On rate-distortion function of Gaussian sources with memory in the presence of side information at the decoder.

S. Sandeep Pradhan
email: pradhan@ifp.uiuc.edu

Project Report

ECE 480

December 12, 1998

1 Introduction

Distributed source coding has been of interest in information theory for the past two decades. It deals with encoding of multiple sources present at remote locations without having to pool their information at any point. Slepian and Wolf introduced this concept in their classic paper [1].

Paraphrased loosely, it tackles the problem of compressing a discrete-alphabet memoryless source X that is correlated with another such source Y. Should X and Y be able to co-operate, then it is clear that the joint description of X and Y can be made theoretically at the minimum cost of $H(X,Y) = H(Y) + H(X|Y)$, where $H(.)$ is the Shannon entropy function. If Y should therefore be available to both the encoder *and* decoder (see Fig. 1(a)), then X can be described in $H(X|Y)$ bits/sample, and there is no way to do better. The Slepian-Wolf theorem says, somewhat surprisingly, that if Y should be known *only to the decoder* but not to the encoder (see Fig. 1(b)), then also X can be described perfectly for the same lowest price of $H(X|Y)$ bits/sample! That is, knowledge of Y does not help X one iota in reducing its bit rate. Of course, as might be expected of information-theoretic results, this is an asymptotic and non-constructive result that requires the use of Shannon-style infinite code-length arguments.

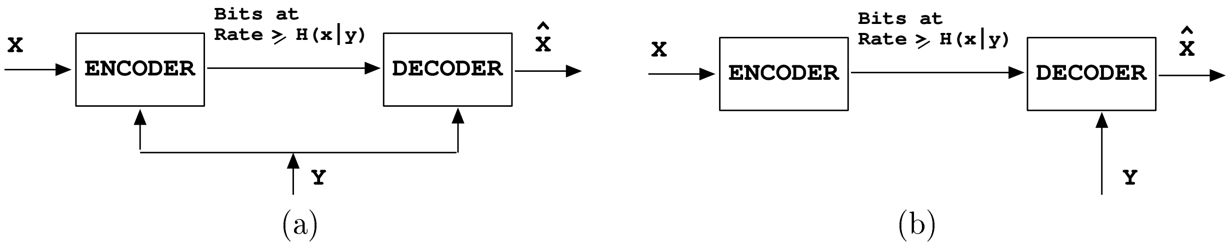

Figure 1: Communication System: (a) Both encoder and decoder have access to the side information Y (which is correlated to X). X can be described with $H(X|Y)$ bits/sample. (b) Only decoder has access to the side information Y (which is correlated to X). The Slepian-Wolf theorem says that X can still be described with $H(X|Y)$ bits/sample.

The Slepian-Wolf theorem has been extended to the lossy coding of continuous-valued sources by Wyner and Ziv [2, 3, 4, 5], who showed that a similar result holds in the case where X and Y are correlated independent identically distributed (*i.i.d.*) Gaussian random variables. If the decoder knows Y (the only caveat is that Y has to be known precisely, that is with no loss), then whether or not X knows Y, the rate-distortion performance for coding X is identical, and knowledge of Y is of no help in improving the rate-distortion performance[1]. As in the lossless case, here too the result is non-constructive and asymptotic in nature A word about the notation. Bold-faced lower and upper case letters represent vectors and matrices respectively. In this project we extend the result of [4] to jointly Gaussian sources having memory. Though it has not been reported in the literature, it has very interesting consequences as shown later in this report.

2 Preliminaries

Let $\mathbf{x}$ and $\mathbf{y}$ be *i.i.d* jointly Gaussian vector sources with dimension k. Let $\mathbf{R_x}$ and $\mathbf{R_y}$ be the autocorrelation matrices of $\mathbf{x}$, $\mathbf{y}$ respectively and let $\mathbf{R_{xy}}$ be the cross correlation matrix of $\mathbf{x}$ and $\mathbf{y}$. $\mathbf{x}$ and $\mathbf{y}$ can be expressed in the following matrix equation.

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \qquad \mathbf{x} = \mathbf{B}\mathbf{y} + \mathbf{m} \tag{1}$$

where $\mathbf{n}$ is a Gaussian random vector which is independent of $\mathbf{x}$ and having autocorrelation matrix $\mathbf{R_n}$ and $\mathbf{m}$ is a Gaussian vector which is independent of $\mathbf{y}$ and having autocorrelation matrix $\mathbf{R_m}$. We have the following relations.

$$\mathbf{R_{yx}} = \mathbf{A}\mathbf{R_x} \qquad \mathbf{R_{xy}} = \mathbf{B}\mathbf{R_y} \tag{2}$$

[1]It is interesting to note that the case where Y is not available losslessly at the decoder is still an open problem in information theory, and no performance bounds are known.

$$\mathbf{R_n} = \mathbf{R_y} - \mathbf{R_{yx}}\mathbf{R_x^{-1}}\mathbf{R_{xy}} \tag{3}$$

$$\mathbf{R_m} = \mathbf{R_x} - \mathbf{R_{xy}}\mathbf{R_y^{-1}}\mathbf{R_{yx}}. \tag{4}$$

Let I($\mathbf{x}$;$\mathbf{y}$) denote the Shannon mutual information [6] between two random vectors (variables) $\mathbf{x}$ and $\mathbf{y}$. We have the following relation [6].

$$I(\mathbf{x};\mathbf{y}) = \frac{1}{2}log\frac{|\mathbf{R_y}|}{|\mathbf{R_n}|} = \frac{1}{2}log\frac{|\mathbf{R_x}|}{|\mathbf{R_m}|} \tag{5}$$

where $|.|$ is the determinant function.

The rate-distortion function of a random variable W, is the minimum rate R, required to represent it so that the reconstruction distortion is less than D. It can be obtained by the following definition [6].

$$R(D) = \inf_{p(\hat{W}|W)} I(W;\hat{W}) \tag{6}$$

such that $E\left[(W-\hat{W})^2\right] \leq D$ and minimization is carried out over all conditional probability density functions $p(\hat{W}|W)$. The intuition behind this definition is that we are finding the test channel $p(\hat{W}|W)$ such that $\hat{W}$ and W are most separated (in the mutual information sense) with the given constraint that $E\left[(W-\hat{W})^2\right] \leq D$. This can be made clear in the following example.

Let W be $i.i.d$ Gaussian random variable with zero mean and variance σ^2, then R(D) is given by

$$R(D) = \frac{1}{2}log\frac{\sigma^2}{D}. \tag{7}$$

The test channel $p(\hat{W}|W)$ is given by the Fig. 2.

We have $\hat{W} = a.(W+\phi)$, where ϕ is a Gaussian random variable independent of W with zero mean and variance given by $\frac{\sigma^2 D}{\sigma^2 - D}$ and a is a constant given by $\frac{\sigma^2 - D}{\sigma^2}$. Let $\mathbf{w}$ be $i.i.d$ Gaussian random vector with dimension k and autocorrelation matrix $\mathbf{R_w}$, then its

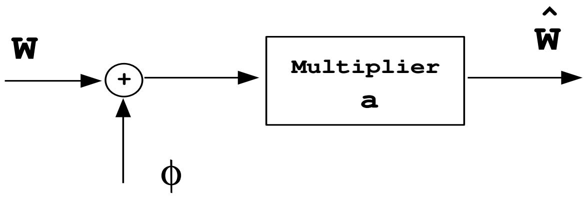

Figure 2: The test channel for the rate-distortion function of W which is Gaussian random variable with zero mean and variance σ^2

rate-distortion function is given by,

$$R(D) = \sum_{i=1}^{k} \frac{1}{2} log \frac{\sigma_i^2}{D_i} \tag{8}$$

where

$$D_i = \begin{cases} \lambda, & if \quad \lambda < \sigma_i^2 \\ \sigma_i^2, & if \quad \lambda \geq \sigma_i^2 \end{cases} \tag{9}$$

and σ_i^2 are eigen values of $\mathbf{R_w}$, for $1 \leq i \leq k$, λ is chosen such that $\sum_{i=1}^{k} D_i = D$. This solution is obtained as we minimize the mutual information given above with the constraint that $\sum_{i=1}^{k} D_i \leq D$ and is solved using Lagrange multipliers and Kuhn-Tucker condition.

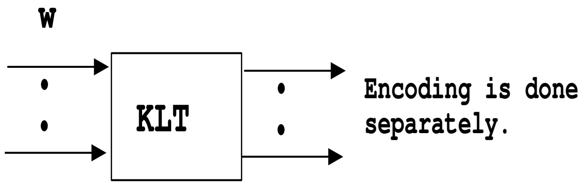

Figure 3: Optimum encoding of a Gaussian random vector $\mathbf{w}$

Thus the encoding is done as shown in Fig. 3. The random vector is passed through a decorrelating Karhunen-Loeve (KL) transform (given by the eigen function matrix of the autocorrelation matrix), and then the independent components are encoded separately, with each component distortion as given by (9).

Consider the communication system given by Fig. 1(a). Let X and Y be jointly Gaussian random variables with zero mean and with σ_x^2, σ_y^2 and ρ as variance X, variance of Y and correlation coefficient of X and Y respectively. Since Y is available to both encoder and decoder, at the encoder, we can obtain the best estimate of X using Y, and encode the error. Since the estimation error is also Gaussian, we have the rate-distortion function $R_{x|y}(D)$ is given by,

$$R_{x|y}(D) = \frac{1}{2} log \frac{\sigma_x^2(1-\rho^2)}{D}. \tag{10}$$

Now consider the communication system as shown in Fig. 1(b). This system corresponds encoding with side information available only at the decoder. Since Y is available only at the decoder, the rate-distortion function is defined [6] as follows:

$$R^*(D) = \inf_{p(W|X)} \inf_{f(Y,W)} [I(X;W) - I(W;Y)] \tag{11}$$

where the minimization is carried out over all borel measurable functions $f : Y \times W \longrightarrow \hat{X}$ such that $\hat{X} = f(Y,W)$ and conditional probability density function $p(W|X)$, such that $Y \longrightarrow X \longrightarrow W$ form a markov chain [6], and $E\left[(X-\hat{X})^2\right] \leq D$.

Clearly it can be seen that for any random variable pair (X,Y), $R_{x|y}(D) \leq R^*(D)$. But when X and Y are jointly Gaussian random variables, it can be shown that $R_{x|y}(D) = R^*(D)$. Thus the presence of Y at the encoder does not help it to reduce the rate of transmission (with Y available at the decoder). The optimum W and function $f(Y,W)$ are given by the following analysis. Consider the test channel for the communication system shown in Fig. 1(a) as shown in Fig. 4.

βY is the optimum linear estimate of X using Y and β is given by $\frac{\rho\sigma_x}{\sigma_y}$. Now Fig. 4 can be redrawn as shown in Fig. 5.

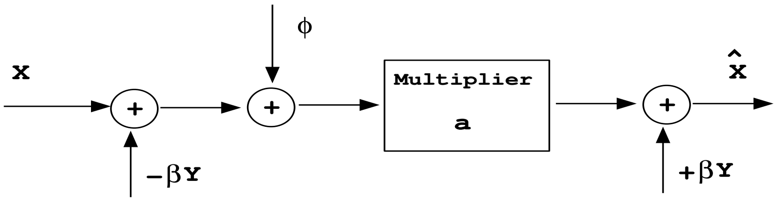

Figure 4: The test channel for the optimum encoding of the random variable X with the availability of the side information at both encoder and decoder.

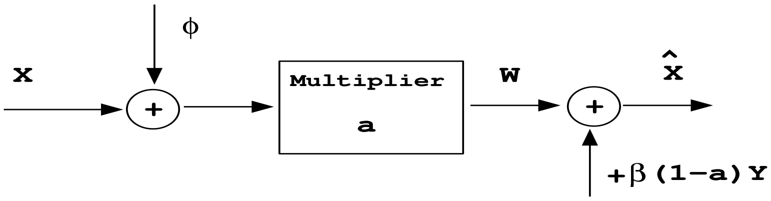

Figure 5: Fig. 4 is redrawn to show the test channel for the optimum encoding of X with the side information available only at the decoder.

Let W be the random variable as shown in the Fig. 5. This can be regarded as the test channel for the system of Fig. 1(b). It can be shown that

$$I(X;W) - I(W;Y) = \frac{1}{2} log \frac{\sigma_x^2(1-\rho^2)}{D} \tag{12}$$

which is the same as $R_{x|y}(D)$. $\hat{X} = f(Y,W)$ is given by $f(Y,W) = W + \beta(1-a).Y$.

3 Extension to sources having memory

Consider the communication system given in Fig.1(a). Let **x** and **y** be jointly Gaussian random vectors of dimension k, with autocorrelation matrix of **x** and **y** given by $\mathbf{R_x}$ and $\mathbf{R_y}$ respectively and cross correlation matrix of **x** and **y** given by $\mathbf{R_{xy}}$. Since **y** is available at both encoder and decoder, we can get the best estimate of **x** using **y** and encode the error vector. The error vector is given by **m**=(**x**-**By**) where **B** is as given by (1),(2) with autocorrelation matrix of **m** given by (4). Now we need to encode **m** optimally. The minimum rate required for optimum encoding is given by

$$R_{x|y}(D) = \sum_{i=1}^{k} \frac{1}{2} log \frac{\sigma_i^2}{D_i} \tag{13}$$

where

$$D_i = \begin{cases} \lambda, & if \quad \lambda < \sigma_i^2 \\ \sigma_i^2, & if \quad \lambda \geq \sigma_i^2 \end{cases} \tag{14}$$

where λ is such that $\sum_{i=1}^{k} D_i = D$ and σ_i^2 are the eigen values of the autocorrelation matrix of $\mathbf{m}$. Without loss of generality, for a given D, let us assume that $\sigma_i^2 > \lambda$, for $1 \leq i \leq l$, for some $l < k$ and $\sigma_i^2 \leq \lambda$, for $(l+1) \leq i \leq k$. Thus we have

$$R_{x|y}(D) = \frac{1}{2} log \frac{\sigma_1^2, \sigma_2^2, \ldots, \sigma_l^2}{\lambda^l} \tag{15}$$

The encoding can be described as given in Fig. 6.

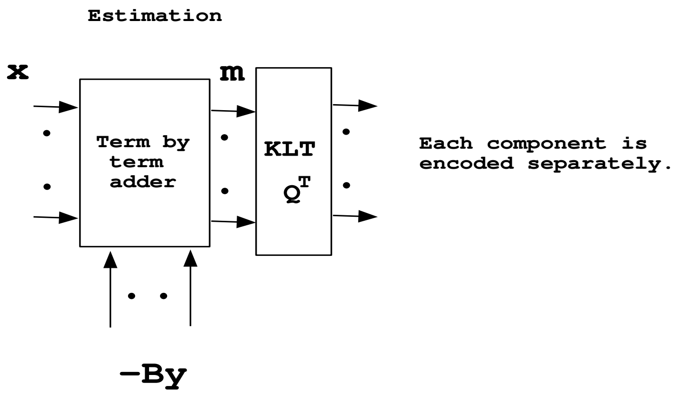

Figure 6: Optimum encoding of a Gaussian random vector $\mathbf{x}$ with the availability of the side information $\mathbf{y}$ at both the encoder and decoder.

Note that the transform $\mathbf{Q}^T$ is such that

$$\mathbf{\Lambda} = \mathbf{Q^T R_m Q}, \tag{16}$$

where $\mathbf{\Lambda} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$. The transform is the KL transform designed for decorrelating $\mathbf{m}$. Since independent components after the transform are encoded separately, the test channel for each of them can be given by the Fig. 7. It can be noted that the last $(k-l)$ components are discarded, as their variances are below the threshold given above. Thus the corresponding $\phi_i' s$ are assumed to have infinitely large variances denoted by say ζ. Thus we have $\hat{\mathbf{x}} = \mathbf{By} + \mathbf{Q\Theta}(\mathbf{Q^T m} + \mathbf{\Phi})$. Here $\mathbf{\Phi}$ is a vector of Gaussian, zero mean, random variables, such that $\mathbf{\Phi} = (\phi_1, \phi_2, \ldots, \phi_k)$ such that for $1 \leq i \leq l$, variance of $\phi_i = \frac{\lambda\sigma_i^2}{\sigma_i^2 - \lambda}$, and for $(l+1) \leq i \leq k$, the variance of $\phi_i = \zeta$. $\mathbf{\Theta}$ is a scaling diagonal matrix given by $\mathrm{diag}(a_1, a_2, \ldots, a_k)$ and for $1 \leq i \leq l$, $a_i = \frac{\sigma_i^2 - \lambda}{\sigma_i^2}$ and for $(l+1) \leq i \leq k$, a_i can be taken to be zero as these components of $\mathbf{m}$ are discarded. Fig. 7 can be redrawn as shown in Fig. 8. Let $\mathbf{w}$ be the random vector as shown in Fig. 8. This can be regarded as the test channel for the communication system of Fig. 1(b). Thus we have,

$$\mathbf{w} = \mathbf{Q\Theta Q^T x} + \mathbf{Q\Theta\Phi} \tag{17}$$

$$\hat{\mathbf{x}} = \mathbf{w} + (\mathbf{I} - \mathbf{Q\Theta Q^T})\mathbf{By}. \tag{18}$$

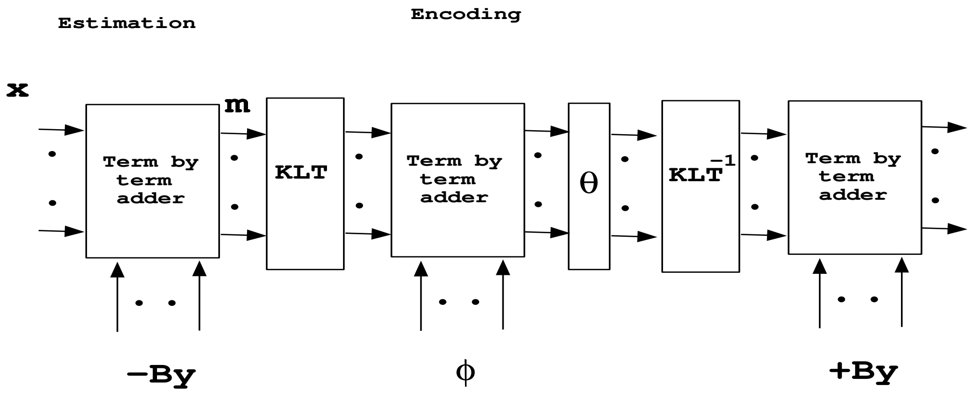

Figure 7: The test channel for the optimum encoding of Gaussian random vector $\mathbf{x}$ with the availability of the side information at both encoder and decoder.

Using (1),(2),(3),(4) and (5), the mutual information between the vectors $\mathbf{x}$ and $\mathbf{w}$

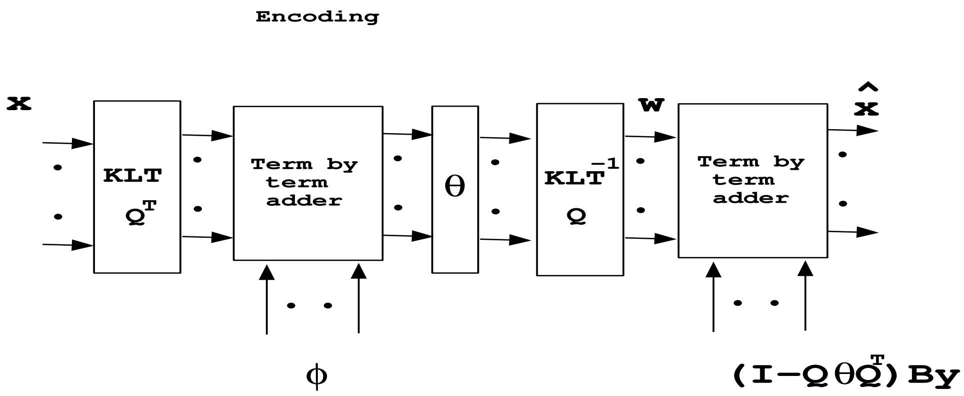

Figure 8: Fig. 7 is redrawn to show the test channel for the optimum encoding of Gaussian vector $\mathbf{x}$ with the availability of the side information only at the decoder.

and between $\mathbf{w}$ and $\mathbf{y}$ can be written as follows.

$$I(\mathbf{w};\mathbf{x}) = \frac{1}{2} log \frac{|\mathbf{R_w}|}{|\mathbf{Q\Theta R_\Phi \Theta Q^T}|} \tag{19}$$

$$I(\mathbf{w};\mathbf{y}) = \frac{1}{2} log \frac{|\mathbf{R_w}|}{|\mathbf{R_w} - \mathbf{R_{wy}R_y^{-1}R_{yw}}|} \tag{20}$$

where $\mathbf{R_w}$ is the autocorrelation matrix of $\mathbf{w}$ and $\mathbf{R_{yw}}$ is the cross correlation matrix of $\mathbf{y}$ and $\mathbf{w}$ and $\mathbf{R}_\phi$ is the autocorrelation matrix of $\mathbf{\Phi}$. It can be easily shown that

$$\mathbf{R_w} = \mathbf{Q\Theta Q^T R_x Q\Theta Q^T} + \mathbf{Q\Theta R_\phi \Theta Q^T} \tag{21}$$

$$\mathbf{R_{wy}R_y^{-1}R_{yw}} = \mathbf{Q\Theta Q^T R_{xy} R_y^{-1} R_{yx} Q\Theta Q^T} \tag{22}$$

We have

$$I(\mathbf{x};\mathbf{w}) - I(\mathbf{w};\mathbf{y}) = \frac{1}{2} log \frac{|\mathbf{Q\Theta(Q^T R_m Q} + \mathbf{R}_\phi)\mathbf{\Theta Q^T}|}{|\mathbf{Q\Theta R_\phi \Theta Q^T}|} \tag{23}$$

which can further be simplified using (16) as follows,

$$I(\mathbf{x};\mathbf{w}) - I(\mathbf{w};\mathbf{y}) = \frac{1}{2} log \frac{|\mathbf{\Lambda} + \mathbf{R}_\phi|}{|\mathbf{R}_\phi|} \tag{24}$$

Using (16) and the expression for $\mathbf{R}_\phi$, we obtain the following expression,

$$I(\mathbf{x};\mathbf{w}) - I(\mathbf{w};\mathbf{y}) = \frac{1}{2} log \frac{\sigma_1^2, \sigma_2^2, \ldots, \sigma_l^2}{\lambda^l} \tag{25}$$

Thus we have (see (15)) $R^*(D) = R_{x|y}(D)$. We have shown that the presence of $\mathbf{y}$ at the encoder does not help the encoder to reduce its rate of transmission (with $\mathbf{y}$ available at the decoder). The test channel shown in Fig. 8. has the following interpretation. We need to transform $\mathbf{x}$ through the KLT designed for the error spectrum $\mathbf{m}$ and need to retain only those components that are having large values along the directions of eigen functions of the autocorrelation matrix of $\mathbf{m}$. These ideas can be used in the constructive design of codes for the representation of $\mathbf{x}$ in the presence of side information $\mathbf{y}$ at the decoder.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[2] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 2–10, January 1974.

[3] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 294–300, May 1975.

[4] 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.

[5] 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, 1978.

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