

On Linear Transforms in Zero-delay Gaussian Source Channel Coding

Emrah Akyol and Kenneth Rose

{eakyol, rose}@ece.ucsb.edu

University of California, Santa Barbara, CA-93106

Abstract—This paper is concerned with the optimal linear transforms in zero-delay source channel coding for Gaussian sources and channels. In preparation for the main results, we first consider the classical problem in the point-to-point setting, which had already been solved by Lee and Petersen, where we provide an alternative proof using majorization. We then analyze the performance of linear source-channel coding in the low signal-to-noise ratio (SNR) regime. We show that at asymptotically low SNR, or equivalently as the power tends to zero, the power-distortion function of zero delay linear coding achieves optimum theoretically achievable performance. Finally, we consider zero-delay source-channel coding with decoder side information. Here, subject to structural constraints on the encoder, we find the optimal encoder-decoder pair in closed form. We analyze two structures: i) the encoder is constrained to be linear where we show that the optimal transform is product of matrices including as factor the conditional Karhunen Loève transform (KLT) of the source given the side information. ii) the encoder consists of linear transformation followed by individual optimal nonlinear mapping of each transform coefficient. Using majorization principles, we show that the optimal transform does not depend on the nonlinearities introduced, as long as they are scale invariant, hence the optimal transform is also a product involving the conditional KLT of the source.

Index Terms—Linear transforms, zero delay source-channel coding

I. INTRODUCTION

Zero delay source channel coding has attracted much recent interest [1]–[6] due to its low complexity and zero delay advantages, and also due to the fact that it provides graceful degradation with channel signal to noise ratio (SNR) mismatch. Most prior efforts were focused on mapping optimization at low dimensions, eg., mapping two source samples to one channel symbol, due to the complexity of such optimizations. In this work, we consider tractable generalization to multi-dimensional settings via structured mappings.

The problem of zero delay linear source-channel codes was previously studied by Lee and Petersen [7], where the optimal encoding transform was found to be the product of three matrices, the Karhunen Loève transform (KLT) of the source, a diagonal scaling matrix, and the KLT of the channel. We first revisit this problem from a different perspective and provide an alternative proof, based on the majorization theory, that gives new insights on this problem and its extensions.

One of our main results pertains to the asymptotic optimality of zero-delay linear coding at the low SNR regime.

This work is supported in part by the NSF under grants CCF-0728986, CCF-1016861 and CCF-1118075.

A motivating factor for considering the low SNR regime in communications, is that the absolute value of the slope of the capacity-cost function is large, which indicates that one achieves the most channel capacity per unit cost at low SNR, as was shown by Verdu [8]. In [9], it was shown that at low SNR, the optimal estimator converges to linear, for a Gaussian channel, regardless of the density of the source. Recently, Ostergaard and Zamir combined the results in [9], and the incremental channel method introduced in [10], to analyze the additive rate distortion function [11] at the low SNR regime. Our main result here establishes the first order optimality of linear coding at low SNR. We show that asymptotically, at low channel SNR, the power-distortion function of the zero delay linear coding approaches the unbeatable power-distortion bound. In other words, zero-delay linear coding is asymptotically (at low SNR) optimal. To the best of our knowledge, this perhaps surprising fact has not been previously known.

Next, we study zero-delay linear source-channel coding within the decoder side information setting. We show that the optimal encoding transform is the product of the “conditional” KLT of the source, a diagonal power allocation matrix and the KLT of the channel. It is a “conditional” relative of the seminal result in the point-to-point setting.

Finally, we analyze the optimal linear transforms for a structured nonlinear zero-delay mapping: a linear code followed by nonlinear “scalar” mappings for each transform coefficient (channel). We show, using majorization tools, that the optimal linear transform remains the same, and hence surprisingly does not depend on the subsequent nonlinearities, as long as these nonlinear mappings are scale invariant.

This paper is organized as follows: In Section II, we provide a brief review of the known results in matrix analysis and majorization theory. In Section III, we present an alternative proof for the well known point-to-point result. In Section IV, we study the optimality of the linear schemes. Section V presents the results in the decoder side information setting. Discussions are presented in Section VI.

II. PRELIMINARIES

A. Notation

In general, lowercase letters (e.g., c) denote scalars, boldface lowercase (e.g., \mathbf{x}) vectors, uppercase (e.g., C, X) matrices and random variables. I denotes the identity matrix. $\mathbb{E}[\cdot]$, R_X , and R_{XZ} denote the expectation, auto-covariance of X and

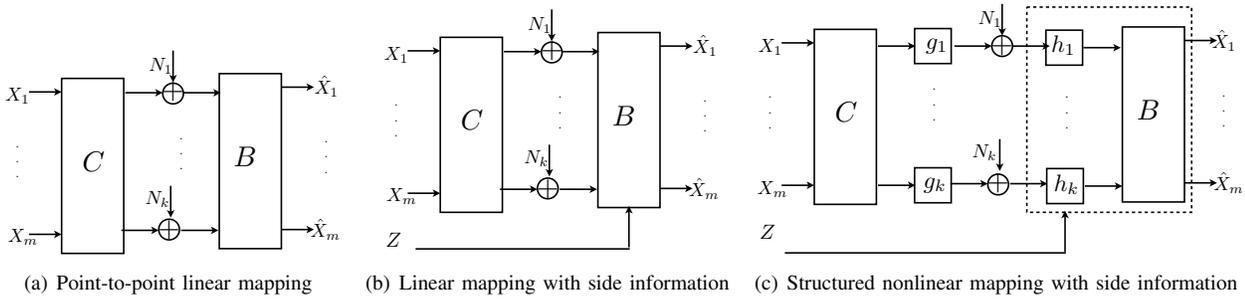


Fig. 1. Problem settings

cross covariance of \mathbf{X} and \mathbf{Z} respectively. A^T denotes the transpose of matrix A . $(x)^+$ denotes the function $\max(0, x)$.

Throughout this paper, we assume that the source is an m -dimensional Gaussian with zero mean and covariance R_X . The channel noise is additive k -dimensional Gaussian, of zero mean and covariance R_N . Covariance matrices R_X and R_N allow the diagonalization

$$R_X = Q_X \Lambda_X Q_X^T, \text{ and } R_N = Q_N \Lambda_N Q_N^T \quad (1)$$

where $Q_X Q_X^T = Q_N Q_N^T = I$ and Λ_X and Λ_N are diagonal matrices, having ordered eigenvalues as entries, i.e., $\Lambda_X = \text{diag}\{\lambda_X(1), \lambda_X(2), \dots, \lambda_X(m)\}$ and $\Lambda_N = \text{diag}\{\lambda_N(1), \lambda_N(2), \dots, \lambda_N(k)\}$ where $\lambda_X(1) \geq \lambda_X(2) \geq \dots \geq \lambda_X(m)$ and $\lambda_N(1) \geq \lambda_N(2) \geq \dots \geq \lambda_N(k)$. Q_X^T and Q_N^T are the KLT matrices of the source and the channel, respectively.

Similarly, let us define the covariance of the prediction error of the optimal linear predictor of \mathbf{X} given observation \mathbf{Z}

$$R_{X|Z} \triangleq R_X - R_{XZ} R_Z^{-1} R_{XZ}^T \quad (2)$$

Since $R_{X|Z}$ is a covariance matrix, it allows the eigendecomposition:

$$R_{X|Z} = Q_{X|Z} \Lambda_{X|Z} Q_{X|Z}^T \quad (3)$$

where $Q_{X|Z} Q_{X|Z}^T = Q_{X|Z}^T Q_{X|Z} = I$ and $\Lambda_{X|Z}$ is a diagonal matrix with ordered eigenvalues as entries. $Q_{X|Z}^T$ is also referred as conditional KLT of $\mathbf{X} \in \mathbb{R}^m$ with respect to $\mathbf{Z} \in \mathbb{R}^k$ [12].

We repeatedly use the simple trace identity

$$\text{Tr}(ABC) = \text{Tr}(CAB) = \text{Tr}(BCA) \quad (4)$$

where A, B, C are matrices of appropriately matching dimensions.

B. Majorization Theory

This paper builds on the results from majorization theory. Let us introduce the basic notion of majorization and state some important auxiliary results that will be of use later (see [13] for complete reference on majorization). Roughly speaking majorization measures how spread out the components of a vector is. Let $\mathbf{x} \in \mathbb{R}^m$ be such that $x(1) \geq x(2) \geq \dots \geq x(m)$.

Majorization: $\mathbf{y} \in \mathbb{R}^m$ majorizes \mathbf{x} if and only if

$$\sum_{i=1}^k y(i) \geq \sum_{i=1}^k x(i), \quad 1 \leq k \leq m \quad (5)$$

$$\sum_{i=1}^m y(i) = \sum_{i=1}^m x(i), \quad (6)$$

and denoted as $\mathbf{y} \succ \mathbf{x}$.

Schur concave functions: A real valued function f is said to be Schur-concave if and only if

$$\mathbf{x} \succ \mathbf{y} \Rightarrow f(\mathbf{x}) \leq f(\mathbf{y}) \quad (7)$$

We reproduce the following well known result, without proof, see eg. [13].

Lemma 1. *Let be A a Hermitian matrix with ordered diagonal elements denoted by the vector \mathbf{a} and ordered eigenvalues denoted by the vector $\boldsymbol{\lambda}$. Then $\boldsymbol{\lambda} \succ \mathbf{a}$.*

C. A Matrix Analysis Result

In this section, we introduce a matrix analysis result.

Lemma 2. *The solution to the minimization problem:*

$$K^* = \min_{\mathcal{K}} \text{Tr} \left\{ \Lambda_X - \Lambda_X K^T (K K^T + I)^{-1} K \right\} \quad (8)$$

where \mathcal{K} is the space of $k \times m$ matrices K that satisfy the constraint

$$\text{Tr} \{ K^T \Lambda_N K \Phi \} \leq P \quad (9)$$

where Φ is a symmetric, positive definite $m \times m$ matrix, is a diagonal $k \times m$ matrix.

Proof: Since variants of this result appeared before [14], we only provide a sketch of the proof for brevity. We first note, using the matrix inversion lemma [15], that

$$I - K^T (K K^T + I)^{-1} K = (I + K^T K)^{-1} \quad (10)$$

Given the fact that $\text{Tr} \{ \Lambda_X (I + K^T K)^{-1} \}$ is a Schur-concave function of the vector consisting the diagonal elements of K and that the constraint is convex, it follows that the objective is minimized when K is diagonal, due to Lemma 1 [14]. ■

III. POINT TO POINT SETTING

The problem setting, depicted in Figure 1-a, can be stated as the following optimization problem: find the encoder and decoder matrices, C and B that minimize distortion, $D = \mathbb{E}\{\|\mathbf{X} - \hat{\mathbf{X}}\|^2\}$, subject to the power constraint $\mathbb{E}\{\|\mathbf{Y}\|^2\} \leq P$ where $\mathbf{Y} = C\mathbf{X}$. Recall that this is the classical problem solved in the seminal paper of Lee and Petersen [7]. Our approach gives an alternative proof as well as some new results. Also note that a general framework for solving such convex problems was presented in [14].

Theorem 1. *The encoding transform that minimizes the MSE distortion subject to the power constraint P is*

$$C = Q_N \Sigma Q_X^T \quad (11)$$

where Σ is diagonal power allocation matrix. Moreover the total distortion is

$$D = \frac{\left(\sum_{i=1}^w (\sqrt{\lambda_X(i)\lambda_N(i)})\right)^2}{P + \sum_{i=1}^w \lambda_N(i)} + \sum_{i=w+1}^m \lambda_X(i) \quad (12)$$

where w is the number of active channels determined by the power P .

Proof: Using orthogonality, $\mathbb{E}\{(\mathbf{X} - \hat{\mathbf{X}})\hat{\mathbf{X}}^T\} = \mathbf{0}$ we have

$$D = \mathbb{E}\left\{Tr\{(\mathbf{X} - \hat{\mathbf{X}})\mathbf{X}^T\}\right\} \quad (13)$$

Noting that the optimal decoder is $\hat{\mathbf{X}} = R_{X\hat{Y}}R_{\hat{Y}}^{-1}\hat{\mathbf{Y}}$. Plugging $\hat{\mathbf{Y}} = C\mathbf{X} + N$, we have

$$R_{X\hat{Y}} = R_X C^T \text{ and } R_{\hat{Y}} = C R_X C^T + R_N \quad (14)$$

Then, the optimal decoder matrix B is

$$B = R_X C^T (C R_X C^T + R_N)^{-1} \quad (15)$$

and the constraint becomes

$$Tr\{C R_X C^T\} \leq P \quad (16)$$

and, we have

$$D = Tr\left\{R_X - R_X C^T (C R_X C^T + R_N)^{-1} C R_X\right\} \quad (17)$$

Plugging (1) in (17) we have

$$D = Tr\left\{\Lambda_X - \Lambda_X Q_X^T C^T (C Q_X \Lambda_X Q_X^T C^T + Q_N \Lambda_N Q_N^T)^{-1} C Q_X \Lambda_X\right\} \quad (18)$$

Let us define

$$K = \Lambda_N^{-1/2} Q_N^T C Q_X \Lambda_X^{1/2} \quad (19)$$

then, our objective becomes minimize

$$D = Tr\left\{\Lambda_X - \Lambda_X K^T (K K^T + I)^{-1} K\right\} \quad (20)$$

which must be minimized subject to

$$Tr\{K^T \Lambda_N K\} \leq P \quad (21)$$

Using Lemma 2, we obtain that K is diagonal, hence $C = Q_N \Sigma Q_X^T$ where Σ is diagonal. The entries of Σ and the distortion, (12), can be obtained applying Kuhn-Tucker necessary conditions for each scalar channel. ■

IV. ON FIRST ORDER OPTIMALITY OF LINEAR CODING

In this section, we derive the conditions under which zero delay linear coding achieves the asymptotic performance bound which overall would require asymptotic delay.

A. Asymptotic Bounds

To obtain asymptotically (in delay) achievable bounds, we have to consider the rate distortion function and channel capacity. In this section, we assume $R_X = I$ for simplicity.

1) *Channel capacity:* The capacity of k -dimensional Gaussian channel with covariance matrix R_N with total power P is given parametrically as

$$C = \sum_{i=1}^k (\log(\frac{\mu}{\lambda_N(i)}))^+, P = \sum_{i=1}^k (\mu - \lambda_N(i))^+ \quad (22)$$

for some $\mu \geq 0$, where we use natural logarithm for convenience.

2) *Rate-distortion:* The rate distortion function for the m -dimensional Gaussian source with covariance matrix R_X under mean square error distortion measure is given by the inverse water-filling formula

$$R = \frac{1}{2} \sum_{i=1}^m (\log(\frac{\lambda_X(i)}{\theta}))^+, D = \sum_{i=1}^m \min(\theta, \lambda_X(i)) \quad (23)$$

for some $\theta \geq 0$. Plugging $R_X = I$ in the above expressions, we get

$$D = m \exp\left\{\frac{-2R}{m}\right\} \quad (24)$$

B. First Order Optimality

Here, we present our main result on first order optimality of zero-delay linear source-channel coding for Gaussian sources and channels.

Theorem 2. *For a m -dimensional Gaussian source with $R_X = I$ and k -dimensional Gaussian channel, as $P \rightarrow 0$, linear coding achieves the (asymptotic) OPTA in first order, i.e.,*

$$\frac{dD_{lin}}{dP}\Big|_{P=0} = \frac{dD_{opta}}{dP}\Big|_{P=0} \quad (25)$$

for any m, k .

Proof: The key observation here is that as $P \rightarrow 0$, the active channels must have the same eigenvalue. Let us call this eigenvalue λ_N . Also, the number of active channels will be identical in asymptotic and zero-delay coding. Let w be the number of active channels, then we have

$$C = w \log\left(1 + \frac{P}{w\lambda_N}\right) \quad (26)$$

using (22). Plugging $R = C$ in (24), we obtain

$$D_{opta} = m \exp\left\{-\frac{w}{m} \log\left(1 + \frac{P}{w\lambda_N}\right)\right\} \quad (27)$$

Plugging $R_X = I$ in (12), we have

$$D_{lin} = m - k + \frac{w}{1 + \frac{P}{w\lambda_N}} \quad (28)$$

Evaluating the slope at $P = 0$, we have

$$\left. \frac{dD_{lin}}{dP} \right|_{P=0} = \left. \frac{dD_{opta}}{dP} \right|_{P=0} = \frac{-1}{\lambda_N} \quad (29)$$

V. SIDE INFORMATION SETTING

In this section, we focus on the settings where side information is available to the decoder. The first result derives the optimal encoding and decoding matrices in the side information setting. Next, we allow scalar nonlinearities per transform coefficient, and obtain the optimal encoding and decoding matrices.

A. Linear Mapping

Consider the setting shown in Figure 1-b. The following theorem states that the optimal encoding transform C is the product of the conditional KLT of the source given the side information, a diagonal power allocation matrix and the KLT of the noise.

Theorem 3. *The encoding transform that minimizes the MSE distortion, subject to a power constraint is*

$$C = Q_N \Sigma Q_{X|Z}^T \quad (30)$$

where Σ is a diagonal power allocation matrix.

Proof: It follows from standard linear estimation principles [16] that

$$\hat{\mathbf{X}} = \begin{bmatrix} R_X C^T \\ R_{XZ} \end{bmatrix} \begin{bmatrix} C R_X C^T + R_N & C R_{XZ} \\ R_{XZ}^T C^T & R_Z \end{bmatrix}^{-1} \begin{bmatrix} C \mathbf{X} + \mathbf{N} \\ \mathbf{Z} \end{bmatrix}$$

and

$$D = \text{Tr} \left\{ R_X - \begin{bmatrix} R_X C^T \\ R_{XZ} \end{bmatrix} \begin{bmatrix} C R_X C^T + R_N & C R_{XZ} \\ R_{XZ}^T C^T & R_Z \end{bmatrix}^{-1} \begin{bmatrix} R_X C^T \\ R_{XZ} \end{bmatrix}^T \right\} \quad (31)$$

We make use of an auxiliary lemma from matrix analysis [15].

Lemma 3 (Schur's complement). *Let U, V, A, B, C be matrices of appropriate dimensions. Then, the following holds:*

$$\begin{bmatrix} U \\ V \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \Psi^T (A - B C^{-1} B^T)^{-1} \Psi$$

where $\Psi = U - B C^{-1} V$.

Using Lemma 3 and (2) in (31),

$$D = \text{Tr} \left\{ R_{X|Z} - R_{X|Z}^T C^T (C R_{X|Z} C^T + R_N)^{-1} C R_{X|Z} \right\} \quad (32)$$

Plugging (3) in (32) we have

$$D = \text{Tr} \left\{ \Lambda_{X|Z} - \Lambda_{X|Z} Q_{X|Z}^T C^T (C Q_{X|Z} \Lambda_{X|Z} Q_{X|Z}^T C^T + Q_N \Lambda_N Q_N^T)^{-1} C Q_{X|Z} \Lambda_{X|Z} \right\}$$

Define

$$K = \Lambda_N^{-1/2} Q_N^T C Q_{X|Z} \Lambda_{X|Z}^{1/2} \quad (33)$$

then, our problem becomes: minimize

$$D = \text{Tr} \left\{ \Lambda_{X|Z} - \Lambda_{X|Z} K^T (K K^T + I)^{-1} K \right\} \quad (34)$$

subject to

$$\text{Tr} \{ K^T \Lambda_N K \Phi \} \leq P \quad (35)$$

where $\Phi = \Lambda_{X|Z}^{-1/2} Q_{X|Z}^T R_X Q_{X|Z} \Lambda_{X|Z}^{-1/2}$ is symmetric and positive definite. Applying Lemma 2, we obtain the result. ■

B. Nonlinear Mappings for Wyner-Ziv Setting

In this section, we allow scalar nonlinearities in the problem definition. Unlike the point-to-point case, the optimal zero-delay scalar mappings for the point-to-point setting are highly nonlinear in the side information setting. In [4], [5] an algorithm was proposed to find such mappings for this setting. Figure 2 presents examples of such encoding mappings. Interestingly, the analog mapping captures the central characteristic observed in digital Wyner-Ziv mappings, in the sense of many-to-one mappings, or multiple source intervals are mapped to the same channel interval, which will potentially be resolved by the decoder given the side information. Within each bin, there is a mapping function which is approximately linear in this case (scalar Gaussian sources and channel).

Here we consider generalizing this highly non-linear scalar mapping to vector spaces. This generalization, albeit being algorithmically infeasible, can be done through a structured combination of linear m to k mapping followed by scalar nonlinear functions, as illustrated in Figure 1-c. The i^{th} transform coefficient is nonlinearly transformed through the mapping g_i and it is estimated through another nonlinear mapping h_i . The problem of interest here is finding the optimal linear transforms, which, in general, might depend on the nonlinearities introduced. For simplicity, we assume $R_N = I$.

The following theorem states that if the functions are scale invariant, the optimal transform is identical to the case where no nonlinearities were present, hence it does not depend on the nonlinearities. Let us define scale invariance in our context.

Let $\{g_1, g_2, \dots\} : \mathbb{R} \rightarrow \mathbb{R}$ be a set of functions. If for all i, j , $g_i(x) = \beta g_j(\alpha x)$, for some $\alpha, \beta \in \mathbb{R}$ then we call these functions as scale invariant. Note that if we use these functions as encoding mappings in conjunction with the associated optimal decoding mappings, the distortion depend on the source variance only through scaling, in other words in the power distortion function can be expressed as $D = \sigma_X^2 f(P)$, where function f is independent of σ_X^2 .

In passing, we note that Ziv showed the optimal encoding mappings cannot be scale invariant for the point-to-point

setting, other than the trivial scalar Gaussian source-channel case [17]. We conjecture, based on the observations from our experimental results in [5], that for Gaussian source, channel and side information, the optimal mappings are scale invariant.

Theorem 4. *In the setting where g_i 's are scale invariant, the encoding transform that minimizes MSE distortion is*

$$C = \Sigma Q_{X|Z}^T \quad (36)$$

where Σ is a diagonal power allocation matrix .

Proof: From matrix differentiation, it follows that the optimal transform must be in the form $C = \Sigma S$ where S is a unitary matrix, and Σ is diagonal. Now, we will show that $S = Q_{X|Z}^T$. Let $\mathbf{P} = [P_1, P_2, \dots, P_k]$ be the power allocation vector. It follows from scale invariance that $D_i = \sigma_{Y_i|Z}^2 f(P_i)$ for each Y_i . Then, the problem is to minimize the function

$$D = \sum_{i=1}^k \sigma_{Y_i|Z}^2 f(P_i) \quad (37)$$

by manipulating $\sigma_{Y_i|Z}^2$ through the choice of S . Consider

$$\boldsymbol{\sigma} = [\sigma_{Y_1|Z}^2, \sigma_{Y_2|Z}^2, \dots, \sigma_{Y_k|Z}^2] = \text{diag}\{SR_{X|Z}S^T\} \quad (38)$$

Note that $SR_{X|Z}S^T$ has eigenvalues $\boldsymbol{\lambda}_{X|Z}$. From Lemma 1, we know that $\boldsymbol{\lambda}_{X|Z}$ majorizes $\boldsymbol{\sigma}$. This implies that $\boldsymbol{\sigma}$ is in the convex hull of $k!$ permutations of $\boldsymbol{\lambda}_{X|Z}$. Hence, our objective is to minimize D over the convex polytope defined by the permutations of $\boldsymbol{\lambda}_{X|Z}$. Note that D is linear in σ_i . It is well known that the corner point minimizes a linear function over a convex polytope, i.e., $\boldsymbol{\sigma} = \boldsymbol{\lambda}_{X|Z}$ will minimize (37). This establishes that the optimal transform is a conditional KLT, i.e., $S = Q_{X|Z}^T$. ■

Remark 1. *This results is perhaps surprising in that although the transform effects the particular nonlinear mapping employed, via changing the correlation between Y_i and Z , optimal transform remains $S = Q_{X|Z}^T$ as in Theorem 3.*

Remark 2. *The linear coding scheme in the previous section is a special case of this setting with the trivial identity mappings, $g_i(x) = h_i(x) = x, \forall i$. Hence, the proof of Theorem 4 also implies Theorem 3 when $R_N = I$.*

Remark 3. *The nonlinear coding scheme can also be extended to the case with $R_N \neq I$, with the help of a linear transformation of the outputs of the scalar nonlinear functions. Such optimal transform would be Q_N . Detailed extensions are omitted due to space constraints.*

VI. CONCLUSION

In this paper, we presented several results regarding linear zero-delay source-channel coding. First, using the majorization tools, we presented an alternative solution to the classical problem in the point-to-point setting. Next, we showed that zero-delay linear source channel coding achieves OPTA in the first order at asymptotically low SNR. We then studied the optimal linear transform for the side information setting and

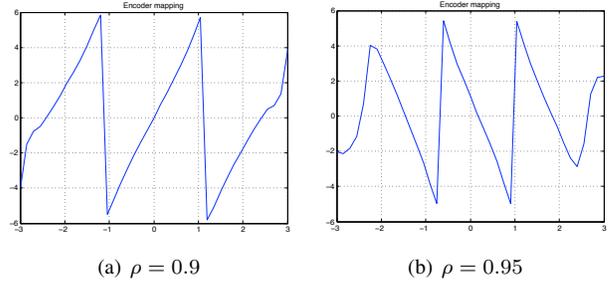


Fig. 2. Example mappings for correlation coefficient $\rho = 0.9$ and $\rho = 0.95$, Gaussian scalar source, channel and side information [5].

showed that, similar to the point-to-point setting, the optimal linear transform is conditional KLT of the source, followed by a diagonal power allocation matrix and the channel noise KLT. Finally, we considered a practical structured nonlinear mapping, as the combination of a linear transform with scalar nonlinearities. We showed that the optimal linear transform is independent of the the scalar nonlinear mappings as long as these mappings are scale invariant.

REFERENCES

- [1] X. Chen and E. Tuncel, "Zero-delay joint source-channel coding for the Gaussian Wyner-Ziv problem," in *IEEE International Symposium on Inf. Theory*, 2011, pp. 2929–2933.
- [2] F. Hekland, P.A. Floor, and T.A. Ramstad, "Shannon-Kotelnikov mappings in joint source-channel coding," *IEEE Trans. on Comm.*, vol. 57, no. 1, pp. 94–105, 2009.
- [3] M. Kleiner and B. Rimoldi, "A tight bound on the performance of a minimal-delay joint source-channel coding scheme," in *IEEE International Symp. on Inf. Theory*, 2010, pp. 136–140.
- [4] J. Karlsson and M. Skoglund, "Optimized low-delay source-channel-relay mappings," *IEEE Trans. on Communications*, vol. 58, no. 5, pp. 1397–1404, 2010.
- [5] E. Akyol, K. Rose, and T. Ramstad, "Optimized analog mappings for distributed source-channel coding," in *IEEE Data Compression Conference*, 2010, pp. 159–168.
- [6] M.N. Khormuji and M. Skoglund, "On instantaneous relaying," *IEEE Trans. on Inf. Theory*, vol. 56, no. 7, pp. 3378–3394, 2010.
- [7] K.H. Lee and D. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. on Communications*, vol. 24, no. 12, pp. 1283–1290, 1976.
- [8] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. on Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, 1990.
- [9] E. Akyol, K. Viswanatha, and K. Rose, "On conditions for linearity of optimal estimation," in *IEEE Inf. Theory Workshop*, 2010, pp. 1–5.
- [10] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. on Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.
- [11] J. Ostergaard and R. Zamir, "Incremental refinement using a Gaussian test channel," in *IEEE International Symposium on Inf. Theory*, 2011, pp. 2233–2237.
- [12] M. Gastpar, P.L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Trans. on Inf. theory*, vol. 52, no. 12, pp. 5177–5196, 2006.
- [13] AW Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.
- [14] D.P. Palomar, J.M. Cioffi, and M.A. Lagunas, "Joint TX-RX beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, 2003.
- [15] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [16] T. Kailath, A.H. Sayed, and B. Hassibi, "Linear Estimation," *Upper Saddle River, NJ*, 2000.
- [17] J. Ziv, "The behavior of analog communication systems," *IEEE Trans. on Inf. Theory*, vol. 16, no. 5, pp. 587–594, 1970.