Minimum Expected Distortion in Gaussian Source Coding with Uncertain Side Information

Chris T. K. Ng*, Chao Tian†, Andrea J. Goldsmith* and Shlomo Shamai (Shitz)‡

*Dept. of Electrical Engineering, Stanford University, Stanford, CA 94305 USA
†School of Computer and Communication Sciences, LICOS, EPFL, Lausanne, Switzerland
‡Dept. of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel
Email: *{ngctk,andrea}@wsl.stanford.edu, †chao.tian@epfl.ch, ‡sshlomo@ee.technion.ac.il

***Abstract*— We consider a layered approach to source coding with side information received over an uncertain channel that minimizes expected distortion. Specifically, we assume a Gaussian source encoder whereby the decoder receives a compressed version of the symbol at a given rate, as well as an uncompressed version over a separate side-information channel with slow fading and noise. The decoder knows the realization of the slow fading but the encoder knows only its distribution. We consider a layered encoding strategy with a base layer describing the source assuming worst-case fading on the side-information channel, and subsequent layers describing the source under better fading conditions. Optimization of the layering scheme utilizes the Heegard-Berger rate-distortion function that describes the rate required to meet a different distortion constraint for each fading state. When the side-information channel has two discrete fading states, we obtain closed-form expressions for the optimal rate allocation between the fading states and the resulting minimum expected distortion. For multiple fading states, the minimum expected distortion is formulated as the solution of a convex optimization problem. Under discretized Rayleigh fading, we show that the optimal rate allocation puts almost all rate into the base layer associated with the worst-case fading. This implies that uncertain side information yields little performance benefit over no side information. Moreover, as the source coding rate increases, the benefit of uncertain side-information decreases.**

I. Introduction

In lossy data compression, side information at the decoder can help reduce the distortion in the reconstruction of the source [1]. However, in scenarios such as distributed compression in a wireless sensor network, the side information may be acquired over an unreliable wireless channel. In this work we consider a Gaussian source where the encoder is subject to a rate constraint and the distortion metric is mean squared error. In addition to the compressed symbol, we assume that the decoder observes the original symbol through a separate analog fading channel. We assume, similar to the approach in [2], that the fading is quasi-static, and that the decoder knows the fading realization but the encode knows only its distribution. The rate-distortion function that dictates the rate required to satisfy the distortion constraint associated with each fading state is given by Heegard and Berger in [3]. In this work we consider a layered encoding strategy based on the uncertain fading in the side-information channel, and optimize the rate allocation among the possible fading states to minimize expected distortion.

When the side-information channel exhibits no fading, the distortion is given by the Wyner-Ziv rate-distortion function [4]. Rate-distortion is considered in [5], [6] when the side information is also available at the encoder, and in [7] when there is a combination of decoder-only and encoder-and-decoder side information. Successive refinement source coding in the presence of side information is considered in [8]. In [9], [10], expected distortion is minimized in the transmission of a Gaussian source over a slowly fading channel in the absence of channel state information at the transmitter (CSIT). Another application of source coding with uncertain side information is in systematic lossy source-channel coding [11] over a fading channel without CSIT: For example, when upgrading legacy communication systems, a digital channel may be added to augment an existing analog channel. In this case the analog reception then plays the role of side information in the decoding of the description from the digital channel.

The remainder of the paper is organized as follows. The system model is presented in Section II. Section III derives the minimum expected distortion when the side-information channel has discrete fading states. Section IV presents numerical results under discretized Rayleigh fading. Section V considers continuous fading distributions, followed by conclusions in Section VI.

II. System Model

A. Source Coding with Uncertain Side Information

Consider the system model shown in Fig. 1. An encoder wishes to describe a real Gaussian source sequence $\{X\}$ under a rate constraint of R_X bits per symbol, where the sequence of random variables are independent identically distributed (iid) with $X \sim \mathcal{N}(0, \sigma_X^2)$. The decoder, in addition to receiving the encoder's description, observes side information Y', where $Y' = \sqrt{S}X + Z$, with $Z \sim$ iid $\mathcal{N}(0, 1)$. Hence the quality of the side information depends on S, the power gain of the side-information channel. We assume S is a quasi-static random variable that

This work was supported by the US Army under MURI award W911NF-05-1-0246, the ONR under award N00014-05-1-0168, DARPA under grant 1105741-1-TFIND, and a grant from Intel.

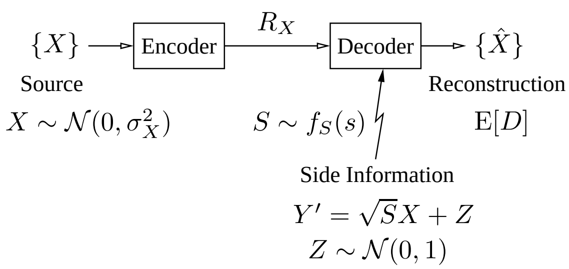

Fig. 1. Source coding with uncertain side information.

is unchanged after its realization. The decoder knows the realization of S, but the encoder knows only its distribution given by the probability density function (pdf) $f_S(s)$. The decoder forms an estimate of the source and reconstructs the sequence $\{\hat{X}\}$. We are interested in minimizing the expected squared error distortion $\mathrm{E}[D]$ of the reconstruction, where $D = (X - \hat{X})^2$.

Suppose the side-information channel has M discrete fading states. Let the probability distribution of S be given as follows:

$$\Pr\{S = s_i\} = p_i, \quad i = 1, \ldots, M; \qquad \sum_{i=1}^{M} p_i = 1, \quad (1)$$

where the s_i's are enumerated in ascending order $s_1 < s_2 < \cdots < s_M$. Let Y_i' denote the side information under fading state s_i:

$$Y_i' \triangleq \sqrt{s_i} X + Z, \qquad i = 1, \ldots, M. \quad (2)$$

Note that the set of side information random variables are stochastically degraded. Let $\hat{X}_i$ be the reconstruction when side information Y_i' is available at the decoder, and D_i be the corresponding squared error distortion. The minimum expected distortion under rate constraint R_X is then given by

$$\mathrm{E}[D]^* = \min_{\mathbf{D}\,:\, R(\mathbf{D}) \le R_X} \mathbf{p}^T \mathbf{D}, \quad (3)$$

where $\mathbf{p} \triangleq \left[p_1 \ldots p_M\right]^T$, $\mathbf{D} \triangleq \left[D_1 \ldots D_M\right]^T$, and $R(\mathbf{D})$ is the rate-distortion function that simultaneously satisfies the distortion set $\mathbf{D}$.

B. Heegard-Berger Rate-Distortion Function

The rate-distortion function that dictates the rate required to simultaneously satisfy a set of distortion constraints associated with a set of degraded side-information random variables is given by Heegard and Berger in [3] (an alternate form for $M = 2$ is described in [12]). When the side information random variables satisfy the degradedness condition $X \leftrightarrow Y_M \leftrightarrow Y_{M-1} \leftrightarrow \cdots \leftrightarrow Y_1$, the rate-distortion function is

$$R_{\mathrm{HB}}(\mathbf{D}) = \min_{W_1^M \in P(\mathbf{D})} \sum_{i=1}^{M} I(X; W_i | Y_i, W_1^{i-1}), \quad (4)$$

where W_1^i denotes the vector $W_1, \ldots, W_i$. The minimization takes place over $P(\mathbf{D})$, the set of all W_1^M jointly distributed with X, Y_1^M such that:

$$W_1^M \leftrightarrow X \leftrightarrow Y_M \leftrightarrow Y_{M-1} \leftrightarrow \cdots \leftrightarrow Y_1, \quad (5)$$

and there exists decoding functions $\hat{X}_i(Y_i, W_1^i)$'s under given distortion measures d_i's that satisfy

$$\mathrm{E}[d_i(X, \hat{X}_i)] \le D_i, \qquad i = 1, \ldots, M. \quad (6)$$

As noted in [3], since $R_{\mathrm{HB}}(\mathbf{D})$ depends on X, Y_1^M only through the marginal distribution $p(x, y_i)$, $i = 1, \ldots, M$, the degradedness of the side information need not be physical. We construct Y_1^M to have the same marginals as ${Y'}_1^M$ by setting $p(y_i|x) = p(y_i'|x)$, $i = 1, \ldots, M$. The rate-distortion function $R(\mathbf{D})$ in (3) is then given by the Heegard-Berger rate-distortion function (4) with squared error distortion measures $d_i(X, \hat{X}_i) = (X - \hat{X}_i)^2$.

III. Minimum Expected Distortion

A. Gaussian Source under Squared Error Distortion

First we consider the case when the side-information channel has only two discrete fading states ($M = 2$). The Heegard-Berger rate-distortion function for this case is

$$R_{\mathrm{HB}}(D_1, D_2) = \min_{W_1, W_2 \in P(D_1, D_2)} \{I(X; W_1|Y_1) + I(X; W_2|Y_2, W_1)\}. \quad (7)$$

For a Gaussian source under a squared error distortion measure, a jointly Gaussian codebook is optimal [3], [13]. When W_1^M, X are jointly Gaussian, the mutual information expressions in (7) evaluate to

$$\begin{aligned} &I(X; W_1|Y_1) + I(X; W_2|Y_2, W_1) \\ &= h(X|Y_1) - h(X|Y_1, W_1) \\ &\quad + h(X|Y_2, W_1) - h(X|Y_2, W_1, W_2) \quad (8) \\ &= \tfrac{1}{2}\log(\mathrm{VAR}[X|Y_1]) - \tfrac{1}{2}\log\frac{\mathrm{VAR}[X|Y_1, W_1]}{\mathrm{VAR}[X|Y_2, W_1]} \\ &\quad - \tfrac{1}{2}\log(\mathrm{VAR}[X|Y_2, W_1, W_2]) \quad (9) \\ &= -\tfrac{1}{2}\log(s_1 + \sigma_x^{-2}) \\ &\quad - \tfrac{1}{2}\log\bigl(1 + (s_2 - s_1)\mathrm{VAR}[X|Y_1, W_1]\bigr) \quad (10) \\ &\quad - \tfrac{1}{2}\log(\mathrm{VAR}[X|Y_2, W_1, W_2]), \end{aligned}$$

where $\log$ is base 2, and (10) follows from expanding the conditional variance expressions by applying Lemma 1 and Corollary 1 as given below.

Lemma 1: Let X, W_1^k be jointly Gaussian random variables. If $Y = \sqrt{s}X + Z$, where $Z \sim \mathcal{N}(0, 1)$ is independent from X, W_1^k, then

$$\mathrm{VAR}[X|Y, W_1^k] = \left(\mathrm{VAR}[X|W_1^k]^{-1} + s\right)^{-1}. \quad (11)$$

Proof: The lemma follows from the minimum mean square error (MMSE) estimate of Gaussian random variables. Let $X, \mathbf{W}$, where $\mathbf{W} \triangleq \left[W_1 \ldots W_k\right]^T$, be distributed as

$$\begin{bmatrix} \mathbf{W} \\ X \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{W}} \\ \mu_X \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{W}} & \boldsymbol{\Sigma}_{\mathbf{W}X} \\ \boldsymbol{\Sigma}_{\mathbf{W}X}^T & \sigma_X^2 \end{bmatrix}\right). \quad (12)$$

The conditional distribution is Gaussian [14], and the corresponding variance is

$$\begin{aligned}\mathrm{VAR}[X|Y,\mathbf{W}] &= \sigma_X^2 \\ &- \begin{bmatrix}\boldsymbol{\Sigma}_{\mathbf{W}X} \\ \sqrt{s}\sigma_X^2\end{bmatrix}^T \begin{bmatrix}\boldsymbol{\Sigma}_{\mathbf{W}} & \sqrt{s}\boldsymbol{\Sigma}_{\mathbf{W}X} \\ \sqrt{s}\boldsymbol{\Sigma}_{\mathbf{W}X}^T & s\sigma_X^2+1\end{bmatrix}^{-1} \begin{bmatrix}\boldsymbol{\Sigma}_{\mathbf{W}X} \\ \sqrt{s}\sigma_X^2\end{bmatrix} \quad (13)\\ &= \frac{\sigma_X^2 - \boldsymbol{\Sigma}_{\mathbf{W}X}^T \boldsymbol{\Sigma}_{\mathbf{W}}{}^{-1}\boldsymbol{\Sigma}_{\mathbf{W}X}}{1+s(\boldsymbol{\Sigma}_{\mathbf{W}X}^T \boldsymbol{\Sigma}_{\mathbf{W}}{}^{-1}\boldsymbol{\Sigma}_{\mathbf{W}X})} \quad (14)\\ &= \left(\mathrm{VAR}[X|\mathbf{W}]^{-1}+s\right)^{-1}. \quad (15)\end{aligned}$$

□

Corollary 1: Let $Y_j = \sqrt{s_j}X + Z$, $Y_i = \sqrt{s_i}X + Z$.

$$\frac{\mathrm{VAR}[X|Y_i, W_1^k]}{\mathrm{VAR}[X|Y_j, W_1^k]} = 1 + (s_j - s_i)\mathrm{VAR}[X|Y_i, W_1^k]. \quad (16)$$

□

We substitute (10) in (7), and minimize over W_1, W_2 to obtain:

$$\begin{aligned}R_{\mathrm{HB}}(D_1,D_2) &= -\tfrac{1}{2}\log(s_1+\sigma_x^{-2}) \\ &+ \min_{W_1}\Big\{-\tfrac{1}{2}\log\big(1+(s_2-s_1)\mathrm{VAR}[X|Y_1,W_1]\big) \quad (17)\\ &+ \min_{W_2}\big\{-\tfrac{1}{2}\log(\mathrm{VAR}[X|Y_2,W_1,W_2])\big\}\Big\}.\end{aligned}$$

In the inner minimization in (17), $R_{\mathrm{HB}}(D_1, D_2)$ is decreasing in $\mathrm{VAR}[X|Y_2, W_1, W_2]$; hence the choice of W_2 is optimal when

$$\begin{aligned}&\max_{W_2} \mathrm{VAR}[X|Y_2,W_1,W_2] \\ &\quad = \min(\mathrm{VAR}[X|Y_2,W_1],\, D_2),\end{aligned} \quad (18)$$

where the first term in the $\min(\cdot)$ expression follows from the non-negativity of mutual information $I(X; W_2|Y_2, W_1)$, and the second one follows from the distortion constraint:

$$\begin{aligned}&\mathrm{VAR}[X|Y_2,W_1,W_2] \\ &\quad = \mathrm{E}\big[\big(X-\hat{X}_2(Y_2,W_1,W_2)\big)^2\big] \le D_2.\end{aligned} \quad (19)$$

Similarly, in the outer minimization in (17), W_1 is optimal when

$$\max_{W_1} \mathrm{VAR}[X|Y_1,W_1] = \min(\mathrm{VAR}[X|Y_1],\, D_1), \quad (20)$$

which follows from the non-negativity of $I(X; W_1|Y_1)$, and the distortion constraint:

$$\mathrm{VAR}[X|Y_1,W_1] = \mathrm{E}\big[\big(X-\hat{X}_1(Y_1,W_1)\big)^2\big] \le D_1. \quad (21)$$

Next, we consider the construction of W_1, W_2 that achieve the rate-distortion function, namely jointly Gaussian random variables with conditional variances that satisfy (18), (20). We construct W_1, W_2 as given by

$$W_1 = a_1X + N_1, \qquad W_2 = a_2X + N_2, \quad (22)$$

where $N_i \sim$ iid $\mathcal{N}(0, 1)$, $i = 1, 2$, is independent from X, Y_1, Y_2. For notational convenience, we define

$$R_1 \triangleq \min_{W_1} I(X;W_1|Y_1) \quad (23)$$

$$R_2 \triangleq \min_{W_2} I(X;W_2|Y_2,W_1). \quad (24)$$

We interpret R_1 as the rate of a source coding base layer that describes X when the side-information quality is that of Y_1 or better. On the other hand, R_2 is the rate of a top layer that describes X only when the decoder has the better side information Y_2.

The rate of the base layer under optimal W_1 is given by

$$\begin{aligned}R_1 &= \min_{W_1}\{h(X|Y_1) - h(X|Y_1,W_1)\} \quad (25)\\ &= \tfrac{1}{2}\log\frac{(\sigma_X^{-2}+s_1)^{-1}}{\tilde{D}_1}, \quad (26)\end{aligned}$$

where

$$\tilde{D}_1 \triangleq \min\big(D_1,\, (\sigma_X^{-2}+s_1)^{-1}\big). \quad (27)$$

The a_1 that achieves (26) is determined from the constraint $\tilde{D}_1 = \mathrm{VAR}[X|Y_1, W_1]$, which evaluates to

$$a_1 = \tilde{D}_1^{-1} - \sigma_X^{-2} - s_1. \quad (28)$$

Similarly, under optimal W_2, the rate of the top layer is

$$\begin{aligned}R_2 &= \min_{W_2}\{h(X|Y_2,W_1) - h(X|Y_2,W_1,W_2)\} \quad (29)\\ &= \tfrac{1}{2}\log\frac{(\tilde{D}_1^{-1}+s_2-s_1)^{-1}}{\tilde{D}_2}, \quad (30)\end{aligned}$$

where

$$\tilde{D}_2 \triangleq \min\big(D_2,\, (\tilde{D}_1^{-1}+s_2-s_1)^{-1}\big). \quad (31)$$

The a_2 that achieves (30) is determined from $\tilde{D}_2 = \mathrm{VAR}[X|Y_2, W_1, W_2]$, which evaluates to

$$a_2 = \tilde{D}_2^{-1} - \tilde{D}_1^{-1} - (s_2 - s_1). \quad (32)$$

Finally, we substitute (26), (30) in (7) to obtain the rate-distortion function:

$$\begin{aligned}R_{\mathrm{HB}}(D_1,D_2) &= R_1 + R_2 \quad (33)\\ &= -\tfrac{1}{2}\log(\sigma_X^{-2}+s_1) - \tfrac{1}{2}\log\tilde{D}_2 \\ &\quad - \tfrac{1}{2}\log\big(1+(s_2-s_1)\tilde{D}_1\big), \quad (34)\end{aligned}$$

where $\tilde{D}_1, \tilde{D}_2$ are as defined in (27), (31). Note that the derivation of (34) depends on the side information only through the marginals $p(y_i|x)$'s; therefore, the rate-distortion function applies as well to the stochastically degraded side information $Y'_M, \ldots, Y'_1$.

B. Optimal Distortion Trade-off and Rate Allocation

Under a source coding rate constraint of R_X, the achievable distortion set is $\{(D_1, D_2) \mid R_{\mathrm{HB}}(D_1, D_2) \le R_X\}$. Setting $R_{\mathrm{HB}}(D_1, D_2) = R_X$, the boundary of $\{(D_1, D_2)\}$ defines the Pareto optimal trade-off curve between the two distortion constraints, which is given by

$$D_2 = \big[2^{2R_X}(\sigma_X^{-2}+s_1)\big(1+(s_2-s_1)D_1\big)\big]^{-1}, \quad (35)$$

over the interval:

$$\big(2^{2R_X}(\sigma_X^{-2}+s_1)\big)^{-1} \le D_1 \le (\sigma_X^{-2}+s_1)^{-1}. \quad (36)$$

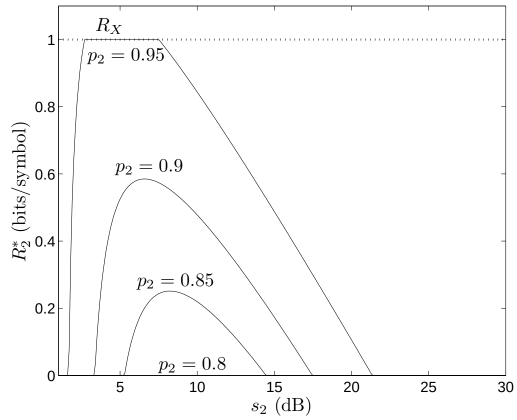

Fig. 2. Optimal rate allocation that minimizes expected distortion. The side-information channel has two discrete fading states ($s_1 = 0$ dB).

We find the optimal operating point on the Pareto curve to minimize the expected distortion:

$$\mathrm{E}[D]^* = \min_{D_1, D_2 :\, R_{\mathrm{HB}}(D_1, D_2) \le R_X} p_1 D_1 + p_2 D_2 \tag{37}$$

After substituting (35) in (37), from the Karush-Kuhn-Tucker (KKT) optimality conditions we obtain the optimal base layer distortion D_1^*:

$$D_1^* = \min\big(\max(D_1^-,\, D_1^\star),\, D_1^+\big), \tag{38}$$

where

$$D_1^- = \left(2^{2R_X}(\sigma_X^{-2} + s_1)\right)^{-1} \tag{39}$$

$$D_1^\star = \frac{1}{s_2 - s_1}\left[\left(2^{2R_X}\frac{\sigma_X^{-2} + s_1}{s_2 - s_1}\frac{p_1}{p_2}\right)^{-1/2} - 1\right] \tag{40}$$

$$D_1^+ = (\sigma_X^{-2} + s_1)^{-1}, \tag{41}$$

and the optimal top layer distortion D_2^*:

$$D_2^* = \min\big(\max(D_2^-,\, D_2^\star),\, D_2^+\big), \tag{42}$$

where

$$D_2^- = \left(2^{2R_X}(\sigma_X^{-2} + s_2)\right)^{-1} \tag{43}$$

$$D_2^\star = \left(2^{2R_X}(\sigma_X^{-2} + s_1)(s_2 - s_1)p_2/p_1\right)^{-1/2} \tag{44}$$

$$D_2^+ = \left(2^{2R_X}(\sigma_X^{-2} + s_1) + s_2 - s_1\right)^{-1}. \tag{45}$$

The corresponding optimal rate allocation R_1^*, R_2^* can be found as given in (26), (30).

The optimal rate allocation is plotted in Fig. 2 for $R_X = 1$, $\sigma_X^2 = 1$, and $s_1 = 0$ dB. Note that R_2^*, the rate allocated to the top layer, is not monotonic with the side-information channel condition. As fading state s_2 improves, R_2^* increases to take advantage of the better side-information quality. However, when s_2 is large, R_2^* begins to decline as the expected distortion is dominated by the worse fading state. Moreover, the optimal rate allocation is heavily skewed towards the lower layer: $R_2^* > 0$ only when p_2 is large.

C. Multiple Discrete Fading States

The rate-distortion function (34) extends directly to the case when the side-information channel has multiple discrete fading states ($M > 2$). Specifically, we construct the random variables W_i's to be

$$W_i = a_i X + N_i, \qquad i = 1, \ldots, M, \tag{46}$$

where $N_i \sim$ iid $\mathcal{N}(0, 1)$, then the rate of the ith layer is

$$R_i \triangleq \min_{W_i} I(X; W_i | Y_i, W_1^{i-1}) \tag{47}$$

$$= \min_{W_i} \{h(X|Y_i, W_1^{i-1}) - h(X|Y_i, W_1^i)\} \tag{48}$$

$$= \tfrac{1}{2}\log\frac{(\tilde{D}_{i-1}^{-1} + s_i - s_{i-1})^{-1}}{\tilde{D}_i}, \tag{49}$$

where

$$\tilde{D}_i \triangleq \min\big(D_i,\, (\tilde{D}_{i-1}^{-1} + s_i - s_{i-1})^{-1}\big). \tag{50}$$

The a_i that achieves (49) is determined from $\tilde{D}_i = \mathrm{VAR}[X|Y_i, W_1^i]$, which evaluates to

$$a_i = \tilde{D}_i^{-1} - \tilde{D}_{i-1}^{-1} - (s_i - s_{i-1}). \tag{51}$$

As $R_{\mathrm{HB}}(\mathbf{D}) = \sum_{i=1}^M R_i$, we substitute (49) in (4) to obtain the rate-distortion function:

$$R_{\mathrm{HB}}(\mathbf{D}) = -\tfrac{1}{2}\log(\sigma_X^{-2} + s_1) - \tfrac{1}{2}\log\tilde{D}_M - \tfrac{1}{2}\sum_{i=1}^{M-1}\log\big(1 + (s_{i+1} - s_i)\tilde{D}_i\big), \tag{52}$$

where $\tilde{D}_i$ is as defined in (50).

Unlike the case when $M = 2$, however, a closed-form expression for the minimum expected distortion $\mathrm{E}[D]^*$ does not appear analytically tractable. Nevertheless, the expected distortion minimization in (3) can be formulated as the following convex optimization problem over the variables $D_i, \tilde{D}_i$ for $i = 1, \ldots, M$:

$$\text{minimize} \quad \mathbf{p}^T\mathbf{D} \tag{53}$$

subject to

$$0 \le D_i, \quad 0 \le \tilde{D}_i, \qquad i = 1, \ldots, M \tag{54}$$

$$-\tfrac{1}{2}\log(\sigma_X^{-2} + s_1) - \tfrac{1}{2}\log\tilde{D}_M - \tfrac{1}{2}\sum_{i=1}^{M-1}\log\big(1 + (s_{i+1} - s_i)\tilde{D}_i\big) \le R_X \tag{55}$$

$$\tilde{D}_1 \le (\sigma_X^{-2} + s_1)^{-1} \tag{56}$$

$$\tilde{D}_i \le (\tilde{D}_{i-1}^{-1} + s_i - s_{i-1})^{-1},\ i = 2, \ldots, M \tag{57}$$

$$\tilde{D}_i \le D_i, \qquad i = 1, \ldots, M. \tag{58}$$

The constraints (56), (57) and (58) derive from expanding the condition in (50). The minimization can be efficiently computed by standard convex optimization techniques [15]. Moreover, convexity implies that a local optimum is globally optimal. Note that under the KKT optimality conditions, $\tilde{D}_i^* = D_i^*$ for $i = 1, \ldots, M$. It implies that the set of distortion inequality constraints (6) are tight when the expected distortion is minimized.

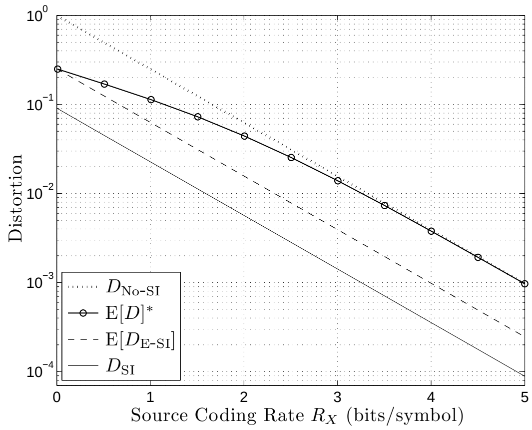

Fig. 3. Minimum expected distortion in source coding with uncertain side information ($\sigma_X^2 = 1$, $\bar{S} = 10$ dB).

IV. Numerical Results

Fig. 3 shows the minimum expected distortion $\mathrm{E}[D]^*$, when the side-information channel is described by a discretized Rayleigh fading distribution. We assume $\sigma_X^2 = 1$, and $\bar{S} \triangleq \mathrm{E}[S] = 10\,\mathrm{dB}$. The Rayleigh distribution is truncated at $2\bar{S}$, and discretized into $M = 20$ fading states with evenly spaced channel power gains $s_1, \ldots, s_M$.

For comparison, along with $\mathrm{E}[D]^*$, in Fig. 3 we also show the distortion under different assumptions on the side information. When no side information is available, the distortion is given by the rate-distortion function for a Gaussian source [16]:

$$D_{\text{No-SI}} = \sigma_X^2 2^{-2R_X}. \tag{59}$$

In the absence of side information, $D_{\text{No-SI}}$ is an upper bound to $\mathrm{E}[D]^*$. Next, if the encoder knows the realization of S, then for each fading state it can achieve the Wyner-Ziv rate-distortion function [4], and the corresponding expected distortion is

$$\mathrm{E}[D_{\text{E-SI}}] = \int_0^\infty f_S(s)(\sigma_X^{-2} + s)^{-1}\, ds\, 2^{-2R_X}. \tag{60}$$

With knowledge of S at the encoder, $\mathrm{E}[D_{\text{E-SI}}]$ is a lower bound to $\mathrm{E}[D]^*$. Finally, when there is no uncertainty in the side-information channel with $S = \bar{S}$, the distortion is given by the Wyner-Ziv rate-distortion function:

$$D_{\text{SI}} = (\sigma_X^{-2} + \bar{S})^{-1} 2^{-2R_X}. \tag{61}$$

By Jensen's inequality, $\mathrm{E}[D_{\text{E-SI}}] > D_{\text{SI}}$; hence uncertainty in the side-information channel always hurts performance.

We observe in Fig. 3 that when R_X is small, $\mathrm{E}[D]^*$ achieves a smaller distortion than when no side information is available. However, when R_X is large, $\mathrm{E}[D]^*$ is almost as large as $D_{\text{No-SI}}$; the uncertain side information is negligibly more useful than having no side information at all. Remarkably, for all R_X (and a wide range of parameters σ_X^2 and $\bar{S}$), the minimum expected distortion is achieved at $R_1^* = R_X$ and $R_i^* = 0$ for $i = 2, \ldots, M$. Hence the optimal rate allocation concentrates at the base layer of the source code.

These numerical results suggest that the optimal rate allocation to minimize the expected distortion is conservative: unless the better fading states are highly probable, no rate is allocated to the source coding layers other than the base layer, since it is not beneficial to dedicate source coding layers to take advantage of these better states. When R_X is small, the reduction in $\mathrm{VAR}[X]$ from the side information at the decoder is significant, thus $\mathrm{E}[D]^*$ outperforms $D_{\text{No-SI}}$. When R_X is large, however, as the better fading states are not exploited, the gap between $\mathrm{E}[D]^*$ and $D_{\text{No-SI}}$ vanishes.

V. Continuous Fading Distribution

A. Continuous Rate Distribution

In this section, we investigate the optimal rate allocation and minimum expected distortion when the side-information channel fading distribution is continuous. We consider the expected-distortion–rate function, i.e., $\mathrm{E}[D]$ as a function of $\mathbf{R} \triangleq \left[R_1 \ldots R_M\right]^T$, since it is continuous and differentiable in the entire nonnegative orthant $\{R_i \geq 0,\ i = 1, \ldots, M\}$. The expected distortion is

$$\mathrm{E}[D] = \mathbf{p}^T \mathbf{D} = p_1 D_1 + p_2 D_2 + \cdots + p_M D_M, \tag{62}$$

where D_i is found by recursively expanding (49):

$$D_i = \left[(((\sigma_X^{-2} + s_1)2^{2R_1} + s_2 - s_1)2^{2R_2} + \cdots)2^{2R_M}\right]^{-1}. \tag{63}$$

Suppose the channel power gains of the fading states start at $s_1 = 0$, and they are evenly spaced with $s_{i+1} - s_i = \Delta s$. In the limit of $\Delta s \to 0$, $M \to \infty$, the fading probability is given by the continuous pdf $f_S(s)$:

$$p_i \cong f_S[i]\Delta s, \quad \text{where } f_S[i] \triangleq f_S((i-1)\Delta s), \tag{64}$$

and the rate allocation is given by the continuous rate distribution function $R(s)$:

$$R_i \cong R[i]\Delta s, \quad \text{where } R[i] \triangleq R((i-1)\Delta s). \tag{65}$$

The expected distortion over $f_S(s)$ is

$$\mathrm{E}[D] = \lim_{\Delta s \to 0} \sum_{i=1}^{\infty} f_S[i] D[i] \Delta s, \tag{66}$$

where

$$D[i] = \left(\sigma_X^{-2} 2^{2\sum_{j=1}^{i} R[j]\Delta s} + \sum_{j=1}^{i-1} \left(2^{2\sum_{k=j}^{i} R[k]\Delta s}\right)\Delta s\right)^{-1}, \tag{67}$$

which follows from substituting $s_{i+1} - s_i = \Delta s$ in (63). Note that $D[i]$ is bounded and monotonically decreasing, thus the expectation in (66) exists when $\mathrm{E}[S]$ is finite. As we assume $R(s)$ is continuous, the Riemann sums in (66), (67) converge to the integral:

$$\mathrm{E}[D] = \int_0^\infty f_S(s) u'(s) \left(\sigma_X^{-2} + u(s)\right)^{-1} ds, \tag{68}$$

where $u(s) \triangleq \int_0^s 2^{-2\int_0^t R(r)\,dr} dt$, with the boundary conditions: $u'(0) = 1$, $u'(\infty) = 2^{-2R_X}$.

Over any interval where the optimal rate distribution $R^*(s)$ is continuous, the corresponding $u(s), u'(s)$ are also continuous, and they have to satisfy the necessary condition for optimality as given by the Euler-Lagrange equation [17]:

$$\frac{d}{ds}\Big(\frac{\partial F}{\partial u'}\Big) - \frac{\partial F}{\partial u} = 0, \tag{69}$$

where $F(s, u, u') \triangleq f_S(s) u'(s) \big(\sigma_X^{-2} + u(s)\big)^{-1}$. Taking the derivatives, (69) evaluates to:

$$\frac{f_S'(s)}{\sigma_X^{-2} + u(s)} = 0. \tag{70}$$

We suppose in general for the given fading distribution, $f_S'(s) \neq 0$; then no $u(s)$ satisfies (70), and the Euler-Lagrange condition does not lead to a continuous solution for the optimal rate distribution. We conjecture that the optimal rate allocation is discrete even when the fading distribution is continuous and smooth.

B. Conservative Rate Allocation

While the Euler-Lagrange equation (69) does not prescribe the necessary conditions for discrete rate allocation, the numerical results in Section IV suggest that under Rayleigh fading the optimal rate allocation may concentrate at the lowest layer, i.e., $R(s) = R_X \delta(s)$. In this section we assume such rate allocation and investigate the expected distortion under Rayleigh fading.

Let the pdf of the Rayleigh fading distribution be

$$f_S(s) = \bar{S}^{-1} e^{-s/\bar{S}}, \quad s \geq 0. \tag{71}$$

When the rate distribution is $R(s) = R_X \delta(s)$, the expected distortion evaluates to

$$\mathrm{E}[D] = \int_0^\infty \frac{f_S(s)}{\sigma_X^{-2} 2^{2R_X} + s}\, ds \tag{72}$$

$$= -\bar{S}^{-1} \exp(\bar{S}^{-1} \sigma_X^{-2} 2^{2R_X})\, \mathrm{Ei}(-\bar{S}^{-1} \sigma_X^{-2} 2^{2R_X}), \tag{73}$$

where $\mathrm{Ei}(\cdot)$ is the exponential integral function: $\mathrm{Ei}(x) \triangleq -\int_{-x}^\infty e^{-t}/t\, dt$. Note that when R_X is large such that $\sigma_X^{-2} 2^{2R_X} \gg K \gg \bar{S}$,

$$\mathrm{E}[D] \approx \int_0^K \frac{f_S(s)}{\sigma_X^{-2} 2^{2R_X} + s}\, ds \approx \int_0^K \frac{f_S(s)}{\sigma_X^{-2} 2^{2R_X}}\, ds \tag{74}$$

$$\approx \sigma_X^2 2^{-2R_X} = D_{\text{No-SI}}. \tag{75}$$

Therefore, as is observed in Fig. 3, under Rayleigh fading with large R_X, we expect uncertain side information is no more useful than no side information.

VI. Conclusion

We considered the problem of Gaussian source coding under squared error distortion with uncertain side information at the decoder. When the side-information channel has two discrete fading states, we derived closed-form expressions for the optimal rate allocation among the fading states and the corresponding minimum expected distortion. The optimal rate allocation is conservative: rate is allocated to the non-base code layer only if the better fading state is highly probable. Otherwise the distortion reduction from utilizing non-base code rate when the better state is realized is not sufficient to compensate for the distortion increase that results from reducing the code rate of the base layer to allocate rate to the higher layer. For multiple discrete fading states, the minimum expected distortion was shown to be the solution of a convex optimization problem. Under discretized Rayleigh fading, it is observed that the optimal rate allocation concentrates at the base layer associated with the worst-case fading state, and the uncertain side information is negligibly more useful than no side information, especially when the source coding rate is large.

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. 22, no. 1, pp. 1–10, Jan. 1976.

[2] S. Shamai (Shitz) and A. Steiner, “A broadcast approach for a single-user slowly fading MIMO channel,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

[3] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 727–734, Nov. 1985.

[4] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder—II: General sources,” *Inform. Contr.*, vol. 38, pp. 60–80, July 1978.

[5] 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. 19, no. 4, pp. 480–489, July 1973.

[6] R. Zamir, “The rate loss in the wyner-ziv problem,” *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.

[7] M. Fleming and M. Effros, “On rate-distortion with mixed types of side information,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1698–1705, Apr. 2006.

[8] Y. Steinberg and N. Merhav, “On successive refinement for the Wyner-Ziv problem,” *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.

[9] C. T. K. Ng, D. Gündüz, A. J. Goldsmith, and E. Erkip, “Recursive power allocation in Gaussian layered broadcast coding with successive refinement,” to appear at *IEEE Internat. Conf. Commun.*, 2007.

[10] ——, “Minimum expected distortion in Gaussian layered broadcast coding with successive refinement,” to appear at *IEEE Int. Symp. Inform. Theory*, 2007.

[11] S. Shamai (Shitz), S. Verdú, and R. Zamir, “Systematic lossy source/channel coding,” *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.

[12] A. H. Kaspi, “Rate-distortion function when side-information may be present at the decoder,” *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.

[13] C. Tian and S. N. Diggavi, “On scalable source coding with decoder side informations,” to appear at *IEEE Int. Symp. Inform. Theory*, 2007.

[14] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Addison-Wesley, 1994.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[17] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Prentice-Hall, Inc., 1963.