Duality Theorems for Joint Source-Channel Coding[†]

Udar Mittal and *Nam Phamdo*

Electrical Engineering Department
State University of New York at Stony Brook
Stony Brook, NY 11794-2350
udar@sbee.sunysb.edu, phamdo@sbee.sunysb.edu

Abstract

We consider joint source-channel coding for a memoryless Gaussian source and an additive white Gaussian noise (AWGN) channel. For a given code defined by an encoder-decoder pair (α, β), its dual code is obtained by interchanging the encoder and decoder: (β, α). It is shown that if a code (α, β) is optimal at rate ρ channel uses per source sample and if it satisfies a certain uniform continuity condition, then its dual code (β, α) is optimal for rate $1/\rho$ channel uses per source sample. Further, it is demonstrated that there is a code which is optimal but its dual code is not optimal. Finally, using random coding, we show that there is an optimal code which has an optimal dual.

Keywords: Joint source-channel coding, duality, blowing-up lemma, memoryless Gaussian source, AWGN Channel.

Submitted to IEEE Transactions on Information Theory, February 1998.

[†] This work is supported in part by Nippon Telegraph and Telephone Corporation and in part by Northrop-Grumman Corporation.

1 Introduction

It is common knowledge in the coding community that there is a duality between source and channel coding. Shannon [1, 2] showed that the minimization problem in the definition of the rate-distortion function is the dual of the maximization problem in the definition of the capacity of a power constraint channel. Cover and Thomas [3] used a geometrical argument to show that a good channel code can be converted to a good quantizer and vice versa. Eyuboğlu and Forney [4] showed that a lattice code used for encoding a Gaussian source at certain signal-to-distortion ratio (SDR) can be suitably modified into a channel transmission code for a Gaussian channel with signal-to-noise ratio (SNR) equal to $1/SDR$. This modification is such that the overload probability for the source quantizer is the same as the probability of error of the channel code. Further, Eyuboğlu and Forney showed the asymptotic optimality of lattice codes. Marcellin and Fischer [5] converted trellis-coded modulation schemes designed by Ungerboeck [6] into trellis-coded quantizers. Laroia et al [7, 8] converted scalar-vector quantizers into optimal shaping schemes for data transmission.

In this paper, a duality in joint source-channel coding is investigated. We consider a memoryless Gaussian source and an additive white Gaussian noise (AWGN) channel. For a given code defined by an encoder-decoder pair, its dual code is obtained by interchanging the roles of the encoder and decoder. Note that if a code operates at a rate ρ channel uses per source sample, then its dual code operates at rate $1/\rho$. We show that if a sequence of codes of rate ρ is asymptotically optimal (in the rate-distortion-capacity sense) and if it satisfies a uniform continuity condition, then its dual code sequence will be asymptotically optimal for rate $1/\rho$.

2 Preliminaries and Notation

Source: Consider a memoryless Gaussian source, $\{X_i\}_{i=1}^{\infty}$, with zero mean and variance σ^2. Thus, $X_i \sim \mathcal{N}(0, \sigma^2)$ and the sequence $\{X_i\}$ is independent and identically distributed (i.i.d.). We assume the source is obtained from uniform sampling of a continuous-time Gaussian process with bandwidth W_s (Hz). Furthermore, the sampling rate is assumed to be $2W_s$ samples per second.

Channel: The source is to be transmitted over an AWGN channel modeled by $Z_i = Y_i + V_i$, where Y_i, Z_i and V_i are the channel input, output and noise, respectively. We assume $E[Y_i^2] \leq P$ and $V_i \sim \mathcal{N}(0, N)$. The channel is derived from a continuous-time AWGN channel with bandwidth W_c (Hz). The discrete-time channel is used at a rate of $2W_c$ channel uses per second.

Coding: The block coding system is depicted in Figure 1. The source samples are grouped into blocks of size n: $X = (X_1, X_2, \ldots, X_n)$. The encoder is a mapping $\alpha : \mathbb{R}^n \mapsto \mathbb{R}^m$ which satisfies the power constraint $E[\|\alpha(X)\|^2] \leq mP$. Define $\rho = m/n = W_c/W_s$. The received signal is given by $Z = Y + V$, where $Y = \alpha(X)$, V is an m-dimensional zero-mean i.i.d. Gaussian random vector with variance N. The decoder is a mapping $\beta : \mathbb{R}^m \mapsto \mathbb{R}^n$. The average squared-error distortion of the coding system is denoted as

$$D_n = \frac{1}{n} E[\|X - \tilde{X}\|^2], \tag{1}$$

where $\tilde{X} = \beta(Z)$. We assume that $\sigma^2 > 0$, $P > 0$, $N > 0$ and $\rho > 0$ (ρ rational) are known.

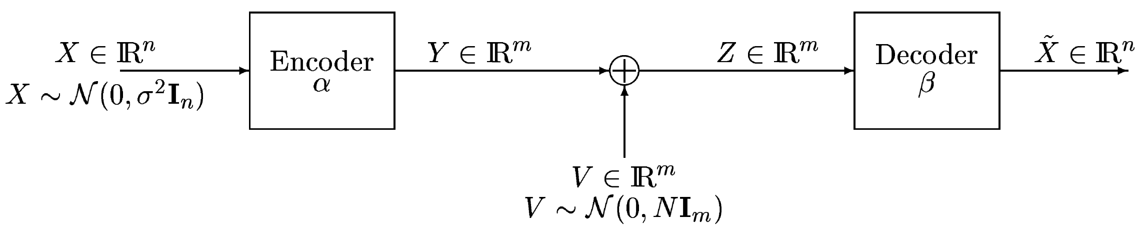

Figure 1: Block Diagram of Joint Source-Channel Coding System.

Theoretical Limit: The rate-distortion function for the memoryless Gaussian source and squared-error distortion measure is given by [2, 3]:

$$R(D) = W_s \log_2(\sigma^2/D) \quad \text{bits/sec.} \tag{2}$$

Similarly, the capacity of the AWGN channel is given by [1, 3]:

$$C = W_c \log_2(1 + P/N) \quad \text{bits/sec.} \tag{3}$$

If a code has average distortion D_n, then $R(D_n) \leq C$. This implies that

$$D_n \geq \frac{\sigma^2}{(1 + P/N)^\rho} \triangleq D_{opt}. \tag{4}$$

According to classical information theory, D_{opt} provides a lower bound on the average squared-error distortion of any coding system. Furthermore, given σ^2, P, N and ρ, there exist codes with average distortion D_n arbitrarily close to D_{opt}.

Dual System: From the block diagram in Figure 1, it can be seen that a joint source-channel coding system can be represented by a seven-tuple $\{\sigma^2,\ P,\ N,\ \alpha,\ \beta,\ \rho,\ D_{opt}\}$. We define the seven-tuple $\{\hat{\sigma}^2 = P + N,\ \hat{P} = \sigma^2 - D_{opt},\ \hat{N} = D_{opt},\ \hat{\alpha} = \beta,\ \hat{\beta} = \alpha,\ \hat{\rho} = 1/\rho,\ \hat{D}_{opt} = N\}$ to be the *dual system*. We use hats to denote the parameters of the dual system. Figure 2 shows the block diagram of the dual system. It can be seen from the block diagram that in the dual system, the source variance is $P + N$, the encoder power constraint is $\sigma^2 - D_{opt}$, the noise power is D_{opt} and the rate is $1/\rho$. Using (4), it can be easily verified that N is the optimum distortion of such system. We call the system in Figure 1 the forward system and the system in Figure 2 the dual system. The pair (β, α) is called the dual code for this dual system.

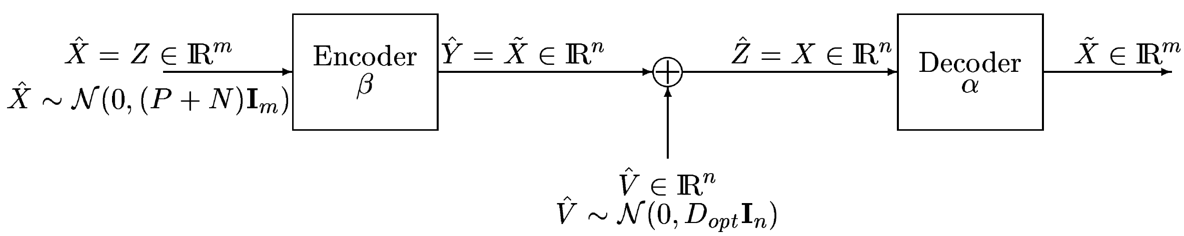

Figure 2: Block Diagram of Dual Joint Source-Channel Coding System.

$\bar{d}$- Distortion: Let $x,\ \tilde{x} \in \mathbb{R}^n$ and define $\bar{d} : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ as the per-letter squared-error distortion between x and $\tilde{x}$:

$$\bar{d}(x, \tilde{x}) = \frac{1}{n}\|x - \tilde{x}\|^2. \tag{5}$$

In subsequent sections we will be interested in a sequence of functions $\alpha_n : \mathbb{R}^n \mapsto \mathbb{R}^m$ for a fixed $\rho = m/n$ but increasing n and m. We need to define continuity of such sequences w.r.t. the above distortion.

Uniform Continuity: A sequence of functions $\{\alpha_n\}$ is uniformly continuous if given $\epsilon > 0$ $\exists$ $\delta > 0$ s.t. $\forall\ x \in \mathbb{R}^n$, $\forall\ n$

$$\bar{d}(x, \tilde{x}) < \delta \Longrightarrow \bar{d}(\alpha_n(x), \alpha_n(\tilde{x})) < \epsilon. \tag{6}$$

Here δ is independent of n and x.

The first theorem of this paper holds provided that the sequence of encoders and decoders are uniformly continuous. However, this assumption can be relaxed to the case where the code sequence is uniformly continuous on sets of probability one. Let

$$B^n_\delta(x) = \{\tilde{x} \in \mathbb{R}^n : \bar{d}(x, \tilde{x}) < \delta\}, \tag{7}$$

and

$$F_\epsilon^n(x) = \{\tilde{x} \in \mathbb{R}^n : \bar{d}(\alpha_n(x), \alpha_n(\tilde{x})) < \epsilon\}, \tag{8}$$

i.e., $B_\delta^n(x)$ is a neighborhood of x and $F_\epsilon^n(x)$ is the inverse image of a neighborhood of $\alpha(x)$. Define a set $C_{\delta,\epsilon}^n = \{x : F_\epsilon^n(x) \supset B_\delta^n(x)\}$.

Uniform Continuity on a Sequence of Sets: A sequence of functions $\{\alpha_n : \mathbb{R}^n \mapsto \mathbb{R}^m\}$, (fixed m/n and n increasing) is uniformly continuous on a sequence of sets $\{S_n\}$ if given $\epsilon > 0$, $\exists \delta > 0$ and n_o s.t. $\forall n > n_o$, $C_{\delta,\epsilon}^n \supset S_n$.

Uniform Continuity in Probability: A sequence of functions $\{\alpha_n : \mathbb{R}^n \mapsto \mathbb{R}^m\}$ is uniformly continuous in probability w.r.t. a sequence of distributions $\{f_n\}$, if there exists a sequence of sets $\{S_n\}$ s.t. $P_{f_n}(S_n) \to 1$ and α_n is uniformly continuous on $\{S_n\}$, where P_{f_n} is the probability under the distribution f_n.

Consider a sequence $\{(\alpha_n, \beta_n)\}$ of the joint source-channel codes for the forward system. Denote

$$\mathcal{N}_x^{(n)}(\mu, \sigma^2) \triangleq (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-(x-\mu)^T(x-\mu)}{2\sigma^2}\right). \tag{9}$$

That is, $\mathcal{N}_x^{(n)}(\mu, \sigma^2)$ is the joint probability distribution function (pdf) of an n-dimensional Gaussian vector with mean vector μ and covariance matrix $\sigma^2 \mathbf{I}_n$. Define

$$f(x, z) \triangleq \mathcal{N}_x^{(n)}(0, \sigma^2)\mathcal{N}_z^{(m)}(\alpha_n(x), N). \tag{10}$$

Note that $f(x, z)$ is the joint pdf of the random vector (X, Z) in Figure 1. Where there is no ambiguity, we will remove the subscript n in (α_n, β_n).

Assumptions

1. We assume that the code sequence $\{(\alpha_n, \beta_n)\}$ is uniformly continuous in probability with respect to $f(x, z)$, i.e., $\{\alpha_n\}$ is uniformly continuous w.r.t. $f(x) = \int f(x, z)dz$ and $\{\beta_n\}$ is uniformly continuous w.r.t. $f(z) = \int f(x, z)dx$.

2. We assume that the encoder α and the decoder β are bounded in the sense that $\forall\ x, z$ and n, $(\alpha^T\alpha)/n \le KP$, and $(\beta^T\beta)/n \le K(\sigma^2 - D_{opt})$, where $K = 2\max(\rho, 1/\rho)$.

Assume that the code sequence is optimal, i.e., D_n converges to D_{opt}. In the subsequent sections we prove that under the above assumptions the average distortion of the dual code $\{(\beta_n, \alpha_n)\}$ converges to N. Thus, if a sequence of code is asymptotically optimal in the forward system, then the dual code is asymptotically optimal for the dual system.

3 Main Results

Definition: A code sequence $\{(\alpha_n, \beta_n)\}$ is said to *satisfy the duality property* if (i) it is asymptotically optimal in the forward system ($D_n \to D_{opt}$) and (ii) its dual code sequence $\{(\beta_n, \alpha_n)\}$ is asymptotically optimal in the dual system ($\hat{D}_n \to N$).

Theorem 1: *If a code sequence $\{(\alpha_n, \beta_n)\}$ satisfies Assumptions 1 and 2 and if it is asymptotically optimal in the forward system, then it satisfies the duality property.*

In the above theorem, we showed that a code sequence, asymptotically optimal in the forward system, has a dual which is asymptotically optimal in the dual system, provided the code sequence satisfies Assumptions 1 and 2. Assumption 1 may be too restrictive in the sense that for $\rho \neq 1$, there may not exist an asymptotically optimal code sequence satisfying this assumption. Furthermore, it is not clear whether it is possible to prove the above theorem without such an assumption. In Theorem 2, we demonstrate that it is not possible to prove the theorem without some sort of assumption on continuity.

Theorem 2: *For $\rho = 1$, there exists a code sequence which is asymptotically optimal, but it does not satisfy the duality property.*

From the above theorems, it can not be inferred that there exists a code sequence which satisfies the duality property. This issue is resolved in Theorem 3.

Theorem 3: *For any values of ρ, σ^2, P and N, there exists a code sequence $\{(\alpha_n, \beta_n)\}$ which satisfies the duality property.*

4 Proof of Theorem 1

4.1 Convergence of Relative Entropy

Let $f(x, z)$ be defined as in (10). Define:

$$\phi(x, z) \triangleq \mathcal{N}_z^{(m)}(0, P + N)\mathcal{N}_x^{(n)}(\beta_n(z), D_{opt}). \tag{11}$$

Observe that in the forward system, $f(x, z)$ is the true joint pdf of (X, Z), whereas $\phi(x, z)$ is the *ideal* joint pdf of (X, Z). That is, under ideal situation, Z should be zero-mean Gaussian with variance $P + N$ and X should be conditionally Gaussian given $\beta_n(Z)$ with variance D_{opt} (the overall distortion should be D_{opt}). Intuitively, we expect that $f(x, z)$ and $\phi(x, z)$ should be close to each other when the code is optimal. Indeed this is true in the relative entropy (Kullback-Leibler distance) sense.

Lemma 1

$$\lim_{n\to\infty}\frac{1}{n}D(f(x,z)||\phi(x,z)) = 0 \Longrightarrow \lim_{m\to\infty}\frac{1}{m}E_f[\alpha(X)^T\alpha(X)] = P. \quad (12)$$

The above lemma implies that the power constraint is asymptotically satisfied if $(1/n)D(f||\phi)$ converges to 0. The proof of this lemma is provided in the appendix.

Lemma 2 $D_n \to D_{opt}$ *if and only if*

$$\lim_{n\to\infty}\frac{1}{n}D(f(x,z)||\phi(x,z)) = 0. \quad (13)$$

Proof: We first show that convergence of D_n to D_{opt} implies convergence of relative entropy to zero. Straightforward evaluation of the relative entropy in (13) results in

$$\begin{aligned}\frac{1}{n}D(f(x,z)||\phi(x,z)) &= \frac{1}{n}E_f[-\frac{X^TX}{2\sigma^2}] + \frac{1}{n}E_f[-\frac{(Z-\alpha(X))^T(Z-\alpha(X))}{2N}] + \\ &\quad \frac{1}{n}E_f[\frac{(X-\beta(Z))^T(X-\beta(Z))}{2D_{opt}}] + \frac{1}{n}E_f[\frac{Z^TZ}{2(P+N)}] \\ &= -\frac{1}{2} - \frac{\rho}{2} + \frac{D_n}{2D_{opt}} + \frac{1}{n}E_f[\frac{(Z^TZ)}{2(P+N)}]. \quad (14)\end{aligned}$$

The channel noise $Z-\alpha(X)$ is zero mean and is independent of the encoder output $\alpha(X)$, hence $E_f[Z^TZ] = E_f[\alpha(X)^T\alpha(X)] + E_f[(Z-\alpha(X))^T(Z-\alpha(X))]$. Thus by the power constraint on the encoder output we have $E_f[Z^TZ] \leq m(P+N)$ and since $D_n \to D_{opt}$, we have

$$\lim_{n\to\infty}\frac{1}{n}D(f(x,z)||\phi(x,z)) \leq 0. \quad (15)$$

The result follows since relative entropy is non-negative. To prove the converse, assume (13) holds. From Lemma 1, we get $\lim_{n\to\infty}(1/n)E_f[\|\alpha(X)\|^2] = \rho P$. Combining this with (13) and (14), we get $\lim_{n\to\infty} D_n = D_{opt}$. □

Observe that in the dual system of Figure 2, $\phi(x,z)$ is the *true* pdf of $(\hat{Z},\hat{X})$ and $f(x,z)$ is the *ideal* pdf. Hence, in the dual system, the roles of $f(x,z)$ and $\phi(x,z)$ are interchanged. We have already shown that the forward code sequence is asymptotically optimal if and only if $(1/n)D(f||\phi) \to 0$. To prove that the dual code sequence is optimal, it suffices to show that $(1/n)D(\phi||f) \to 0$.

Lemma 3

$$\lim_{n\to\infty} \frac{1}{n} D(f||\phi) = 0 \Longrightarrow \lim_{n\to\infty} E_f \left| \frac{1}{n} \log\left(\frac{f}{\phi}\right) \right| = 0. \tag{16}$$

Proof: By Equation 2.4.10 of [13]

$$E_f \left(\left| \log \frac{f}{\phi} \right| \right) \leq D_f(f||\phi) + \frac{2}{e}. \tag{17}$$

Dividing both sides by n we get

$$E_f \left(\frac{1}{n} \left| \log \frac{f}{\phi} \right| \right) \leq \frac{1}{n} D_f(f||\phi) + \frac{2}{ne}. \tag{18}$$

Taking the limit as $n \to \infty$ we get the required result. $\square$

4.1.1 Typical Set

A set A_ϵ^n is a typical set if it satisfies the following properties:

1. $(x,z) \in A_\epsilon^n$, if $\left| \frac{1}{n} \log(\frac{f(x,z)}{\phi(x,z)}) \right| \leq \epsilon$.

2. $(x,z) \in A_\epsilon^n$, if $\frac{z^T z}{n} \leq K(P+N) + \epsilon$, $\frac{x^T x}{n} \leq K\sigma^2 + \epsilon$.

i.e., $A_\epsilon^n = \{(x,z) : \left| \frac{1}{n} \log(\frac{f(x,z)}{\phi(x,z)}) \right| \leq \epsilon, \frac{z^T z}{n} \leq K(P+N) + \epsilon, \frac{x^T x}{n} \leq K\sigma^2 + \epsilon\}$, where K is defined in Assumption 2.

Lemma 4 *$\forall\ \epsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} D(f||\phi) = 0 \Longrightarrow \lim_{n\to\infty} P_f(A_\epsilon^n) = 1. \tag{19}$$

Proof: By Lemma 3, $|\frac{1}{n} \log \frac{f}{\phi}|$ converges in the mean to 0, hence it converges in probability to 0, i.e.,

$$\lim_{n\to\infty} P_f \left(\left| \frac{1}{n} \log \frac{f}{\phi} \right| > \epsilon \right) = 0.$$

As X is iid Gaussian under f, hence by the weak law of large numbers (WLLN),

$$\lim_{n\to\infty} P_f \left(\frac{X^T X}{n} \leq K\sigma^2 + \epsilon \right) = 1, \quad K > 1.$$

Moreover,

$$\frac{Z^T Z}{n} = \frac{(Z - \alpha(X))^T (Z - \alpha(X))}{n} + \frac{\alpha(X)^T \alpha(X)}{n} + 2\frac{\alpha(X)^T (Z - \alpha(X))}{n}.$$

As $Z - \alpha(X)$ is iid Gaussian under f, hence by the WLLN,

$$\lim_{n\to\infty} P_f \left(\frac{(Z - \alpha(X))^T (Z - \alpha(X))}{n} \leq KN + \epsilon \right) = 1, \tag{20}$$

and by the bounded assumption of the encoder $\alpha(X)$, we have

$$\lim_{n\to\infty} P_f \left(\frac{\alpha(X)^T \alpha(X)}{n} \leq KP + \epsilon \right) = 1. \tag{21}$$

As $Z - \alpha(X)$ is iid Gaussian with mean zero and $\alpha(X)$ is independent of $Z - \alpha(X)$ in the distribution $f(x, z)$, hence by WLLN,

$$\lim_{n\to\infty} P_f \left(\frac{2\alpha(X)^T (Z - \alpha(X))}{n} \leq \epsilon \right) = 1. \tag{22}$$

Thus

$$\lim_{n\to\infty} P_f \left(\frac{Z^T Z}{n} \leq K(P + N) + \epsilon \right) = 1. \tag{23}$$

The set $A_\epsilon^n = \{|\frac{1}{n}\log(\frac{f}{\phi})| \leq \epsilon\} \cap \{\frac{x^T x}{n} \leq K\sigma^2 + \epsilon\} \cap \{\frac{z^T z}{n} \leq K(P+N) + \epsilon\}$ is an intersection of three sets of probability approaching one. Hence,

$$\lim_{n\to\infty} P_f(A_\epsilon) = 1. \tag{24}$$

This can be shown by taking the complement of each set and applying the union bound. □

By the continuity assumption, there exists a sequence of sets $\{S_n^x\}$ and $\{S_n^z\}$ s.t. $P_f(S_n^x \times S_n^z) \to 1$ and the encoder and the decoder sequences are uniformly continuous on S_n^x and S_n^z, respectively. Define $S_n = S_n^x \times S_n^z$. Thus, given $\epsilon > 0$, $\exists\ n_o$ s.t. $\forall\ n > n_o$

$$P_f(A_\epsilon^n \cap S_n) > 1 - \epsilon. \tag{25}$$

Define

$$B_\epsilon^n = A_\epsilon^n \cap S_n. \tag{26}$$

This set has probability approaching one under f. In the following, we show that under ϕ its probability does not approach zero too fast. We then blow up B_ϵ^n and show that the blown up set has probability approaching one under ϕ.

Lemma 5 *Given* $\delta > 0$, $\epsilon > 0$, $\exists \delta_1 > 0$ *and* n_o *s.t.* $\forall\ n > n_o$

$$\frac{1}{n} D(f||\phi) \leq \delta_1 \Longrightarrow P_\phi(B_\epsilon^n) > e^{-n\delta}. \tag{27}$$

Proof: We will prove this Lemma by contradiction. Consider a transformation T on (x, z) such that $T(x, z) = 1$, if $(x, z) \in B_\epsilon^n$ and zero otherwise. By the chain rule on relative entropy [3], we have

$$D(f||\phi) \geq D(T_f||T_\phi), \tag{28}$$

where T_f (T_ϕ) is the distribution on $T(X, Z)$ when (X, Z) is distributed as f (ϕ).
Since

$$\frac{1}{n} D(T_f||T_\phi) = \frac{1}{n} P_f(B_\epsilon^n) \log \frac{P_f(B_\epsilon^n)}{P_\phi(B_\epsilon^n)} + \frac{1}{n} P_f((B_\epsilon^n)^c) \log \frac{P_f((B_\epsilon^n)^c)}{P_\phi((B_\epsilon^n)^c)}, \tag{29}$$

from (28) and (29)

$$\frac{1}{n} D(f||\phi) \geq -\frac{1}{n} H(P_f(B_\epsilon^n)) - \frac{1}{n} P_f(B_\epsilon^n) \log(P_\phi(B_\epsilon^n)) - \frac{1}{n} P_f((B_\epsilon^n)^c) \log(P_\phi((B_\epsilon^n)^c)), \tag{30}$$

where $H(p)$ is the binary entropy function. Since $-\frac{1}{n} P_f((B_\epsilon^n)^c) \log(P_\phi((B_\epsilon^n)^c)) \geq 0$ and $\frac{1}{n} H(P_f(B_\epsilon^n)) \leq \frac{1}{n}$, hence

$$\frac{1}{n} D(f||\phi) \geq -\frac{1}{n} - \frac{1}{n} P_f(B_\epsilon^n) \log(P_\phi(B_\epsilon^n)). \tag{31}$$

Now, if $P_\phi(B_\epsilon^n) \leq e^{-n\delta}$ then

$$\frac{1}{n} D(f||\phi) \geq -\frac{1}{n} + P_f(B_\epsilon^n)\delta. \tag{32}$$

By Lemma 4, we can choose a large enough n such that $P_f(B_\epsilon^n) > \frac{1}{2}$ and $\frac{1}{n} < \frac{\delta}{4}$. Thus there exists n_o s.t. $\forall\ n > n_o$

$$\frac{1}{n} D(f||\phi) > \frac{\delta}{4}, \tag{33}$$

which is a contradiction. □

4.2 Blowing Up Property

Consider a set $A \subset \mathbb{R}^n$, and let $b_\epsilon(A)$ denote the ϵ-neighborhood of A:

$$b_\epsilon(A) = \{y \in \mathbb{R}^n : \bar{d}(x,y) < \epsilon \ for \ some \ x \in A\}. \tag{34}$$

In subsequent sections, the set A will be called the original set and the set $b_\epsilon(A)$ will be called the blown up set.

Definition: A process $\{X_i\}$, or a sequence of distribution P_n, has a blowing up property [9, 10, 11] if for any $\epsilon > 0$, $\exists \delta > 0$ and n_o s.t. $\forall n > n_o$ and $\forall$ measurable set $A \subset \mathbb{R}^n$

$$P_n(A) \geq e^{-n\delta} \Longrightarrow P_n(b_\epsilon(A)) \geq 1 - \epsilon. \tag{35}$$

Consider a transformation $T_1 : \mathbb{R}^{m+n} \mapsto \mathbb{R}^{m+n}$ on (x,z) s.t. $(u,z) = T_1(x,z) = (\sqrt{(P+N)/D_{opt}}$ $(x - \beta(z)), z)$. It can be easily seen that the above transformation is invertible if $\beta(.)$ is known. Let $\psi(u,z)$ be the distribution of the transformed random vectors induced by $\phi(x,z)$.

Lemma 6

$$\psi(u,z) = \mathcal{N}_u^{(n)}(0, P+N)\mathcal{N}_z^{(m)}(0, P+N). \tag{36}$$

Lemma 7 *$\psi(u,z)$ has a blowing up property.*

Proof: From Lemma 6, $\psi(u,z)$ is an i.i.d. Gaussian process. From [9] and [10], an iid Gaussian process has the blowing up property, hence $\psi(u,z)$ has the blowing up property. □

4.3 Blowing up a Typical Set

In Lemma 5, we have shown that for large n, $P_\phi(B_\delta^n) > e^{-n\delta}$ for any given $\delta > 0$. Since the transformation T_1 defined in the previous section is invertible, the typical set or any set in (x,z), can be transformed into a set in (u,z) and vice-versa. The transformed set is the image of a given set under the transformation T_1. Let B be a set in (x,z) and denote $T_1(B)$ as the image of that set under T_1. Clearly, for any set B, $P_\psi(T_1(B)) = P_\phi(B)$. In this section, we will use the blowing up property of $\psi(u,z)$, to show that the typical set has probability close to one under ϕ.

Lemma 8 *Given $\epsilon_1 > 0$, $\epsilon_2 > \epsilon_1$, $\exists \delta > 0$ and n_o s.t. $\forall\ n > n_o$*

$$T_1(A^n_{\epsilon_2}) \supset b_\delta(T_1(B^n_{\epsilon_1})), \tag{37}$$

where $A^n_{\epsilon_2}$ is defined in Section 4.1.1, $B^n_{\epsilon_1}$ is defined in (26) and the blowing up is done in (u, z).

The proof of this Lemma is provided in the appendix.

Lemma 9 *Assume $\lim_{n\to\infty} \frac{1}{n} D(f||\phi) = 0$, given $\epsilon > 0$, $\exists n_o$ s.t. $\forall\ n > n_o$, $P_\phi(A^n_\epsilon) > 1 - \epsilon$.*

Proof: Let $\epsilon_1 = \epsilon/2$. By Lemma 8, given $\epsilon > \epsilon/2$, $\exists \delta$ s.t. for large n, $T_1(A^n_\epsilon) \supset b_\delta(T_1(B^n_{\epsilon/2}))$. Since $\psi(u, z)$ has a blowing up property, given $\epsilon, \delta > 0$ $\exists$ $\delta_1 > 0$ and n_o s.t. $\forall n > n_o$, $P_\psi(T_1(B^n_{\epsilon/2})) > e^{-n\delta_1} \Longrightarrow P_\psi(b_\delta(T_1(B^n_{\epsilon/2}))) > 1 - \epsilon$. By Lemma 5, given $\epsilon, \delta_1 > 0$ $\exists$ n_o s.t. $\forall$ $n > n_o$, $P_\psi(T_1(B^n_{\epsilon/2})) = P_\phi(B^n_{\epsilon_1}) > e^{-n\delta_1}$.

Thus, there exists n_o s.t. $\forall n > n_o$, $P_\psi(b_\delta(T_1(B^n_{\epsilon/2}))) > 1-\epsilon$ and moreover since $P_\psi(T_1(A^n_\epsilon)) \geq P_\psi(b_\delta(T_1(B^n_{\epsilon/2})))$ and $P_\psi(T_1(A^n_\epsilon)) = P_\phi(A^n_\epsilon)$, the above result follows. $\square$

From the above Lemma it follows that $\left|\frac{1}{n}\log(\frac{f(X,Z)}{\phi(X,Z)})\right|$ converges in probability to 0 w.r.t. distribution ϕ.

4.4 Uniform Integrability

Lemma 10 *If α and β satisfy Assumption 2, then $\left|\frac{1}{n}\log(\frac{f(X,Z)}{\phi(X,Z)})\right|$ is uniformly integrable w.r.t. $\phi(x, z)$.*

Proof: Let $h(X, Z) = \frac{1}{n}\log(\frac{f}{\phi})$. Expanding $h(X, Z)$ we get

$$h(X,Z) = -\frac{X^TX}{2n\sigma^2} - \frac{(Z-\alpha(X))^T(Z-\alpha(X))}{2nN} + \frac{Z^TZ}{2n(P+N)} + \frac{(X-\beta(Z))^T(X-\beta(Z))}{2nD_{opt}}. \tag{38}$$

Z and $X - \beta(Z)$ are independent Gaussian random vectors under ϕ, hence $Z^TZ/n(P+N)$ and $(X - \beta(Z))^T(X - \beta(Z))/2nD_{opt}$ are uniformly integrable.
By Assumption 2, $\beta(Z)^T\beta(Z)/n < K(\sigma^2 - D)$, hence it is uniformly integrable. By Cauchy-Schwartz inequality, we have

$$\left|\frac{(X-\beta(Z))^T\beta(Z)}{n}\right| \leq \sqrt{\frac{(X-\beta(Z))^T(X-\beta(Z))}{n}}\sqrt{\frac{\beta(Z)^T\beta(Z)}{n}}. \tag{39}$$

Uniform integrability of $(X-\beta(Z))^T\beta(Z)/n$ now follows from Assumption 2, uniform integrability of $(X-\beta(z))^T(X-\beta(Z))/n$ and the fact that L^2 norm is greater than L^1 norm.
Note that

$$\frac{X^TX}{2n\sigma^2} = \frac{(X-\beta(Z)+\beta(Z))^T(X-\beta(Z)+\beta(Z))}{2n\sigma^2}$$
$$= \frac{(X-\beta(Z))^T(X-\beta(Z))}{2n\sigma^2} + 2\frac{(X-\beta(Z))^T\beta(Z)}{2n\sigma^2} + \frac{\beta(Z)^T\beta(Z)}{2n\sigma^2}, \quad (40)$$

is the sum of three uniformly integrable random variables, hence it is uniformly integrable. Similarly, we can prove uniform integrability of $(Z-\alpha(X))^T(Z-\alpha(X))/m$. Thus, all the summands of (38) are uniformly integrable and hence $h(X,Z)$ is uniformly integrable. □

Since, $\left|\frac{1}{n}\log(\frac{f(X,Z)}{\phi(X,Z)})\right|$ converges in probability to 0 wrt ϕ and is uniformly integrable [12], hence it converges in the mean to 0 wrt ϕ [12].

4.5 Wrap up of Proof

In the previous subsections we have shown that if a joint source-channel code sequence $\{(\alpha_n,\beta_n)\}$ satisfies the continuity and boundedness assumptions and if this code sequence is asymptotically optimal for the forward system, then the random variables $\left|\frac{1}{n}\log(\frac{f(X,Z)}{\phi(X,Z)})\right|$ converges in the mean to 0 wrt both distributions f and ϕ. Thus $(1/n)D(\phi||f)$ also converges to 0.

Lemma 11 *Assume Assumptions 1 and 2 hold, then*

$$\lim_{n\to\infty}\frac{1}{n}D(f(x,z)||\phi(x,z)) = 0 \Longrightarrow \lim_{n\to\infty}\frac{1}{n}D(\phi(x,z)||f(x,z)) = 0. \quad (41)$$

Note that in the dual system, ϕ is the true pdf while f is the ideal pdf. Lemma 1 implies that in the dual system, the encoder sequence β_n asymptotically satisfies the power constraint. Thus, by the converse in Lemma 2, the dual code sequence $\{(\beta_n,\alpha_n)\}$ is asymptotically optimal for the dual system.

5 Proof of Theorem 2

Let $\delta_k\downarrow 0$ be a sequence of positive real numbers. For each δ_k choose n s.t. $P[\frac{1}{n}\sum_{i=1}^n Z_i^2 \le P+N-\frac{\delta_k}{2}] > 1-\delta_k$, when Z_i's are i.i.d. $\mathcal{N}(0,P+N-\delta_k)$, and $P[\frac{1}{n}\sum_{i=1}^n Z_i^2 \le P+N-\frac{\delta_k}{2}] < \delta_k$, when Z_i's are i.i.d. $\mathcal{N}(0, P+N)$. Note that given $\delta_k > 0$, there exists n which satisfies the above two conditions. Let $\alpha_n : \mathbb{R}^n \mapsto \mathbb{R}^n$ be $\alpha_n(X) = \sqrt{(P-\delta_k)/\sigma^2}X$, $X \in \mathbb{R}^n$, and $\beta_n(Z) : \mathbb{R}^n \mapsto \mathbb{R}^n$ be

$$\beta_n(Z) = \begin{cases} \frac{\sqrt{\sigma^2(P-\delta_k)}}{P+N} Z & \text{if } \frac{1}{n}\sum_{i=1}^n Z_i^2 \le P+N-\frac{\delta_k}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

Note that this sequence of decoders does not satisfy Assumption 1 of Section 2. We will show that for the code sequence $\{(\alpha_n, \beta_n)\}$, the average distortion D_n converges to $\sigma^2 N/(P+N)$ in the forward system, and it is greater than N for the dual system. Note that the dimension n depends on δ_k. In the forward system, the source X is iid $\mathcal{N}(0, \sigma^2)$ and the noise V is iid $\mathcal{N}(0, N)$, hence the channel output Z is iid $\mathcal{N}(0, (P+N-\delta_k))$. Let $\mathcal{P}_n$ be the event that $\frac{1}{n}\sum_{i=1}^n Z_i^2 \le P+N-\frac{\delta_k}{2}$. $D_n \to D_{opt}$ is equivalent to saying that for a given $\epsilon > 0$, there exists n_o s.t. $\forall\ n \ge n_o$, $D_n \le D_{opt} + \epsilon$. Choose γ s.t. $E(X_i^2 \mathbf{1}_{(X_i^2 > \gamma)}) < \epsilon/3$, where $\mathbf{1}_{\mathcal{E}}$ is the indicator function for event $\mathcal{E}$. Since $\delta_k \downarrow 0$, there exists r_1 s.t. $\forall\ k \ge r_1$, $\delta_k \gamma < \epsilon/3$. Now

$$\begin{aligned} D_n &= \frac{1}{n}E\left[\|X-\beta(Z)\|^2 \mathbf{1}_{\mathcal{P}_n}\right] + \frac{1}{n}E\left[\|X-\beta(Z)\|^2 \mathbf{1}_{\mathcal{P}_n^c}\right] \\ &\le \frac{1}{n}E\left[\|X - \frac{\sqrt{\sigma^2(P-\delta_k)}}{P+N}Z\|^2\right] + \frac{1}{n}E\left[\|X-\beta(Z)\|^2 \mathbf{1}_{\mathcal{P}_n^c}\right] \\ &= \frac{N\sigma^2}{P+N} + \frac{\delta_k \sigma^2 N}{(P+N)^2} + \frac{\delta_k^2 \sigma^2}{(P+N)^2} + \frac{1}{n}E\left[\|X-\beta(Z)\|^2 \mathbf{1}_{\mathcal{P}_n^c}\right]. \end{aligned} \tag{43}$$

Since $\delta_k \downarrow 0$, there exists r_2 s.t. $\forall\ k \ge r_2$, $\frac{\delta_k \sigma^2 N}{(P+N)^2} + \frac{\delta_k^2 \sigma^2}{(P+N)^2} \le \epsilon/3$. Fix $k = \max(r_1, r_2)$ and choose n as above. Now consider the fourth term in (43). Conditioned on $\mathcal{P}_n^c$, $\beta(Z) = 0$. Thus

$$\begin{aligned} \frac{1}{n}E\left[\|X-\beta(Z)\|^2 \mathbf{1}_{\mathcal{P}_n^c}\right] &= \frac{1}{n}E\left[\sum_{i=1}^n X_i^2 \mathbf{1}_{\mathcal{P}_n^c}\right] \\ &= \frac{1}{n}\sum_{i=1}^n E(X_i^2 \mathbf{1}_{\mathcal{P}_n^c}) \\ &= \frac{1}{n}\sum_{i=1}^n \left[E(X_i^2 \mathbf{1}_{(X_i^2 \le \gamma)} \mathbf{1}_{\mathcal{P}_n^c}) + E(X_i^2 \mathbf{1}_{(X_i^2 > \gamma)} \mathbf{1}_{\mathcal{P}_n^c})\right] \\ &\le \gamma P(\mathcal{P}_n^c) + \frac{\epsilon}{3} \end{aligned}$$

$$\leq \frac{\epsilon}{3} + \frac{\epsilon}{3}. \tag{44}$$

The last inequality follows from $P(\mathcal{P}_n^c) < \delta_k$ (by choice of n). Thus $D_n \leq N\sigma^2/(P+N) + \epsilon$. Hence, the average mean squared distortion for the code sequence $\{(\alpha_n, \beta_n)\}$ converges to D_{opt}.

In the dual system the source Z is iid $\mathcal{N}(0, P+N)$ and the noise $\hat{V}$ is iid $\mathcal{N}(0, \sigma^2 N/(P+N))$. Let $\tilde{Z}$ be the decoded value. The distortion $\hat{D}_n$ is given by

$$\begin{aligned}
\hat{D}_n &= \frac{1}{n}E\left[\|Z-\tilde{Z}\|^2 \mathbf{1}_{\mathcal{P}_n}\right] + \frac{1}{n}E\left[\|Z-\tilde{Z}\|^2 \mathbf{1}_{\mathcal{P}_n^c}\right] \\
&\geq \frac{1}{n}E\left[\|Z-\tilde{Z}\|^2 \mathbf{1}_{\mathcal{P}_n^c}\right] \\
&= \frac{1}{n}E\left[\|Z-\tilde{Z}\|^2 \middle| \mathcal{P}_n^c\right] P(\mathcal{P}_n^c).
\end{aligned} \tag{45}$$

Note that conditioned on $\mathcal{P}_n^c$, $\tilde{Z} = \alpha(\hat{V})$ is independent of Z. Hence, for large n,

$$\hat{D}_n \geq \left((P+N-\frac{\delta_k}{2}) + \frac{(P-\delta_k)N}{(P+N)} \right)(1-\delta_k). \tag{46}$$

Taking the limit on k, we get

$$\limsup \hat{D}_n \geq \liminf \hat{D}_n \geq P+N+\frac{PN}{P+N} > N. \tag{47}$$

Thus $\hat{D}_n$ does not converge to the optimal distortion of the dual system.□

6 Proof of Theorem 3

We use the random coding argument to show the existence of a joint source-channel code sequence satisfying the duality property. Let $x^1, \tilde{x}^1 \in \mathbb{R}$ (the superscript referring to the fact that these are one-dimensional variables) and define

$$f_1(x^1, \tilde{x}^1) \triangleq \mathcal{N}_{\tilde{x}^1}^{(1)}(0, \sigma^2 - D_{opt} - D_{opt}\delta_2)\mathcal{N}_{x^1}^{(1)}(\tilde{x}^1, D_{opt} + D_{opt}\delta_2) \tag{48}$$

$$f_2(x^1, \tilde{x}^1) \triangleq \mathcal{N}_{\tilde{x}^1}^{(1)}(0, P - N\delta_1)\mathcal{N}_{x^1}^{(1)}(\tilde{x}^1, N) \tag{49}$$

$$f_3(x^1, \tilde{x}^1) \triangleq \mathcal{N}_{\tilde{x}^1}^{(1)}(0, P - N\delta_1)\mathcal{N}_{x^1}^{(1)}(\tilde{x}^1, N + N\delta_1) \tag{50}$$

$$f_4(x^1, \tilde{x}^1) \triangleq \mathcal{N}_{\tilde{x}^1}^{(1)}(0, \sigma^2 - D_{opt} - D_{opt}\delta_2)\mathcal{N}_{x^1}^{(1)}(\tilde{x}^1, D_{opt}) \tag{51}$$

We define four distortion typical sets $A_{1,\epsilon}^{(n)}$, $A_{2,\epsilon}^{(m)}$, $A_{3,\epsilon}^{(m)}$ and $A_{4,\epsilon}^{(n)}$ as follows. The set $A_{2,\epsilon}^{(m)}$ is a collection of all m-tuple pairs $(x, \tilde{x}) = ((x_1^1, x_2^1, \ldots, x_m^1), (\tilde{x}_1^1, \tilde{x}_2^1, \ldots, \tilde{x}_m^1))$ which satisfy

1. $|-\frac{1}{m}\log f_2(x) - \frac{1}{2}\log(2\pi e(P + N - N\delta_1))| < \epsilon$.

2. $|-\frac{1}{m}\log f_2(\tilde{x}) - \frac{1}{2}\log(2\pi e(P - N\delta_1))| < \epsilon$.

3. $|-\frac{1}{m}\log f_2(x, \tilde{x}) - \frac{1}{2}\log((2\pi e)^2 N(P - N\delta_1))| < \epsilon$.

4. $|\bar{d}(\tilde{x}, x) - N| < \epsilon$.

The distributions in 1-3 are *product* distributions, e.g., $f_2(x, \tilde{x}) = \prod_{i=1}^m f_2(x_i^1, \tilde{x}_i^1)$. $A_{2,\epsilon}^{(m)}$ is called the *distortion typical set* w.r.t. f_2 [3] (with distortion N). Similarly $A_{1,\epsilon}^{(n)}$, $A_{3,\epsilon}^{(m)}$ and $A_{4,\epsilon}^{(n)}$ are the distortion typical sets w.r.t. f_1, f_3 and f_4, respectively (with distortions $D_{opt} + D_{opt}\delta_2$, $N + N\delta_1$, and D_{opt}, respectively). Choose $\delta > 0$. We need to show that there exists a code such that the distortion for the forward system is less than $D_{opt} + \delta$ and its dual code has distortion less than $N + \delta$ in the dual system. Fix $\delta_1 = \delta_2/\rho < \min(\delta/N, \delta/(D_{opt}\rho))$.

Generation of Codebook. Generate a codebook $\mathcal{C}$ having two constituent codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$. Codebook $\mathcal{C}_1$ comprises of 2^{nR}, n-dimensional i.i.d. Gaussian vectors of zero mean and variance $\sigma^2 - D_{opt} - D_{opt}\delta_2$. Codebook $\mathcal{C}_2$ (generated independently of $\mathcal{C}_1$) consists of $2^{nR} = 2^{mR'}$, m-dimensional i.i.d. Gaussian vectors of zero mean and variance $P - N\delta_1$. The rate R will be decided later. Enumerate both codebooks from 1 to 2^{nR}. A special codeword consisting of all-zero elements is added to both codebooks and is referred to as the 0^{th} codeword. The codebook $\mathcal{C}$ is revealed to the encoder and decoder of the forward and dual systems.

Encoding. A two-stage encoding is performed in the forward (dual) system. For an incoming source vector X, choose a codeword $\tilde{X} \in \mathcal{C}_1$ ($\mathcal{C}_2$) s.t. $(X, \tilde{X}) \in A_{4,\epsilon}^{(n)}$ ($A_{2,\epsilon}^{(m)}$). If more than one such codeword exist, choose one of them randomly. If no such $\tilde{X}$ exists then choose a codeword (randomly if more than one exist) $\tilde{X} \in \mathcal{C}_1$ ($\mathcal{C}_2$) s.t. $(X, \tilde{X}) \in A_{1,\epsilon}^{(n)}$ ($A_{3,\epsilon}^{(m)}$). If no such codeword exists then select and send the 0^{th} codeword in $\mathcal{C}_2$ ($\mathcal{C}_1$). If the i^{th} codeword in codebook $\mathcal{C}_1$ ($\mathcal{C}_2$) is chosen, then select the corresponding i^{th} codeword in the m- (n-)dimensional codebook $\mathcal{C}_2$ ($\mathcal{C}_1$). This i^{th} codeword is sent over the channel provided it satisfies the power constraint of the forward (dual) system. Otherwise send an all-zero codeword over the channel. However, we still say in this case that the i^{th} codeword has been *selected*.

Decoding. A two-stage decoding is performed in the forward (dual) system. For the channel output Z, choose a codeword $\tilde{Y} \in \mathcal{C}_2$ ($\mathcal{C}_1$) s.t. $(Z, \tilde{Y}) \in A_{2,\epsilon}^{(m)}$ ($A_{4,\epsilon}^{(n)}$). If more than one such codeword exist, choose one of them randomly. If no such $\tilde{Y}$ exists, then choose a codeword (randomly if more than one exist) $\tilde{Y} \in \mathcal{C}_2$ ($\mathcal{C}_1$) s.t. $(Z, \tilde{Y}) \in A_{3,\epsilon}^{(m)}$ ($A_{1,\epsilon}^{(n)}$). If no such codeword exists then select and decode the 0^{th} codeword in $\mathcal{C}_1$ ($\mathcal{C}_2$). If the i^{th} codeword in codebook $\mathcal{C}_2$ ($\mathcal{C}_1$) is chosen, select the corresponding i^{th} codeword in the n- (m-)dimensional codebook $\mathcal{C}_1$ ($\mathcal{C}_2$). Decode the i^{th} codeword provided it satisfies the power constraint of the dual (forward) system, i.e., if $\|\tilde{X}\|^2 \leq n(\sigma^2 - D_{opt})$ ($\|\tilde{X}\|^2 \leq mP$). Otherwise decode an all-zero codeword but say that i^{th} codeword has been *selected*.

Note that the encoder in the forward system is same as the decoder in the dual system and vice versa. A two-stage encoding and decoding is done since in the forward (dual) system $(X, \tilde{X}) \in A_{1,\epsilon}^{(n)}$ ($A_{3,\epsilon}^{(m)}$) and $(Z, \tilde{Y}) \in A_{2,\epsilon}^{(m)}$ ($A_{4,\epsilon}^{(n)}$) with probability one.

Calculation of Distortion. We calculate the average distortion as the sum of average distortions of the forward system and the dual system. The expected distortion is calculated over the random codebook $\mathcal{C}$. Let X (Z) denote a source vector in the forward (dual) system and $\tilde{X}$ ($\tilde{Y}$) be the decoded value. Define

$$\bar{D} = E_{\mathcal{C}}(E(d(X, \tilde{X})|\mathcal{C}) + E_{\mathcal{C}}(E(d(Z, \tilde{Y})|\mathcal{C}), \tag{52}$$

and

$$\bar{D}_1 = E_{\mathcal{C}}(E(d(X, \tilde{X})|\mathcal{C})). \tag{53}$$

Denote $d(X, \tilde{X})$ as d and consider the forward system. Let $\mathcal{E}_{enc}$ be the event that the 0^{th} codeword is selected by the encoder (encoding error event). Let $\mathcal{E}_{dec}$ be the event that the selected codeword of the decoder is *not* the same as the selected codeword of the encoder (decoding error event).

$$\begin{aligned} \bar{D}_1 = E_{\mathcal{C}}[E(d|\mathcal{C})] &= E_{\mathcal{C}}\left[E\left(d|\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}^c \cap \mathcal{C}\right) P\left(\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}^c|\mathcal{C}\right)\right] \\ &+ E_{\mathcal{C}}\left[E\left(d|\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec} \cap \mathcal{C}\right) P\left(\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}|\mathcal{C}\right)\right] \\ &+ E_{\mathcal{C}}\left[E\left(d|\mathcal{E}_{enc} \cap \mathcal{C}\right) P\left(\mathcal{E}_{enc}|\mathcal{C}\right)\right] \end{aligned} \tag{54}$$

Note that in the event $\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}^c$, the selected codeword in $\mathcal{C}_1$ must satisfy the dual system power constraint. Hence, the first term in the above summation is less than $D + \delta + \epsilon$. Now consider the second term in the above summation. Since the source vector belongs to $A_{4,\epsilon}^{(n)} \cup A_{1,\epsilon}^{(n)}$ and since the mean squared value of the decoder output is always bounded by $\sigma^2 - D_{opt} - D_{opt}\delta_2 + \epsilon$,

$$E(d|\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec} \cap \mathcal{C})) \leq 4(\sigma^2 + \epsilon). \tag{55}$$

Applying Bayes rule, we get

$$P\left(\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}|\mathcal{C}\right) \leq P\left(\mathcal{E}_{dec}|\mathcal{E}_{enc}^c \cap \mathcal{C}\right). \tag{56}$$

Let I denotes the codeword *selected* by the encoder:

$$P(\mathcal{E}_{dec}|\mathcal{E}_{enc}^c \cap \mathcal{C}) = \sum_{i=1}^{2^{nR}} P(I = i|\mathcal{E}_{enc}^c \cap \mathcal{C}) P(\mathcal{E}_{dec}|(I = i) \cap \mathcal{C}). \tag{57}$$

Since the probability that $I = i$ depends only on the $\mathcal{C}_1$ codebook and the conditional error event $\mathcal{E}_{dec}$ given $I = i$ depends only on the codebook $\mathcal{C}_2$, hence

$$E_{\mathcal{C}}(P(\mathcal{E}_{dec}|\mathcal{E}_{enc}^c \cap \mathcal{C})) = \sum_{i=1}^{2^{nR}} E_{\mathcal{C}_1}\left(P(I = i|\mathcal{E}_{enc}^c \cap \mathcal{C}_1)\right) E_{\mathcal{C}_2}\left(P(\mathcal{E}_{dec}|(I = i) \cap \mathcal{C}_2)\right). \tag{58}$$

The symmetry of codebook generation and the random selection of codewords when more than one codewords are distortion typical with the source vectors implies that $P(I = i|\mathcal{E}_{enc}^c \cap \mathcal{C}_1)$ averaged over all codebooks is equal for all i, hence $E_{\mathcal{C}_1}(P(I = i|\mathcal{E}_{enc}^c \cap \mathcal{C}_1)) = 1/2^{nR}$. Thus

$$E_{\mathcal{C}}(P(\mathcal{E}_{dec}|\mathcal{E}_{enc}^c \cap \mathcal{C})) = \sum_{i=1}^{2^{nR}} \frac{1}{2^{nR}} E_{\mathcal{C}_2}\left(P(\mathcal{E}_{dec}^i|(I = i) \cap \mathcal{C}_2)\right), \tag{59}$$

where $\mathcal{E}_{dec}^i = \mathcal{E}_{dec} \cap (I = i)$. Let $\mathcal{P}_i$ be the event that the i^{th} codeword of $\mathcal{C}_2$ satisfies the power constraint. Let $\mathcal{B}_j$ denotes the event that $(Z, \tilde{Y}_j) \in A_{2,\epsilon}^{(m)}$, where Z is the decoder input and $\tilde{Y}_j$ is the j^{th} codeword of $\mathcal{C}_2$. Note that the event

$$\mathcal{E}_{dec}^i \subset \mathcal{P}_i^c \cup \left(\left(\mathcal{B}_i^c \cup \left(\bigcup_{j \neq i} \mathcal{B}_j\right)\right) \cap \mathcal{P}_i\right). \tag{60}$$

Applying the union bound we get

$$P(\mathcal{E}_{dec}^i|(I = i) \cap \mathcal{C}_2) \leq P(\mathcal{P}_i^c|(I = i) \cap \mathcal{C}_2) + P(\mathcal{B}_i^c \cap \mathcal{P}_i|(I = i) \cap \mathcal{C}_2) + \sum_{j \neq i} P(\mathcal{B}_j \cap \mathcal{P}_i|(I = i) \cap \mathcal{C}_2). \tag{61}$$

Taking the expectation over the random codebook $\mathcal{C}_2$, using same approach as in [3], we get

$$\begin{aligned} E_{\mathcal{C}_2}[P(\mathcal{E}^i_{dec}|(I=i)\cap\mathcal{C}_2)] &\leq P(\mathcal{P}_i)+P(\mathcal{B}_i^c\cap\mathcal{P}_i)+\sum_{j\neq i}P(\mathcal{B}_j\cap\mathcal{P}_i) \\ &\leq \epsilon+\epsilon+2^{n(R-3\epsilon-\rho I(Z;\tilde{Y}))}, \end{aligned} \tag{62}$$

where

$$I(Z;\tilde{Y})=\frac{1}{2}\log\left(1+\frac{P-N\delta_1}{N}\right). \tag{63}$$

Thus if $R<\frac{\rho}{2}\log\left(1+\frac{P-N\delta_1}{N}\right)$ then for large n, the second term of (54) is less than $\epsilon'=(4\sigma^2+\epsilon)3\epsilon$.

Now consider third term in (54). Let X be the source vector and $\tilde{X}$ be the decoded vector. Since $\|\tilde{X}\|/\sqrt{n}$ is bounded by $\sqrt{\sigma^2-D_{opt}}$,

$$\begin{aligned} d(X,\tilde{X}) &\leq \frac{1}{n}(\|X\|^2+\|\tilde{X}\|^2+2\|X\|\|\tilde{X}\|) \\ &\leq \frac{\|X\|^2}{n}+K_1+K_2\frac{\|X\|}{\sqrt{n}}, \end{aligned} \tag{64}$$

where $K_1=K_2^2=\sigma^2-D_{opt}$. Thus

$$E(d|\mathcal{E}_{enc}\cap\mathcal{C})\leq E\left[\frac{\|X\|^2}{n}+K_1+K_2\frac{\|X\|}{\sqrt{n}}\middle|\mathcal{E}_{enc}\cap\mathcal{C}\right]. \tag{65}$$

Let $V(X,\mathcal{C})$ be an indicator function for $\mathcal{E}_{enc}$, i.e., $V(X,\mathcal{C})=1$, if X does not belong to $A^{(n)}_{4,\epsilon}\cup A^{(n)}_{1,\epsilon}$ for some codeword in $\mathcal{C}_1$ and zero otherwise. Note that

$$E\left[\frac{1}{n}\|X\|^2\middle|(V(X,\mathcal{C}_1)=1)\right]P(V(X,\mathcal{C}_1)=1|\mathcal{C}_1)=E\left[\frac{1}{n}V(X,\mathcal{C}_1)\|X\|^2\middle|\mathcal{C}_1\right]\triangleq D^*(\mathcal{C}_1) \tag{66}$$

Let $d_i=X_i^2$, thus

$$D^*(\mathcal{C}_1)=\frac{1}{n}\sum_{i=1}^{n}E(V(X,\mathcal{C}_1)X_i^2|\mathcal{C}_1). \tag{67}$$

Let $\gamma>0$ be a number to be chosen later. Define

$$U_i=\begin{cases}1 & \text{if } X_i^2\leq\gamma\\ 0 & \text{otherwise}\end{cases} \tag{68}$$

Note that $U_i X_i^2 \leq \gamma$. Let $V = V(X, \mathcal{C}_1)$

$$\begin{aligned} E(V X_i^2 | \mathcal{C}_1) &= E(U_i V X_i^2 | \mathcal{C}_1) + E((1-U_i) V X_i^2 | \mathcal{C}_1) \\ &\leq \gamma E(V|\mathcal{C}_1) + E(X_i^2 (1-U_i)|\mathcal{C}_1). \end{aligned} \tag{69}$$

Taking the expectation over $\mathcal{C}_1$ we get

$$E(V X_i^2) \leq \gamma P(V=1) + E(X_i^2(1-U_i)). \tag{70}$$

Since $E(X_i^2) = \sigma^2 \leq \infty$, $\exists$ a sufficiently large γ such that $E(X_i^2(1-U_i)) \leq \epsilon$. Now following proof in [3] for rate distortion function we get that for this γ we can choose a sufficiently large n s.t. $\gamma P(V=1) \leq \epsilon$ provided $R > \frac{1}{2}\log(\sigma^2/(D_{opt} + D_{opt}\delta_2)) + 3\epsilon$. Further

$$\begin{aligned} K_2 E_{\mathcal{C}}\left[E\left[\frac{\|X\|}{\sqrt{n}} V \middle| \mathcal{C}\right]\right] &\leq K_2 E_{\mathcal{C}}\left[\left(E\left[\frac{\|X\|^2}{n} V \middle| \mathcal{C}\right]\right)^{\frac{1}{2}}\right] \\ &\leq K_2 \left(E_{\mathcal{C}}\left[E\left[\frac{\|X\|^2}{n} V \middle| \mathcal{C}\right]\right]\right)^{\frac{1}{2}} \\ &\leq \epsilon \end{aligned} \tag{71}$$

This proves that the first term in (52) is bounded by $D_{opt} + \delta$ provided

$$\frac{1}{2}\log\left(\sigma^2/(D_{opt} + D_{opt}\delta_2)\right) + 3\epsilon < R < \frac{\rho}{2}\log\left(1 + \frac{P - N\delta_1}{N}\right) - 3\epsilon. \tag{72}$$

Similarly, we can prove that the second term in (52) is bounded by $N + \delta$ provided

$$\frac{\rho}{2}\log\left((P+N)/(N + N\delta_1)\right) + 3\epsilon < \rho R' = R < \frac{1}{2}\log\left(1 + \frac{\sigma^2 - D_{opt} - D_{opt}\delta_2}{D}\right) - 3\epsilon. \tag{73}$$

We need to show that there exists R which satisfies (72) and (73).

Lemma 12 *Given $\delta_1 > 0$, $\delta_2 = \rho\delta_1$ and $\delta_1 << 1$, there exists R which satisfies (72) and (73).*

Proof: Since ϵ in these equations is arbitrary hence such an R exists iff

$$\frac{1}{2}\log\left(\frac{\sigma^2}{D_{opt}(1+\delta_2)}\right) < \frac{\rho}{2}\log\left(1 + \frac{P - N\delta_1}{N}\right) \tag{74}$$

$$\frac{\rho}{2}\log\left(\frac{P+N}{N(1+\delta_1)}\right) < \frac{1}{2}\log\left(1+\frac{\sigma^2-D_{opt}(1+\delta_2)}{D_{opt}}\right) \tag{75}$$

$$\frac{1}{2}\log\left(\frac{P+N}{N(1+\delta_1)}\right) < \frac{1}{2}\log\left(1+\frac{P-N\delta_1}{N}\right) \tag{76}$$

$$\frac{1}{2}\log\left(\frac{\sigma^2}{D_{opt}(1+\delta_2)}\right) < \frac{1}{2}\log\left(1+\frac{\sigma^2-D_{opt}(1+\delta_2)}{D_{opt}}\right) \tag{77}$$

For δ_1, $\delta_2 << 1$, (74) is equivalent to

$$\delta_2 > \frac{\rho\delta_1}{1+\frac{P}{N}}. \tag{78}$$

Similarly, (75) is equivalent to

$$\rho\delta_1 > \frac{\delta_2}{\frac{\sigma^2}{D_{opt}}} \tag{79}$$

Note that the above two inequalities holds for $\delta_2 = \rho\delta_1$. Inequality (77) is equivalent to

$$0 < \delta_2\left(\left(1-\frac{D_{opt}}{\sigma^2}\right)-\delta_2\frac{D_{opt}}{\sigma^2}\right), \tag{80}$$

which is true for $0 < \delta_2 < \frac{\sigma^2}{D} - 1$. Similarly (76) holds if $0 < \delta_1 < \frac{P}{N}$. This proves that there exists R, which satisfies (72) and (73). $\square$

Thus we have shown that averaged over all such codebooks the sum of distortions of the forward and dual codes is less than $D_{opt} + N + \delta$. By the converse of the joint source-channel coding theorem, for all codebooks, the distortion for the forward system is $\geq D_{opt}$ and the distortion for the dual code is $\geq N$. Thus there exists a codebook for which the distortion for the forward system is less than $D_{opt} + \delta$ and the distortion for the dual system is less than $N + \delta$. $\square$

7 Discussion of Open Problems

In this paper we proved a joint source-channel duality theorem for an iid Gaussian source and an AWGN channel assuming boundedness and uniform continuity of code sequence (α_n, β_n). There is an open question as to whether these assumptions are necessary or are too restrictive. It can be shown that any optimal code sequence can be modified into an optimal code sequence which satisfies the boundedness assumption. The serious open question has to do with the uniform continuity assumption. Many communication systems consists of a quantizer and a modulator. All quantizers-modulators pairs are inherently discontinuous. However, it is not clear whether there are sequences of quantizer-modulator pairs which are uniformly continuous in probability. Furthermore, if they are uniformly continuous, can they be optimal?

Another open problem is the following. Suppose that (α_n, β_n) are an optimal code sequence. Are there uniformly continuous approximations of this code sequence which are also optimal?

In the case of $\rho = 1$, some of these questions can be answered. For example the encoder sequence $\alpha_n(x) = \sqrt{P}x$, and the decoder sequence $\beta_n(z) = (\sqrt{P}/(P+N))z$, satisfy uniform continuity assumption and $D_n = D_{opt}$ for this code sequence.

The result of this paper can be applied to a binary iid source with $Pr\{X_i = 1\} = 1 - Pr\{X_i = 0\} = p$ and Hamming distortion measure and binary symmetric channel (BSC) with crossover probability q. Assume the channel can be used ρ times per source sample. Let $W(x)$ denotes the Hamming weight of a binary vector x and $H(p)$ be the binary entropy function. Let

$$\mathcal{B}_x^n(\mu, p) \triangleq p^{W(x \oplus \mu)}(1-p)^{(n - W(x \oplus \mu))}. \tag{81}$$

Denote

$$f(x, z) = \mathcal{B}_x^n(0, p)\mathcal{B}_z^m(\alpha(x), q), \tag{82}$$

$$\phi(x, z) = \mathcal{B}_z^m(0, \frac{1}{2})\mathcal{B}_x^n(\beta(z), e), \tag{83}$$

where e is the minimum value for which $H(e) \geq H(p) + \rho H(q) - \rho$. In this case we define the dual system to be a binary iid source with probability $1/2$ and a BSC with crossover probability e with $1/\rho$ channel uses per source sample.

Using the same approach with these f and ϕ, we can prove that if Assumption 1 holds and if the code sequence (α_n, β_n) is asymptotically optimal for the forward system, then the dual code sequence (β_n, α_n) is asymptotically optimal for the dual system.

8 Conclusion

We proved that in a joint source-channel coding system there is a duality in the encoder-decoder pair. In order to prove this duality we used the blowing up property [9, 10] of probability distributions and assumed uniform-continuity of the encoder-decoder pair. It is also shown that if this condition does not hold, the dual system might not be optimal. The existence of a code sequence having such duality is also shown.

Appendix

Proof of Lemma 1: By the chain rule of relative entropy, $\frac{1}{n}D(f(x,z)||\phi(x,z)) \to 0$ implies that $\frac{1}{n}D(f(z)||\phi(z)) \to 0$ [3]. Assume that $E_f[Z^TZ] = mS_m$ and let $h(f)$ represents the differential entropy of $f(z)$. Let C be the covariance matrix of Z under f and let λ_i, $i = 1, \ldots, m$ be the eigenvalues of C. By assumption, trace$(C) = mS_m$ and hence $\sum_{i=1}^m \lambda_i = mS_m$. By the Gaussian bound on differential entropy for a given covariance matrix we have $h(f) \leq (1/2)\sum_{i=1}^m \log(2\pi e \lambda_i)$. Further, a product of m positive numbers which sum to a constant is maximized if all the numbers are equal. Hence $h(f) \leq (1/2)m\log(2\pi e S_m)$.
Now Consider

$$\begin{aligned}
\frac{1}{n}D(f(z)||\phi(z)) &= -\frac{1}{n}h(f) - \frac{1}{n}E_f(\log(\phi(z))) \\
&\geq -\frac{\rho}{2}\log(2\pi e S_m) + \frac{\rho}{2}\frac{S_m}{P+N} + \frac{\rho}{2}\log(2\pi(P+N)) \\
&= \frac{\rho}{2}(\log\frac{P+N}{S_m} + \frac{S_m}{P+N} - 1). \qquad (84)
\end{aligned}$$

Note that $g(x) = \log(x) + 1/x - 1 \geq 0$ with equality iff $x = 1$. Further the slope of $g(x)$ is negative for $x < 1$ and positive for $x > 1$, hence the function increases as we go away from $x = 1$ in either direction and is continuous at $x = 1$. This implies that S_m converges to $P + N$ as m approaches infinity. By the independence of channel noise and encoder output under $f(x,z)$, we have $E_f[Z^TZ] = E_f[\alpha(X)^T\alpha(X)] + E_f[(Z-\alpha(X))^T(Z-\alpha(X))]$, hence $\frac{1}{m}E_f[\alpha(X)^T\alpha(X)]$ converges to P. $\Box$

Define $l = \sqrt{\frac{D_{opt}}{P+N}}$. $T_1(B)$ is a set of $(m+n)$-dimensional points (u,z), where u is n-dimensional and z is m-dimensional. For notational convenience, we say that $u \in T_1(B)$ if u belongs to the n-dimensional projection of $T_1(B)$ and we say that $z \in T_1(B)$ if z belongs to m-dimension projection of $T_1(B)$.

Proof of Lemma 8: Let $h(x,z) = \frac{1}{n}\log\frac{f(x,z)}{\phi(x,z)}$. Define $g(u,z) = h(\beta(z) + lu, z)$. Note that $\forall\ (u,z) \in T_1(B_{\epsilon_1}^n)$, $|g(u,z)| < \epsilon_1$. We need to show that $\exists n_o$ s.t. $\forall\ n > n_o$ and $\forall (u,z) \in b_\delta(T_1(B_{\epsilon_1}^n))$, $|g(u,z)| < \epsilon_2$. Let $\epsilon = \epsilon_2 - \epsilon_1$. Let $(u_2, z_2) \in b_\delta(T_1(B_{\epsilon_1}^n))$ and let $(u_1, z_1) \in T_1(B_{\epsilon_1}^n)$

be s.t. $\bar{d}((u_2, z_2), (u_1, z_1)) < \delta$. By the definition of the blown-up set, $\forall\ (u_2, z_2) \in b_\delta(T_1(B^n_{\epsilon_1}))\ \exists$ $(u_1, z_1) \in T_1(B^n_{\epsilon_1})$ s.t. $\bar{d}((u_1, z_1), (u_2, z_2)) \leq \delta$. Thus it is sufficient to prove uniform continuity of $g(u, z)$ on the set $T_1(B^n_{\epsilon_1})$, i.e., to show that given $\epsilon > 0$, $\exists\delta$ and n_o s.t. $\forall\ n > n_o$ and $\forall$ $(u_1, z_1) \in T_1(B^n_{\epsilon_1})$

$$\bar{d}((u_1, z_1), (u_2, z_2)) < \delta \Longrightarrow |g(u_1, z_1) - g(u_2, z_2)| \leq \epsilon. \tag{85}$$

Expanding $g(u, z)$ we get

$$g(u, z) = -\frac{(lu + \beta(z))^T(lu + \beta(z))}{2n\sigma^2} - \frac{(z - \alpha(\beta(z) + lu))^T(z - \alpha(\beta(z) + lu))}{2nN} + \frac{z^T z}{2n(P + N)} + \frac{u^T u}{2n(P + N)}. \tag{86}$$

If each term of (86) is uniformly continuous on $T_1(B^n_{\epsilon_1})$, then $g(u, z)$ is uniformly continuous on $T_1(B^n_{\epsilon_1})$. Since $x^T x/n$ is bounded on $B^n_{\epsilon_1}$ and $\beta(z)$ satisfies Assumption 2, $u^T u/n$ is bounded on $T_1(B^n_{\epsilon_1})$.

Let $v = u_2 - u_1$. Now consider

$$\left|\frac{u_1^T u_1}{n} - \frac{u_2^T u_2}{n}\right| = \left|\frac{(u_1 - u_2)^T(u_1 + u_2)}{n}\right| \leq \frac{\|v\|\|2u_1 + v\|}{n}, \tag{87}$$

where the inequality follows from Cauchy-Schwartz. Since $\bar{d}(u_1, u_2) < \delta$, $\|v\|/\sqrt{n} < \sqrt{\delta}$. Applying the triangular inequality to $\|2u_1 + v\|$ we get the required continuity. Similarly we can prove uniform continuity of $z^T z/m$ on $T_1(B^n_{\epsilon_1})$. This proves uniform continuity of 3^{rd} and 4^{th} term of (86). Now, consider 1^{st} term of (86)

$$\begin{aligned}
&\left|\frac{(lu_1 + \beta(z_1))^T(lu_1 + \beta(z_1))}{n} - \frac{(lu_2 + \beta(z_2))^T(lu_2 + \beta(z_2))}{n}\right| \\
&= \left(\frac{l(u_1 + u_2) + \beta(z_1) + \beta(z_2)}{\sqrt{n}}\right)^T \left(\frac{l(u_1 - u_2) + \beta(z_1) - \beta(z_2)}{\sqrt{n}}\right) \\
&\leq \frac{\|l(u_1 + u_2) + \beta(z_1) + \beta(z_2)\|}{\sqrt{n}} \frac{\|l(u_1 - u_2) + \beta(z_1) - \beta(z_2)\|}{\sqrt{n}} \\
&\leq M \frac{\|l(u_1 - u_2) + \beta(z_1) - \beta(z_2)\|}{\sqrt{n}}
\end{aligned}$$

$$\leq M\left(l\sqrt{\bar{d}(u_1,u_2)}+\sqrt{\bar{d}(\beta(z_1),\beta(z_2))}\right)$$
$$\leq \epsilon, \tag{88}$$

where $M = \|l(u_1+u_2)+\beta(z_1)+\beta(z_2)\|/\sqrt{n}$ is a finite constant depending on σ^2, D_{opt} and K. The above inequalities follows from Cauchy-Schwartz, the fact that $\beta(z_i)^T\beta(z_i)/n$ and $u_i^Tu_i/n$ are bounded on $b_\delta(T_1(B_{\epsilon_1}^n))$, triangular inequality, and uniform continuity of $\beta(z)$ on $T_1(B_{\epsilon_1}^n)$.

Uniform continuity of $(z-\alpha(\beta(z)+lu))^T(z-\alpha(\beta(z)+lu))/n$ can be similarly proved provided we show that $\alpha(\beta(z)+lu)$ is uniformly continuous on $T_1(B_{\epsilon_1}^n)$. By uniform continuity of $\beta(z)$ on $T_1(B_{\epsilon_1}^n)$, it follows that given $\delta_1>0$, $\exists\delta>0$ and n_1 s.t. $n>n_1$,

$$\bar{d}((u_1,z_1),(u_2,z_2))<\delta \Longrightarrow \bar{d}(lu_1+\beta(z_1),lu_2+\beta(z_2))<\delta_1 \tag{89}$$

Let $x_1 = lu_1+\beta(z_1)$ and $x_2 = lu_2+\beta(z_2)$. Note that $x_1\in B_{\epsilon_1}^n$ and since α is uniformly continuous on $B_{\epsilon_1}^n$, given $\epsilon>0$ $\exists$ $\delta_1>0$ and n_2 s.t. $\forall n>n_2$, $\bar{d}(\alpha(x_1),\alpha(x_2))<\epsilon$. Pick this δ_1 and n_2. Choose δ and n_1 in (89) for this δ_1 and let $n_o=\max(n_1,n_2)$. Thus given $\epsilon>0$, $\exists\delta$, n_o s.t $\forall n>n_o$

$$\bar{d}((u_1,z_1),(u_2,z_2))<\delta \Longrightarrow \bar{d}(lu_1+\beta(z_1),lu_2+\beta(z_2))<\delta_1 \Longrightarrow \bar{d}(\alpha(lu_1+\beta(z_1)),\alpha(lu_2+\beta(z_2)))<\epsilon. \tag{90}$$

This proves uniform continuity of $\alpha(lu+\beta(z))$.

Thus we have shown that $g(u,z)$ is uniformly continuous on $T_1(B_{\epsilon_1}^n)$. Hence, $\exists$ $\delta>0, n_o$ s.t. $\forall n>n_o$, $\forall$ $(u_2,z_2)\in b_\delta(T_1(B_{\epsilon_1}^n))$, $|g(u_2,z_2))|<\epsilon_2=\epsilon_1+\epsilon$. Similarly we can show that $\forall(u_2,z_2)\in b_\delta(T_1(B_{\epsilon_1}^n))$, Property 2 (Section 4.1.1) holds. Thus $b_\delta(T_1(B_{\epsilon_1}^n))\subset T_1(A_{\epsilon_2}^n)$. $\square$

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[2] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pp. 142–163, Mar. 1959.

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

[4] M. V. Eyuboğlu and G. D. Forney, Jr., "Lattice and trellis quantization with lattice- and trellis-bounded codebooks–High rate theory for memoryless sources," *IEEE Trans. on Information Theory*, pp. 46–59, Jan. 1993.

[5] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. on Commun.*, pp. 82–93, Jan. 1990.

[6] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. on Information Theory*, pp. 55-67, Jan. 1982.

[7] R. Laroia, N. Farvardin and S. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Trans. on Information Theory*, pp. 1044–1056, July 1994.

[8] R. Laroia and N. Farvardin, "A structured fixed-rate vector quantizer derived from a variable-length scalar quantizer: Part I–Memoryless sources," *IEEE Trans. on Information Theory*, pp. 851–867, May 1993.

[9] K. Marton, "Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration," *The Annals of Probability*, vol. 24, pp. 857–866, 1996.

[10] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product space," *Publ. IHES.* 81, 73–205, 1995.

[11] H. Shalaby and A. Papamarcou, "Error exponents for distributed detection of Markov sources," *IEEE Trans. on Information Theory*, pp. 397-408, March 1994.

[12] A. F. Karr, *Probability*, New York: Springer Texts in Statistics, 1992.

[13] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, San Francisco: Holden-Day, Inc. 1964.