The Posterior Matching Feedback Scheme for Joint Source-Channel Coding with Bandwidth Expansion

Ofer Shayevitz
Information Theory & Applications Center
University of California, San Diego
La Jolla, CA 92093, USA
ofersha@ucsd.edu

Meir Feder
Dept. of Electrical Engineering - Systems
Tel Aviv University
Tel Aviv, Israel 69978
meir@eng.tau.ac.il

Abstract

When transmitting a Gaussian source over an AWGN channel with an input power constraint and a quadratic distortion measure, it is well known that optimal performance can be obtained using an analog joint source-channel scalar scheme which merely scales the input and output signals. In the case of bandwidth expansion, such a joint source-channel analog scheme attaining optimal performance is no longer simple. However, when feedback is available a simple and sequential analog linear procedure based on the Schalkwijk-Kailath scheme for communication, is optimal. Recently, we have introduced a fundamental feedback communication scheme, termed *posterior matching*, which generalizes the Schalkwijk-Kailath scheme to arbitrary memoryless channels and input distributions. In this paper, we show how the posterior matching scheme can be adapted to the joint source-channel coding setting with bandwidth expansion and a general distortion measure, when feedback is available.

I Introduction

Suppose a memoryless source is to be transmitted over a memoryless channel with some input constraint, and the quality of the source's reconstruction at the receiving terminal is measured via some given distortion measure. In this case, the separation principle [1] states that the best achievable average distortion D is given by $R(D) = C$, where $R(\cdot)$ is the rate-distortion function of the source relative to the distortion measure, and C is the capacity of the channel under the input constraint. This guaranteed performance can be obtained by first optimally compressing the source with average distortion D using $R(D)$ bits per source symbol, and then using a good channel code to reliably send the compressed bit-stream over the channel. However, this method is usually very complicated and requires coding over long blocks, which results in long delays.

Optimal performance can be obtained in many cases without the artificial separation into source coding and channel coding, using a much simpler *joint-source channel* *coding* (JSCC) scheme. Perhaps the most classical example is the case of transmitting a Gaussian source $\sim \mathcal{N}(0, \mathrm{P_s})$ over an additive white Gaussian noise (AWGN) channel with noise $\mathcal{N}(0, \mathrm{N})$, an input power constraint P, and a quadratic distortion measure. In this case, the optimal performance guaranteed by the separation principle $D = \mathrm{P_s} \cdot (1 + \frac{\mathrm{P}}{\mathrm{N}})^{-1}$ can be reaped by a remarkably simple JSCC scheme. The source is linearly scaled by a factor of $\sqrt{\frac{\mathrm{P}}{\mathrm{P_s}}}$ to satisfy the input power constraint, and then simply transmitted over the channel. On the receiving end, the reconstruction is set to be the minimum mean square error (MMSE) estimator of the source given the channel's output, which by Gaussianity is simply a linear scaling of the output by the factor $\frac{\sqrt{\mathrm{P} \cdot \mathrm{P_s}}}{\mathrm{P}+\mathrm{N}}$. Unfortunately, such a simple JSCC scheme exists in general only when a special relation between the source, the channel and the distortion measure is satisfied [2].

In the above discussion we have implicitly assumed that the source emits one sample per every channel use, which is of course not the case in general. The case where multiple channel uses are available per each source sample is referred to as *bandwidth expansion*. In this paper we assume an integer bandwidth expansion factor (BEF) $m > 1$, although the discussion is easily generalized to any rational factor. Let us return to the Gaussian source/AWGN channel example, this time with bandwidth expansion. In this case, the optimal performance guaranteed by the separation principle is given by $D = \mathrm{P_s} \cdot (1 + \frac{\mathrm{P}}{\mathrm{N}})^{-m}$, i.e., the distortion decays exponentially with the BEF. A trivial generalization of the aforementioned JSCC scheme to this case would be to simply retransmit the scaled source over the m slots, perform a coherent summation of the outptus, and employ an MMSE estimator to reconstruct the source. However, this approach is strictly suboptimal since it results in $D = m^{-1} \cdot \mathrm{P_s} \cdot (1 + \frac{\mathrm{P}}{\mathrm{N}})^{-1}$, i.e., the distortion decays only polynomially with the BEF. In fact, in this case a simple JSCC scheme attaining optimal performance is not known, and does not even exist in high resolution [3].

However, if noiseless feedback is available then optimal performance can nevertheless be obtained in the above Gaussian case [4, 5]. The idea is simple, essentially based on the Schalkwijk-Kailath scheme for communication with feedback over an AWGN channel [6, 7]. On the first time slot, the scaled source is transmitted as before. The transmitter then calculates the MMSE estimate of the input given that output (which is available via feedback), and on the second slot sends the error term scaled to satisfy the input power constraint. This procedure is then repeated recursively until all m slots are used. This linear *zoom in* process results in an exponential decay of the distortion with m, which turns out to be optimal. Unfortunately, once again this elegant scheme is specifically tailored to the Gaussian case, and such a "no coding" scheme does not generally exist unless some unique situation occurs [8].

In a recent development [9, 10], we have identified a fundamental principle of feedback communication and used it to describe a generic feedback transmission scheme, termed *posterior matching*, which is suitable for essentially any memoryless channel and any desired input distribution, achieving the corresponding mutual information in a simple and sequential "no coding" fashion. In particular, when specialized to the AWGN channel with a Gaussian input, the posterior matching scheme reduces to the Schalkwijk-Kailath scheme. This fact together with the elegant use of the Schalkwijk-Kailath scheme in a JSCC setting with bandwidth expansion and feedback described above [5], motivates the study of a similar usage of the posterior matching scheme to the same end, over other memoryless channels with a different distortion measure and different input constraints. Although one cannot hope to be generally optimal in this way [8], it is still interesting to explore the performance guaranteed by such a simple and attractive scheme.

II Notations and Background

Random variables (r.v's) are denoted by upper-case letters, their realizations by corresponding lower-case letters. A real r.v. X is associated with a probability distribution $P_X(\cdot)$ and a *cumulative distribution function* (c.d.f.) given by $F_X(x) = P_X\big((-\infty, x\,]\big)$. We also assume that X admits a *probability density function* (p.d.f.) $f_X(x)$. The *support* of X is denoted $\mathrm{supp}(X)$. The *tail function* $\mathscr{T}_X : \mathbb{R}^+ \mapsto [0,1]$ of X is defined by

$$\mathscr{T}_X(\ell) \;\stackrel{\triangle}{=}\; \inf\big\{p \,:\, P_X\big((x_0,x_1)\big) = 1-p\,,\, x_1 - x_0 = \ell\big\}$$

Namely, $\mathscr{T}_X(\ell)$ is the minimal probability which P_X assigns outside a length ℓ interval. We use $|\Delta|$ for the length of an interval $\Delta \subseteq \mathbb{R}$, log for $\log_2$, and $\circ$ for function composition. The indicator function over a set A is denoted by $\mathbb{1}_A(\cdot)$.

II.1 Reversed Iterated Function System

Let $\mathcal{F} \subseteq \mathbb{R}$, and let $\omega : \mathbb{R}\times\mathcal{F} \mapsto \mathcal{F}$ be some measurable function, where we write $\omega_y(\cdot) \stackrel{\triangle}{=} \omega(y,\cdot)$ for short. Let $\{Y_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of real r.v's. A *Reversed Iterated Function System (RIFS)* $\{S_n(s)\}_{n=1}^{\infty}$ is a stochastic process over $\mathcal{F}$, defined by

$$S_1 = s \in \mathcal{F}\,, \quad S_{n+1}(s) = \omega_{Y_1}\circ\omega_{Y_2}\circ\cdots\circ\omega_{Y_n}(s) \tag{1}$$

We say that the RIFS is *generated* by the RIFS *kernel* $\omega_y(\cdot)$, *controlled* by the sequence $\{Y_n\}_{n=1}^{\infty}$, and s is its initial point. We now state a useful convergence Lemma for the RIFS. First we need the following definition:

$$D_{s,t}(h) \;\stackrel{\triangle}{=}\; \frac{|h(s)-h(t)|}{|s-t|} \tag{2}$$

for any $h : \mathbb{R}\mapsto\mathbb{R}$ and $s,t\in\mathbb{R}$.

Lemma 1 (from [10]). *Consider the RIFS in (1) with the i.i.d. control sequence* $\{Y_n\}_{n=1}^{\infty}$*, and suppose the following conditions hold for some* $q>0$*:*

$$r \;\stackrel{\triangle}{=}\; \sup_{s\neq t\in\mathcal{F}} \mathbb{E}\,[D_{s,t}(\omega_{Y_1})]^q < 1 \tag{3}$$

Then for any $\varepsilon > 0$

$$\mathbb{P}\left(|S_n(s)-S_n(t)| > \varepsilon\right) \;\leq\; \varepsilon^{-q}|s-t|^q r^n \qquad s,t\in\mathcal{F}$$

II.2 The Communication Problem

A *memoryless channel* is defined by a conditional probability distribution $P_{Y|X}$ on $\mathbb{R}$. The *input alphabet* $\mathcal{X}$ of the channel is the set of all $x \in \mathbb{R}$ for which the distribution $P_{Y|X}(\cdot|x)$ is defined, the output alphabet of the channel is the set $\mathcal{Y} \triangleq \bigcup_{x\in\mathcal{X}} \mathrm{supp}(Y|X = x) \subseteq \mathbb{R}$. A sequence of real r.v. pairs $\{(X_n, Y_n)\}_{n=1}^{\infty}$ is said to be an *input/output sequence* for the channel $P_{Y|X}$ if

$$P_{Y_n|X^n Y^{n-1}}(\cdot|x^n, y^{n-1}) = P_{Y|X}(\cdot|x_n)\,, \qquad n \in \mathbb{N} \tag{4}$$

A probability distribution P_X is said to be a (memoryless) *input distribution* for the channel $P_{Y|X}$ if $\mathrm{supp}(X) \subseteq \mathcal{X}$. The pair $(P_X, P_{Y|X})$ induces an *output distribution* P_Y over the output alphabet, a joint input/output distribution P_{XY}, and an *inverse channel* $P_{X|Y}$.

Let Θ_0 be a random *message point* uniformly distributed over the unit interval, its binary expansion representing an infinite independent-identically-distributed (i.i.d) binary sequence to be reliably transmitted over the channel $P_{Y|X}$. A *transmission scheme* (with feedback) is a sequence of *transmission functions* $\{g_n : (0,1) \times \mathbb{R}^{n-1} \mapsto \mathbb{R}\}_{n=1}^{\infty}$, so that the input to the channel $P_{Y|X}$ at time n is given by

$$X_n = g_n(\Theta_0, Y^{n-1}) \tag{5}$$

A *decoding rule* is a sequence of mappings $\{\Delta_n : \mathbb{R}^n \mapsto \mathcal{E}\}_{n=1}^{\infty}$, where $\mathcal{E}$ is the set of all open intervals in (0,1). The *error probability* at time n associated with a transmission scheme and a decoding rule, is defined as

$$p_e(n) \triangleq \mathbb{P}(\Theta_0 \notin \Delta_n(Y^n)) \tag{6}$$

and the corresponding *rate* at time n is defined to be

$$R_n \triangleq -\frac{1}{n} \log |\Delta_n(Y^n)| \tag{7}$$

We say that a transmission scheme together with a decoding rule *achieve* a rate R over a channel $P_{Y|X}$ if

$$\lim_{n\to\infty} \mathbb{P}(R_n < R) = 0\,, \qquad \lim_{n\to\infty} p_e(n) = 0$$

The rate is achieved *within an input constraint* (η, u), if in addition

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \eta(X_k) \leq u \quad \text{a.s.} \tag{8}$$

where $\eta : \mathbb{R} \mapsto \mathbb{R}$ and $u \in \mathbb{R}$.

II.3 The Posterior Matching Scheme

Recently, we have introduced the following *posterior matching* transmission scheme for communication using feedback over a memoryless channel [10]:

$$X_{n+1} = F_X^{-1} \circ F_{\Theta_0|Y^n}(\Theta_0|Y^n) \tag{9}$$

This scheme corresponds to the transmission functions $g_{n+1}(\theta, y^n) = F_X^{-1} \circ F_{\Theta_0|Y^n}(\theta|y^n)$. The inputs are produced in two steps: First, the information missing at the receiver is "extracted" from the a-posteriori distribution, by generating a r.v. independent of past observations, that together with them uniquely determines Θ. Then, the distribution of that r.v. is "matched" to the channel by transforming it into the desired input distribution P_X. This results in $X_n \sim P_X$ independent of an i.i.d output sequence Y^{n-1}. It turns out that the posterior matching scheme admits a simple recursive form.

Lemma 2 (from [10]). *If the input/channel pair $(P_X, P_{Y|X})$ induces a joint p.d.f. f_{XY}, then the posterior matching scheme is also given by*

$$X_1 = F_X^{-1}(\Theta_0)\,, \quad X_{n+1} = F_X^{-1} \circ F_{X|Y}(X_n|Y_n) \tag{10}$$

In [10], the posterior matching scheme together with an optimal decoding rule are shown to achieve any rate below the mutual information $I(X;Y)$, within the input constraint $(\eta, \mathbb{E}\eta(X))$. In particular, the posterior matching scheme reduces to the Schalkwijk-Kailath scheme in the AWGN channel case with an input power constraint.

III Joint Source-Channel Coding

III.1 The Setting

Let A be some r.v. taking values over a (possibly infinite) interval $\mathcal{A} \subseteq \mathbb{R}$, let $P_{Y|X}$ be a memoryless channel with an input alphabet $\mathcal{X} \subseteq \mathbb{R}$ and output alphabet $\mathcal{Y} \subseteq \mathbb{R}$, and let $d : \mathcal{A} \times \mathcal{A} \mapsto \mathbb{R}^+$ be a distortion measure. A JSCC scheme with feedback and BEF m is a sequence $\{g_n : \mathcal{A} \times \mathbb{R}^{n-1} \mapsto \mathbb{R}\}_{n=1}^m$ of transmission functions, so that the input to the channel $P_{Y|X}$ at time n is given by

$$X_n = g_n(A, Y^{n-1}) \tag{11}$$

a *reconstruction rule* is a function $\widehat{A} : \mathcal{Y}^m \mapsto \mathcal{A}$. The (average) *distortion* associated with the transmission scheme and the reconstruction rule is

$$D = \mathbb{E}\, d(A, \widehat{A}(Y^m)) \tag{12}$$

The distortion D is achieved *within an average input constraint* (η, u), if in addition

$$\mathbb{E}\eta(X_k) \le u\,, \qquad k = 1, \ldots, m \tag{13}$$

where $\eta : \mathbb{R} \mapsto \mathbb{R}$ and $u \in \mathbb{R}$.

III.2 Main Result

Let P_X be some input distribution such that the pair $(P_X, P_{Y|X})$ induces a joint p.d.f. f_{XY} which is continuous over a convex support. P_X is a design parameter, incorporating the input constraints. Consider the following JSCC scheme with a BEF of m:

$$X_1 = F_X^{-1} \circ F_A(A)\,, \qquad X_{n+1} = F_X^{-1} \circ F_{X|Y}(X_n|Y_n)\,, \qquad n = 1 \ldots m-1 \tag{14}$$

We call this scheme the *posterior matching JSCC (PM-JSCC) scheme.*

Example 1. It is easily verified that in the Gaussian source/AWGN channel/quadratic distortion setting with an input power constraint P, setting $P_X \sim \mathcal{N}(0, \mathrm{P})$ results a PM-JSCC scheme that coincides with the optimal linear scheme [10].

Let us now make two simplifying assumptions:

(i) The distortion measure is bounded, i.e., $\sup_{a,b\in\mathcal{A}} d(a,b) = d_{\max} < \infty$.

(ii) An input distribution P_X is selected such that $F_A^{-1} \circ F_X$ is Lipschitz, i.e.,

$$\sup_{s\neq t\in\mathcal{X}} D_{s,t}\left(F_A^{-1}\circ F_X\right) = M < \infty \tag{15}$$

The following Theorem provides an upper bound on the distortion achieved by the PM-JSCC scheme under the above assumptions, and the basic idea behind it is simple. Say the receiver has some estimate $\widehat{X}_m$ for X_m. Then $(\widehat{X}_m, Y^{m-1})$ corresponds to a unique estimate $\widehat{X}_1$ of X_1 which is recovered by *reversing* the transmission scheme, i.e., running a RIFS over $\mathcal{X}$ generated by the kernel $\omega_y(\cdot) \triangleq F^{-1}_{X|Y}(\cdot|y)\circ F_X$, controlled by the output sequence $\{Y_k\}_{k=1}^m$, and initialized at $\widehat{X}_m$. This estimate in turn corresponds to a unique estimate $\widehat{A}$ of A. Similarly, one can tentatively select an interval in which X_m lies with high probability, then roll it back via the RIFS to recover an exponentially smaller interval w.r.t. X_1, and by the Lipschitz assumption, a corresponding exponentially small interval w.r.t. A. Since A lies in this small interval with high probability and the distortion measure is bounded, the distortion will be composed of a small term corresponding to the probability that A lies outside this interval, and another small term pertaining to the maximal distortion within the interval.

Theorem 1. *Let $\omega_y(\cdot) \triangleq F^{-1}_{X|Y}(\cdot|y)\circ F_X$ and define for any $q,\varepsilon > 0$*

$$r_q \triangleq \sup_{s\neq t\in\mathrm{supp}(X)} \mathbb{E}\left[D_{s,t}(\omega_Y)\right]^q\,, \quad \bar{d}_\varepsilon \triangleq \sup_{(a,b)\subseteq\mathcal{A}\,,|b-a|\leq\varepsilon} d(a,b) \tag{16}$$

If there exists $q^* > 0$ such that $r_{q^*} < 1$, then the PM-JSCC scheme in (14) achieves an average distortion upper bounded by

$$D \leq \inf_{0<q<q^*,\varepsilon,\ell>0}\left\{d_{\max}\left(\mathscr{T}_X(\ell) + \frac{(M\varepsilon^{-1}\ell)^q}{1-\mathscr{T}_X(\ell)}\,r_q^m\right) + \bar{d}_\varepsilon\right\} \tag{17}$$

within any input constraint of the form $(\eta, \mathbb{E}\eta(X))$.

Proof. Consider the following suboptimal reconstruction rule. Set an interval $\mathscr{J}_{m+1} = (s,t) \subseteq \mathcal{X}$ pertaining to X_{m+1}, where X_{m+1} is defined by the same recursion rule (14), although it is not actually transmitted. Let $\{S_n(s)\}_{n=1}^m$ be the RIFS generated by $\omega_y(\cdot)$ with the control sequence Y^m, and define $\mathscr{J}_1 = (S_m(s), S_m(t)) \subseteq \mathcal{X}$ to be the corresponding (random) interval pertaining to X_1. Now, set $\widehat{X}_1$ to be any point within

$\mathscr{J}_1$, say its middle point, and set $\widehat{A} = F_A^{-1} \circ F_X(\widehat{X}_1)$. Denoting $\mathscr{A} = F_A^{-1} \circ F_X(\mathscr{J}_1)$, we have

$$\begin{aligned}\mathbb{P}(A \in \mathscr{A}) &= \mathbb{P}(X_1 \in \mathscr{J}_1) = \mathbb{E}\mathbb{P}(X_1 \in \mathscr{J}_1|Y^m) = \mathbb{E}\mathbb{P}(X_{m+1} \in \mathscr{J}_{m+1}|Y^m)\\ &= \mathbb{P}(X_{m+1} \in \mathscr{J}_{m+1}) = P_X(\mathscr{J}_{m+1})\end{aligned}$$

where we have used the fact that by construction, $X_{m+1} \sim P_X$ and is independent of Y^m. Moreover, we can apply Lemma 1 to the edges of $\mathscr{J}_1$ to obtain

$$\mathbb{P}(|\mathscr{J}_1| > \varepsilon) \le \varepsilon^{-q}|\mathscr{J}_{m+1}|^q r_q^m \tag{18}$$

and so by assumption (ii) we also have

$$\mathbb{P}(|\mathscr{A}| > \varepsilon) \le \mathbb{P}(|\mathscr{J}_1| > M^{-1}\varepsilon) \le (M\varepsilon^{-1})^q|\mathscr{J}_{m+1}|^q r_q^m \tag{19}$$

Thus, the average distortion induced by the reconstruction $\widehat{A}$ is upper bounded by

$$\begin{aligned}\mathbb{E}\, d(A, \widehat{A}) &\le \mathbb{E}(d(A, \widehat{A})|A \notin \mathscr{A})\mathbb{P}(A \notin \mathscr{A}) + \mathbb{E}(d(A, \widehat{A})|A \in \mathscr{A})\\ &\le d_{\max}(1 - P_X(\mathscr{J}_{m+1})) + \mathbb{E}(d(A, \widehat{A})|A \in \mathscr{A}, |\mathscr{A}| > \varepsilon)\mathbb{P}(|\mathscr{A}| > \varepsilon|A \in \mathscr{A})\\ &\qquad + \mathbb{E}(d(A, \widehat{A})|A \in \mathscr{A}, |\mathscr{A}| \le \varepsilon)\mathbb{P}(\mathscr{A} \le \varepsilon|A \in \mathscr{A})\\ &\le d_{\max}\left(1 - P_X(\mathscr{J}_{m+1}) + \frac{\mathbb{P}(|\mathscr{A}| > \varepsilon)}{\mathbb{P}(A \in \mathscr{A})}\right) + \bar{d}_\varepsilon\\ &\le d_{\max}\left(1 - P_X(\mathscr{J}_{m+1}) + \frac{(M\varepsilon^{-1})^q|\mathscr{J}_{m+1}|^q r_q^m}{P_X(\mathscr{J}_{m+1})}\right) + \bar{d}_\varepsilon\end{aligned}$$

To tighten the bound, for any desired value of $|\mathscr{J}_{m+1}|$ we can select $\mathscr{J}_{m+1}$ such that $P_X(\mathscr{J}_{m+1})$ is maximized, namely such that $1-P_X(\mathscr{J}_{m+1}) = \mathscr{T}_X(|\mathscr{J}_{m+1}|)$. Therefore, for any $0 < \varepsilon < |\mathcal{A}|, 0 < \ell < |\mathcal{X}|, q > 0$

$$\mathbb{E}\, d(A, \widehat{A}) \le d_{\max}\left(\mathscr{T}_X(\ell) + \frac{(M\varepsilon^{-1}\ell)^q r_q^m}{1 - \mathscr{T}_X(\ell)}\right) + \bar{d}_\varepsilon$$

and the result follows by taking the infimum over the parameters. □

Example 1 (continued). In this setting assumption (ii) is satisfied with $M = \sqrt{\frac{\mathrm{P}_s}{\mathrm{P}}}$, but assumption (i) does not hold (the distortion is unbounded), and so Theorem 1 does not apply to the corresponding PM-JSSC scheme. However, in this special simple case where the PM-JSCC scheme reduces to the linear Schalkwijk-Kailath based scheme, it is easy to tailor the proof and recover the well known optimality result.

Corollary 1. *Under the assumptions of Theorem 1, suppose that for any $\alpha > 0$ small enough*

$$\limsup_{n\to\infty} -\frac{1}{n}\log \bar{d}_{\alpha^n} > 0 \tag{20}$$

and that $\mathscr{T}_X(\ell) = \mathrm{O}(\ell^{-\beta})$ for some $\beta > 0$. Then the distortion attained by the PM-JSCC scheme decays exponentially with the BEF m. In particular, (20) is satisfied for $d(a, b) = |a - b|^\gamma$ with any $\gamma > 0$.

Proof. Since $r_{q^*} < 1$, it is easy to see that one can set $\varepsilon = \alpha^m$ for some $0 < \alpha < 1$ small enough and $\ell = \ell_m$ to grow exponentially with m yet slowly enough, so that $\ell_m^{q^*} r_{q^*}^m \varepsilon^{-q^*}$ decays exponentially to zero with m. Moreover, since $\mathscr{T}_X(\ell) = \mathrm{O}(\ell^{-\beta})$ we have that $\mathscr{T}_X(\ell_m)$ decays exponentially to zero with m as well. By (20), if we choose α to be sufficiently small then $\bar{d}_\varepsilon$ will also decay exponentially to zero with m, hence the result. □

Example 2. Let $A \sim U(0,1)$, i.e., uniformly distributed over the unit interval, and let the distortion measure be $d(a,b) = |a-b|^\gamma$ for some $\gamma > 0$. This results in $d_{\max} = 1$ and $\bar{d}_\varepsilon = \varepsilon^\gamma$. Let $P_{Y|X}$ be an additive noise channel with noise uniformly distributed over the unit interval as well, and set the input $X \sim U(0,1)$ which results in a Lipschitz factor $M = 1$. Let us derive the PM-JSCC scheme in this case. It is easy to verify that the inverse channel's p.d.f. is given by

$$f_{X|Y}(x|y) = \begin{cases} y^{-1}\mathbb{1}_{(0,y)}(x) & y \in (0,1] \\ (2-y)^{-1}\mathbb{1}_{(y-1,1)}(x) & y \in (1,2) \end{cases}$$

Since the conditions of Lemma 2 are satisfied, we can use the recursive representation. It is readily verified that

$$F_X^{-1} \circ F_{X|Y}(x|y) = F_{X|Y}(x|y) = \begin{cases} \frac{x}{y} \cdot \mathbb{1}_{(0,y)}(x) + \mathbb{1}_{[y,\infty)}(x) & y \in (0,1] \\ \frac{x-y+1}{2-y} \cdot \mathbb{1}_{(y-1,1)}(x) + \mathbb{1}_{[1,\infty)}(x) & y \in (1,2) \end{cases} \tag{21}$$

Now, since the input distribution is $U(0,1)$, the initialization step is trivial and therefore the PM-JSCC scheme is given by

$$X_1 = A\,, \qquad X_{n+1} = \frac{X_n}{Y_n} \cdot \mathbb{1}_{(0,1]}(Y_n) + \frac{X_n - Y_n + 1}{2 - Y_n} \cdot \mathbb{1}_{(1,2)}(Y_n) \tag{22}$$

The above has in fact a very simple interpretation. We start by transmitting the source sample itself $X_1 = A$. Then, given Y_1 we determine the range of inputs that could have generated this output value, and find an affine transformation that stretches this range to fill the entire unit interval. Applying this transformation to X_1 generates X_2. We now determine the range of possible inputs given Y_2, and apply the corresponding affine transformation to X_2, and so on. This is intuitively appealing since what we do in each iteration is just *zoom-in* on the remaining uncertainty region for A. Since the posterior distribution is always uniform, this zooming-in is linear.

Now, the output p.d.f. is given by

$$f_Y(y) = y\mathbb{1}_{(0,1]}(y) + (2-y)\mathbb{1}_{(1,2)}(y) \tag{23}$$

The RIFS kernel is given by

$$\begin{aligned} \omega_y(s) &= F_{X|Y}^{-1}(F_X(s)|y) = \Big(y\mathbb{1}_{(0,1]}(y) + (2-y)\mathbb{1}_{(1,2)}(y)\Big) \cdot s + (y-1) \cdot \mathbb{1}_{(1,2)}(y) \\ &= f_Y(y) \cdot s + (y-1) \cdot \mathbb{1}_{(1,2)}(y) \end{aligned}$$

Therefore $D_{s,t}(\omega_y(s)) = f_Y(y)$ and so

$$r_q = \sup_{s \neq t \in \mathrm{supp}(X)} \mathbb{E}\left[D_{s,t}(\omega_Y)\right]^q = \mathbb{E}\left(f_Y(Y)\right)^q = \frac{1}{1+\frac{q}{2}} < 1 \tag{24}$$

for any $q > 0$, so the conditions of Theorem 1 are satisfied. Since X has a bounded support, the conditions of Corollary 1 are also satisfied and we have an exponential decay of the distortion with m. However in this case we can obtain a more explicit bound on the distortion using Theorem 1. We can set $\ell = 1$ which results in $\mathscr{T}(\ell) = 0$, and therefore

$$D \le \varepsilon^{-q} r_q^m + \varepsilon^\gamma \tag{25}$$

Optimizing for ε we get $\varepsilon = \left(\frac{q}{\gamma}\right)^{\frac{1}{q+\gamma}} r_q^{\frac{m}{q+\gamma}}$ and substituting into the above using (24), we finally have for any $q > 0$,

$$D \le \left(\left(\frac{\gamma}{q}\right)^{\frac{q}{q+\gamma}} + \left(\frac{q}{\gamma}\right)^{\frac{\gamma}{q+\gamma}}\right) \cdot \left(\frac{1}{1+\frac{q}{2}}\right)^{\frac{m\gamma}{q+\gamma}} \tag{26}$$

and so the distortion decays exponentially with the BEF, with an exponent of at least

$$\lim_{m\to\infty} -\frac{1}{m}\log D \ge \frac{\gamma}{q+\gamma}\log\left(1+\frac{q}{2}\right) \tag{27}$$

The best lower bound can now be found by optimizing over $q > 0$. For a quadratic distortion measure ($\gamma = 2$) the optimum is given by $q = 2(e-1)$, which results in an exponent of $\frac{\log e}{e}$. This should be compared with the best achievable exponent promised by the separation principle, which in this case is $\log e$.

IV Discussion

A simple and sequential JSCC scheme with feedback was presented for the case of transmitting a general memoryless source over a general memoryless channel with bandwidth expansion, under an arbitrary distortion measure and input constraint. The scheme coincides with a known optimal scheme in the Gaussian case, and was otherwise shown to yield an average distortion decaying exponentially fast with the BEF, under general conditions.

The results of [8] imply that in many cases optimal performance cannot be achieved by a (finite dimensional) JSCC scheme, hence by the PM-JSCC scheme in particular. However, it would be interesting to examine the behavior of the scheme in the special cases where optimality can nevertheless be achieved. Moreover, a cleaner analysis doing without the superfluous assumptions of a bounded distortion and possibly the Lipschitz behavior, should be pursued. Finally, the robustness of the PM-JSCC scheme to channel variations should be further explored. A prominent merit of the scheme in the Gaussian case is that of *graceful degradation*, i.e., optimal performance is still guaranteed even if the channel SNR deteriorates, assuming the receiver is aware of that. It is unlikely that the performance of the PM-JSCC scheme will not be affected by channel variations in the general case, even if the receiver is assumed to have full channel state information. Yet, it seems plausible that an exponential decay of the distortion with the BEF could still be obtained.

V Acknowledgments

The authors would like to thank Bixio Rimoldi for drawing their attention to the discussion of feedback in [8], and the reviewers for their helpful comments.

References

[1] T.M. Cover and J.A Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[2] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Transactions on Informormation Theory*, vol. 49, pp. 1147– 1158, May 2003.

[3] A. Ingber, I. Leibowitz, R. Zamir, and M. Feder, "Distortion lower bounds for finite dimensional joint source-channel coding," in *Proc. of the International Symposium on Information Theory*, July 2008, pp. 1183–1187.

[4] P. Elias, "Channel capacity without coding," *MIT Research Laboratory of Electronics, Quarterly Progress Report*, pp. 90–93, Oct. 1956.

[5] J. Schalkwijk and L. Bluestein, "Transmission of analog waveforms through channels with feedback (corresp.)," *IEEE Transactions on Information Theory*, vol. 13, no. 4, pp. 617–619, Oct 1967.

[6] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback part I: No bandwidth constraint," *IEEE Trans. Info. Theory*, vol. IT-12, pp. 172 – 182, 1966.

[7] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback part II: Band-limited signals," *IEEE Trans. Info. Theory*, vol. IT-12, pp. 183 – 189, 1966.

[8] M. Gastpar, *To Code or not to Code*, doctoral dissertation, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland, 2002.

[9] O. Shayevitz and M. Feder, "The posterior matching feedback scheme: Capacity achieving and error analysis," in *Proc. of the International Symposium on Information Theory, ISIT 2008.*

[10] O. Shayevitz and M. Feder, "The posterior matching scheme for memoryless channels with feedback," IEEE Trans. Info. Theory, to be submitted.