Optimal Multiple Description Transform Coding of Gaussian Vectors

Vivek K Goyal
Dept. of Elec. Eng. & Comp. Sci.
University of California, Berkeley
v.goyal@ieee.org

Jelena Kovačević
Bell Laboratories
Murray Hill, NJ
jelena@bell-labs.com

Abstract

Multiple description coding (MDC) is source coding for multiple channels such that a decoder which receives an arbitrary subset of the channels may produce a useful reconstruction. Orchard *et al.* [1] proposed a transform coding method for MDC of pairs of independent Gaussian random variables. This paper provides a general framework which extends multiple description transform coding (MDTC) to any number of variables and expands the set of transforms which are considered. Analysis of the general case is provided, which can be used to numerically design optimal MDTC systems. The case of two variables sent over two channels is analytically optimized in the most general setting where channel failures need not have equal probability or be independent. It is shown that when channel failures are equally probable and independent, the transforms used in [1] are in the optimal set, but many other choices are possible. A cascade structure is presented which facilitates low-complexity design, coding, and decoding for a system with a large number of variables.

1 Introduction

For decades after the inception of information theory, techniques for source and channel coding developed separately. This was motivated both by Shannon's famous "separation principle" and by the conceptual simplicity of considering only one or the other. Recently, the limitations of separate source and channel coding has lead many researchers to the problem of designing joint source-channel (JSC) codes. An examination of Shannon's result leads to the primary motivating factor for constructing joint source-channel codes: The separation theorem is an asymptotic result which requires infinite block lengths (and hence infinite complexity and delay) at both source coder and channel coder; for a particular finite complexity or delay, one can often do better with a JSC code. JSC codes have also drawn interest for being robust to channel variation.

Multiple description transform coding is a technique which can be considered a JSC code for erasure channels. The basic idea is to introduce correlation between transmitted coefficients in a known, controlled manner so that erased coefficients can be statistically estimated from received coefficients. This correlation is used at the

decoder at the coefficient level, as opposed to the bit level, so it is fundamentally different from schemes that use information about the transmitted data to produce likelihood information for the channel decoder. The latter is a common element of JSC coding systems.

Our general model for multiple description coding is as follows: A source sequence $\{x_k\}$ is input to a coder, which outputs m streams at rates R_1, R_2, $\ldots R_m$. These streams are sent on m separate channels. There are many receivers, and each receives a subset of the channels and uses a decoding algorithm based on which channels it receives. Specifically, there are $2^m - 1$ receivers, one for each distinct subset of streams except for the empty set, and each experiences some distortion. (This is equivalent to communicating with a single receiver when each channel may be working or broken, and the status of the channel is known to the decoder but not to the encoder.)

This is a reasonable model for a lossy packet network.[1] Each "channel" corresponds to a packet or set of packets. Some packets may be lost, but because of header information it is known which packets are lost. An appropriate objective is to minimize a weighted sum of the distortions subject to a constraint on the total rate.

When $m = 2$, the situation is that studied in information theory as the *multiple description* problem [2, 3, 4]. Denote the distortions when both channels are received, only channel 1 is received, and only channel 2 is received by D_0, D_1, and D_2, respectively. The classical problem is to determine the achievable $(R_1, R_2, D_0, D_1, D_2)$-tuples. A complete characterization is known only for an i.i.d. Gaussian source and squared-error distortion [3].

This paper considers the case where $\{x_k\}$ is an i.i.d. sequence of zero-mean *jointly Gaussian vectors* with a known correlation matrix $R_x = E[x_k x_k^T]$.[2] Distortion is measured by the mean-squared error (MSE). The technique we develop is based on square, linear transforms and simple scalar quantization, and the design of the transform is paramount. Rather dissimilar methods have been developed which use nonsquare transforms [5]. The problem could also be addressed with an emphasis on quantizer design [6, 7].

2 Proposed Coding Structure

Since the source is jointly Gaussian, we can assume without loss of generality that the components are independent. If not, one can use a Karhunen-Loève transform of the source at the encoder and the inverse at each decoder. We propose the following steps for multiple description transform coding (MDTC) of a source vector x:

1. x is quantized with a uniform scalar quantizer with stepsize Δ: $x_{q_i} = [x_i]_\Delta$, where $[\cdot]_\Delta$ denotes rounding to the nearest multiple of Δ.

2. The vector $x_q = [x_{q_1},\ x_{q_2},\ \ldots x_{q_n}]^T$ is transformed with an invertible, discrete transform $\hat{T} : \Delta \mathbb{Z}^n \to \Delta \mathbb{Z}^n$, $y = \hat{T}(x_q)$. The design and implementation of $\hat{T}$ are described below.

---

[1]For example, the internet, when UDP is used as opposed to TCP.
[2]The vectors can be obtained by blocking a scalar Gaussian source.

3. The components of y are independently entropy coded.

4. If $m < n$, the components of y are grouped to be sent over the m channels.

When all the components of y are received, the reconstruction process is to (exactly) invert the transform $\hat{T}$ to get $\hat{x} = x_q$. The distortion is precisely the quantization error from Step 1. If some components of y are lost, they are estimated from the received components using the statistical correlation introduced by the transform $\hat{T}$. The estimate $\hat{x}$ is then generated by inverting the transform as before.

Starting with a linear transform T with determinant one, the first step in deriving a discrete version $\hat{T}$ is to factor T into "lifting" steps [8]. This means that T is factored into a product of upper and lower triangular matrices with unit diagonals $T = T_1 T_2 \cdots T_k$. The discrete version of the transform is then given by

$$\hat{T}(x_q) = \left[T_1 \left[T_2 \ldots [T_k x_q]_\Delta \right]_\Delta \right]_\Delta . \tag{1}$$

The lifting structure ensures that the inverse of $\hat{T}$ can be implemented by reversing the calculations in (1):

$$\hat{T}^{-1}(y) = \left[T_k^{-1} \ldots \left[T_2^{-1} \left[T_1^{-1} y \right]_\Delta \right]_\Delta \right]_\Delta .$$

The factorization of T is not unique; for example,

$$\begin{bmatrix} a & b \\ c & \frac{1+bc}{a} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1+bc-a}{ab} & 1 \end{bmatrix} \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{a-1}{b} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{a-1}{c} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1+bc-a}{ac} \\ 0 & 1 \end{bmatrix} . \tag{2}$$

Different factorizations yield different discrete transforms, except in the limit as Δ approaches zero.

The coding structure proposed here is a generalization of the method proposed by Orchard, *et al.* [1]. In [1], only 2×2 transforms implemented in two lifting steps were considered. (By fixing $a = 1$ in (2), both factorizations reduce to having two nonidentity factors.)

It is very important to note that we first quantize and then use a (discrete) transform. If we were to apply a (continuous) transform first and then quantize, the use of a nonorthogonal transform would lead to noncubic partition cells, which are inherently suboptimal among the class of partition cells obtainable with scalar quantization [9]. The present configuration allows one to use discrete transforms derived from *nonorthogonal* linear transforms, and thus obtain better performance [1].

3 Analysis of an MDTC System

The analysis and optimizations presented in this paper are based on fine quantization approximations. Specifically, we make three assumptions which are valid for small Δ: First, we assume that the scalar entropy of $y = \hat{T}([x]_\Delta)$ is the same as that of $[Tx]_\Delta$. Second, we assume that the correlation structure of y is unaffected by the

quantization. Finally, when at least one component of y is lost, we assume that the distortion is dominated by the effect of the erasure, so quantization can be ignored.

Denote the variances of the components of x by σ_1^2, σ_2^2, ..., σ_n^2 and denote the correlation matrix of x by $R_x = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. Let $R_y = TR_xT^T$. In the absence of quantization, R_y would be exactly the correlation matrix of y. Under our fine quantization approximations, we will use R_y in the estimation of rates and distortions.

Estimating the rate is straightforward. Since the quantization is fine, y_i is approximately the same as $[(Tx)_i]_\Delta$, $i.e.$, a uniformly quantized Gaussian random variable. If we treat y_i as a Gaussian random variable with power $\sigma_{y_i}^2 = (R_y)_{ii}$ quantized with bin width Δ, we get for the entropy of the quantized coefficient [10, Ch. 9]

$$H(y_i) \approx \frac{1}{2}\log 2\pi e \sigma_{y_i}^2 - \log \Delta = \frac{1}{2}\log \sigma_{y_i}^2 + \frac{1}{2}\log 2\pi e - \log \Delta = \frac{1}{2}\log \sigma_{y_i}^2 + k_\Delta,$$

where $k_\Delta \triangleq (\log 2\pi e)/2 - \log \Delta$ and all logarithms are base-two. Notice that k_Δ depends only on Δ. We thus estimate the total rate as

$$R = \sum_{i=1}^{n} H(y_i) = nk_\Delta + \frac{1}{2}\log \prod_{i=1}^{n} \sigma_{y_i}^2. \tag{3}$$

The minimum rate occurs when $\prod_{i=1}^{n} \sigma_{y_i}^2 = \prod_{i=1}^{n} \sigma_i^2$ and at this rate the components of y are uncorrelated. Interestingly, $T = I$ is not the only transform which achieves the minimum rate. In fact, an arbitrary split of the total rate among the different components of y is possible. This is a justification for using a total rate constraint in our following analyses. However, we will pay particular attention to the case where the rates sent across each channel are equal.

We now turn to the distortion, and first consider the average distortion due only to quantization. Since the quantization noise is approximately uniform, this distortion is $\Delta^2/12$ for each component. Thus the distortion when no components are erased is given by

$$D_0 = \frac{n\Delta^2}{12} \tag{4}$$

and is independent of T.

Now consider the case when $\ell > 0$ components are lost. We first must determine how the reconstruction should proceed. By renumbering the variables if necessary, assume that $y_1, y_2, \ldots, y_{n-\ell}$ are received and $y_{n-\ell+1}, \ldots, y_n$ are lost. Partition y into "received" and "not received" portions as $y = [\tilde{y}_\mathrm{r}, \tilde{y}_\mathrm{nr}]^T$ where $\tilde{y}_\mathrm{r} = [y_1, y_2, \ldots, y_{n-\ell}]^T$ and $\tilde{y}_\mathrm{nr} = [y_{n-\ell+1}, \ldots, y_{n-1}, y_n]^T$. The minimum MSE estimate of x given $\tilde{y}_\mathrm{r}$ is $E[x|\tilde{y}_\mathrm{r}]$, which has a simple closed form because x is a jointly Gaussian vector. Using the linearity of the expectation operator gives the following sequence of calculations:

$$\begin{aligned}
\hat{x} &= E[x|\tilde{y}_\mathrm{r}] = E[T^{-1}Tx|\tilde{y}_\mathrm{r}] = T^{-1}E[Tx|\tilde{y}_\mathrm{r}] \\
&= T^{-1}E\left[\left.\begin{bmatrix} \tilde{y}_\mathrm{r} \\ \tilde{y}_\mathrm{nr} \end{bmatrix}\right| \tilde{y}_\mathrm{r}\right] = T^{-1}\begin{bmatrix} \tilde{y}_\mathrm{r} \\ E[\tilde{y}_\mathrm{nr}|\tilde{y}_\mathrm{r}] \end{bmatrix}.
\end{aligned} \tag{5}$$

4

If the correlation matrix of y is partitioned in a way compatible with the partition of y as

$$R_y = TR_xT^T = \begin{bmatrix} R_1 & B \\ B^T & R_2 \end{bmatrix},$$

then it can be shown that $\tilde{y}_{\mathrm{nr}}|\tilde{y}_{\mathrm{r}}$ is Gaussian with mean $B^T R_1^{-1} \tilde{y}_{\mathrm{r}}$ and correlation matrix $A \triangleq R_2 - B^T R_1^{-1} B$. Thus $E[\tilde{y}_{\mathrm{nr}}|\tilde{y}_{\mathrm{r}}] = B^T R_1^{-1} \tilde{y}_{\mathrm{r}}$, and $\eta \triangleq \tilde{y}_{\mathrm{nr}} - E[\tilde{y}_{\mathrm{nr}}|\tilde{y}_{\mathrm{r}}]$ is Gaussian with zero mean and correlation matrix A. η is the error in predicting $\tilde{y}_{\mathrm{nr}}$ from $\tilde{y}_{\mathrm{r}}$ and hence is the error caused by the erasure. However, because we have used a nonorthogonal transform, we must return to the original coordinates using T^{-1} in order to compute the distortion. Substituting $\tilde{y}_{\mathrm{nr}} + \eta$ for $E[\tilde{y}_{\mathrm{nr}}|\tilde{y}_{\mathrm{r}}]$ in (5) gives

$$\hat{x} = T^{-1} \begin{bmatrix} \tilde{y}_{\mathrm{r}} \\ \tilde{y}_{\mathrm{nr}} + \eta \end{bmatrix} = x + T^{-1} \begin{bmatrix} 0 \\ \eta \end{bmatrix}, \quad \text{so} \quad \|x - \hat{x}\|^2 = \left\| T^{-1} \begin{bmatrix} 0 \\ \eta \end{bmatrix} \right\|^2 = \eta^T U^T U \eta,$$

where U is the last ℓ columns of T^{-1}. Finally,

$$E\|x - \hat{x}\|^2 = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (U^T U)_{ij} A_{ij}. \tag{6}$$

We denote the distortion with ℓ erasures by D_ℓ. To determine D_ℓ we must now average (6) over the $\binom{n}{\ell}$ possible erasures of ℓ components, weighted by their probabilities if necessary.

Our final distortion criterion is a weighted sum of the distortions incurred with different numbers of channels available:

$$\bar{D} = \sum_{\ell=0}^{n} \alpha_\ell D_\ell.$$

For the case where each channel has an outage probability of p and the channel outages are independent, the weighting $\alpha_\ell = \binom{n}{\ell} p^\ell (1-p)^{n-\ell}$ makes $\bar{D}$ the overall expected MSE. However, there are certainly other reasonable choices for the weights. Consider an image coding scenario when an image is split over ten packets. One might want acceptable image quality as long as eight or more packets are received. In this case, one should set $\alpha_3 = \alpha_4 = \cdots = \alpha_{10} = 0$.

For a given rate R, our goal is to minimize $\bar{D}$. The expressions given in this section can be used to numerically determine transforms to realize this goal. Analytical solutions are possible in certain special cases. Some of these are given in the following section.

4   Sending Two Variables Over Two Channels

General case   Let us now apply the analysis of the previous section to find the best transforms for sending $n = 2$ variables over $m = 2$ channels. In the most general situation, channel outages may have unequal probabilities and may be dependent. Suppose the probabilities of the combinations of channel states are given by the following table:

		Channel 1	
		broken	working
Channel 2	broken	$1 - p_0 - p_1 - p_2$	p_1
	working	p_2	p_0

Let $T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, normalized so that $\det T = 1$. Then $T^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ and

$$R_y = T R_x T^T = \begin{bmatrix} a^2\sigma_1^2 + b^2\sigma_2^2 & ac\sigma_1^2 + bd\sigma_2^2 \\ ac\sigma_1^2 + bd\sigma_2^2 & c^2\sigma_1^2 + d^2\sigma_2^2 \end{bmatrix}.$$

By (3), the total rate is given by

$$R = 2k_\Delta + \frac{1}{2}\log(R_y)_{11}(R_y)_{22} = 2k_\Delta + \frac{1}{2}\log(a^2\sigma_1^2 + b^2\sigma_2^2)(c^2\sigma_1^2 + d^2\sigma_2^2). \qquad (7)$$

Minimizing (7) over transforms with determinant one gives a minimum possible rate of $R^* = 2k_\Delta + \log\sigma_1\sigma_2$. We refer to $\rho = R - R^*$ as the *redundancy* [1], *i.e.*, the price we pay in rate in order to potentially reduce the distortion when there are erasures.

In order to evaluate the overall average distortion, we must form a weighted average of the distortions for each of the four possible channel states. If both channels are working, the distortion (due to quantization only) is $D_0 = \Delta^2/6$. If neither channel is working, the distortion is $D_2 = \sigma_1^2 + \sigma_2^2$. The remaining cases require the application of the results of the previous section. We first determine $D_{1,1}$, the MSE distortion when y_1 is received but y_2 is lost. Substituting in (6),

$$D_{1,1} = \underbrace{\left\| \begin{bmatrix} -b \\ a \end{bmatrix} \right\|^2}_{(U^T U)_{1,1}} \cdot \underbrace{\left((R_y)_{22} - \frac{(R_y)_{12}^2}{(R_y)_{11}} \right)}_{A_{1,1}} = (a^2 + b^2) \cdot \frac{\sigma_1^2 \sigma_2^2}{a^2\sigma_1^2 + b^2\sigma_2^2},$$

where we have used $\det T = ad - bc = 1$ in the simplification. Similarly, the distortion when y_2 is received but y_1 is lost is $D_{1,2} = (c^2 + d^2)\sigma_1^2\sigma_2^2/(c^2\sigma_1^2 + d^2\sigma_2^2)$. The overall average distortion is

$$\begin{aligned}
\bar{D} &= p_0 \cdot D_0 + p_1 \cdot D_{1,1} + p_2 \cdot D_{1,2} + (1 - p_0 - p_1 - p_2) \cdot D_2 \\
&= \left[p_0 \frac{\Delta^2}{6} + (1 - p_0 - p_1 - p_2)(\sigma_1^2 + \sigma_2^2) \right] + p_2 \underbrace{\left[\frac{p_1}{p_2} D_{1,1} + D_{1,2} \right]}_{\bar{D}'},
\end{aligned}$$

where the first bracketed term is independent of T. Thus our optimization problem is to minimize $\bar{D}'$ for a given redundancy ρ.

If the source has a circularly symmetric probability density, *i.e.*, $\sigma_1 = \sigma_2$, then $\bar{D}' = (1 + p_1/p_2)\sigma_1^2$ independent of T. Henceforth we assume $\sigma_1 > \sigma_2$.

After eliminating d through $d = (1 + bc)/a$, one can show that the optimal transform will satisfy

$$|a| = \frac{\sigma_2}{2c\sigma_1}\left[\sqrt{2^{2\rho} - 1} + \sqrt{2^{2\rho} - 1 - 4bc(bc + 1)} \right].$$

Furthermore, $\bar{D}'$ depends only on the product $b \cdot c$, not on the individual values of b and c. The optimal value of bc is given by

$$(bc)_{\text{optimal}} = -\frac{1}{2} + \frac{1}{2}\left(\frac{p_1}{p_2} - 1\right)\left[\left(\frac{p_1}{p_2} + 1\right)^2 - 4\left(\frac{p_1}{p_2}\right)2^{-2\rho}\right]^{-1/2}.$$

It is easy to check that $(bc)_{\text{optimal}}$ ranges from -1 to 0 as p_1/p_2 ranges from 0 to ∞. The limiting behavior can be explained as follows: Suppose $p_1 \gg p_2$, *i.e.*, channel 1 is much more reliable than channel 1. Since $(bc)_{\text{optimal}}$ approaches 0, *ad* must approach 1, and hence one optimally sends x_1 (the larger variance component) over channel 1 (the more reliable channel), and vice-versa. This is the intuitive, layered solution. The multiple description approach is most useful when the channel failure probabilities are comparable, but this demonstrates that the multiple description framework subsumes layered coding.

Equal channel failure probabilities   If $p_1 = p_2$, then $(bc)_{\text{optimal}} = -1/2$, independent of ρ. The optimal set of transforms is described by

$$
\begin{aligned}
a &\neq 0 \text{ (but otherwise arbitrary)} & c &= -1/2b \\
b &= \pm(2^\rho - \sqrt{2^{2\rho} - 1})\sigma_1 a/\sigma_2 & d &= 1/2a
\end{aligned}
\tag{8}
$$

and using a transform from this set gives

$$D_1 = \frac{1}{2}(D_{1,1} + D_{1,2}) = \sigma_1^2 - \frac{1}{2 \cdot 2^\rho \left(2^\rho - \sqrt{2^{2\rho} - 1}\right)}(\sigma_1^2 - \sigma_2^2). \tag{9}$$

This relationship is plotted in Figure 1(a). Notice that, as expected, D_1 starts at a maximum value of $(\sigma_1^2 + \sigma_2^2)/2$ and asymptotically approaches a minimum value of $\sigma_2^2/2$. By combining (3), (4), and (9), one can find the relationship between R, D_0, and D_1. For various values of R, the trade-off between D_0 and D_1 is plotted in Figure 1(b).

The solution for the optimal set of transforms (8) has an interesting property that after fixing ρ, there is an "extra" degree of freedom which does not affect the ρ vs. D_1 performance. This degree of freedom can be used to control the partitioning of the rate between channels or to give a simplified implementation.

Optimality of Orchard *et al.* transforms   In [1] it is suggested to use transforms of the form $\begin{bmatrix} 1 & b \\ -1/(2b) & 1/2 \end{bmatrix}$. As a result of our analysis we conclude that these transforms in fact lie in the optimal set of transforms. The "extra" degree of freedom has been used by fixing $a = 1$, which yields a transform which can be factored into two lifting steps; in the general case three lifting steps are needed.

Optimal transforms that give balanced rates   The transforms of [1] do not give channels with equal rate (or, equivalently, power). In practice, this can be remedied
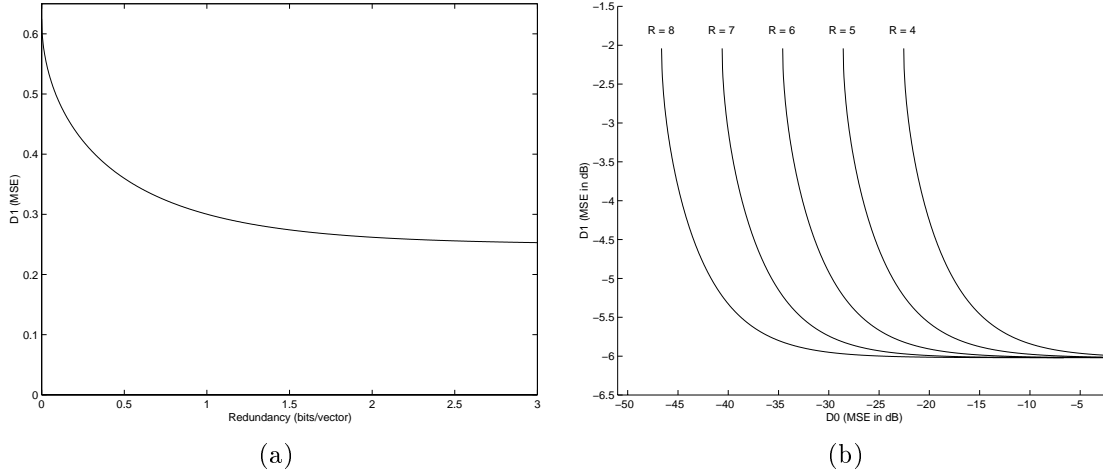
(a)   (b)

Figure 1: Optimal R-D_0-D_1 trade-offs for $\sigma_1 = 1$, $\sigma_2 = 0.5$: (a) Relationship between redundancy ρ and D_1; (b) Relationship between D_0 and D_1 for various rates.

through time-multiplexing. An alternative is to use the "extra" degree of freedom to make $R_1 = R_2$. Doing this is equivalent to requiring $|a| = |c|$ and $|b| = |d|$, and yields

$$a = \pm\sqrt{\frac{\sigma_2}{2\sigma_1\left(2^\rho - \sqrt{2^{2\rho} - 1}\right)}}, \qquad b = \pm\frac{1}{2a} = \pm\sqrt{\frac{\sigma_1\left(2^\rho - \sqrt{2^{2\rho} - 1}\right)}{2\sigma_2}}.$$

In the next section, when we apply a two-by-two correlating transform, we will assume a balanced-rate transform. Specifically, we will use $T_a \triangleq \begin{bmatrix} a & 1/(2a) \\ -a & 1/(2a) \end{bmatrix}$.

Geometric interpretation The transmitted representation of x is given by $y_1 = \langle x, \varphi_1 \rangle$ and $y_2 = \langle x, \varphi_2 \rangle$, where $\varphi_1 = [a, \ b]^T$ and $\varphi_2 = [c, \ d]^T$. In order to gain some insight into the vectors φ_1 and φ_2 that result in an optimal transform, let us neglect the rate and distortion that are achieved, and simply consider the transforms described by $ad = 1/2$ and $bc = -1/2$. We can show that φ_1 and φ_2 form the same (absolute) angles with the positive x_1-axis (see Figure 2(a)). For convenience, suppose $a > 0$ and $b < 0$. Then $c, d > 0$. Let θ_1 and θ_2 be the angles by which φ_1 and φ_2 are below and above the positive x_1-axis, respectively. Then $\tan\theta_1 = -b/a = d/c = \tan\theta_2$. If we assume $\sigma_1 > \sigma_2$, then the maximum angle (for $\rho = 0$) is $\arctan(\sigma_1/\sigma_2)$ and the minimum angle (for $\rho \to \infty$) is zero. This has the nice interpretation of emphasizing x_1 over x_2—because it has higher variance—as the coding rate is increased (see Figure 2(b)).

5   Three or More Variables

Three variables over three channels Applying the results of Section 3 to the design of 3×3 transforms is considerably more complicated than what has been
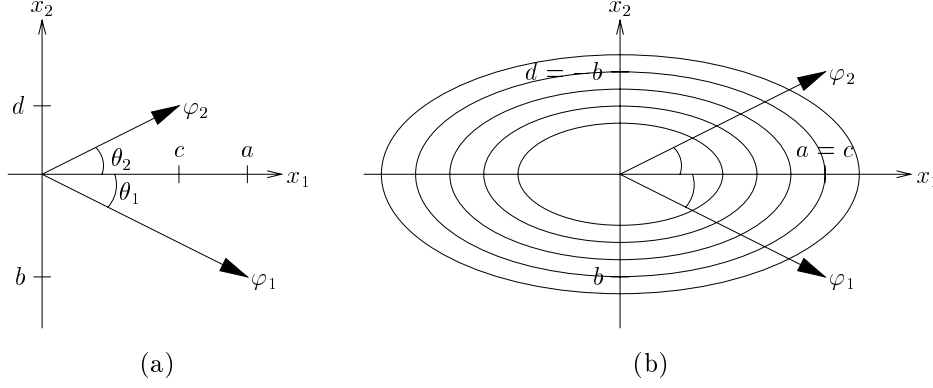
8

Figure 2: Geometric interpretations. (a) When $\sigma_1 > \sigma_2$, the optimality condition ($ad = 1/2$, $bc = -1/2$) is equivalent to $\theta_1 = \theta_2 < \theta_{\max} = \arctan(\sigma_1/\sigma_2)$. (b) If in addition to the optimality condition we require the output streams to have equal rate, the analysis vectors are symmetrically situated to capture the dimension with greatest variation. At $\rho = 0$, $\theta_1 = \theta_2 = \theta_{\max}$; as $\rho \to \infty$, φ_1 and φ_2 close on the x_1-axis.

presented thus far. Even in the case of equal channel failures, a closed form solution will be much more complicated that (8). When $\sigma_1 > \sigma_2 > \sigma_3$ and erasure probabilities are equal and small, a set of transforms which gives near optimal performance is described by

$$
\begin{bmatrix}
a & -\dfrac{\sqrt{3}\sigma_1 a}{\sigma_2} & -\dfrac{\sigma_2}{6\sqrt{3}\sigma_1^2 a^2} \\[2ex]
2a & 0 & \dfrac{\sigma_2}{6\sqrt{3}\sigma_1^2 a^2} \\[2ex]
a & \dfrac{\sqrt{3}\sigma_1 a}{\sigma_2} & -\dfrac{\sigma_2}{6\sqrt{3}\sigma_1^2 a^2}
\end{bmatrix}.
$$

A derivation of this set must be omitted for lack of space.

Four and more variables   For sending any even number of variables over two channels, Orchard *et al.* [1] have suggested the following: form pairs of variables, add correlation within each pair, and send one variable from each pair across each channel. A necessary condition for optimality is that all the pairs are operating at the same distortion-redundancy slope. If T_α is used to transform variables with variances σ_1^2 and σ_2^2 and T_β is used to transform variables with variances σ_3^2 and σ_4^2, then the equal-slope condition implies that we should have

$$
\beta^4 = \frac{\gamma(16\alpha^8\sigma_1^4 - \sigma_2^4) + \sqrt{\gamma^2(16\alpha^8\sigma_1^4 - \sigma_2^4)^2 + 64\alpha^8\sigma_3^4\sigma_4^4}}{32\alpha^4\sigma_3^4}, \text{ where } \gamma = \frac{\sigma_3^2\sigma_4^2(\sigma_3^2 - \sigma_4^2)}{\sigma_1^2\sigma_2^2(\sigma_1^2 - \sigma_2^2)}.
$$

Finding the optimal transform under this pairing constraint still requires finding the optimal pairing.

Cascade Structures   In order to extend these schemes to an arbitrary number of channels while maintaining reasonable ease of design, we propose the cascade use of
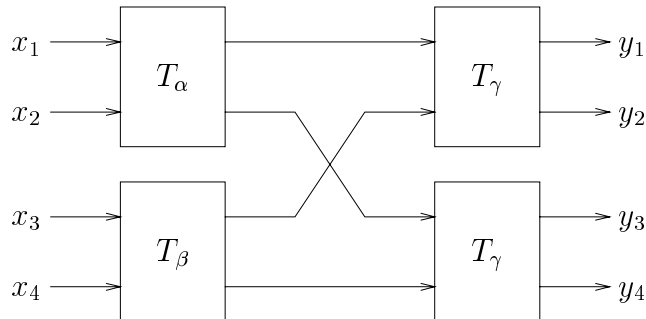
9

Figure 3: Cascade structure allows efficient MDTC of large vectors.

pairing transforms as shown in Figure 3. The cascade structure simplifies the encoding, decoding (both with and without erasures), and design when compared to using a general $n \times n$ transform. Empirical evidence suggests that for $n = m = 4$ and considering up to one component erasure, there is no performance penalty in restricting consideration to cascade structures. This phenomenon is under investigation. As expected, there is great improvement over simple pairing of coefficients; pairing can not be expected to be near optimal for m larger than two.

References

[1] M. T. Orchard, Y. Wang, V. Vaishampayan, and A. R. Reibman. Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms. In *Proc. IEEE Int. Conf. Image Proc. 1997*.

[2] J. K. Wolf, A. D. Wyner, and J. Ziv. Source coding for multiple descriptions. *Bell Syst. Tech. J.*, 59(8):1417–1426, 1980.

[3] L. Ozarow. On a source-coding problem with two channels and three receivers. *Bell Syst. Tech. J.*, 59(10):1909–1921, 1980.

[4] A. A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Info. Theory*, IT-28(6):851–857, 1982.

[5] V. K Goyal, J. Kovačević, and M. Vetterli. Multiple description transform coding: Robustness to erasures using tight frame expansions. In *Proc. IEEE Int. Symp. Info. Th.*, August 1998. Submitted.

[6] V. A. Vaishampayan. Design of multiple description scalar quantizers. *IEEE Trans. Info. Theory*, 39(3):821–824, 1993.

[7] J.-C. Batllo and V. A. Vaishampayan. Asymptotic performance of multiple description transform codes. *IEEE Trans. Info. Theory*, 43(2):703–707, 1997.

[8] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. Technical report, Bell Laboratories, Lucent Technologies, September 1996.

[9] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Acad. Pub., Boston, MA, 1992.

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