Optimal Mappings for Joint Source Channel Coding

Emrah Akyol, Kenneth Rose
Department of Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106, US
Email:{eakyol, rose}@ece.ucsb.edu

Tor Ramstad
Department of Electronics and Telecommunications
Norwegian University of Science and Technology
Trondheim, NO7491, Norway
Email: ramstad@iet.ntnu.no

***Abstract*— What is the optimal source-channel communication system for a given finite block length? The problem of obtaining the vector transformations that optimally map between the m-dimensional source space and the k-dimensional channel space is considered under a given channel power constraint and mean square error distortion measure. Closed form necessary conditions for optimality of the encoder and decoder mappings are derived. The optimal mappings are obtained using an iterative algorithm that updates encoder and decoder mappings according to optimality conditions at each iteration. Such mappings are used in a practical analog joint source channel system that transmits a continuous alphabet discrete time source over a noisy channel. Numerical results are presented for several source-channel distributions and it is shown that the optimal mappings outperform the previous heuristic mappings for both bandwidth expansion and compression.**

I. Introduction

One of the fascinating results in information theory is that uncoded transmission of Gaussian samples over an additive white Gaussian noise (AWGN) channel is optimal in the sense that it yields the minimum mean square error between source and reconstruction [1]. This result demonstrates the potential of joint source channel coding: Such a simple scheme with no delay provides the performance of the asymptotically optimal separate source-channel coding system, without recourse to complex compression and channel coding schemes that require asymptotically long delays. However, it is well known that, in general, the best source channel coding system with fixed finite delay may not achieve Shannon's asymptotic coding bound (see eg. [2], [3]). Nevertheless, the problem of obtaining the optimal scheme for a given finite delay is an important open problem with considerable practical implications.

In the practical problem of transmitting a discrete time continuous alphabet source over a discrete time additive analog channel, there are two main approaches: "analog communication" through direct amplitude modulation, and "digital communication" which typically consists of quantization, error control coding and digital modulation. The main advantage of digital over analog communication is due to advanced quantization and error control schemes. However, there are two notable shortcomings: First, error control coding (and to some extent also source coding) usually incurs substantial delay to achieve good performance. The other problem is the level off effect due to underlying quantization. The performance saturates as channel signal to noise ratio (CSNR) increases above the threshold. Analog systems offer the potential to avoid these problems. However, there are no known explicit methods to obtain such mappings for a general source and channel, nor is the best mapping known for other than the trivial one of the scalar Gaussian source-channel pair. Among the few practical analog coding schemes that have appeared in the literature are those based on the use of space-filling curves for bandwidth compression, originally proposed more than 50 years ago by Shannon [4] and Kotelnikov [5], and recently extended in the work of Fuldseth and Ramstad [6], Chung [7], Ramstad [8] and Hekland et.al. [9]. Spiral-like curves are explored for transmission of Gaussian sources over AWGN channels. It is also noteworthy that, a similar problem was solved in [10] with the constraint that both encoder and decoder are linear. A similar problem for digital systems was also studied by Fine [11]. Certain extensions of Fine's work can be found in [12].

In this paper, we investigate the problem of obtaining the vector transformations that optimally map between the m-dimensional source space and the k-dimensional channel space under a given channel power constraint and mean square error distortion measure. We provide necessary conditions for optimality of the mappings used at the encoder and the decoder. Note that virtually any source-channel communication system (including digital communication) is a special case of such mappings, as shown in Figure 1. A digital system including quantization, error correction and modulation boils down to a mapping from the source space $\mathbb{R}^m$ to the channel space $\mathbb{R}^k$. Hence the derived optimality conditions are generally valid and allow for digital communications as the extreme special case. Based on the optimality conditions we derive, we propose an iterative algorithm to optimize the mappings for any given m, k (i.e., for both bandwith expansion or compression) and for any given source-channel distribution. To our knowledge, this problem has not been solved, even for the scalar source-channel pair, except when both source and channel are scalar and Gaussian. We provide examples of such $m : k$ mappings for source-channel pairs (other than the well known scalar Gaussian example) and construct a source-channel system that performs such mappings that outperform those obtained in [6], [7], [8], [9].

In Section II, we formulate the problem. In Section III, we derive the necessary conditions for optimality. The iterative algorithm is presented in Section IV. We provide example

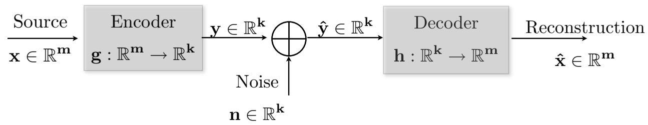


Fig. 1. A general scheme for a block based communication system

mappings and comparative results in Section V. We conclude the paper in Section VI.

II. Problem Formulation

A. Preleminaries and Notation

We consider the general communication system whose block diagram is shown in Figure 1. An m-dimensional vector source $\mathbf{x} \in \mathbb{R}^m$ is transformed into a k-dimensional vector $\mathbf{y} \in \mathbb{R}^{\mathbf{k}}$ by a nonlinear function $g : \mathbb{R}^m \to \mathbb{R}^k$ and transmitted over an additive noise channel. The received output $\hat{\mathbf{y}} = \mathbf{y}+\mathbf{n}$ is transformed to the estimate $\hat{\mathbf{x}}$ through a nonlinear function $h : \mathbb{R}^k \to \mathbb{R}^m$. The noise $\mathbf{n}$ is assumed to be independent of the source $\mathbf{x}$. The m-fold source density is denoted as $f_X(\mathbf{x})$ and the k-fold noise density is $f_N(\mathbf{n})$. Let $\mathbf{G}$ and $\mathbf{H}$ denote the set of all square integrable functions $\{\mathbf{g} : \mathbb{R}^{\mathbf{m}} \to \mathbb{R}^{\mathbf{k}}\}$ and $\{\mathbf{h} : \mathbb{R}^{\mathbf{k}} \to \mathbb{R}^{\mathbf{m}}\}$ respectively. Bold letters denote vectors or matrices.

B. Problem Statement

Given an i.i.d., zero-mean vector source $\mathbf{x} \in \mathbb{R}^{\mathbf{m}}$ with probability density $f_X(\mathbf{x})$, and an additive noise $\mathbf{n} \in \mathbb{R}^{\mathbf{k}}$ with $f_N(\mathbf{n})$, we want to minimize MSE $\mathbb{E}(|\mathbf{x} - \hat{\mathbf{x}}|^{\mathbf{2}})$ subject to the average power constraint

$$P[\mathbf{g}] = \int_{\mathbb{R}^m} \mathbf{g}(\mathbf{x})^{\mathbf{T}}\mathbf{g}(\mathbf{x})f_X(\mathbf{x})\mathbf{dx} \leq P, \quad (1)$$

where P is the allowed power, by suitably selecting the encoder $\mathbf{g} \in \mathbf{G}$ and decoder $\mathbf{h} \in \mathbf{H}$. Bandwidth compression-expansion is determined by the setting of the source and channel dimensions, k/m. The power constraint limits the choice of encoder function $\mathbf{g}$. Note that, without a constraint on $\mathbf{g}$, the channel can be made effectively noise free.

C. Asymptotical Bounds

From Shannon's lossy joint source channel coding theorem (see eg.[2]) for the memoryless Gaussian source-channel pair, it is known that if a code has asymptotic distortion D and the additive noise has power σ_n^2 then $R(D) \leq \frac{k}{m}C(\sigma_n^2)$ must hold, where $R(D)$ and $C(\sigma_n^2)$ denote the rate-distortion function and channel capacity per channel use, respectively. By letting $R(D) = \frac{k}{m}C(\sigma_n^2)$, a lower bound on the asymptotic distortion of any code can be obtained. The rate-distortion function for the memoryless Gaussian source under the squared-error distortion measure is given by

$$R(D) = \max(0, \frac{1}{2}\log\frac{\sigma_x^2}{D}). \quad (2)$$

for any distortion value $D \geq 0$. The capacity of the AWGN channel with input power constraint P and noise variance σ_n^2 is given by

$$C(\sigma_n^2) = \frac{1}{2}\log(1 + \frac{P}{\sigma_n^2})(\text{bits/channel use}). \quad (3)$$

Equating $R(D) = \frac{k}{m}C(\sigma_n^2)$ we reach the optimal performance theoretically attainable (OPTA). It is given by

$$D_{OPTA} = \frac{\sigma_x^2}{(1 + \frac{P}{\sigma_n^2})^{\frac{k}{m}}}. \quad (4)$$

Note that OPTA is derived without any delay constraints and the optimal delay constrained mapping may not reach OPTA. There is no achievable theoretical bound for joint source channel coding with limited delay.

III. Optimality Conditions

We proceed to develop the necessary conditions for optimality of the encoder and decoder subject to the average power constraint (1).

A. Optimal Decoder Given Encoder

Assume that the encoder $\mathbf{g}$ is fixed. Then the optimal decoder is the minimum mean square error estimator (MMSE) of $\mathbf{x}$ given $\hat{\mathbf{y}}$, i.e.,

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

Plugging the expressions for expectation, we obtain

$$\mathbf{h}(\hat{\mathbf{y}}) = \int \mathbf{x}\, f_{X|\hat{Y}}(\mathbf{x}|\hat{\mathbf{y}})\, \mathbf{dx}. \quad (6)$$

Applying Bayes' rule

$$f_{X|\hat{Y}}(\mathbf{x}|\hat{\mathbf{y}}) = \frac{f_X(\mathbf{x})f_{\hat{Y}|X}(\hat{\mathbf{y}}|\mathbf{x})}{\int f_X(\mathbf{x})\, f_{\hat{Y}|X}(\hat{\mathbf{y}}|\mathbf{x})\, \mathbf{dx}} \quad (7)$$

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

$$\mathbf{h}(\hat{\mathbf{y}}) = \frac{\displaystyle\int \mathbf{x}\, f_X(\mathbf{x})\, f_N[\hat{\mathbf{y}} - \mathbf{g}(\mathbf{x})]\, \mathbf{dx}}{\displaystyle\int f_X(\mathbf{x})\, f_N[\hat{\mathbf{y}} - \mathbf{g}(\mathbf{x})]\, \mathbf{dx}}. \quad (8)$$

B. Optimal Encoder Given Decoder

Assume that the decoder $\mathbf{h}$ is fixed. Our goal is to minimize MSE subject to the average power constraint. Let us write MSE explicitly as a functional of $\mathbf{g}$

$$D[\mathbf{g}] = \int\int [\mathbf{x} - \mathbf{h}(\mathbf{g}(\mathbf{x}) + \mathbf{n})]^{\mathbf{T}}[\mathbf{x} - \mathbf{h}(\mathbf{g}(\mathbf{x}) + \mathbf{n})]\, f_X(\mathbf{x})f_N(\mathbf{n})\mathbf{dxdn}. \quad (9)$$

We construct the Lagrangian cost functional to minimize

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

over the mapping $\mathbf{g}$. To obtain necessary conditions we apply the standard method in variational calculus [13]:

$$\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} J\left[\mathbf{g}(\mathbf{x})+\epsilon\eta(\mathbf{x})\right] = 0 \tag{11}$$

for all admissible variation functions $\eta(\mathbf{x})$. Note that, since the power constraint is considered in the cost function, the variation function $\eta(\mathbf{x})$ does not need to satisfy the power constraint (all continous differentiable functions $\eta : \mathbb{R}^{\mathbf{m}} \to \mathbb{R}^{\mathbf{k}}$ are admissible). Applying the above condition, we get

$$\int \left\{\lambda\mathbf{g}(\mathbf{x}) - \int \mathbf{h}'(\mathbf{g}(\mathbf{x})+\mathbf{n})[\mathbf{x}-\mathbf{h}(\mathbf{g}(\mathbf{x})+\mathbf{n})]f_N(\mathbf{n})\mathbf{dn})\right\} \eta(\mathbf{x})f_X(\mathbf{x})\mathbf{dx} = 0, \tag{12}$$

where $\mathbf{h}'$ denotes the Jacobian of the vector valued function $\mathbf{h}$. The equality for all admissible variation functions, $\eta(\mathbf{x})$, requires the expression in braces to be zero (more formally the Frechet derivative [13] should be zero to have an extremum point of the functional J [13]). This gives the necessary condition for optimality as

$$\nabla J[\mathbf{g}] = 0, \tag{13}$$

where

$$\nabla J[g] = \lambda f_X(\mathbf{x})\mathbf{g}(\mathbf{x}) - \int \mathbf{h}'(\mathbf{g}(\mathbf{x})+\mathbf{n})[\mathbf{x}-\mathbf{h}(\mathbf{g}(\mathbf{x})+\mathbf{n})] f_N(\mathbf{n})f_X(\mathbf{x})\mathbf{dn}. \tag{14}$$

Unlike the decoder, the optimal encoder is not in closed form but a necessary condition for optimality is given.

IV. Algorithm Design

The basic idea is to iteratively solve the necessary conditions for optimality, successively decreasing the total Lagrangian cost. Iterations are performed until the algorithm reaches a stationary point where the total cost does not decrease anymore. Solving for the optimal decoder is straightforward since the decoder can be expressed as closed form functional of known quantities, $\mathbf{g}$, $\mathbf{f_X}$ and $\mathbf{f_N}$. Since the encoder cannot be expressed as a closed form, we perform steepest descent search on the direction of the Frechet derivative of the total cost function with respect to the encoder mapping $\mathbf{g}$. At each iteration of $\mathbf{g}$ total cost decreases, iterations are kept till convergence. We keep updating $\mathbf{g}$ according to

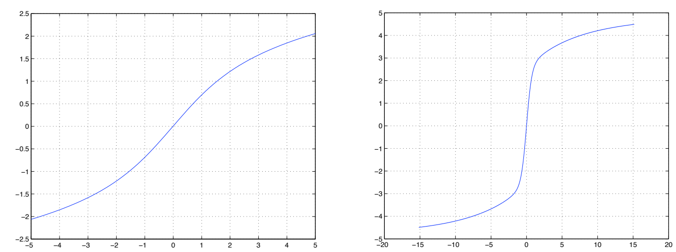
$$\mathbf{g}_{i+1}(\mathbf{x}) = \mathbf{g}_i(\mathbf{x}) - \mu\nabla\mathbf{J}[\mathbf{g}], \tag{15}$$

where i is the iteration index and μ is the step size. At each iteration i, total cost decreases monotonically and iterations are kept until convergence. As the initial condition for the encoder mapping $\mathbf{g_{init}}$, previously proposed suboptimal mappings[8], [7] can be used.

Note that, like every iterative algorithm of this type, there is no guarantee that the algorithm 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 or may apply more structured solutions such as deterministic annealing [14]. In this paper we used the previously proposed suboptimal mappings that are known to perform relatively well, as initial condition to avoid poor local minima.

Algorithm 1 Encoder and Decoder Iterations

Initialization Set $\mathbf{g}(\mathbf{x}) = \mathbf{g_{init}}(\mathbf{x})$, $i = 0$
  Find the optimal decoder using (8)
  Set cost
  while $cost_i < cost_{i-1}$ **do**
    while $cost_i < cost_{i-1}$ **do**
      $i \to i+1$
      Update the encoder according to (15)
      Evaluate the total cost according to (10)
    end while
    $i \to i+1$
    Update the decoder according to (8)
    Evaluate the total cost according to (10)
  end while
return($\mathbf{g}(\mathbf{x})$, $\mathbf{h}(\mathbf{x})$)



(a) Encoder mapping (b) Decoder mapping

Fig. 2. Example mappings for bi-modal GMM source, Gaussian channel, modes at 3 and -3

V. Results

We implemented the above algorithm by numerically calculating the integrals needed. For that purpose, we sampled the distribution on the uniform grid. We also assumed bounded support (-5σ to 5σ) for the infinite support distributions used in the examples.

A. *Scalar Mappings ($m = 1, k = 1$) for Gaussian Mixture Source - Gaussian Channel*

To demonstrate the use of nonlinear mappings consider the Gaussian mixture source with distribution $f_x(x) = \frac{1}{2\sqrt{2\pi}}\left\{e^{\frac{-(x-3)^2}{2}} + e^{\frac{-(x+3)^2}{2}}\right\}$ and unit variance Gaussian noise. The encoder and decoder mappings for this source-channel setting are given in Figure 2. As intuitively expected, since the two modes of the Gaussian mixture are well separated, each mode locally behaves as Gaussian. Hence the curve is roughly piece-wise linear, deviating significantly from a truly linear mapping. This suggests that for most distributions, considerable gains can be obtained by applying nonlinear mappings.

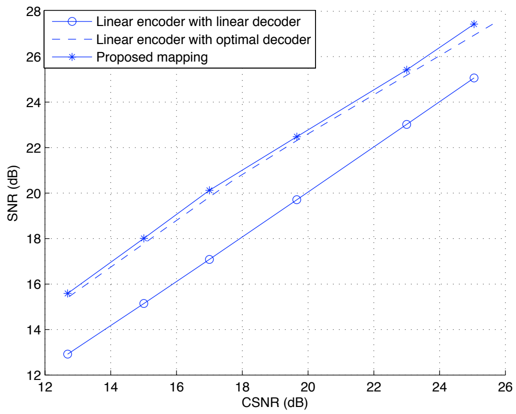


Fig. 3. Comparative results for 1:1 (scalar) mappings, GMM source-Gaussian channel

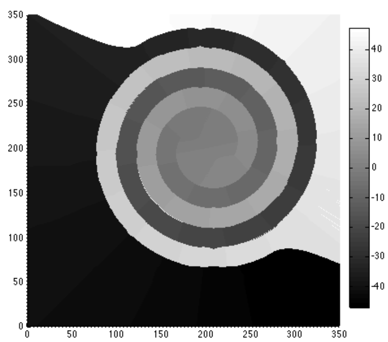


Fig. 4. Obtained mapping for 2:1 Gaussian, CSNR=40dB, SNR=19.41dB. The axes show the two dimensional input ($\mathbf{x}$) and the function value ($\mathbf{g}(\mathbf{x})$) is shown as the intensity level.

We also compare the performance of the proposed mappings to linear encoder and decoder and the intermediate option of linear encoder with the optimal decoder, as shown in Figure 3. The proposed mapping outperforms the other mappings for the shown range of CSNR values. Also note that optimizing only the decoder improves the performance significantly compared to a linear decoder.

B. ($m = 2, k = 1$) Gaussian source-channel mapping

In this section, we present a bandwidth compression example with 2:1 mappings for Gaussian source and channel. We compare the proposed mapping to the asymptotical bound (OPTA) and prior work [9]. We also compare the optimal encoder-decoder pair to the setting where only the decoder is optimized and encoder is fixed. In prior work [7], [8], [9], Archimedian spiral is found to perform well for Gaussian 2:1 mappings, and used for encoding and decoding with maximum likelihood criteria. We also initialize our algorithm with Archimedian spiral (i.e., we set the initial encoder mapping ($g_{init}(\mathbf{x})$ to Archimedian spiral). For details of Archimedian spiral and its settings, see eg. [9] and the references therein.

The obtained mapping is shown in Figure 4. As can be shown, the mapping obtained by our algorithm resembles the spiral although there is still a significant difference which can be seen in the performance results. Note that the encoding scheme is also different from prior work, as we continuously map the source to channel signal where two dimensional source is mapped to the closet point on the space filling spiral. The obtained mapping is also spiral shaped but the points between spiral arms are not mapped to exactly the same value, unlike the prior mappings used in the literature.

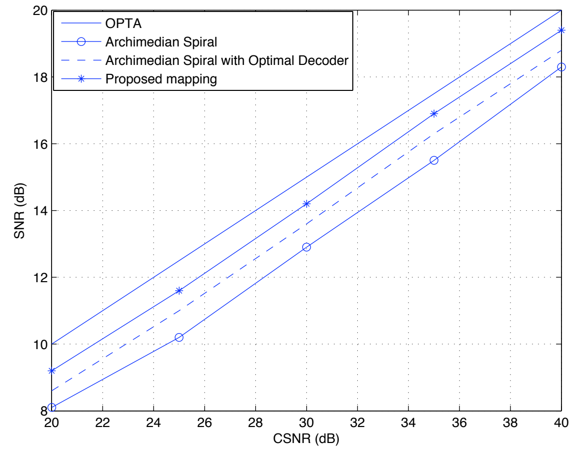


Fig. 5. Comparative results for Gaussian source-channel, 2:1 mapping

The comparative performance results are shown in Figure 5. The proposed mapping outperforms the Archimedian spiral [9] for the entire range of CSNR values. It is notable that intermediate option of optimizing the decoder improves the performance significantly compared to using the inverse spiral with maximum likelihood decoding.

C. ($m = 1, k = 2$) Gaussian source-channel mapping

In this section we compare the proposed mappings for bandwidth expansion of Gaussian scalar source transmitting over a vector Gaussian channel with two dimensions. We compare the obtained mapping to prior work and OPTA. Similarly, we use the prior work (inverse spiral) as the initial condition. The results are presented in Figure 6. The proposed mapping outperforms the inverse of Archimedian spiral [9] for the whole range of CSNR values. Note the gap between OPTA and and the achieved performance by our mappings is significantly greater than that of in 2:1 mappings case. There might be two possible reasons: i) The actual gap between theoretically achievable performance with zero delay and OPTA might be larger compared to 2:1 case. ii) Our mappings might have converged to a local minimum that is significantly greater than the global minimum. Currently, we are investigating this problem.

D. ($m = 2, k = 2$) Laplacian source - Gaussian channel mapping

In this part, we compare the mapping obtained through our method, to optimal two dimensional vector quantization followed by optimal channel coding that achieves the capacity, for a two dimensional Laplacian vector source and two dimensional Gaussian channel. The reason that we select a Laplacian

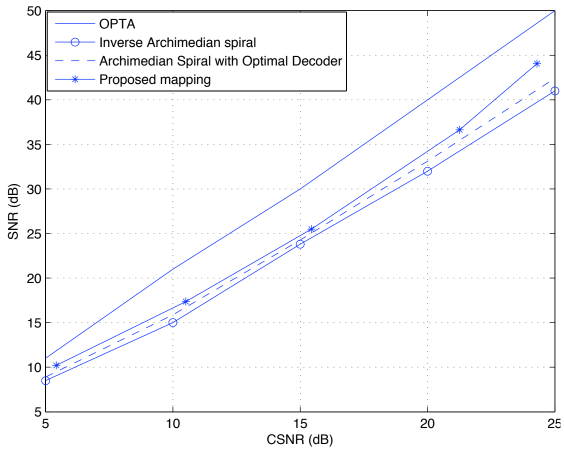


Fig. 6. Comparative results for Gaussian source- channel, 1:2 mapping

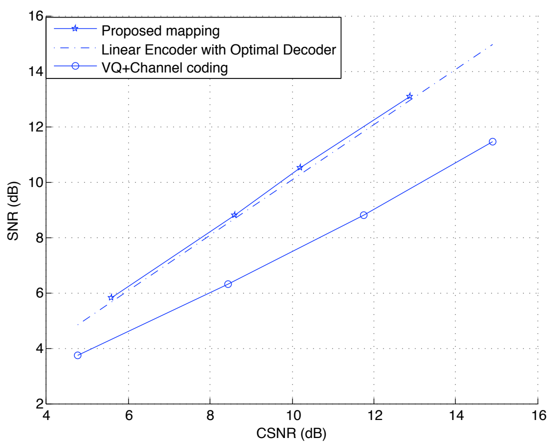


Fig. 7. Comparative results for Laplacian source-Gaussian channel, 2:2 mapping

source instead of a Gaussian is that, optimal mapping for a two dimensional Gaussian source over a Gaussian vector channel is a straightforward extension of that of the scalar Gaussian source over scalar Gaussian channel, whose solution is well known to be linear.

Note that, comparing our mapping to this method is unfair to our method because while our method is zero-delay, the capacity achieving channel codes would require long (possibly infinite) delay. Still, proposed mappings outperform vector quantization followed by channel coding significantly. For the vector quantization rate-distortion performance we use the results reported in [15], [16]. The comparative performance results are shown in Figure 7.

VI. Discussion and Future Work

In this paper, we derived the necessary conditions of optimality for a given source-channel system. Based on the necessary conditions, we derived an iterative algorithm which generates the locally optimal analog mappings. Comparative results and example mappings are provided and it is shown that the proposed method improves upon prior work. Note that like every iterative algorithm of this type, there is no guarantee that the algorithm will converge to the globally optimal solution. This problem can be solved by using a deterministic annealing approach, which is left as a future work.

VII. Acknowledgements

This work is supported by the NSF under grant CCF-0728986.

References

[1] T. Goblick Jr, "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, no. 4, pp. 558–567, 1965.
[2] T. Cover and J. Thomas, *Elements of Information Theory*. J.Wiley New York, 1991.
[3] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Transactions on Information Theory,*, vol. 49, no. 5, pp. 1147–1158, 2003.
[4] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.
[5] V. Kotelnikov, *The theory of optimum noise immunity*. New York: McGraw-Hill, 1959.
[6] A. Fuldseth and T. Ramstad, "Bandwidth compression for continuous amplitude channels based onvector approximation to a continuous subset of the source signal space," in *IEEE International Conference on Acoustics, Speech, and Signal Processing,*, vol. 4, 1997.
[7] S. Chung, "On the construction of some capacity-approaching coding schemes," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.
[8] T. Ramstad, "Shannon mappings for robust communication," *Telektronikk*, vol. 98, no. 1, pp. 114–128, 2002.
[9] F. Hekland, G. Oien, and T. Ramstad, "Using 2: 1 Shannon mapping for joint source-channel coding," in *Data Compression Conference, 2005. Proceedings. DCC 2005*, 2005, pp. 223–232.
[10] K. Lee and D. Petersen, "Optimal linear coding for vector channels," *IEEE Transactions on Communications*, vol. 24, no. 12, pp. 1283–1290, 1976.
[11] T. Fine, "Properties of an optimum digital system and applications," *IEEE Transactions on Information Theory*, vol. 10, no. 4, pp. 287–296, 1964.
[12] J. Gibson and T. Fischer, "Alphabet-constrained data compression," *IEEE Transactions on Information Theory*, vol. 28, no. 3, pp. 443–457, 1982.
[13] D. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons Inc, 1969.
[14] 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.
[15] T. Fischer and R. Dicharry, "Vector quantizer design for memoryless Gaussian, gamma, and Laplacian sources," *IEEE Transactions on Communications,*, vol. 32, no. 9, pp. 1065–1069, 1984.
[16] T. Lookabaugh and R. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 1020–1033, 1989.