

Optimized Analog Mappings for Distributed Source-Channel Coding *

Emrah Akyol, Kenneth Rose

Dept. of Electrical & Computer Engineering
University of California, Santa Barbara, CA 93106, USA
Email:{eakyol, rose}@ece.ucsb.edu

Tor Ramstad

Dept. of Electronics and Telecommunications
Norwegian University of Science and Tech.
Trondheim, NO7491, Norway
Email: ramstad@iet.ntnu.no

Abstract

This paper focuses on optimal analog mappings for zero-delay, distributed source-channel coding. The objective is to obtain the optimal vector transformations that map between m -dimensional source spaces and k -dimensional channel spaces, subject to a prescribed power constraint and assuming the mean square error distortion measure. Closed-form necessary conditions for optimality of encoding and decoding mappings are derived. An iterative design algorithm is proposed, which updates encoder and decoder mappings by sequentially enforcing the complementary optimality conditions at each iteration. The obtained encoding functions are shown to be a continuous relative of, and in fact subsume as a special case, the Wyner-Ziv mappings encountered in digital distributed source coding systems, by mapping multiple source intervals to the same channel interval. Example mappings and performance results are presented for Gaussian sources and channels.

1 Introduction

Analog mappings for joint source channel coding have been studied since Shannon's seminal work [1] where the use of space filling curves was suggested for transmission of Gaussian sources over higher dimensional Gaussian channels (i.e., with bandwidth expansion) [2, 3, 4, 5]. Although analog mappings with finite delay will not, in general,

*The work is supported in part by the NSF under grant CCF-0728986

reach the asymptotic coding performance bounds, except for a few special cases (such as Gaussian source and channel of the same dimension [6]), they do possess desirable properties: low complexity, zero delay, and better robustness to varying channel SNR.

In our prior work [7], we presented the optimality conditions for such analog mappings for point to point scenarios, along with an algorithm to obtain the locally optimal mappings. In this paper, we extend the work to distributed coding scenarios. We focus on two settings: The first one involves source-channel coding when side information is available to the decoder. This setting has been studied extensively both theoretically [8, 9] and in practice [10]. The second setting involves distributed source-channel coding where multiple correlated sources are encoded separately and transmitted over different channels. An important practical motivation for this setup is sensor networks where sensor measurements are correlated but are not encoded jointly due to physical constraints. This problem has also been studied extensively, especially for Gaussian sources [11].

The derivation of the optimality conditions is a direct extension of our prior work, but the distributed nature of the setting results in complex mappings that are highly nontrivial. Numerical optimization of such mappings heavily depends on initial conditions due to numerous local minima of the cost functional. Note in particular that in the case of Gaussian sources and channels, linear encoders and decoder satisfy the necessary conditions of optimality while, as we will see, careful optimization obtains better mappings that are far from linear.

In Section II, we formulate the problem, present the necessary conditions for optimality and proposed an iterative design algorithm. We provide example mappings and comparative performance results in Section III. We conclude in Section IV.

2 Problem and Proposed Solution

2.1 Problem Formulation

The first setup of source-channel coding with side information is shown in Figure 1, and the second setup of distributed coding is depicted in Figure 2. In both cases, there are two correlated vector sources $\mathbf{X}_1 \in \mathbb{R}^{m_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{m_2}$. The noise variables, $\mathbf{N}, \mathbf{N}_1, \mathbf{N}_2$ are assumed to be independent of the sources $\mathbf{X}_1, \mathbf{X}_2$. The m_1, m_2 -fold source density is denoted as $f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2)$.

In the side information setup, \mathbf{X}_2 is available to the decoder, and \mathbf{X}_1 is transformed into $\mathbf{Y} \in \mathbb{R}^k$ by a function $g : \mathbb{R}^{m_1} \rightarrow \mathbb{R}^k$ and transmitted over an additive noise channel where $N \in \mathbb{R}_k$, with k -fold noise density is $f_N(\mathbf{n})$, is independent of $\mathbf{X}_1, \mathbf{X}_2$. The received output $\hat{\mathbf{Y}} = \mathbf{Y} + \mathbf{N}$ is transformed to the estimate $\hat{\mathbf{X}}$ through a function $\mathbf{h} : \mathbb{R}^k \times \mathbb{R}^{m_2} \rightarrow \mathbb{R}_1^m$. The m_1 -fold source density is denoted as $f_X(\mathbf{x})$ and k -fold noise density is $f_N(\mathbf{n})$. The problem is to find optimal mapping functions \mathbf{g}, \mathbf{h} that minimize distortion

$$\mathbb{E}(|\mathbf{X}_1 - \hat{\mathbf{X}}_1|^2), \quad (1)$$

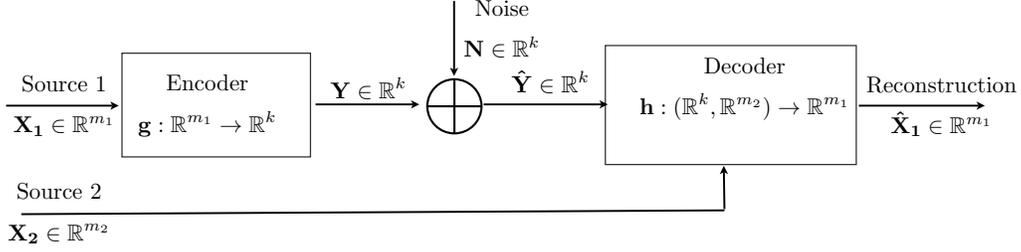


Figure 1: Block diagram: mapping for source-channel coding with side information

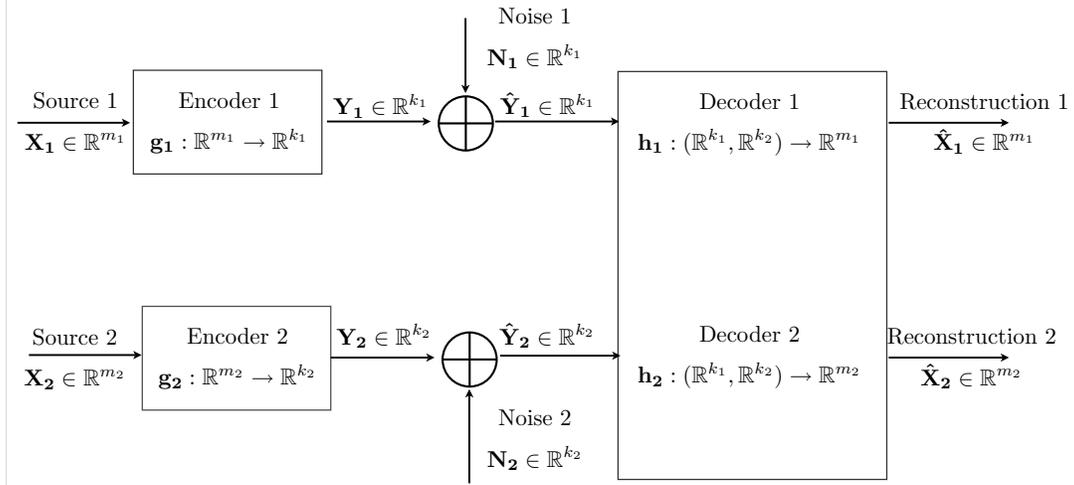


Figure 2: Block diagram: mapping for distributed source-channel coding

subject to average power constraint

$$P[\mathbf{g}] = \int_{\mathbb{R}^{m_1}} \mathbf{g}(\mathbf{x})^T \mathbf{g}(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} \leq P. \quad (2)$$

In the second case, both sources are transformed through $g_1 : \mathbb{R}^{m_1} \rightarrow \mathbb{R}^{k_1}$ and $g_2 : \mathbb{R}^{m_2} \rightarrow \mathbb{R}^{k_2}$ and transmitted over the noisy channel with k_1 -fold noise density, $f_{N_1}(\mathbf{n}_1)$ and k_2 -fold noise density $f_{N_2}(\mathbf{n}_2)$. At the decoder, $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$ are generated by $\mathbf{h}_1 : \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \rightarrow \mathbb{R}^{m_1}$ and $\mathbf{h}_2 : \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \rightarrow \mathbb{R}^{m_2}$. The problem is to find optimal mapping functions $\mathbf{g}_1, \mathbf{g}_2, \mathbf{h}_1, \mathbf{h}_2$ that minimize distortion

$$\mathbb{E}(|\mathbf{X}_1 - \hat{\mathbf{X}}_1|^2 + |\mathbf{X}_2 - \hat{\mathbf{X}}_2|^2), \quad (3)$$

subject to the average power constraint

$$P[\mathbf{g}_1, \mathbf{g}_2] = \int_{\mathbb{R}^{m_1}, \mathbb{R}^{m_2}} (\mathbf{g}_1(\mathbf{x})^T \mathbf{g}_1(\mathbf{x}) + \mathbf{g}_2(\mathbf{x})^T \mathbf{g}_2(\mathbf{x})) f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \leq P. \quad (4)$$

Bandwidth compression-expansion is determined by the setting of the source and channel dimensions, k/m_1 for case-1 and $k_1/m_1, k_2/m_2$ for case-2.

2.2 Optimal Decoding Given Encoder(s)

Case-1: Assume that the encoder \mathbf{g} is fixed. Then the optimal decoder is the minimum mean square error estimator (MMSE) of \mathbf{X}_1 given $\hat{\mathbf{Y}}$ and \mathbf{X}_2 , i.e.,

$$\mathbf{h}(\hat{\mathbf{y}}, \mathbf{x}_2) = \mathbb{E}[\mathbf{X}_1 | \hat{\mathbf{y}}, \mathbf{x}_2]. \quad (5)$$

Plugging the expressions for expectation, applying Bayes' rule and noting that $f_{\hat{\mathbf{Y}}|\mathbf{X}_1}(\hat{\mathbf{y}}|\mathbf{x}_1) = f_N[\hat{\mathbf{y}} - \mathbf{g}(\mathbf{x}_1)]$, the optimal decoder can be written, in terms of known quantities, as

$$\mathbf{h}(\hat{\mathbf{y}}) = \frac{\int \mathbf{x}_1 f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) f_N[\hat{\mathbf{y}} - \mathbf{g}(\mathbf{x}_1)] d\mathbf{x}_1}{\int f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) f_N[\hat{\mathbf{y}} - \mathbf{g}(\mathbf{x}_1)] d\mathbf{x}_1}. \quad (6)$$

Case-2: Here we assume that the two encoders are fixed. Then, the optimal decoders are $\mathbf{h}_1(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2) = \mathbb{E}[\mathbf{X}_1 | \hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2]$ and $\mathbf{h}_2(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2) = \mathbb{E}[\mathbf{X}_2 | \hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2]$ and can be written as

$$\mathbf{h}_1(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2) = \frac{\int \mathbf{x}_1 \mathbf{f}_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) \mathbf{f}_{N_1, N_2}[\hat{\mathbf{y}}_1 - \mathbf{g}(\mathbf{x}_1), \hat{\mathbf{y}}_2 - \mathbf{g}(\mathbf{x}_2)] d\mathbf{x}_1 d\mathbf{x}_2}{\int \mathbf{f}_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) \mathbf{f}_{N_1, N_2}[\hat{\mathbf{y}}_1 - \mathbf{g}(\mathbf{x}_1), \hat{\mathbf{y}}_2 - \mathbf{g}(\mathbf{x}_2)] d\mathbf{x}_1 d\mathbf{x}_2} \quad (7)$$

$$\mathbf{h}_2(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2) = \frac{\int \mathbf{x}_2 \mathbf{f}_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) \mathbf{f}_{N_1, N_2}[\hat{\mathbf{y}}_1 - \mathbf{g}(\mathbf{x}_1), \hat{\mathbf{y}}_2 - \mathbf{g}(\mathbf{x}_2)] d\mathbf{x}_1 d\mathbf{x}_2}{\int \mathbf{f}_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) \mathbf{f}_{N_1, N_2}[\hat{\mathbf{y}}_1 - \mathbf{g}(\mathbf{x}_1), \hat{\mathbf{y}}_2 - \mathbf{g}(\mathbf{x}_2)] d\mathbf{x}_1 d\mathbf{x}_2}. \quad (8)$$

2.3 Optimal Encoding Given Decoder(s)

Case-1: Assume that the decoder \mathbf{h} is fixed. Our goal is to minimize the distortion subject to the average power constraint. The distortion is expressed as a functional of \mathbf{g} :

$$D[\mathbf{g}] = \int [\mathbf{x}_1 - \mathbf{h}(\mathbf{g}(\mathbf{x}_1) + \mathbf{n}, \mathbf{x}_2)]^T [\mathbf{x}_1 - \mathbf{h}(\mathbf{g}(\mathbf{x}_1) + \mathbf{n}, \mathbf{x}_2)] f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) f_N(\mathbf{n}) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{n}. \quad (9)$$

We construct the Lagrangian cost functional to minimize over the mapping \mathbf{g} ,

$$J[\mathbf{g}] = D[\mathbf{g}] + \lambda \{P[\mathbf{g}] - P\}. \quad (10)$$

To obtain necessary conditions we apply the standard method in variational calculus: The Frechet derivative must vanish at an extremum point of the functional J [12]. This gives the necessary condition for optimality as

$$\nabla J[\mathbf{g}](x_1, x_2) = 0, \quad \forall x_1, x_2 \quad (11)$$

where

$$\nabla J[g](x_1, x_2) = \lambda f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) \mathbf{g}(\mathbf{x}_1) - \int \mathbf{h}'(\mathbf{g}(\mathbf{x}_1) + \mathbf{n}, \mathbf{x}_2) [\mathbf{x} - \mathbf{h}(\mathbf{g}(\mathbf{x}) + \mathbf{n}, \mathbf{x}_2)] f_N(\mathbf{n}) f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{n}. \quad (12)$$

Case-2: Here we assume the decoders are fixed and we find the encoders $\mathbf{g}_1, \mathbf{g}_2$ that minimize the total cost $J = D[\mathbf{g}_1, \mathbf{g}_2] + \lambda \{P[\mathbf{g}_1, \mathbf{g}_2] - P\}$ where

$$D[\mathbf{g}_1, \mathbf{g}_2] = \int [\mathbf{x}_1 - \mathbf{h}_1(\mathbf{g}_1(\mathbf{x}_1) + \mathbf{n}_1, \mathbf{g}_2(\mathbf{x}_2) + \mathbf{n}_2)]^T [\mathbf{x}_1 - \mathbf{h}_1(\mathbf{g}_1(\mathbf{x}_1) + \mathbf{n}_1, \mathbf{g}_2(\mathbf{x}_2) + \mathbf{n}_2)] \\ + [\mathbf{x}_2 - \mathbf{h}_2(\mathbf{g}_1(\mathbf{x}_1) + \mathbf{n}_1, \mathbf{g}_2(\mathbf{x}_2) + \mathbf{n}_2)]^T [\mathbf{x}_2 - \mathbf{h}_2(\mathbf{g}_1(\mathbf{x}_1) + \mathbf{n}_1, \mathbf{g}_2(\mathbf{x}_2) + \mathbf{n}_2)] \\ f_{X_1, X_2}(\mathbf{x}_1, \mathbf{x}_2) f_{N_1, N_2}(\mathbf{n}_1, \mathbf{n}_2) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{n}_1 d\mathbf{n}_2. \quad (13)$$

The necessary conditions are derived by requiring

$$\nabla J[\mathbf{g}_1](x_1, x_2) = \nabla J[\mathbf{g}_2](x_1, x_2) = 0 \quad \forall x_1, x_2, \quad (14)$$

where $\nabla J[\mathbf{g}_1](x_1, x_2) = \frac{\partial J}{\partial g_1}$ and $\nabla J[\mathbf{g}_2](x_1, x_2) = \frac{\partial J}{\partial g_2}$ [12]. (Explicit expressions are omitted for brevity, but can be derived similar to Case-1). Note that, unlike the decoder result, the optimal encoder is not specified in closed form.

2.4 Algorithm Sketch

The basic idea is to alternate between enforcing the complementary necessary conditions for optimality. Iterations are continued until the algorithm reaches a stationary point where the total cost does not decrease anymore (in practice, until incremental improvements fall below a convergence threshold). Since the encoder(s) condition is not in closed form, we perform steepest descent search on the direction of the Frechet derivative of the Lagrangian with respect to the encoder mapping(s) \mathbf{g} for Case-1 and $\mathbf{g}_1, \mathbf{g}_2$ for Case-2. By design, the Lagrangian cost decreases monotonically as the algorithm proceeds iteratively. The various encoders are updated as given generically in (15), where i is the iteration index and μ is the step size.

$$\mathbf{g}_{i+1}(\mathbf{x}) = \mathbf{g}_i(\mathbf{x}) - \mu \nabla \mathbf{J}[\mathbf{g}] \quad (15)$$

Note that there is no guarantee that an iterative descent algorithms of this type will converge to the globally optimal solution. The algorithm will converge to a local minimum, which may not be unique. To avoid poor local minima, one can run the algorithm several times with different initial conditions, and in the case of highly complex cost surfaces it may be necessary to apply more structured solutions such as deterministic annealing [13]. The scope of this paper and the simulations are restricted to use a more modest means by applying a noisy channel relaxation variant to avoid local minima following the proposal in [14, 15].

3 Results

In this section, we demonstrate the use of the proposed algorithm by focusing on specific scenarios. Note the algorithm is general and directly applicable to any choice

of dimensions as well as source and noise distributions, and also allows for correlations across noise dimensions. However, for conciseness of the results section we assume that sources are jointly Gaussian scalars with correlation coefficient ρ , and are identically distributed. We also assume the noise is scalar and Gaussian.

3.1 Initial Conditions

An important observation is that the linear encoder-decoder pair satisfies the necessary conditions of optimality, although, as we will illustrate, there are other mappings that perform better. Hence, initial conditions have paramount importance in obtaining better results. In an attempt to circumvent the poor local minima problem, we embed in the solution the noisy relaxation method of [14, 15]. We initialize the encoding mapping(s) with random initial conditions and run the algorithm for a very noisy channel (high Lagrangian parameter λ). Then, we gradually increase channel SNR (CSNR) (decrease λ) while using the prior mapping as initial condition.

We implemented the above algorithm by numerically calculating the derived integrals. For that purpose, we sampled the distribution on a uniform grid. We also imposed bounded support (-3σ to 3σ) i.e., neglected tails of the infinite support distributions in the examples.

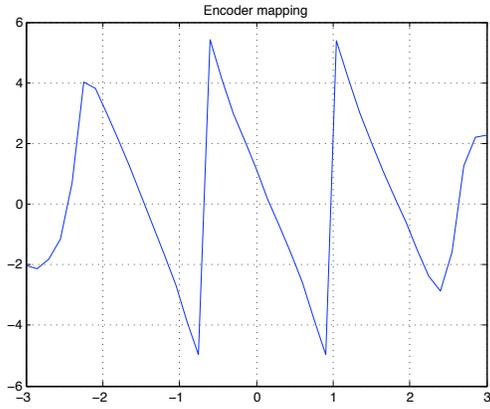
3.2 Case-1: JSCC with Side information

Figure 3 presents an example of encoding mapping obtained for correlation coefficient $\rho = 0.97$. 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). To see the effect of correlation on the encoding mappings, we present another example in Figure 4. In this case $\rho = 0.9$ and, as intuitively expected, the side information is less reliable and source points mapped to the same channel representation grow further apart from each other. Comparative performance results are shown in Figure 5. The proposed mapping outperforms the linear mapping for the entire range of CSNR values significantly.

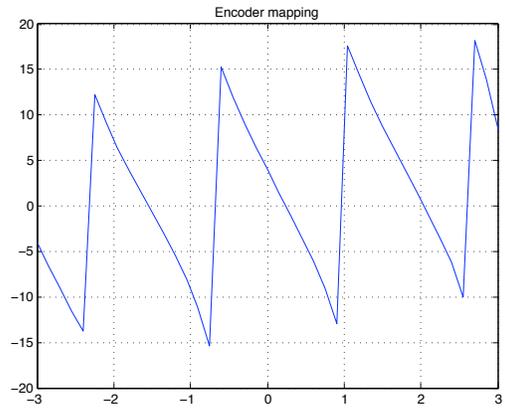
3.3 Case-2: Distributed JSCC

Here we consider jointly Gaussian sources that are transmitted separately over independent channels. Note that our algorithm and derived necessary conditions allow the channels to be correlated, but for the sake of simplicity we assume they are independent, additive Gaussian channels.

Figure 6 presents an example of encoding mappings for correlation coefficient $\rho = 0.95$. The comparative performance results are shown in Figure 7. The proposed mapping outperforms the linear mapping for the entire range of CSNR values. Note that encoders are highly asymmetric in the sense that one is Wyner-Ziv like and

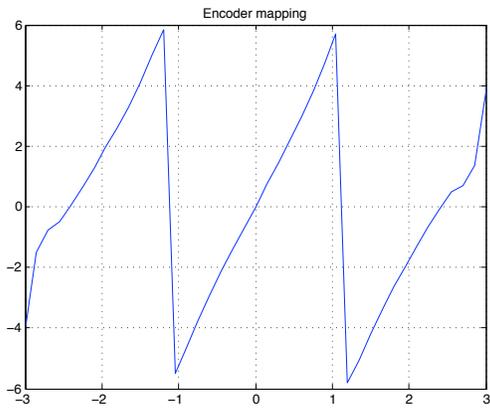


(a) $CSNR = 10$

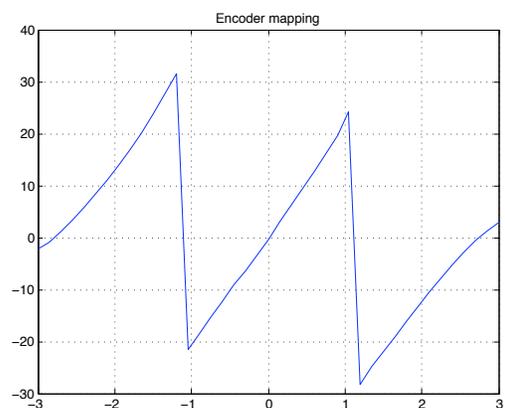


(b) $CSNR = 22$

Figure 3: Example encoder mappings at correlation coefficient $\rho = 0.97$, Gaussian scalar source, channel and side information



(a) $CSNR = 10$



(b) $CSNR = 23$

Figure 4: Example encoder mappings at correlation coefficient $\rho = 0.9$, Gaussian scalar source, channel and side information

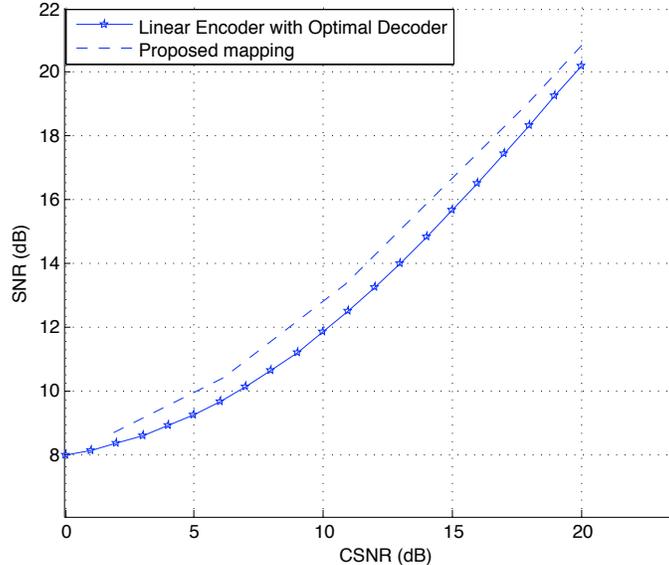


Figure 5: Comparative results for correlation coefficient $\rho = 0.9$, Gaussian scalar source, channel and side information

maps several source intervals to the same channel interval, whereas other encoder is almost a monotonic function. While we can offer tentative explanation to justify the form of this solution, we above all observe that it does outperform the symmetric linear solution. This demonstrates that by breaking from the linear solution and re-optimizing we do obtain an improved solution. Moreover, it suggests that some symmetric highly non-linear solution may exist whose discovery would need more powerful optimization tools – a direction we are currently pursuing.

4 Discussion and Future Work

In this paper, we derived the necessary conditions of optimality of the distributed analog mappings. Based on these necessary conditions, we derived an iterative algorithm to generate locally optimal analog mappings. Comparative results and example mappings are provided and it is shown that the proposed method provides significant gains over simple linear encoding. As expected the obtained mappings resembles Wyner-Ziv mappings in the sense that multiple source intervals are mapped to the same channel interval. The algorithm does not guarantee a globally optimal solution. This problem can be largely mitigated by using more powerful optimization, in particular a deterministic annealing approach, which is left as future work.

References

- [1] CE Shannon, “Communication in the presence of noise,” *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

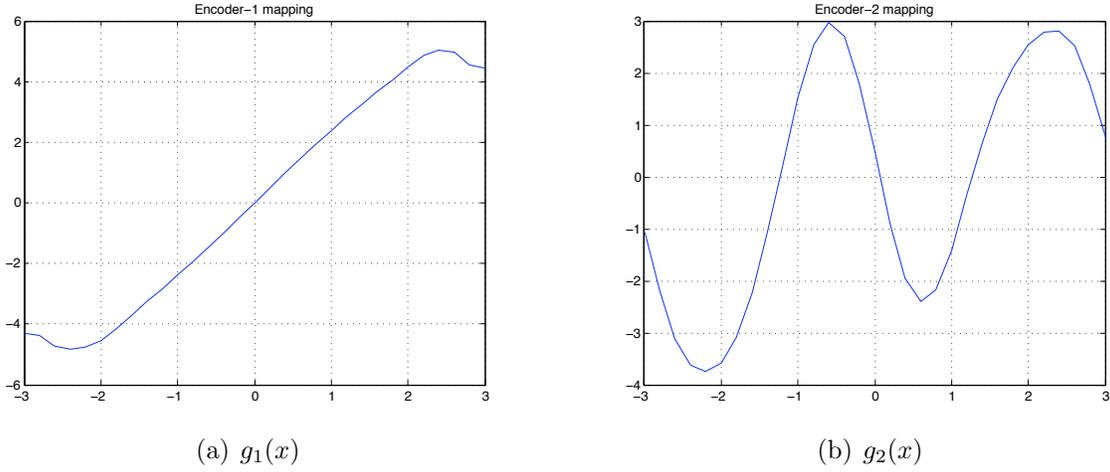


Figure 6: Example encoder mappings for correlation coefficient $\rho = 0.95$, Gaussian scalar sources and channels, $CSNR = 10$

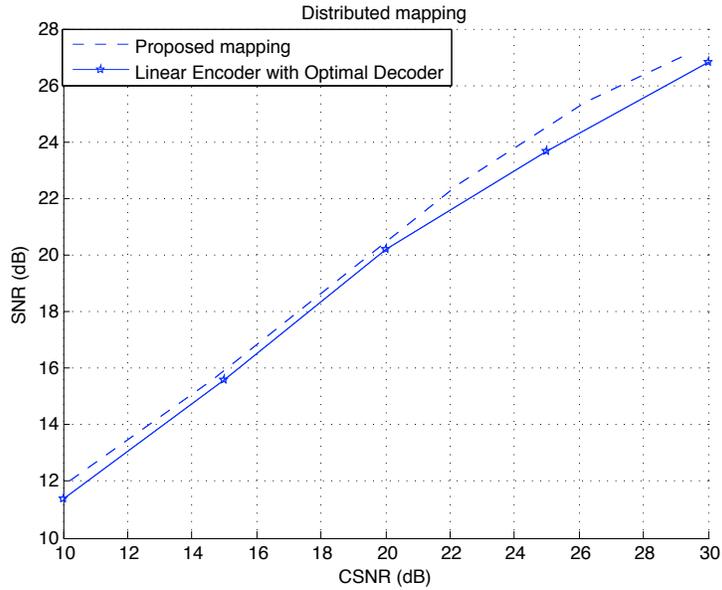


Figure 7: Comparative results for correlation coefficient $\rho = 0.95$, Gaussian scalar sources and channels

- [2] VA Kotelnikov, *The theory of optimum noise immunity*, New York: McGraw-Hill, 1959.
- [3] S.Y. Chung, *On the construction of some capacity-approaching coding schemes*, Ph.D. thesis, Massachusetts Institute of Technology, 2000.
- [4] T.A. Ramstad, “Shannon mappings for robust communication,” *Teletronikk*, vol. 98, no. 1, pp. 114–128, 2002.
- [5] F. Hekland, GE Oien, and TA Ramstad, “Using 2: 1 Shannon mapping for joint source-channel coding,” in *IEEE Data Compression Conference Proceedings*, 2005, pp. 223–232.
- [6] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, J.Wiley New York, 1991.
- [7] E. Akyol, K. Rose, and TA Ramstad, “Optimal mappings for joint source channel coding,” in *IEEE Information Theory Workshop, ITW 2010*, to appear.
- [8] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [9] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [10] SS Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): design and construction,” *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, 2003.
- [11] A.B. Wagner, S. Tavildar, and P. Viswanath, “Rate region of the quadratic Gaussian two-encoder source-coding problem,” *IEEE Transactions on Information Theory*, vol. 54, no. 5, 2008.
- [12] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons Inc, 1969.
- [13] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.
- [14] S. Gadkari and K. Rose, “Robust vector quantizer design by noisy channel relaxation,” *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1113–1116, 1999.
- [15] P. Knagenhjelm, “A recursive design method for robust vector quantization,” in *Proc. Int. Conf. Signal Processing Applications and Technology*, pp. 948–954.