On Zero Delay Source-Channel Coding: Functional Properties and Linearity Conditions

Emrah Akyol, Kumar Viswanatha, Kenneth Rose
Dep. of Electrical & Computer Eng.
UC Santa Barbara, CA, US
Email: {eakyol, kumar, rose}@ece.ucsb.edu

Tor Ramstad
Dep. of Electronics & Telecommunications
Norwegian U. of Science and Tech.,Trondheim, Norway,
Email: ramstad@iet.ntnu.no

***Abstract*—In this paper, we study the zero-delay source-channel coding problem, and specifically 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 transmission power constraint and for the mean square error distortion. We first study the functional properties of this problem and show that the objective is concave in the source and noise densities and convex in the density of the input to the channel. We then present the necessary conditions for optimality of the encoder and decoder mappings. A well known result in information theory pertains to the linearity of optimal encoding and decoding mappings in the scalar Gaussian source and channel setting, at all channel signal-to-noise ratios (CSNRs). In this paper, we study this result more generally, beyond the Gaussian source and channel, and derive the necessary and sufficient condition for linearity of optimal mappings, given a noise (or source) distribution, and a specified power constraint. We also prove that the Gaussian source-channel pair is unique in the sense that it is the only source-channel pair for which the optimal mappings are linear at more than one CSNR value. Moreover, we show the asymptotic linearity of optimal mappings for low CSNR if the channel is Gaussian regardless of the source and, at the other extreme, for high CSNR if the source is Gaussian, regardless of the channel.**

I. Introduction

The zero delay source-channel coding problem has recently gained revived interest [1]–[5]. In this paper, we study the functional properties of this problem and the conditions for linearity of the optimal mappings, building on our prior work [6], [7].

II. Problem Definition

We consider the general communication system with the block diagram as shown in Figure 1. Let $\mathcal{S}_m^k$ denote the set of Borel measurable, square integrable functions $\{\boldsymbol{f}: \mathbb{R}^m \rightarrow \mathbb{R}^k\}$. An m-dimensional zero mean[1] vector source $\boldsymbol{X} \in \mathbb{R}^m$ is mapped into a k-dimensional vector $\boldsymbol{Y} \in \mathbb{R}^k$ by the function $\boldsymbol{g} \in \mathcal{S}_m^k$, and transmitted over an additive noise channel. The received vector $\hat{\boldsymbol{Y}} = \boldsymbol{Y} + \boldsymbol{Z}$ is mapped by the decoder to the estimate $\hat{\boldsymbol{X}}$ via the function $\boldsymbol{h} \in \mathcal{S}_k^m$. The zero mean noise $\boldsymbol{Z}$ is assumed to be independent of the source $\boldsymbol{X}$. The m-fold source density is denoted $f_X(\cdot)$ and the k-fold noise density is $f_Z(\cdot)$ with characteristic functions $F_X(\omega)$ and $F_Z(\omega)$, respectively.

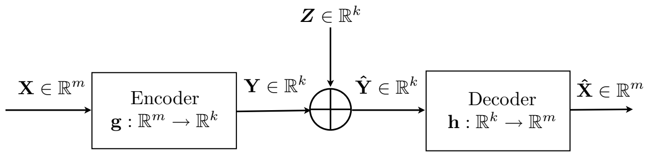


Fig. 1. A general block-based point-to-point communication system

The objective is to minimize, over the choice of encoder $\boldsymbol{g} \in \mathcal{S}_m^k$ and decoder $\boldsymbol{h} \in \mathcal{S}_k^m$, the distortion

$$D(\boldsymbol{g}, \boldsymbol{h}) = \mathbb{E}\{||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2\}, \tag{1}$$

subject to the average power constraint,

$$P(\boldsymbol{g}) = \mathbb{E}\{||\boldsymbol{g}(\boldsymbol{X})||^2\} \leq P_T, \tag{2}$$

where P_T is the specified transmission power level. Bandwidth compression-expansion is determined by the setting of the source and channel dimensions, k/m. To impose the power constraint, we construct the Lagrangian cost functional:

$$J(\boldsymbol{g}, \boldsymbol{h}) = D(\boldsymbol{g}, \boldsymbol{h}) + \lambda P(\boldsymbol{g}) \tag{3}$$

to minimize over the mappings $\boldsymbol{g}(\cdot)$ and $\boldsymbol{h}(\cdot)$.

III. Functional Properties of Zero-Delay Source-Channel Coding Problem

In this section, we study the functional properties of the optimal zero-delay source-channel coding problem. These properties are not only important in their own right, but also, as we will show in the following sections, enable the derivation of several subsequent results. Let us restate the Lagrangian cost (3), as $J(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{g}, \boldsymbol{h})$ which makes explicit its dependence on the source and channel noise $\boldsymbol{X} \sim f_X(\cdot)$ and $\boldsymbol{Z} \sim f_Z(\cdot)$ beside the deterministic mappings $\boldsymbol{g}(\cdot)$ and $\boldsymbol{h}(\cdot)$ as:

$$J(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{g}, \boldsymbol{h}) = \mathbb{E}\left\{||\boldsymbol{X} - \boldsymbol{h}(\boldsymbol{g}(\boldsymbol{X}) + \boldsymbol{Z})||^2\right\} + \lambda \mathbb{E}\left\{||\boldsymbol{g}(\boldsymbol{X})||^2\right\} \tag{4}$$

The minimum achievable cost is

$$J_m(\boldsymbol{X}, \boldsymbol{Z}) \triangleq \inf_{\boldsymbol{g}, \boldsymbol{h}} J(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{g}, \boldsymbol{h}) \tag{5}$$

Similarly, conditioned on another random variable $\boldsymbol{U}$, $J_m(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{U})$ denotes the overall cost when $\boldsymbol{U}$ is available

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

[1]The zero mean assumption is not necessary, but it considerably simplies the notation. Therefore, it is made throughout the paper.

to both encoder and decoder. We define J_r as the value of overall cost as a function of $\boldsymbol{g}(\cdot)$, when $\boldsymbol{h}(\cdot)$ has already been optimized for $\boldsymbol{g}(\cdot)$:

$$J_r(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{g}) \triangleq \inf_{\boldsymbol{h}} J(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{g}, \boldsymbol{h}) \tag{6}$$

A. Concavity of J_m in $f_X(\cdot)$ and $f_Z(\cdot)$

In this section, we show the concavity of the minimum cost, J_m in the source density $f_X(\cdot)$ and in the channel noise density $f_Z(\cdot)$. Similar results were derived for minimum mean squared error estimation in the scalar setting, in [8], where no encoder is present in the problem formulation. We start with the following lemma.

Lemma 1: Conditioning cannot increase the overall cost, i.e., $J_m(\boldsymbol{X}, \boldsymbol{Z}) \geq J_m(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{U})$ for any $\boldsymbol{U}$.

Proof: The knowledge of $\boldsymbol{U}$ cannot increase the total cost, since we can always ignore U and use the $\boldsymbol{g}(\cdot), \boldsymbol{h}(\cdot)$ pair that is optimal for $J_m(\boldsymbol{X}, \boldsymbol{Z})$ to achieve $J_m(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{U}) = J_m(\boldsymbol{X}, \boldsymbol{Z})$. ∎

Using Lemma 1, we prove the following theorem, which states the concavity of the minimum cost $J_m(\boldsymbol{X}, \boldsymbol{Z})$.

Theorem 1: J_m is concave in $f_X(\cdot)$ and $f_Z(\cdot)$.

Proof: Let $\boldsymbol{X}$ has the density $f_X = pf_{X_1} + (1-p)f_{X_2}$, where f_{X_1} and f_{X_2} respectively denote the densities of random variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Then, $\boldsymbol{X}$ can be expressed, in terms of a time sharing random variable U which takes values in the alphabet $\{1, 2\}$, with $\mathbb{P}\{U = 1\} = p$: $\boldsymbol{X} = \boldsymbol{X}_U$. Then,

$$\begin{aligned} J_m(\boldsymbol{X}, \boldsymbol{Z}) \geq & J_m(\boldsymbol{X}, \boldsymbol{Z}|U) \\ = & pJ_m(\boldsymbol{X}_1, \boldsymbol{Z}) + (1-p)J_m(\boldsymbol{X}_2, \boldsymbol{Z}) \end{aligned} \tag{7}$$

which proves the concavity of $J_m(\boldsymbol{X}, \boldsymbol{Z})$ for fixed f_Z. Similar arguments on $\boldsymbol{Z}$ prove that $J_m(\boldsymbol{X}, \boldsymbol{Z})$ is also concave in f_Z for fixed f_X. ∎

B. Convexity of J_r in $f_Y(\cdot)$

Here, we assume that the source and the channel are scalar, i.e., $m = k = 1$, for simplicity, although our results can be extended to higher matched dimensions, $m = k, \forall k \in \mathbb{N}$. We show the convexity of $J_r(X, Z, g)$ in the channel input density $f_Y(\cdot)$ of $Y = g(X)$. An important distinction to make is that convexity in $g(\cdot)$ is not implied. A trivial example to demonstrate non-convexity in $g(\cdot)$ is the scalar Gaussian source and channel setting, where both $Y = \sqrt{\frac{P_T}{\sigma_X^2}}X$ and $Y = -\sqrt{\frac{P_T}{\sigma_X^2}}X$ are optimal (when used in conjunction with their respective optimal decoders). This example also leads to the intuition that the cost functional may be "essentially" convex (i.e., convex up to the sign of $g(\cdot)$), although it is clearly not convex in the strict sense. It turns out that this intuition is correct: $J_r(X, Z, g)$ is convex in $f_Y(\cdot)$.

Towards showing convexity, we first introduce the idea of probabilistic (random) mappings, similar, in spirit, to the random encoders used in the coding theorems [9]. We reformulate the mapping problem by allowing random mappings, i.e., we relax the mapping from a deterministic function to $Y = g(X)$ to a probabilistic transformation, expressed as $f_{Y|X}(x, y)$. Note that similar relaxation to stochastic settings have been used in the literature, e.g. recently in [10]. We define this "generalized" mapping problem as: minimize $J_{gen}(X, Y, Z)$ over the conditional density $f_{Y|X}$ where the cost functional J_{gen} is defined as

$$J_{gen}(X, Y, Z) \triangleq \inf_h \mathbb{E}\{(X - h(Y+Z))^2\} + \lambda\mathbb{E}\{Y^2\}. \tag{8}$$

We first need to show that this relaxation does not change the solution space.

Lemma 2: $Y \sim f_Y(\cdot)$ which minimizes (8) is a deterministic function of the input $Y = g(X)$, i.e., $J_m(X, Z) = \inf_g J_r(X, Z, g) = \inf_{f_{Y|X}} J_{gen}(X, Y, Z)$.

Proof: Let us first define an auxiliary function

$$G(X, Y, Z) \triangleq (X - h(Y+Z))^2 + Y^2 \tag{9}$$

Next, we observe that

$$\begin{aligned} &\inf_h \inf_{f_{Y|X}} J_{gen}(X, Y, Z) \\ &= \inf_h \int f_Z(z) \int f_X(x) \inf_{f_{Y|X}} \left\{ \int G(x, y, z) f_{Y|X}(x, y) dy \right\} dx dz \end{aligned} \tag{10}$$

The minimization in (10) can be done, for a fixed $h(\cdot)$, by choosing the y that minimizes $G(x, y, z)$ for each x. Hence, for any fixed $h(\cdot)$, the minimizing $f_{Y|X}$ is deterministic. Using the optimal $h(\cdot)$ as the fixed $h(\cdot)$ in (10), we show that the optimal $Y \sim f_Y(\cdot)$ is a deterministic function: $Y = g(X)$. ∎

Next, we proceed to show that the generalized mapping problem is convex in $f_Y(\cdot)$. To this aim, we show that J_{gen} can be written in terms of a known metric in probability theory, the Wasserstein metric [11] and use its functional properties. The Wasserstein metric is a metric defined on the quadratic Wasserstein space $\mathcal{P}_2(\mathbb{R})$, defined for $S, Q \in \mathcal{P}_2(\mathbb{R})$ as

$$W_2(S, Q) = \inf \{||X - Y||_2 : X \sim S, Y \sim Q\} \tag{11}$$

where $||X - Y||_2 \triangleq \sqrt{\mathbb{E}\{(X - Y)^2\}}$ and the infimum is over the joint distribution of X and Y. The following properties of this metric will be used to derive the subsequent results.

Lemma 3 ([11]): $W_2(S, Q)$ satisfies the following properties:

1) The metric $W_2(S, Q)$ is lower semi-continuous in both S and Q.

2) For a given S, $W_2^2(S, Q)$ is convex in Q.

Next, we present our main result in this section. In the derivation of the subsequent results, we limit the space of decoding functions to monotone increasing, without any loss of generality (see e.g.. [8]).

Theorem 2: J_r is convex in $f_Y(\cdot)$ and hence the solution to the mapping problem is unique in $f_Y(\cdot)$.

Proof: We will first express $J_m(X, Y)$ as a minimization over $f_Y(\cdot)$. Let us define $V = h(Y+Z)$ for a fixed $h(\cdot)$. Then, using Lemma 2, $J_m(X, Y)$ can be re-written as

$$\begin{aligned} J_m(X, Y) &= \inf_h \inf_{f_{V|X}} \mathbb{E}\{(X - V)^2\} + \inf_{f_{Y|X}} \lambda\mathbb{E}\{Y^2\} \\ &= \inf_h \inf_{f_V} \left\{W_2^2(f_X, f_V)\right\} + \lambda \inf_{f_Y} \mathbb{E}\{Y^2\}. \end{aligned} \tag{12}$$

The first term in the right hand side of (12) is convex in f_V since, $W_2^2(f_X, f_V)$ is convex in f_V (due to Lemma 3-property 2) when f_X is fixed, and the pointwise minimizer of a convex function is convex. Since Y and V are related in a one-to-one manner through $V = h(Y+Z)$ and $h(\cdot) \in \mathcal{S}^+$, this term is convex also in f_Y. Since $\mathbb{E}\{Y^2\}$ is linear in $f_Y(\cdot)$, we conclude that J_m is the infimum of a convex functional of $f_Y(\cdot)$, where the infimum is taken over $f_Y(\cdot)$, which implies the solution is unique in $f_Y(\cdot)$. Given that the solution is unique, we can express J_m as

$$J_m(X,Z) = \inf_{f_Y} J_r(X,Z,g) \tag{13}$$

almost everywhere (*a.e.*) in X and Z, where $Y = g(X)$. Hence, the functional we are interested in is indeed $J_r(X,Z,g)$ which is convex $f_Y(\cdot)$. ■

A practically important consequence of Theorem 2 is stated in the following corollary.

Corollary 1: J_r is convex in $g(\cdot)$ where $g(\cdot) \in \mathcal{G}^+$.

Proof: There is one-to-one mapping between Y and the encoder $g(\cdot)$ as $\mathcal{F}_X(X) = \mathcal{F}_Y(g(X))$ where $\mathcal{F}_X$ and $\mathcal{F}_Y$ denote the cumulative distribution functions of X and Y respectively. It follows from Theorem 2 that for any f_{Y_1} and f_{Y_2} and $1 \geq \alpha \geq 0$

$$\alpha J_r(f_{Y_1}) + (1-\alpha) J_r(f_{Y_2}) \geq J_r(\alpha f_{Y_1} + (1-\alpha) f_{Y_2}) \tag{14}$$

Since $J_r(f_Y)$ is achieved by a unique $g(\cdot) \in \mathcal{G}^+$, we have

$$\alpha J_r(f_{g_1}) + (1-\alpha) J_r(f_{g_2}) \geq J_r(\alpha f_{g_1} + (1-\alpha) f_{g_2}) \tag{15}$$

which shows the convexity of J_r in $g(\cdot)$, where $g(\cdot) \in \mathcal{G}^+$. ■

Remark 1: The optimal mappings, i.e., the mappings that achieve the infimum in (8) exist. To see this, we use the semi-lower continuity property of the $W_2(S,Q)$ in both S and Q as given in Lemma 3. The set of Y is compact since $\mathbb{E}\{Y^2\} \leq P_T$, hence the infimum in the problem definition is achievable.

C. Optimality Conditions

The necessary conditions for optimality, in the general setting of $m, k \in \mathbb{N}$, were derived in [6]. Here, we present them in the following theorem.

Theorem 3 ([6]): Given source and noise densities, a coding scheme $(\boldsymbol{g}(\cdot), \boldsymbol{h}(\cdot))$ is optimal *only if*

$$\boldsymbol{g}(\boldsymbol{x}) = \frac{1}{\lambda} \int \boldsymbol{h}'(g(\boldsymbol{x})+z)[\boldsymbol{x} - \boldsymbol{h}(\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{z})] f_Z(\boldsymbol{z}) \mathrm{d}\boldsymbol{z}, \tag{16}$$

$$\boldsymbol{h}(\hat{\boldsymbol{y}}) = \frac{\int \boldsymbol{x}\, f_X(\boldsymbol{x})\, f_Z[\hat{\boldsymbol{y}} - \boldsymbol{g}(\boldsymbol{x})]\, \mathrm{d}\boldsymbol{x}}{\int\, f_X(\boldsymbol{x})\, f_Z\, [\hat{\boldsymbol{y}} - \boldsymbol{g}(\boldsymbol{x})]\, \mathrm{d}\boldsymbol{x}}, \tag{17}$$

where varying λ provides solutions at different levels of the power constraint P_T. In fact, λ is the slope of the distortion-power curve: $\lambda = -\frac{dD}{dP_T}$.
The necessary conditions in Theorem 3 are not sufficient in general, as is demonstrated by the following corollary.

Corollary 2: For Gaussian source and channel, the necessary conditions of Theorem 3 are satisfied by linear mappings $\boldsymbol{g}(\boldsymbol{x}) = K_e \boldsymbol{x}$ and $\boldsymbol{h}(\boldsymbol{y}) = K_d \boldsymbol{y}$ for some $K_e \in \mathbb{R}^{m\times k}, K_d \in \mathbb{R}^{k \times m}$ for any $m, k \in \mathbb{N}$.

Proof: Linear mappings satisfy (16), regardless of the source and channel densities. Optimal decoder is linear in the Gaussian source and channel setting, hence the linear encoder-decoder pair satisfies both of the necessary conditions. ■
Although linear mappings satisfy the necessary conditions of optimality for the Gaussian case, they are known to be highly suboptimal when dimensions of source and channel do not match, i.e., $m \neq k$, see e.g. [12].

IV. On Linearity of Optimal Mappings

In this section, we address the problem of "linearity" of optimal encoding and decoding mappings, where we focus on the scalar setting, $m = k = 1$.

A. Gaussian Source and Channel

We briefly revisit the special case in which both X and Z are Gaussian, $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Z \sim \mathcal{N}(0, \sigma_Z^2)$. It is well known that the optimal mappings are linear, i.e., $g(X) = k_e X$ and $h(Y) = k_d Y$ where k_e and k_d are given by

$$k_e = \sqrt{\frac{P_T}{\sigma_X^2}}, \quad k_d = \frac{1}{k_e}\left(\frac{P_T}{P_T + \sigma_Z^2}\right). \tag{18}$$

B. On Simultaneous Linearity of Optimal Mappings

Here, we show, in two steps that optimality requires that either both mappings be linear or that they both be nonlinear.

Lemma 4: The optimal encoder is linear *a.e.* if the optimal decoder is linear.

Proof: Let us plug $h(y) = k_d y$ for some $k_d \in \mathbb{R}$ in (16). Noting that $h'(y) = k_d$ *a.e.* in y, we have

$$\lambda g(x) = k_d \int (x - k_d g(x) - k_d z) f_Z(z) dz \tag{19}$$

a.e. in x. Evaluating the integral and noting that $\mathbb{E}\{Z\} = 0$, we have

$$\lambda g(x) = k_d (x - k_d g(x)) \tag{20}$$

a.e. and hence $g(x) = \frac{k_d}{\lambda + k_d^2} x \triangleq k_e x$. ■

Lemma 5: The optimal decoder is linear *a.e.* if the optimal encoder is linear.

Proof: Plugging $g(x) = k_e x$ for some $k_e \in \mathbb{R}$ in (16), we obtain

$$\lambda k_e x = \int (x - h(k_e x + z)) h'(k_e x + z) f_Z(z) dz \tag{21}$$

a.e. in x. Since $h(\cdot)$ is a function $\mathbb{R} \to \mathbb{R}$, Weierstrass theorem [13] guarantees that we can uniformly approximate $h(\cdot)$ arbitrarily closely by a polynomial

$$h(y) = \sum_{r=0}^{\infty} \alpha_r y^r \tag{22}$$

a.e. in y. Plugging (22) in (21) and interchanging the summation and integration we obtain

$$\begin{aligned} &- \lambda k_e x + x - \sum_{i=0}^{\infty} i\alpha_i \int (k_e x + z)^{i-1} f_Z(z) dz \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i\alpha_i\alpha_j \int (k_e x + z)^{i-1} (k_e x + z)^j f_Z(z) dz. \end{aligned} \tag{23}$$

Note that the above equation must hold *a.e.* in x, hence the coefficients of x^r must be identical for all r. Opening up the expressions $(k_e x + z)^{i-1}$ and $(k_e x + z)^j$ via binomial expansion, we have the following set of equations

$$\begin{aligned}&\sum_{i=r+1}^{\infty} i\binom{i-1}{r}\alpha_i \mathbb{E}\{Z^{i-1-r}\}\\ =&\sum_{i=0}^{\infty}\sum_{i,j=0}^{\infty}\sum_{l=0}^{i-1}\sum_{p=r-l+1}^{j-1}\binom{j}{p}\binom{i-1}{l} i\alpha_i\alpha_j\mathbb{E}\{Z^{i+j-1-p-l}\},\end{aligned} \tag{24}$$

which must hold for all $r \geq 2$. We note that every equation introduces a new variable α_r, so each new equation is linearly independent of its predecessors. Next, we solve these equations recursively, starting from $r = 1$. At each r, we have one unknown (α_r) which is related "linearly" to known constants. Since the number of linearly independent equations is equal to the number of unknowns for each r, there must exist a unique solution. We know that $\alpha_r = 0$, for all $r \geq 2$ is a solution to (24), so it is the only solution.

■

Next, we summarize our main result pertaining to the simultaneous linearity of optimal encoder and decoder.

Theorem 4: The optimal mappings are either both linear or they are both nonlinear.

Proof: The proof follows from Lemma 4 and Lemma 5. ■

C. Conditions for Linearity of Optimal Mappings

In this section, we study conditions for linearity of optimal encoder or decoder. Towards obtaining our main result, we will use the following auxiliary lemma.

Lemma 6: The linear encoder and decoder in (18) satisfy the first of the necessary conditions of optimality (16) regardless of the source and channel densities.

Proof: The proof follows from substitution of (18) in (16). ■

The following theorem presents the necessary and sufficient condition for linearity of optimal mappings.

Theorem 5: For a given power limit P_T, noise Z with variance σ_Z^2 and characteristic function $F_Z(\omega)$, source X with variance σ_X^2 and characteristic function $F_X(\omega)$, mappings $g(X) = k_e X$ or $h(\hat{Y}) = k_d \hat{Y}$ are optimal *if and only if*

$$F_X(\alpha\omega) = F_Z^{\gamma}(\omega), \tag{25}$$

where $\gamma = \frac{P_T}{\sigma_Z^2}$ and $\alpha = \sqrt{\frac{P_T}{\sigma_X^2}}$.

Proof: Theorem 4 states that the optimal encoder is linear if and only if the optimal decoder is linear. Hence, we will only focus on the case where encoder and decoder are simultaneously linear. The first necessary condition is satisfied by Lemma 6, hence only the second necessary condition, (17) remains to be verified. Plugging $g(X) = k_e X$ and $h(\hat{Y}) = k_d \hat{Y}$ in (17), we have

$$k_d \hat{y} = \frac{\int x\, f_X(x)\, f_Z(\hat{y} - k_e x)\mathrm{d}x}{\int\, f_X(x)\, f_Z(\hat{y} - k_e x)\,\mathrm{d}x}. \tag{26}$$

Expanding (26), we obtain

$$k_d\hat{y}\int\, f_X(x)\, f_Z(\hat{y} - k_e x)\,\mathrm{d}x = \int x\, f_X(x)\, f_Z\left(\hat{y} - k_e x\right)\,\mathrm{d}x.$$

Taking the Fourier transform if both sides and via change of variables $u \triangleq \hat{y} - k_e x$, we have

$$\begin{aligned}\int\int k_d(u + k_e x) f_X(x) f_Z(u)\exp(-j\omega(u + k_e x))dxdu\\ = \int\int x f_X(x) f_Z(u)\exp(-j\omega(u + k_e x))dxdu,\end{aligned}$$

and rearranging the terms, we obtain

$$\left(\frac{1 - k_e k_d}{k_e k_d}\right) F_Z(\omega) F_X'(k_e\omega) = F_X(k_e\omega) F_Z'(\omega). \tag{27}$$

Noting that

$$\gamma = \frac{k_e k_d}{1 - k_e k_d} = \frac{P_T}{\sigma_Z^2}, \tag{28}$$

we have

$$\frac{F_X'(k_e\omega)}{F_X(k_e\omega)} = \gamma\frac{F_Z'(\omega)}{F_Z(\omega)}, \tag{29}$$

which implies

$$(\log F_X(k_e\omega))' = (\log F_Z^{\gamma}(\omega))'. \tag{30}$$

The solution to this differential equation is

$$\log F_X(k_e\omega) = \log F_Z^{\gamma}(\omega) + C, \tag{31}$$

where C is constant. Noting that $F_X(0) = F_Z(0) = 1$, we determine $C = 0$ and hence

$$F_X(k_e\omega) = F_Z^{\gamma}(\omega). \tag{32}$$

Since the solution is essentially unique, due to Corollary 1, (32) is not only necessary but also sufficient. ■

D. Implications of the Matching Conditions

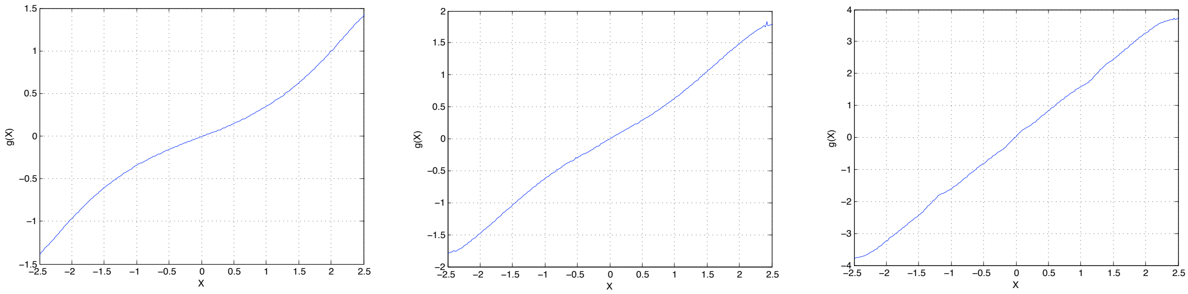
In this section, we explore some special cases obtained by varying CNSR (i.e., γ).

Theorem 6: Given a source and noise of equal variance identical to the power limit ($\sigma_X^2 = \sigma_Z^2 = P_T$), the optimal mappings are linear *if and only if* $f_X(x) = f_Z(x), a.e.$ and in which case, the optimal encoder is $g(X) = X$ and the optimal decoder is $h(\hat{Y}) = \frac{1}{2}\hat{Y}$.

Proof: It is straightforward to see from (25) that, at $\gamma = 1$, the characteristic functions must be identical. Since the characteristic function uniquely determines the distribution [14], $f_X(x) = f_Z(x)$, *a.e.* ■

Remark 2: Theorem 6 holds irrespective of the source (or channel) density, which demonstrates the generality of linearity of the optimal mappings beyond the well known example of scalar Gaussian source and channel.

Next, we investigate the asymptotic behavior of optimal encoding and decoding functions at low and high CSNR. The results of our asymptotic analysis are of practical importance since they justify, under certain conditions, the use of linear mappings without recourse to complexity arguments.



(a) CSNR=-5.70dB (b) CSNR=-1.69dB (c) CSNR=5.69dB

Fig. 2. This figure shows the optimal encoder at various CSNR values when $X \sim \mathcal{N}(0,1)$ and Z is distributed uniformly on the interval $[-1, 1]$.

Theorem 7: In the limit $\gamma \to 0$, the optimal encoding and decoding functions are asymptotically linear if the channel is Gaussian, regardless of the source. Similarly, as $\gamma \to \infty$, the optimal mappings are asymptotically linear if the source is Gaussian, regardless of the channel.

Proof: The proof follows from applying the central limit theorem (CLT) [14] to the matching condition (25). CLT states that as $\gamma \to \infty$, for any finite variance noise Z, the characteristic function of the matching source $F_X(\omega) = F_Z^{\gamma}(\omega/k_e)$ converges to the Gaussian characteristic function. Hence, at asymptotically high CSNR, any noise distribution is matched by the Gaussian source. Similarly, as $\gamma \to 0$ and for any $F_X(\omega)$, $F_X^{\frac{1}{\gamma}}(k_e\omega)$ converges to the Gaussian characteristic function, hence the optimal mappings are asymptotically linear. ■

Let us consider a numerical example that illustrates the findings in Theorem 7. Consider a setting where X is Gaussian $X \sim \mathcal{N}(0,1)$ and Z is uniform over $[-1, 1]$. We change γ (CSNR) by varying allowed power P_T, and observe how the optimal mappings[2] behave for different γ. Figure 2 demonstrates that the optimal encoder mapping converges to linear as CSNR increases, as anticipated by Theorem 7.

Let us next consider a setup with given source and noise variables and a given power which may be scaled to vary the CSNR, γ. Can the optimal mappings be linear at multiple values of γ (i.e., at different power)? This question is motivated by the practical setting where γ is not known in advance or may vary. Below, we show that the Gaussian source-Gaussian noise is the only pair for which the optimal mappings are linear at multiple CSNRs.

Theorem 8: The optimal encoding mapping is linear at two different power levels P_1 and P_2 *if and only if* source and noise are both Gaussian.

Proof: The proof follows from the proof of Theorem 4 in [7] with straightforward adaptation to this setting. ■

E. On the Existence of Matching Source and Channel

Given a valid characteristic function $F_Z(\omega)$, and for some $\gamma \in \mathbb{R}^+$, the function $F_Z^{\gamma}(\omega)$ may or may not be a valid characteristic function, which determines the existence of a matching source. For example, matching is guaranteed for integer γ and it is also guaranteed for infinitely divisible Z. Such matching conditions were studied detail in [7], to which we refer for brevity.

[2]We numerically calculated the mappings using the algorithm in [6].

V. CONCLUSION

In this paper, we studied the functional properties of the zero-delay source channel coding problem and conditions for linearly of optimal mappings. We showed the cost functional is concave in the source and noise densities and convex in the density of the input to the channel. We then derived the necessary and sufficient condition for linearity of optimal mappings, given noise and source distributions, and a power constraint. The matching condition has several implications, one of which is on the asymptotic linearity of optimal mappings.

REFERENCES

[1] F. Hekland, P. Floor, and T. Ramstad, “Shannon-Kotelnikov mappings in joint source-channel coding,” *IEEE Trans. on Comm.*, vol. 57, no. 1, pp. 94–105, 2009.
[2] Y. Hu, J. Garcia-Frias, and M. Lamarca, “Analog joint source-channel coding using non-linear curves and mmse decoding,” *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3016–3026, 2011.
[3] V. Vaishampayan and S. Costa, “Curves on a sphere, shift-map dynamics, and error control for continuous alphabet sources,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1658–1672, 2003.
[4] X. Chen and E. Tuncel, “Zero-delay joint source-channel coding for the Gaussian Wyner-Ziv problem,” in *Proc. IEEE Int. Symp. on Inf. Theory*, 2011, pp. 2929–2933.
[5] A. Ingber, I. Leibowitz, R. Zamir, and M. Feder, “Distortion lower bounds for finite dimensional joint source-channel coding,” in *Proc. of IEEE Int. Symp. on Inf. Th.*, 2008, pp. 1183–1187.
[6] E. Akyol, K. Rose, and T. Ramstad, “Optimal mappings for joint source channel coding,” in *Proc. of IEEE Inf. Theory Workshop*, 2010.
[7] E. Akyol, K. Viswanatha, and K. Rose, “On conditions for linearity of optimal estimation,” *IEEE Trans.on Inf. Th.*, vol. 58, no. 6, pp. 3497 –3508, 2012.
[8] Y. Wu and S. Verdú, “Functional properties of minimum mean-square error and mutual information,” *IEEE Trans. on Inf. Th.*, vol. 58, no. 3, pp. 1289–1301, 2012.
[9] C. Shannon, “The mathematical theory of information,” *Bell System Technical Journal*, vol. 27, no. 6, pp. 379 –423, 1949.
[10] Y. Wu and S. Verdú, “Witsenhausens counterexample: A view from optimal transport theory,” in *IEEE Conf. on Dec. and Control*, 2011.
[11] C. Villani, *Optimal transport: old and new*. Springer Verlag, 2009.
[12] T. Başar, B. Sankur, and H. Abut, “Performance bounds and optimal linear coding for discrete-time multichannel communication systems,” *IEEE Trans. on Inf. Th.*, vol. 26, no. 2, pp. 212–217, 1980.
[13] R. Dudley, *Real Analysis and Probability*. Cambridge U. Press, 2002.
[14] P. Billingsley, *Probability and Measure*. John Wiley & Sons Inc., 2008.