}, s) = (x - \hat{x})^2$, the optimal conditional distribution (see Fig. 10(b) with S taken as Gaussian) satisfying (37) for $D < N$ is $U = \alpha X + q$, where $\alpha = \frac{N-D}{N}$ and $p(q) = \mathcal{N}(0, \frac{D(N-D)}{N})$, i.e., $p^*(u|x) = \mathcal{N}\left(\alpha x, \frac{D(N-D)}{N}\right)$, the reconstruction is given by $\hat{X} = U + \frac{D}{N}S$. The minimum of $[I(U; X) - I(U; S)] = \frac{1}{2}\log\frac{N}{D}$ and it can be shown that $X \to (\hat{X}, S) \to U$. There is no rate-loss compared to its corresponding SCSIB. $p^*(x|\hat{x}, s) = \mathcal{N}(\hat{x}, D)$ implying $p^*(u|s) = \mathcal{N}(\alpha s, (N - D))$.

Dual CCSI: given an additive memoryless Gaussian channel $p^*(x|\hat{x}, s) = \mathcal{N}(\hat{x}, D)$, with Gaussian side information $\bar{p}(s)$, a cost measure $\omega : \hat{\mathcal{X}} \times \mathcal{S} \to \mathbb{R}^+$ and a cost constraint W; $p^*(u|s)$ and $\hat{x} = f^*(u, s)$ maximize $[I(U; X) - I(U; S)]$ over all $p(u|s)$, $\hat{x} = f(u, s)$ such that $X \to (\hat{X}, S) \to U$, $E\omega(\hat{X}, S) \leq W$, where (see Fig. 10(a) with S taken as Gaussian) the channel input is given by $\hat{X} = U + \frac{D}{N}S$ and the cost measure is

$$\omega(\hat{x}, s) = c_1 D\left(p^*(x|\hat{x}, s) \| \bar{p}(x|s)\right) + \theta(s) = (\hat{x} - s)^2, \quad W = N - D, \tag{48}$$

if we choose $c_1 = 2N$, $\theta(s) = c_1\left[\frac{1}{2}\log\frac{eD}{N} - \frac{D}{2N}\right]$. Note again the dual roles of the encoder and the decoder of these dual SCSI and CCSI problems.

3.3.4 Example of discrete SCSI and CCSI

Here we consider a discrete example which has the following interesting characteristics:

- we find a pair of SCSIB and CCSIB (given by $SCSIB_1$ and $CCSIB_1$) to be dual problems in the sense of Theorem 2a, but their corresponding SCSI and CCSI (given by $SCSI_1$ and $CCSI_1$), although looking similar, are not duals in the sense of Theorem 3a and 3b;

- $SCSI_1$ and $CCSI_1$ have rate-loss (unlike the previous example) compared to their corresponding $SCSIB_1$ and $CCSIB_1$ respectively;
- the optimal distribution in $SCSI_1$ satisfies the two Markov chains and we find its dual CCSI given by $CCSI_2$.

$$\begin{array}{ccccc} & & SCSIB_1 & \overset{\text{dual}}{\longleftrightarrow} & CCSIB_1 \\ & & \text{rate-loss} \Downarrow & & \Downarrow \text{rate-loss} \\ CCSI_2 & \overset{\text{dual}}{\longleftrightarrow} & SCSI_1 & \overset{\text{NOT dual}}{\longleftrightarrow} & CCSI_1 \end{array}$$

$CCSI_1$: given a binary channel $\bar{p}(x|\hat{x}, s)$, characterized by $X = \hat{X} \oplus S \oplus Z$ where $\hat{X}, S$ and $Z \in \{0,1\}$, $p(S = 1) = 0.5$, $p(Z = 1) = p$ for some $0 \leq p \leq 0.5$, Z and S independent. Let the cost measure be $\omega(\hat{x}, s) = \hat{x}$ (equivalent to a constraint on the duty cycle of $\hat{X}$). The encoder wishes to communicate with $E\omega(\hat{X}, S) \leq W$ for some $0 \leq W \leq 0.5$. The capacity of this channel is given by:

Lemma 2: The capacity of above channel is $C^*(W) = k^*(W)$, where

$$k(W) = \begin{cases} h(W) - h(p), & p \leq W \leq 0.5 \\ 0, & W \leq p \end{cases} \quad \text{and } k^*(W) = \sup_{\theta, \beta} \left[\theta k(\beta)\right], \tag{49}$$

for $0 \leq \theta \leq 1$, $p \leq \beta \leq 0.5$ and $W = \theta\beta$.

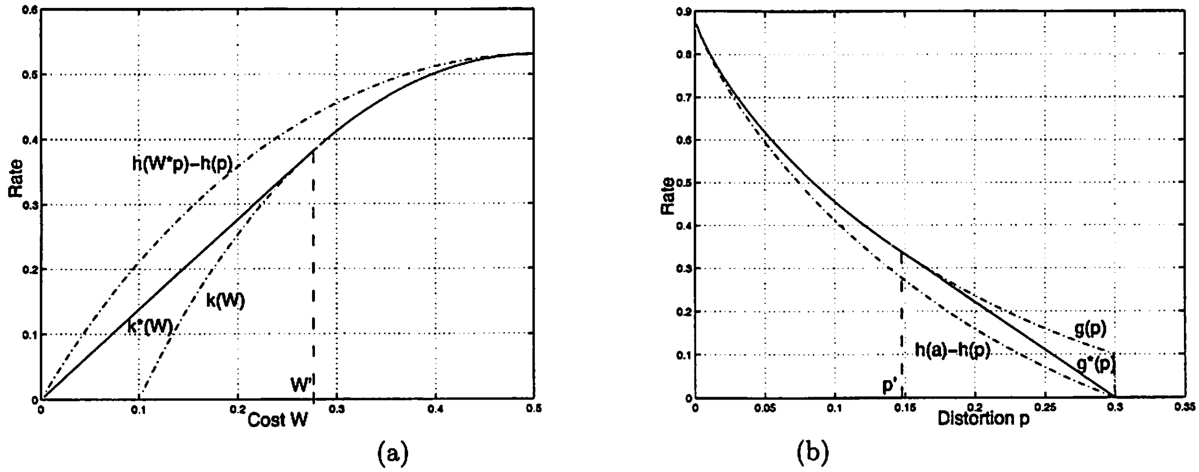
Proof: See Appendix

This is essentially the concave hull of the function $k(W)$. The capacity is plotted as a function of W in Fig. 11(a). When $W \geq W'$ (see Fig. 11(a)), $k^*(W) = k(W)$ and we have $U = \hat{X} \oplus S$, $p^*(\hat{X} = 1) = W$ and $\hat{X}$ is independent of S thus implying $\hat{X} = U \oplus S$. For $W < W'$, we need time-sharing to achieve the optimal points. The joint distribution *does not* satisfy $S \to X \to U$. Thus it is not possible to find a dual SCSI problem.

$CCSIB_1$: The capacity of the corresponding channel when both the encoder and the decoder have access to the side information is given by $C_{x|s}(W) = h(p * W) - h(p)$.

$SCSI_1$ [1]: Consider [1] a doubly symmetric binary source $\{X, S\}$ such that $X, S \in \{0,1\}$, $p(S = 1) = 0.5$, and $X = S \bigoplus Z$ where $Z \in \{0,1\}$ and independent of X and $P[Z = 1] = a$ for some $0 \leq a \leq 0.5$. The encoder wishes to represent X such that $E[d(X, \hat{X}, S)] \leq p$, where $0 \leq p \leq 0.5$ and $d(x, \hat{x}, s) = x \oplus \hat{x}$. The rate-distortion function [1] is $R^*(p) = g^*(p)$ where

$$g(p) = \begin{cases} h(a * p) - h(p) & 0 \leq p < a \\ 0 & p \geq a \end{cases} \quad \text{and } g^*(p) = \inf_{\theta, \beta} \left[\theta g(\beta)\right], \tag{50}$$



(a) (b)

Figure 11: (a) Capacity-cost function for the discrete channel with side information with $p = 0.1$. Solid line: capacity of $CCSI_1$. Upper dotted line: capacity of $CCSIB_1$. (b) Rate-distortion function of doubly symmetric binary random variable with $a = 0.3$. Solid line: rate of $SCSI_1$. Lower dotted line: rate of $SCSIB_1$.

for $0 \leq \theta \leq 1$ and $0 \leq \beta < a$ and $p = \theta\beta + (1-\theta)a$. Thus the rate-distortion function is the convex hull of the function $g(p)$. The rate-distortion function is plotted as a function of p in Fig. 11(b). When $p \leq p'$ (see Fig. 11(b)), $g^*(p) = g(p)$ and we have $U = X \oplus Q$ where $p^*(Q = 1) = p$ and Q is independent of X and $\hat{X} = U$. Further, the joint distribution satisfies $X \to (\hat{X}, S) \to U$.

$SCSIB_1$: The rate-distortion function when the side information is available at both the encoder and decoder is given by $R_{x|s}(p) = h(a) - h(p)$ if $0 \leq p \leq a$ and 0 if $p > a$.

Note that for a channel noise p, and cost W, the capacity-cost function of $CCSI_1$ is $h(W) - h(p)$ (which for $CCSIB_1$ is $h(p * W) - h(p)$) and for a distortion p and correlation noise $a = p * W$, the rate-distortion bound for $SCSI_1$ is $h(p*p*W) - h(p)$ (which for $SCSIB_1$ is $h(p*W) - h(p)$). Thus $SCSIB_1$ and $CCSIB_1$ are duals in the sense of Theorem 2a but the corresponding $SCSI_1$ and $CCSI_1$ are not duals in the sense of Theorem 3a.

$CCSI_2$: The given $SCSI_1$ does satisfy $X \to (\hat{X}, S) \to U$, and we can associate with it a dual CCSI as given by Theorem 3a with the same joint distributions: with the channel given by $p^*(x|\hat{x}, s)$. We consider a numerical example for illustration.

Example 4: Let $a = 0.3$ and $p = 0.1$ and the minimization in $SCSI_1$ induces the following distribution: $p^*(U = 1|S = 0) = p^*(U = 0|S = 1) = 0.34$, and $p^*(X = 0|\hat{X}, S = 0, 0) = 1 - p^*(X = 0|\hat{X}, S = 1, 1) = 0.95$, $p^*(X = 0|\hat{X}, S = 0, 1) = 1 - p^*(X = 0|\hat{X}, S = 1, 0) = 0.79$. The cost measure (using Theorem 3a) is $\omega(\hat{x}, s) = \hat{x} \oplus s$, where we choose $c_1 = 2.222$ and $\theta(s) = -0.642$. Note that the Hamming distortion measure $(x \oplus \hat{x})$ in $SCSI_1$ corresponds to the Hamming cost measure $(\hat{x} \oplus s)$ in the dual $CCSI_1$. The capacity is 0.456 bits/sample, which is equal to the rate of transmission for dual $SCSI_1$. Further, by following Remark 7, for

$SCSI_1$ with the same source $\bar{p}(x|s)$ and side information $\bar{p}(s)$ we can find a distortion measure d' (which results in the same joint distribution $p^*(x,s,\hat{x},u)$) such that there is no rate-loss. By choosing $c_2 = 1$ and $d_0(x,s) = 0$ in (46), this is given as follows: $d'(x,\hat{x},s) = d'(1\oplus x, 1\oplus\hat{x}, 1\oplus s)$ and

$x\hat{x}s$	$d'(x,\hat{x},s)$	$x\hat{x}s$	$d'(x,\hat{x},s)$	$x\hat{x}s$	$d'(x,\hat{x},s)$	$x\hat{x}s$	$d'(x,\hat{x},s)$
000	0.067	001	0.333	010	2.280	011	4.459

4 Conclusions

We have considered a mathematical formulation of duality between source and channel coding with side information. Our notion of duality is in the functional sense, i.e. where the encoder/decoder functional mappings can be exactly swapped in the dual problems. We have, as a necessary first step, first addressed the well-known duality between conventional source and channel coding and then extended this notion to the more general case of coding with side information. We have given several examples for illustration. By using this concept we have obtained a generalization of the Wyner-Ziv result relating to no rate-loss in source coding with side information (with respect to the side information being available at both ends) from Gaussian to arbitrary distributions. A key observation that follows from our duality treatment is that contrary to popular opinion, duality does not equate to no rate-loss in coding with side information. The main motivation for understanding the duality between these two important problems is to enable efficient constructions of the encoder and the decoder based on the framework of structured coset codes. Our goal is to inspire dual constructions for these two problems that can leverage this functional encoder-decoder duality, and allow progress made in one field to apply naturally to the other as well. This would have an impact on a wide range of emerging applications including sensor networks, data hiding, broadcast and multi-antenna communication systems.

Acknowledgments

We would like to thank D. Tse, P. Viswanath and M. Gastpar for inspiring discussions and comments regarding the paper.

Appendix

Proof of Theorem 3a: The outcome of the given source coding problem is $p^*(u|x)$ and $\hat{x} = f^*(u,s)$ such that $S \to X \to U$ and $X \to (\hat{X}, S) \to U$, which defines a joint distribution $p^*(x,s,u,\hat{x})$. Consider the channel $p^*(x|\hat{x},s)$ (derived from the joint distribution) with input $\hat{X}$ and output X, and side information S present at both the encoder and the decoder (CCSIB). For this channel, given the input distribution $p^*(\hat{x}|s)$ (derived from the above joint distribution), $\exists$ a cost measure (using Lemma 1a) given by: $\omega(\hat{x}, s) = c_1 D(p^*(x|\hat{x}, s)\|\bar{p}(x|s)) + \theta(s)$, such that

$$p^*(\hat{x}|s) = \underset{p(\hat{x}|s) : (X|\hat{X}, S) \sim p^*(x|\hat{x}, s), S \sim \bar{p}(s), E\omega(\hat{X}, S) \leq W}{\text{argmax}} I(X; \hat{X}|S), \tag{51}$$

where $E_{p^*(\hat{x}|s)\bar{p}(s)}(\omega(\hat{X}, S)) = W$. Now clearly, the joint distribution $p^*(x, s, \hat{x})$ (obtained from the above $p^*(x, s, \hat{x}, u)$) can be represented in the form $p^*(x, s, \hat{x}, u)$ such that $\hat{x} = f^*(u, s)$ and $S \rightarrow X \rightarrow U$ and $X \rightarrow (\hat{X}, S) \rightarrow U$, the condition required for no rate-loss. Thus the capacity of this channel coding problem is the same even if the side information is present only at the encoder. Hence, $p^*(u|s)$ and $\hat{x} = f^*(u, s)$ derived from the above joint distribution maximize $[I(U; X) - I(U; S)]$ under the constraint $(X|\hat{X}, S) \sim p^*(x|\hat{x}, s)$, $S \sim \bar{p}(s)$, $X \rightarrow (\hat{X}, S) \rightarrow U$ and $E\omega(\hat{X}, S) \leq W$.

Markov constraint in Section 3.3.2 [19]: Using the induced joint distributions, we note that $[I(U; X) - I(U; S)] = h(\hat{X}|S) - h(\hat{X} + \alpha S|X)$. Now $h(\hat{X} + \alpha S|X) = h(\hat{X} - \alpha(\hat{X} + Z)|X)$. Since $\hat{X} - \alpha(\hat{X} + Z)$ is independent of $\hat{X} + Z$ and S, we can see that $h(\hat{X} + \alpha S|X) = h(\hat{X}|S, X)$ implying that $[I(U; X) - I(U; S)] = I(\hat{X}; X|S)$. Thus using this and the Markov condition in the definition of the problem, $U \rightarrow (\hat{X}, S) \rightarrow X$, we have $I(U; X) - I(U; S) = I(U; X|S)$ which implies $S \rightarrow X \rightarrow U$.

Proof of Lemma 2 : Clearly $k(W)$ is a monotone non-decreasing function of W. For $p \leq W \leq 0.5$, it is a concave function of W. First we prove that $C^*(W) \geq k^*(W)$. Let $U = \hat{X} \bigoplus S$ where $\hat{X}$ and S are independent and $P[\hat{X} = 1] = \beta$ which implies that $\hat{X} = U \bigoplus S$. Now

$$I(U, X) - I(U, S) = (1 - h(p)) - (1 - h(\beta)) = h(\beta) - h(p). \tag{52}$$

Thus $C^*(\beta) \geq h(\beta) - h(p)$ for $p \leq \beta \leq 0.5$. It can be noted that $C^*(0) \geq 0$. Let $0 \leq W \leq 0.5$ be given, and θ and β are given such that $W = \theta\beta$ and $0 \leq \theta \leq 1$. Thus

$$C^*(W) = C^*(\theta\beta) \geq \theta C^*(\beta) + (1 - \theta)C^*(0) \geq \theta[h(\beta) - h(p)]. \tag{53}$$

We maximize the right hand side to get the tightest lower bound.

Now we prove that $C^*(W) \leq k^*(W)$. We use techniques similar to those given in [1]. Let U, S be the random variables and $\hat{X} = f(U, S)$ such that their joint distribution satisfy the constraint: $E[\omega(\hat{X}, S)] \leq W$. Let the set A be defined as: $A = \{u : f(0, u) = f(1, u)\}$. Now let $d_u = E[\hat{X}|U = u] = P[\hat{X} = 1|U = u]$ for $u \in U$. Consider $u \in A^c$, if $f(0, u) = 0$ and $f(1, u) = 1$, then $X = 0 \bigoplus Z = Z$ and if $f(0, u) = 1$ and $f(1, u) = 0$ then $X = 1 \bigoplus Z$.

Thus for a $u \in A^c$, $H(X|U=u) = h(d_u)$ and $H(X|U=u) = h(p)$. Now consider $u \in A$. Let β_u be such that $0 \le \beta_u \le 1$ and $h(\beta_u) = H(S|U=u)$. If $f(0,u) = f(1,u) = 0$ then $X = S \bigoplus Z$ and if $f(0,u) = f(1,u) = 1$ then $X = 1 \bigoplus S \bigoplus Z$. Thus $H(X|U=u) = h(\beta_u * p)$. Since $H(S) = 1$, we have

$$I(U;X) - I(U;S) = H(X) - H(S) - H(X|U) + H(S|U) \le H(S|U) - H(X|U) \tag{54}$$

$$= \sum_{u \in A} P[U=u]\,[H(S|U=u) - H(X|U=u)] + \sum_{u \in A^c} P[U=u]\,[H(S|U=u) - H(X|U=u)] \tag{55}$$

$$= (1-\theta) \sum_{u \in A^c} \lambda_u [h(d_u) - h(p)] + \sum_{u \in A} P[U \in u][h(\beta_u) - h(\beta_u * p)] \le (1-\theta) \sum_{u \in A^c} \lambda_u [h(d_u) - h(p)], \tag{56}$$

where $\theta = P[U \in A]$ and $\lambda_u = P[U=u]/(1-\theta)$ and $h(\beta_u) - h(\beta_u * p) \le 0$ for $1 \ge p \ge 0$. Since $0 \le \lambda_u \le 1$ and $\sum_{u \in A^c} \lambda_u = 1$, we have,

$$I(U;X) - I(U;S) \le (1-\theta) h \left[\sum_{u \in A^c} d_u \lambda_u \right] - (1-\theta) h(p) = (1-\theta)[h(\beta) - h(p)], \tag{57}$$

where $\beta = \sum_{u \in A^c} d_u \lambda_u$. Clearly $0 \le I(U;X) - I(U;S)$, which implies that $\beta \ge p$ as $0 \le p \le 0.5$. We have

$$W' = (1-\theta)\beta = \sum_{u \in A^c} P[U=u] P[\hat{X} = 1 | U = u] \le E[\hat{X}] \le W. \tag{58}$$

Thus we have shown that $I(U;X) - I(U;S) \le \sup_{\theta,\beta} \theta[h(\beta) - h(p)] = k^*(W')$, such that $p \le \beta \le 0.5$ and $W' = \theta\beta$ Since $k^*(W)$ is monotone in W, we have proved that $C^*(W) \le k^*(W)$.

References

[1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, January 1976.

[2] D. P. Bertsekas, *Nonlinear Programming.* Athena Scientific, 1999.

[3] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *In IRE Nat. Conv. Rec.*, vol. Part 4, pp. 142–163, 1959.

[4] T. M. Cover and J. A. Thomas, *Elements of information theory.* New York: Wiley, 1991.

[5] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, July 1972.

[6] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. on Comm.*, vol. 38, pp. 82–93, Jan. 1990.

[7] M. V. Eyuboglu and G. D. Forney, "Lattice and trellis quantization with lattice and trellis-bounded codebooks- High rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 46–59, Jan. 1993.

[8] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. on Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[9] R. Laroia, N. Farvardin, and S. A. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Trans. on Inform. Theory*, vol. IT-40, pp. 1044–1056, July 1994.

[10] R. Laroia and N. Farvardin, "A structured fixed-rate vector quantizer derived from a variable length scalar quantizer part I: memoryless sources," *IEEE Trans. on Inform. Theory*, vol. IT-39, pp. 851–867, May 1993.

[11] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, pp. 19–31, 1980.

[12] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code," *Proc. IEEE International Symposium on Information Theory (ISIT), Sorrento, Italy*, June 2000.

[13] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code," *submitted to IEEE Trans. Inform. Theory, (preprint)*, May 2001.

[14] T. Berger, "Multiterminal source coding," *in The information theory approach to communication, (CISM Courses and Lecture Notes No. 229), G. Longo, Ed., Wien and New York: Springer-Verlag.*, 1977.

[15] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, March 1998.

[16] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger, "Information theoretic analysis of steganography," *Proc. Int. Symp. on information Theory (ISIT), Cambridge, MA*, August 1998.

[17] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. on Inform. Theory*, vol. IT-25, pp. 306–311, May 1979.

[18] M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, pp. 439–441, May 1983.

[19] A. Cohen and A. Lapidoth, "The Gaussian watermarking game- Part I," *submitted to IEEE Trans. Inform. Theory*, 2000.

[20] J. Chou, S. S. Pradhan, and K. Ramchandran, "On the duality between distributed source coding and data hiding," *Proc. of 33rd Asilomar Conf. on Signals, Systems and Computers*, November 1999.

[21] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. on Information Theory*, vol. 29, pp. 731–739, September 1983.

[22] B. Chen and G. W. Wornell, "An information-theoretic approach to the design of robust digital watermarking systems," *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, March 1999.

[23] M. Kesal, M. K. Mihcak, R. Kotter, and P. Moulin, "Iteratively decodable codes for watermarking applications," *Proc. 2nd Int. symp. on Turbo codes and related topics, Brest, France*, September 2000.

[24] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," in *ISITA 2000, Honalulu, Hawaii*, November 2000.

[25] R. Zamir and S. Shamai, "Nested linear/ lattice codes for Wyner-Ziv encoding," in *IEEE Information theory workshop, Killarney, Ireland*, 1998.

[26] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS) : Design and construction," *Proc. IEEE Data Compression Conf., Snowbird, UT*, March 1999.

[27] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 480–489, July 1973.

[28] M. Chiang and T. M. Cover, "Unified duality between channal capacity and rate distortion with side information," in *Proc. Int. Symp. on Info. Theory (ISIT) 2001, Washington D.C,*, June 2001.

[29] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electr. Let.*, vol. 7, pp. 138–139, Mar. 1971.