Gaussian Codes and Shannon Bounds for Multiple Descriptions

Ram Zamir

Dept. of Electrical Engineering - System
Tel-Aviv University, Tel-Aviv 69978, Israel

June 30, 1998

Abstract

A well known inequality due to Shannon, upper / lower bounds the rate distortion function of a real source by the rate distortion function of the Gaussian source with the same variance / entropy. We extend these bounds to multiple descriptions, a problem for which a general "single letter" solution is not known. We show that the set $\mathcal{D}_X(R_1, R_2)$ of achievable marginal (d_1, d_2) and central (d_0) mean squared errors in decoding X from two descriptions at rates R_1 and R_2 satisfies

$$\mathcal{D}^*(\sigma_x^2, R_1, R_2) \subseteq \mathcal{D}_X(R_1, R_2) \subseteq \mathcal{D}^*(P_x, R_1, R_2)$$

where σ_x^2 and P_x are the variance and the entropy-power of X, respectively, and $\mathcal{D}^*(\sigma^2, R_1, R_2)$ is the distortion region for a Gaussian source with variance σ^2 found by Ozarow. We further show that like in the single description case, a Gaussian random code achieves the *outer* bound in the limit as $d_1, d_2 \to 0$, thus the outer bound is asymptotically tight at high resolution conditions.

Key Words: multiple descriptions, Gaussian codes, Shannon lower bound, high resolution.

I. Introduction

The multiple description problem [5] arises in communication of analog source information (speech, image, video) via lossy packet networks. In this increasingly frequent scenario, a source code is broken into a few packets, some of which may not arrive to the destination. The decoder wishes to achieve a certain basic reproduction quality if a small subset of the packets arrives, and an improved quality if more packets or the whole source code arrives. Thus, portions of various size of the code should contain individually good, complementary descriptions of the source.

The basic formulation of the multiple description problem in the information theoretic literature involves two (lossless) sub-channels of rates R_1 and R_2, corresponding to two "packets", and three receivers. Each receiver corresponds to a possible case of packet arrival, the first arrived, the second arrived or both arrived. In response to a source block $\boldsymbol{x} = (x_1, \dots x_n)$, the encoder generates two code words (indices) $f_i(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ at rates

$$R_i = \tfrac{1}{n} \log |f_i(\cdot)| \qquad i = 1, 2$$

where $|f_i|$ denotes the size of code $f_i\ (\cdot)$, and transmits codeword f_i through sub-channel $i, i = 1, 2$. The two individual ("marginal") receivers and the combined ("central") receiver then generate reconstructions $\widehat{\boldsymbol{x}}_1, \widehat{\boldsymbol{x}}_2$ and $\widehat{\boldsymbol{x}}_0$, respectively, using the decoding functions g_1, g_2 and g_0, i.e.

$$\begin{aligned}
\widehat{\boldsymbol{x}}_1 &= g_1\big(f_1(\boldsymbol{x})\big) \\
\widehat{\boldsymbol{x}}_2 &= g_2\big(f_2(\boldsymbol{x})\big) \\
\widehat{\boldsymbol{x}}_0 &= g_0\big(f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\big)\ .
\end{aligned}$$

Assume that $X_i \in \mathrm{R}^1, i = 1, \dots, n$ is a real memoryless source, with a generating random variable X, and we wish to satisfy mean square errors d_1, d_2 and d_0 in the three non void cases of packet arrival, i.e.,

$$\begin{aligned}
\tfrac{1}{n} E \|\widehat{\boldsymbol{X}}_1 - \boldsymbol{X}\|^2 &\le d_1 \\
\tfrac{1}{n} E \|\widehat{\boldsymbol{X}}_2 - \boldsymbol{X}\|^2 &\le d_2 \\
\tfrac{1}{n} E \|\widehat{\boldsymbol{X}}_0 - \boldsymbol{X}\|^2 &\le d_0
\end{aligned}$$

where the reconstructions $\widehat{\boldsymbol{X}}_0, \widehat{\boldsymbol{X}}_1, \widehat{\boldsymbol{X}}_2 \in \mathcal{R}^n$. Let

$$\mathcal{R}_X(d_0, d_1, d_2) = \{(R_1, R_2) : d_0, d_1, d_2 \quad \text{are achievable for some } n\} \tag{1}$$

be the set of achievable rate pairs, or inversely let

$$\mathcal{D}_X(R_1, R_2) = \{(d_0, d_1, d_2) \quad \text{are achievable with rates } (R_1 R_2)\}$$

be the set of achievable distortions for a given rate pair (R_1, R_2). The multiple description problem differs from the classical rate-distortion problem, since it is usually impossible to satisfy the Shannon limits $R_1 = R_X(d_1), R_2 = R_X(d_2)$, where $R_X(d)$ denotes the rate distortion function of the source, and at the same time obtain from the central receiver distortion d_0 which satisfies the Shannon limit $R_1 + R_2 = R_X(d_0)$.

A "single letter" characterization of the set of achievable distortions $\mathcal{D}_X(R_1, R_2)$ for the case of a Gaussian source was found by Ozarow [8]. This set, denoted $\mathcal{D}^*(\sigma^2, R_1, R_2)$, and its inverse set $\mathcal{R}^*(\sigma^2, d_0, d_1, d_2)$, where σ^2 is the source variance, are described in the next section. For a general source, however, the complete solution for $\mathcal{R}_X(d_0, d_1, d_2)$ is still unknown. El-Gamal and Cover [5] characterize an inner bound for $\mathcal{R}_X(d_0, d_1, d_2)$, i.e., a set of achievable rates; see Section IV. We do not know of any *outer* bound for $\mathcal{R}_X(d_0, d_1, d_2)$ for general *real* sources. Only the special case of no excess rate sum, i.e., $R_1 + R_2 = R_X(d_0)$, was solved completely by Ahlswede [1]. For the case of a *discrete* source, Zhang and Berger [11, 12] provide inner and outer bounds which are relatively tight for the case of no excess marginal rate, i.e., $R_i = R_X(d_i), i = 1, 2$.

In this paper we develop an outer bound and an inner bound for the multiple description rate-distortion region for a general memoryless real source. We also discuss efficient coding schemes which operate close to these bounds. Our bounds have the form

$$\mathcal{R}^*(\sigma_x^2, d_0, d_1, d_2) \subseteq \mathcal{R}_X(d_0, d_1, d_2) \subseteq \mathcal{R}^*(P_x, d_0, d_1, d_2) \ , \tag{2}$$

and they parallel Shannon's lower and upper bounds for the rate-distortion function subject to a mean squared error constraint [2]

$$\frac{1}{2} \log \left(\frac{\sigma_x^2}{d} \right) \geq R_X(d) \geq \frac{1}{2} \log \left(\frac{P_x}{d} \right) \ , \tag{3}$$

where $\sigma_x^2, h(X)$ and $P_x = 2^{2h(X)}/2\pi e$ are the variance, the differential entropy and the entropy-power of X. Equality in (2) and (3) holds for a Gaussian source X^*, in which case $P_x = \sigma_x^2$. The tightness of (2) and (3) for a general source X depends on the closeness of the source to Gaussianity, e.g., as measured by the divergence [4, 10]

$$\mathcal{D}(X; X^*) = \frac{1}{2} \log(\sigma_x^2 / P_x) \ . \tag{4}$$

We state the new outer and inner bounds in Theorem 1 and Theorem 2 in Sections III and IV, respectively, after briefly describing Ozarow's solution for a Gaussian source in Section II. In Section V we show that for a "smooth" source at high resolution conditions, i.e., in the limit as d_1 and d_2 go to zero, the outer bound above becomes tight, and is realized by the optimum Gaussian "test channel" [5]. This result parallels the well known asymptotic tightness of the *Shannon lower bound* (right hand side of (3)) at high resolution; see [2, 7]. Finally, in Section VI we consider multiple description coding schemes induced by the Gaussian "test channel".

II. Ozarow's Solution of the Gaussian Case

Ozarow [8] proved that the set of achievable distortions $\mathcal{D}^*(\sigma^2, R_1, R_2)$ for a Gaussian source with variance σ^2 is the union of all triplets d_0, d_1, d_2 satisfying

$$d_1 \geq \sigma^2 \cdot 2^{-2R_1} \tag{5a}$$

$$d_2 \geq \sigma^2 \cdot 2^{-2R_2} \tag{5b}$$

$$d_0 \geq \frac{\sigma^2 \cdot 2^{-2(R_1+R_2)}}{1-\left(\sqrt{\pi}-\sqrt{\triangle}\right)^2} \tag{5c}$$

where $\pi = (1-d_1/\sigma^2)(1-d_2/\sigma^2)$ and $\triangle = d_1 d_2/\sigma^4 - 2^{-2(R_1+R_2)}$. In the sequel we will need also the inverse function of (5). Given a triplet of mean square errors d_0, d_1, d_2, the set $\mathcal{R}^*(\sigma^2, d_0, d_1, d_2)$ of admissible rate pairs is the union of all (R_1, R_2) satisfying

$$R_1 \geq \tfrac{1}{2}\log\left(\frac{\sigma^2}{d_1}\right) \tag{6a}$$

$$R_2 \geq \tfrac{1}{2}\log\left(\frac{\sigma^2}{d_2}\right) \tag{6b}$$

$$R_1 + R_2 \geq \tfrac{1}{2}\log\left(\frac{\sigma^4}{d_1 d_2}\right) + \delta \tag{6c}$$

where $\delta = \delta(\sigma^2, d_0, d_1, d_2)$ is defined by

$$\delta = \begin{cases} \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right), & d_0 \leq d_{0max} \\ 0, & d_0 \geq d_{0max}\ , \end{cases} \tag{6d}$$

$$\rho = -\frac{\sqrt{\pi\epsilon_0^2+\gamma}-\sqrt{\pi\epsilon_0^2}}{(1-\epsilon_0)\sqrt{\epsilon_1\epsilon_2}}\ , \quad \gamma = (1-\epsilon_0)[(\epsilon_1-\epsilon_0)(\epsilon_2-\epsilon_0)+\epsilon_0\epsilon_1\epsilon_2-\epsilon_0^2] \tag{6e}$$

$$\pi = (1-\epsilon_1)(1-\epsilon_2)\ , \qquad \epsilon_i = d_i/\sigma^2 \quad \text{for} \quad i=0,1,2 \tag{6f}$$

and

$$d_{0max} = \frac{1}{\frac{1}{d_1}+\frac{1}{d_2}-\frac{1}{\sigma^2}} = \frac{d_1 d_2}{d_1+d_2-d_1 d_2/\sigma^2}\ .$$

Note that $\delta, \gamma \geq 0$ and $-1 \leq \rho \leq 0$ for all $d_1, d_2 \leq \sigma^2$ and $d_0 \leq d_{0max}$. Note also that δ depends on d_0, d_1, d_2 and σ^2 only through the ratios d_i/σ^2. As we will see in Section IV, ρ has the meaning of a correlation coefficient in the Gaussian optimal test channel. The right hand sides of (6a)-(6b) are the rate-distortion functions of the Gaussian source $X^* \sim \mathcal{N}(0, \sigma^2)$ at distortion levels d_1 and d_2 respectively. Thus, the quantity $\delta = \delta_1 + \delta_2 \geq 0$ above represents the *total excess marginal rate* (TEMR) in the Gaussian case, where $\delta_i \triangleq R_i - \frac{1}{2}\log(\sigma^2/d_i)$, $i = 1, 2$ [1]. In the case of no excess marginal rate[2], i.e., $\delta = \triangle = 0$, we have the maximum central distortion $d_0 = d_{0max}$. As δ increases from zero to infinity (ρ varies from zero to -1), the central distortion d_0 decreases from d_{0max} to zero as $O(2^{-2\delta})$.

The case of no excess rate *sum*, i.e., $R_1 + R_2 = R_x(d_0) = \frac{1}{2}\log(\sigma^2/d_0)$ ($\pi = \Delta$ in (5)), whose solution was a breakthrough in the research of multiple descriptions [1], turns out to be not interesting practically in the Gaussian case. That is since no excess rate sum happens when $d_1 + d_2 = \sigma^2(1 + 2^{-2(R_1+R_2)}) = \sigma^2 + d_0$ (see [8]), i.e., when at least one of the two marginal receivers suffers from distortion higher than $\sigma^2/2$, meaning signal-to-noise ratio that is worse than 3dB. Having both types of excess rate (marginal-rate and rate-sum) zero is possible only in the trivial case where one of the sub-channels is disabled (e.g. $R_2 = 0$ and $d_2 = \sigma^2$).

Another special case in which we will be interested in the sequel is that of high resolution coding, i.e., the limit as d_1/σ^2 and d_2/σ^2 go to zero. In this limit the TEMR and the maximum central distortion do not depend on the variance of the source. Specifically we have

$$\begin{aligned}\delta_{HR}(d_0, d_1, d_2) &\triangleq \lim_{\sigma \to \infty} \delta(\sigma^2, d_0, d_1, d_2) \triangleq \lim_{\lambda \to 0} \delta(\sigma^2, \lambda d_0, \lambda d_1, \lambda d_2) \\ &= \frac{1}{2}\log\left(\frac{1}{1 - \rho_{HR}^2}\right) \end{aligned} \tag{7}$$

where

$$\rho_{HR} = -\frac{\sqrt{d_1/d_0 - 1}\sqrt{d_2/d_0 - 1} - 1}{\sqrt{d_1 d_2/d_0^2}} \tag{8}$$

and

$$\lim_{\sigma \to \infty} d_{0max} = \frac{d_1 d_2}{d_1 + d_2} \; . \tag{9}$$

III. A "Shannon Lower Bound" for Multiple Descriptions

In this section we show the following outer bound for the set of achievable distortions.

[1]We borrow the term "excess marginal rate" from [12].

[2]Both δ and Δ measure the excess of $R_1 + R_2$ over $\frac{1}{2}\log(\sigma^2/d_1) + \frac{1}{2}\log(\sigma^2/d_2)$; δ is their direct difference and Δ is their exponent difference.

Theorem 1 *For any real source X,*

$$\mathcal{D}_X(R_1, R_2) \subseteq \mathcal{D}^*(P_x, R_1, R_2) \tag{10}$$

where $P_x = (2\pi e)^{-1} 2^{2h(x)}$ and $h(X)$ are the entropy-power and the differential entropy of the source, respectively, and $\mathcal{D}^*(P_x, R_1, R_2)$ is the set defined in (5) (here at $\sigma^2 = P_x$) of all triplets (d_0, d_1, d_2) satisfying

$$d_1 \geq P_x \cdot 2^{-2R_1} \tag{11a}$$

$$d_2 \geq P_x \cdot 2^{-2R_2} \tag{11b}$$

$$d_0 \geq \frac{P_x \cdot 2^{-2(R_1+R_2)}}{1 - (\sqrt{\tilde{\pi}} - \sqrt{\tilde{\Delta}})^2} \tag{11c}$$

where $\tilde{\pi} = (1 - d_1/P_x)(1 - d_2/P_x)$ and $\tilde{\Delta} = d_1 d_2 / P_x^2 - 2^{-2(R_1+R_2)}$.

Proof: The proof follows the line of the proof of the converse in [8], except that at few points it uses the Shannon lower bound instead of the explicit rate distortion function, and at one point it uses the entropy-power inequality instead of the exact value of the entropy.

Assume that some multiple description scheme encodes the source $\mathbf{X} \in \mathrm{R}^n$ with rates R_1, R_2. We will show that the resulting distortions d_0, d_1, d_2 satisfy (11). Recall that the encoder transmits the codewords $f_1(\boldsymbol{X})$ and $f_2(\boldsymbol{X})$. Thus we have

$$nR_1 \geq H(f_1(\boldsymbol{X})) = I(\mathbf{X}; f_1(\boldsymbol{X})) \tag{12}$$

$$nR_2 \geq H(f_2(\boldsymbol{X})) = I(\mathbf{X}; f_2(\boldsymbol{X})) \tag{13}$$

where $H()$ and $I(\ ;\ )$ denote entropy and mutual information, respectively. In the sequel we use the following identity which follows by applying twice the chain rule

$$I(A; BC) = I(A; B) + I(A; C) + I(B; C|A) - I(B; C) \ . \tag{14}$$

Applying (14) to (12) and (13) we obtain

$$\begin{aligned} n(R_1 + R_2) &\geq I(\boldsymbol{X}; f_1(\boldsymbol{X})) + I(\boldsymbol{X}; f_2(\boldsymbol{X})) \\ &= I(\boldsymbol{X}; f_1(\boldsymbol{X}), f_2(\boldsymbol{X})) + I(f_1(\boldsymbol{X}); f_2(\boldsymbol{X})) - I(f_1(\boldsymbol{X}); f_2(\boldsymbol{X})|\boldsymbol{X}) \\ &= I(\boldsymbol{X}; f_1(\boldsymbol{X}), f_2(\boldsymbol{X})) + I(f_1(\boldsymbol{X}); f_2(\boldsymbol{X})) \ . \end{aligned} \tag{15}$$

Since $\widehat{\boldsymbol{X}}_1 = g_1(f_1(\boldsymbol{X})), \widehat{\boldsymbol{X}}_2 = g_2(f_2(\boldsymbol{X}))$ and $\widehat{\boldsymbol{X}}_0 = g_0(f_1(\boldsymbol{X}), f_2(\boldsymbol{X}))$, the data processing theorem implies from (12), (13) and (15)

$$R_1 \geq \tfrac{1}{n} I(\boldsymbol{X}; \widehat{\boldsymbol{X}}_1) \tag{16}$$

$$R_2 \geq \tfrac{1}{n} I(\boldsymbol{X}; \widehat{\boldsymbol{X}}_2) \tag{17}$$

$$R_1 + R_2 \geq \tfrac{1}{n} I(\boldsymbol{X}; \widehat{\boldsymbol{X}}_0) + \tfrac{1}{n} I(\widehat{\boldsymbol{X}}_1; \widehat{\boldsymbol{X}}_2) \tag{18}$$

$$R_1 + R_2 \geq \tfrac{1}{n} I(\boldsymbol{X}; \widehat{\boldsymbol{X}}_1, \widehat{\boldsymbol{X}}_2) + \tfrac{1}{n} I(\widehat{\boldsymbol{X}}_1; \widehat{\boldsymbol{X}}_2)\ . \tag{19}$$

By the definition of the rate-distortion function, and since $\frac{1}{n} E\|\widehat{\boldsymbol{X}}_1 - \boldsymbol{X}\|^2 \leq d_1$, $\frac{1}{n} E\|\widehat{\boldsymbol{X}}_2 - \boldsymbol{X}\|^2 \leq d_2$ and $\frac{1}{n} E\|\boldsymbol{X}_0 - \boldsymbol{X}\|^2 \leq d_0$, we have from (16)–(18)

$$R_1 \ \geq \ \tfrac{1}{n} I(\boldsymbol{X}; \widehat{\boldsymbol{X}}_1) \geq R_X(d_1) \tag{20a}$$

$$R_2 \ \geq \ \tfrac{1}{n} I(\boldsymbol{X}; \widehat{\boldsymbol{X}}_2) \geq R_X(d_2) \tag{20b}$$

$$R_1 + R_2 \ \geq \ \tfrac{1}{n} I(\boldsymbol{X}; \widehat{\boldsymbol{X}}_0) + \tfrac{1}{n} I(\widehat{\boldsymbol{X}}_1; \widehat{\boldsymbol{X}}_2) \geq R_X(d_0) + \tfrac{1}{n} I(\widehat{\boldsymbol{X}}_1; \widehat{\boldsymbol{X}}_2)\ . \tag{20c}$$

Using the Shannon lower bound $R_X(d) \geq \frac{1}{2}\log(\frac{P_x}{d})$ (right side of (3)) in (20a) and (20b) we get (11a) and (11b) immediately. For the central distortion we obtain from (20c) and (3)

$$d_0 \geq P_x \cdot 2^{-2(R_1+R_2)} \cdot 2^{\frac{2}{n} I(\widehat{\mathbf{X}}_1; \widehat{\mathbf{X}}_2)}\ . \tag{21}$$

Our next task is to lower bound $I(\widehat{\boldsymbol{X}}_1; \widehat{\boldsymbol{X}}_2)$ in terms of d_1, d_2, R_1 and R_2. For that we introduce, like in [8], an auxiliary random vector $\boldsymbol{Y} \in \mathrm{R}^n$, where

$$Y_j = X_j + Z_j \qquad j = 1, \ldots, n$$

where the Z_j's are independent zero mean Gaussian variables, jointly independent of $(\boldsymbol{X}, \widehat{\boldsymbol{X}}_1, \widehat{\boldsymbol{X}}_2)$, having variance σ_z^2. (Later in the proof we pick a specific value for σ_z^2.) Using identity (14) with $A = \boldsymbol{Y}, B = \widehat{\boldsymbol{X}}_1$ and $C = \widehat{\boldsymbol{X}}_2$, we have

$$\begin{aligned} I(\widehat{\boldsymbol{X}}_1; \widehat{\boldsymbol{X}}_2) &= I(\boldsymbol{Y}; \widehat{\boldsymbol{X}}_1) + I(\boldsymbol{Y}; \widehat{\boldsymbol{X}}_2) + I(\widehat{\boldsymbol{X}}_1; \widehat{\boldsymbol{X}}_2 | \boldsymbol{Y}) - I(\boldsymbol{Y}; \widehat{\boldsymbol{X}}_1 \widehat{\boldsymbol{X}}_2) \\ &\geq I(\boldsymbol{Y}; \widehat{\boldsymbol{X}}_1) + I(\boldsymbol{Y}; \widehat{\boldsymbol{X}}_2) - I(\boldsymbol{Y}; \widehat{\boldsymbol{X}}_1 \widehat{\boldsymbol{X}}_2) \end{aligned} \tag{22}$$

where the inequality follows from the nonnegativity of the mutual information. To lower bound the first two terms in (22) we use the definition of the rate-distortion function of Y_i, denoted $R_Y(d)$, and the fact that

$$\tfrac{1}{n} E\|\widehat{\boldsymbol{X}}_i - \boldsymbol{Y}\|^2 = \tfrac{1}{n} E\|\widehat{\boldsymbol{X}}_i - \boldsymbol{X}\|^2 + \tfrac{1}{n} E\|\mathbf{Z}\|^2 \leq d_i + \sigma_z^2$$

and obtain

$$\tfrac{1}{n}I(\boldsymbol{Y};\widehat{\boldsymbol{X}}_i) \geq R_Y(d_i+\sigma_z^2) \geq \tfrac{1}{2}\log\left(\frac{P_y}{d_i+\sigma_z^2}\right) \qquad i=1,2 \tag{23}$$

where $P_y = (2\pi e)^{-1}2^{2h(Y)}$ is the entropy power of $Y = X+Z$ (where $Z \sim \mathcal{N}(0,\sigma_z^2)$) and the second inequality follows from the Shannon lower bound (3) for the rate distortion function of Y. We upper bound the third term in (22) as follows

$$\begin{aligned} I(\boldsymbol{Y};\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2) &= h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2) && \text{(24a)}\\ &= \tfrac{n}{2}\log(2\pi e P_y) - h(\boldsymbol{X}+\boldsymbol{Z}|\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2) && \text{(24b)}\\ &\leq \tfrac{n}{2}\log(2\pi e P_y) - \tfrac{n}{2}\log 2\pi e\left(\frac{2^{\frac{2}{n}h(\mathbf{X}|\widehat{\mathbf{X}}_1\widehat{\mathbf{X}}_2)}}{2\pi e}+\sigma_z^2\right) && \text{(24c)} \end{aligned}$$

where in the last step we used the fact that $\boldsymbol{Z}$ is independent of $(\boldsymbol{X},\widehat{\boldsymbol{X}}_1,\widehat{\boldsymbol{X}}_2)$ and applied the conditional version of the entropy-power inequality [8, 3] to lower bound the conditional entropy of the sum of $\boldsymbol{X}$ and $\boldsymbol{Z}$ given $(\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2)$. Finally we lower bound the conditional entropy $h(\boldsymbol{X}|\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2)$ in (24c) as follows. From (19) we have

$$\begin{aligned} n(R_1+R_2) &\geq I(\boldsymbol{X};\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2) + I(\widehat{\boldsymbol{X}}_1;\widehat{\boldsymbol{X}}_2)\\ &= h(\boldsymbol{X}) - h(\boldsymbol{X}|\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2) + I(\widehat{\boldsymbol{X}}_1;\widehat{\boldsymbol{X}}_2) \end{aligned}$$

Thus (using $h(\boldsymbol{X}) = \frac{n}{2}\log(2\pi e P_x)$)

$$h(\boldsymbol{X}|\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2) \geq \tfrac{n}{2}\log(2\pi e P_x) - n(R_1+R_2) + I(\widehat{\boldsymbol{X}}_1;\widehat{\boldsymbol{X}}_2)\ . \tag{25}$$

Combining (24c) and (25) we upperbound $I(\boldsymbol{Y};\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2)$ by

$$I(\boldsymbol{Y};\widehat{\boldsymbol{X}}_1\widehat{\boldsymbol{X}}_2) \leq \tfrac{n}{2}\log(2\pi e P_y) - \tfrac{n}{2}\log 2\pi e\left[P_x\cdot 2^{-2(R_1+R_2)}\cdot 2^{\frac{1}{n}I(\widehat{\mathbf{X}}_1;\widehat{\mathbf{X}}_2)}+\sigma_z^2\right] \tag{26}$$

Combining (22), (23) and (26), we thus have

$$\begin{aligned} \tfrac{1}{n}I(\widehat{\boldsymbol{X}}_1;\widehat{\boldsymbol{X}}_2) &\geq \tfrac{1}{2}\log\left(\frac{P_y}{d_1+\sigma_z^2}\right) + \tfrac{1}{2}\log\left(\frac{P_y}{d_2+\sigma_z^2}\right) - \tfrac{1}{2}\log(2\pi e P_y) +\\ &\quad + \tfrac{1}{2}\log 2\pi e\left[P_x\cdot 2^{-2(R_1+R_2)}\cdot 2^{\frac{2}{n}I(\widehat{\mathbf{X}}_1;\widehat{\mathbf{X}}_2)}+\sigma_z^2\right]\\ &= \tfrac{1}{2}\log\left[\left(\frac{P_y}{(d_1+\sigma_z^2)(d_2+\sigma_z^2)}\right)\left(P_x\cdot 2^{-2(R_1+R_2)}\cdot 2^{\frac{2}{n}I(\widehat{\mathbf{X}}_1;\widehat{\mathbf{X}}_2)}+\sigma_z^2\right)\right] \end{aligned} \tag{27}$$

Define $t = 2^{\frac{2}{n}I(\widehat{\mathbf{X}}_1;\widehat{\mathbf{X}}_2)}$. We have from (27)

$$t \geq \frac{(P_x+\sigma_z^2)(P_x\cdot 2^{-2(R_1+R_2)}t+\sigma_z^2)}{(d_1+\sigma_z^2)(d_2+\sigma_z^2)} \tag{28}$$

where we have also used the entropy-power inequality for the independent sum $Y = X + Z$

$$P_y \geq P_x + P_z = P_x + \sigma_z^2 \ .$$

Define $\widetilde{\epsilon} = \sigma_z^2/P_x$, $\widetilde{\pi} = (1 - d_1/P_x)(1 - d_2/P_x)$ and $\widetilde{\Delta} = d_1 d_2/P_x^2 - 2^{-2(R_1+R_2)}$. Isolating t in (16), we get

$$t \geq \frac{\widetilde{\epsilon}(1+\widetilde{\epsilon})}{\widetilde{\epsilon}^2 + \widetilde{\epsilon}(1+\widetilde{\Delta}-\widetilde{\pi}) + \widetilde{\Delta}} \tag{29}$$

which holds for any choice of $\widetilde{\epsilon} \geq 0$. Substituting $\widetilde{\epsilon} = \frac{\sqrt{\widetilde{\Delta}}}{\sqrt{\widetilde{\pi}} - \sqrt{\widetilde{\Delta}}}$ (the value that maximizes (29)), we obtain (see [8, eq. (11)])

$$t \geq \frac{1}{1 - (\sqrt{\widetilde{\pi}} - \sqrt{\widetilde{\Delta}})^2} \ ,$$

which together with (21) shows (11c), and the Theorem is proved. Ξ

IV. An Inner Bound via a Gaussian Random Code

In this section we show that the El-Gamal-Cover construction for a Gaussian source (see [8, 5]) provides a sub-optimal multiple description coding scheme for a non-Gaussian source. We start by stating an achievability result for multiple descriptions under a mean square error constraint which follows from [5]. A triplet of mean square errors (d_0, d_1, d_2) is achievable from two descriptions of a source X with rates R_1 and R_2 if (but not necessarily only if!) there exist random variables U and V jointly distributed with X, and three functions $\widehat{X}_1 = \widehat{X}_1(U), \widehat{X}_2 = \widehat{X}_2(V)$ and $\widehat{X}_0 = \widehat{X}_0(U,V)$, such that[3]

$$R_1 \geq I(X;U) \tag{30a}$$

$$R_2 \geq I(X;V) \tag{30b}$$

$$R_1 + R_2 \geq I(X;UV) + I(U;V) \tag{30c}$$

$$d_i \geq E(X - \widehat{X}_i)^2 \qquad i = 0, 1, 2 \ . \tag{30d}$$

The El-Gamal Cover construction is optimal for a Gaussian source; the entire rate region $R^*(\sigma_x^2, d_0, d_1, d_2)$ characterized in (6) is realized in this case by choosing

$$U = X + N_1 \qquad V = X + N_2 \tag{31a}$$

$$\widehat{X}_1 = \alpha_1 U \qquad \widehat{X}_2 = \alpha_2 V \qquad \widehat{X}_0 = \beta_1 U + \beta_2 V \tag{31b}$$

[3]The characterization of the achievable rate region in (30) might be looser than that of [5], but it suffices for our purpose.

where (N_1, N_2) are correlated Gaussians, independent of X, with covariance matrix

$$\mathrm{COV}(N_1, N_2) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{31c}$$

whose parameters ρ, σ_1 and σ_2 are chosen so that $\mathrm{VAR}(X|U) = d_1$, $\mathrm{VAR}(X|V) = d_2$ and $\mathrm{VAR}(X|UV) = d_0$, where $\mathrm{VAR}(A|B) \stackrel{\triangle}{=} E[A - E(A|B)]^2$ denotes conditional variance. The scaling factors $\alpha_1, \alpha_2, \beta_1$ and β_2 are chosen to yield the conditional means $\widehat{X}_1 = E(X|U), \widehat{X}_2 = E(X|V)$ and $\widehat{X}_0 = E(X|UV)$, which are linear functions since (X, U, V) are jointly Gaussian. The above implies that $\sigma_i^2 = \sigma_x^2 d_i/(\sigma_x^2 - d_i)$, $i = 1, 2$, and that ρ is given by the expression in (6e). See [8] for explicit characterization of $\alpha_1, \alpha_2, \beta_1, \beta_2$ as a function of $\sigma_x^2, d_0, d_1, d_2$.

For an *arbitrary* real source X with variance σ_x^2, we refer to the relation (31) (or to the random code induced from it) as the *Gaussian multiple description "test channel"* (or the *Gaussian multiple description random code*). Since the relations in (31) are linear, this test channel yields the *same* distortion triplet (d_0, d_1, d_2) in (30d) as in the Gaussian source case. Substituting (31) in (30) we obtain the rate region

$$\begin{aligned}
R_1 &\geq I(X; X + N_1) && (32a)\\
R_2 &\geq I(X; X + N_2) && (32b)\\
R_1 + R_2 &\geq I(X; X + N_1, X + N_2) + I(X + N_1; X + N_2) && (32c)\\
&= I(X; X + N_1) + I(X; X + N_2) + I(N_1; N_2) && (32d)\\
&= I(X; X + N_1) + I(X; X + N_2) + \tfrac{1}{2}\log\left(\frac{1}{1-\rho^2}\right) && (32e)
\end{aligned}$$

where ρ is given in (6e), eq. (32d) follows using identity (14), and in (32e) we used the formula for the mutual information between correlated Gaussians. We denote the rate region in (32) by $\mathcal{R}_X^G(d_0, d_1, d_2)$. For a Gaussian source $X^* \sim \mathcal{N}(0, \sigma_x^2)$ we have

$$I(X^*; X^* + N_i) = \tfrac{1}{2}\log\left(1 + \frac{\sigma_x^2}{\sigma_i^2}\right) = \tfrac{1}{2}\log\left(\frac{\sigma_x^2}{d_i}\right),$$

and $\mathcal{R}_X^G(d_0, d_1, d_2)$ coincides with the Gaussian rate region $\mathcal{R}^*(\sigma_x^2, d_0, d_1, d_2)$ defined in (6). In this case the total excess marginal rate (TEMR) is given directly by

$$\delta = I(N_1; N_2) = \tfrac{1}{2}\log\left(\frac{1}{1-\rho^2}\right). \tag{33}$$

From the discussion above it is easy to prove the following inner bound for the rate region in multiple description coding of an arbitrary (not necessarily Gaussian) real source.

Theorem 2 *For any real source X with variance σ_x^2,*

$$\mathcal{R}_X(d_0, d_1, d_2) \supseteq \mathcal{R}^*(\sigma_x^2, d_0, d_1, d_2) \ .$$

Proof: Since $\widehat{X}_0, \widehat{X}_1, \widehat{X}_2$ of (31b) satisfy the distortion constraints for any X with variance σ_x^2, the El-Gamal-Cover Theorem [5] implies that the rate region $\mathcal{R}_X^G(d_0, d_1, d_2)$ inner bounds $\mathcal{R}_X(d_0, d_1, d_2)$. Now, since N_1 and N_2 are Gaussian, the quantities $I(X; X+N_1)$ and $I(X; X+N_2)$ in (32) are upper bounded by their respective values for a Gaussian source, (see [4, pp. 263]), i.e.

$$I(X; X+N_i) \leq I(X^*; X^*+N_i) = \tfrac{1}{2}\log\left(\frac{\sigma_x^2}{d_i}\right) . \tag{34}$$

In view of (30), (32), (34) and the tightness of (32) for X^*, we thus have

$$\mathcal{R}^*(\sigma_x^2, d_0, d_1, d_2) \subseteq R_X^G(d_0, d_1, d_2) \subseteq \mathcal{R}_X(d_0, d_1, d_2) \tag{35}$$

and the theorem is proved. Ξ

V. Multiple Descriptions with High Resolution

In this section we show that the outer bound of Theorem 1 is asymptotically tight at high resolution, i.e., when d_0, d_1 and d_2 are small compared to the variance of the source. The only requirement is that the source satisfies a simple "smoothness" condition. Furthermore, we show that this outer bound is asymptotically attainable by the Gaussian multiple description random code described in the previous section. In the end of the section we consider a certain modification of this random code.

Theorem 3 *Let X be a source with finite entropy-power $P_x > 0$ and finite variance. Let d_0, d_1, d_2 be fixed, and let $(d_0', d_1', d_2') = (\lambda d_0, \lambda d_1, \lambda d_2)$, . For any $\epsilon > 0$, if $\lambda > 0$ is small enough, then*

$$\mathcal{R}_X(d_0', d_1', d_2') \supseteq \mathcal{R}_X^G(d_0', d_1', d_2') \supseteq \mathcal{R}^*(P_x, d_0', d_1', d_2')^{(+\epsilon)} \tag{36}$$

where $\mathcal{R}^G$ is defined in (32), and $\mathcal{R}^{(+\epsilon)}$ denotes $\{(R_1, R_2) :\ (R_1 - \epsilon, R_2 - \epsilon) \in \mathcal{R}\}$.

In other words, at high resolution conditions the rate regions $R_X^G(d_0, d_1, d_2)$ and $\mathcal{R}_X(d_0, d_1, d_2)$ approach the outer bound $\mathcal{R}^*(P_x, d_0, d_1, d_2)$, i.e., they contain all rate pairs satisfying

$$R_1 \geq \tfrac{1}{2}\log\left(\frac{P_x}{d_1}\right) + \epsilon \tag{37a}$$

$$R_2 \geq \tfrac{1}{2}\log\left(\frac{P_x}{d_2}\right) + \epsilon \tag{37b}$$

$$R_1 + R_2 \geq \tfrac{1}{2}\log\left(\frac{P_x}{d_1 d_2}\right) + \delta_{HR}(d_0, d_1, d_2) + \epsilon \tag{37c}$$

where $\delta_{HR}(d_0, d_1, d_2)$ is the TEMR at high resolution defined in (7). Note that the "smoothness" condition $P_x > 0$ in Theorem 3 is equivalent to having finite differential entropy.

Before proving Theorem 3 we show the following lemma.

Lemma 1 *Let $(N_1^\lambda, N_2^\lambda)$ denote the noise pair in (31) associated with distortions $\lambda d_0, \lambda d_1$ and λd_2 and source variance σ_x^2. If $P_x > 0$ and $\sigma_x^2 < \infty$ then for $i = 1, 2$*

$$\lim_{\lambda \to 0} \left[I(X; X + N_i^\lambda) - \tfrac{1}{2} \log\left(\frac{P_x}{\lambda d_i}\right) \right] = 0 \ . \tag{38}$$

Proof Decompose the mutual information into a difference between differential entropies. Since $h(X) = \frac{1}{2}\log 2\pi e P_x$ and $h(N_i^\lambda) = \frac{1}{2} \log 2\pi e \mathrm{VAR}(N_i^\lambda)$, we have for $i = 1, 2$

$$I(X; X + N_i^\lambda) - \frac{1}{2}\log\left(\frac{P_x}{\lambda d_i}\right) = \Big[h(X + N_i^\lambda) - h(X)\Big] - \frac{1}{2}\log\left(\frac{VAR(N_i^\lambda)}{\lambda d_i}\right) . \tag{39}$$

From (31) we have $\frac{1}{\lambda}\mathrm{VAR}(N_i^\lambda) = \frac{1}{\lambda}\frac{\sigma_x^2 \lambda d_i}{\sigma_x^2 - \lambda d_i} \longrightarrow d_i$ as $\lambda \to 0$. Thus the second term in (39) is asymptotically zero. Furthermore, since $\sigma_x^2 < \infty$ and $P_x > 0$, and since N_i^λ is Gaussian whose variance $\to 0$, it follows from [7] that $h(X + N_i^\lambda) \to h(X)$, and (38) follows. Ξ

Proof of Theorem 3 Substituting the limits (38) and (7) in (32), we see that $\mathcal{R}_X^G(d_0', d_1', d_2')$ satisfies (37a) for small enough λ, and the proof follows. Ξ

A Robust Gaussian Test Channel

Finally, we point out a simple modification of the Gaussian multiple description test channel which in some sense is more robust. Consider the limiting structure, as $\sigma^2 \to \infty$, of the test channel in (31). In this limit the covariance of the additive noises becomes

$$\mathrm{COV}(N_1, N_2) = \begin{pmatrix} d_1 & \rho_{HR}\sqrt{d_1 d_2} \\ \rho_{HR}\sqrt{d_1 d_2} & d_2 \end{pmatrix}$$

where $\rho_{HR} \stackrel{\Delta}{=} \rho_{HR}(d_0/d_1, d_0/d_2)$ is given in (8), and the scaling parameters become $\alpha_1 = \alpha_2 = 1$, $\beta_1 = (|\rho_{HR}|\sqrt{d_1 d_2} + d_2)/(d_1 + |\rho_{HR}|\sqrt{d_1 d_2}) + d_2)$, and $\beta_2 = 1 - \beta_1$. Suppose we use the El-Gamal-Cover random code associated with these parameters (which by Theorem 3 is optimal at high resolution conditions) for coding a source with a *finite* variance. It is easy to verify that the distortion triplet of this code is still (d_0, d_1, d_2), i.e., *independently* of the variance of the source. The reason is that the parameters above correspond to estimation of X without a priori knowledge, i.e., as if X were a "parameter". This property is in the spirit of "universal quantization"; see [13] and the discussion below.

VI. Discussion: Gaussian Codes

We have shown that: (1) Ozarow's formula for the multiple description rate region of a Gaussian source provides an inner (outer) bound for the multiple description rate region of a general source when substituting the variance (entropy-power) of the source. (2) The El-Gamal-Cover optimum test channel for a Gaussian source $X^* \sim \mathcal{N}(0, \sigma_x^2)$, when applied to a general source X with variance σ_x^2, induces a random code whose distortion triplet d_0, d_1, d_2 is the same as for X^*, and whose rate region, $\mathcal{R}_X^G(d_0, d_1, d_2)$, lies between the inner and outer bounds. (3) At high resolution, i.e., when the source is "smooth" and $d_0, d_1, d_2 \to 0$, the "Gaussian" rate region $\mathcal{R}_X^G(d_0, d_1, d_2)$ approaches the *outer* bound.

An interesting consequence of these bounds is that a Gaussian multiple description codebook provides an efficient mean for encoding sources which are not too far from Gaussianity (e.g., in the sense of (4)). For *fixed rate* encoding, the result of Lapidoth [6] can be generalized to show that a Gaussian multiple description codebook guarantees achieving the Gaussian distortion region $\mathcal{D}^*(\sigma_x^2, R_1, R_2)$ for any real source with variance σ_x^2. See e.g. [9] for a geometric treatment of such a coding scheme. For *variable rate* encoding, we expect the *better* rate region, $\mathcal{R}_X^G(d_0, d_1, d_2)$, to be approached by generalizing the concept of entropy-coded dithered lattice quantization [10]. This generalization will be considered in a forthcoming paper.

References

[1] R. Ahlswede. The rate-distortion region for multiple descriptions without excess rate. *IEEE Trans. Information Theory*, IT-31:721–726, Nov. 1985.

[2] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[3] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Information Theory*, IT-11:267–271, 1965.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[5] A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Information Theory*, IT-28:851–857, Nov. 82.

[6] A. Lapidoth. On the role of missmatch in rate-distortion theory. *IEEE Trans. Information Theory*, 43:38–47, Jan. 1997.

[7] T. Linder and R. Zamir. On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Information Theory*, IT-40:2026–2031, Nov. 1994.

[8] L. H. Ozarow. On the source coding problem with two channels and three receivers. *Bell Sys. Tech. Jour.*, 59:1909–1922, 1980.

[9] V. A. Vaishampayan. Multiple description source encoding of a Gaussian random variable: A geometric treatment. In *Proceedings of the Int. Symposium on Info. Theory*, page 273, 1997.

[10] R. Zamir and M. Feder. Information rates of pre/post filtered dithered quantizers. *IEEE Trans. Information Theory*, pages 1340–1353, September 1996.

[11] Z. Zhang and T. Berger. New results in binary multiple descriptions. *IEEE Trans. Information Theory*, IT-33:502–521, July 1987.

[12] Z. Zhang and T. Berger. Multiple description source coding with no excess marginal rate. *IEEE Trans. Information Theory*, IT-41:349–357, March 1995.

[13] J. Ziv. On universal quantization. *IEEE Trans. Information Theory*, IT-31:344–347, May 1985.