The Posterior Matching Approach in Feedback Communication

Ofer Shayevitz

Information Theory and Applications Center

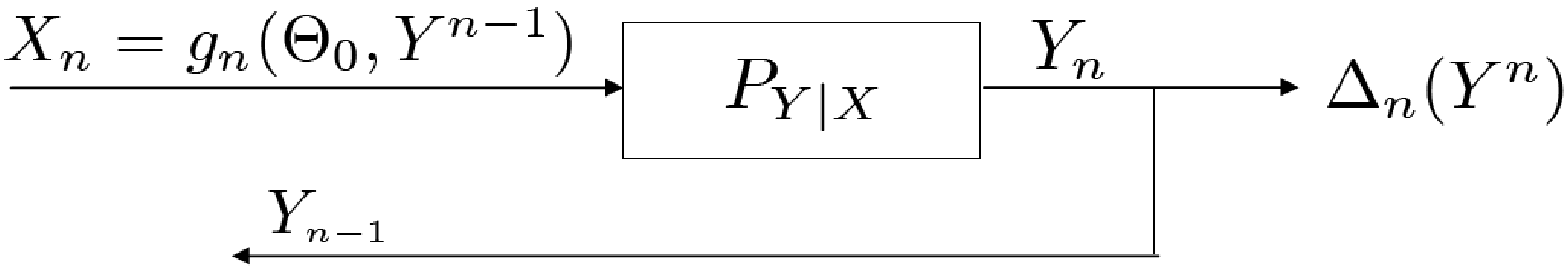
University of California, San Diego

UIUC, May 2009

Joint work with Meir Feder, Tel Aviv University

Preliminaries

Feedback Communication Setting



- Memoryless channel $P_{Y|X}$
- Instantaneous noiseless feedback
- Message point representation $\Theta_0 \sim \mathrm{Unif}(0,1)$
- A general *transmission scheme*:
  - *Transmission functions* $\{g_n : (0,1) \times \mathbb{R}^{n-1} \mapsto \mathbb{R}\}_{n=1}^{\infty}$
  - *Decoding rules* $\{\Delta_n : \mathbb{R}^n \mapsto \{(a,b) | (a,b) \subseteq (0,1)\}\}_{n=1}^{\infty}$

Definitions

- *Error probability* $p_e(n) = \mathbb{P}(\Theta_0 \notin \Delta_n(Y^n))$
- *Instantaneous rate* $R_n = -\frac{1}{n}\log|\Delta_n(Y^n)|$
- A transmission scheme *achieves* a rate R (possibly within input constraints (η, μ)) if

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

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \eta(X_k) < \mu \quad \text{a.s.}$$

- Implies achievability in the “standard sense”

Optimal Decoding Rules

- Calculate the posterior distribution $P_{\Theta_0|Y^n}$
- *Fixed rate*
  - Set the desired rate $R_n = R$
  - Select $|\Delta_n| = 2^{-nR}$ to maximize $P_{\Theta_0|Y^n}(\Delta_n|Y^n)$
- *Variable rate*
  - Set a target error probability $p_e(n) = \varepsilon_n \to 0$
  - Select Δ_n with minimal size such that $P_{\Theta_0|Y^n}(\Delta_n|Y^n) \geq 1 - \varepsilon_n$

Background

Feedback – What is it Good for?

- Feedback cannot increase capacity of memoryless channels

  [Shannon'56] [Kadota&Ziv'71]

- Nevertheless, feedback can sometimes
  - Boost reliability
  - Allow rate adaptation to cope with unknown channels
  - Significantly reduce complexity, attain capacity "without coding"

Feedback – What is it Good for?

- Feedback cannot increase capacity of memoryless channels

  [Shannon'56] [Kadota&Ziv'71]

- Nevertheless, feedback can sometimes
  - Boost reliability
  - Allow rate adaptation to cope with unknown channels
  - *Significantly reduce complexity, attain capacity "without coding"*

Previous Work

- Several “no coding” feedback schemes have been suggested for specific channels
- *The Horstein Scheme* (1963)
  - Conjectured to achieve the BSC capacity $1 - h_b(p)$
  - Send $0/1$ according to whether Θ_0 lies to the left/right of the posterior’s median
- *The Schalkwijk-Kailath (SK) Scheme* (1966)
  - Achieves the AWGN capacity $C = \frac{1}{2}\log(1 + \mathrm{SNR})$
  - Send Θ_0, receive Θ_0 with a Gaussian bias Z
  - Find MMSE estimate for Z, send amplified error term
  - Repeat last step

Motivation for this Work

- Horstein/SK share many similarities!
  - Message point representation
  - Simple, sequential, “no coding” schemes
  - “Steering” the receiver in the right direction
- *However...*
  - Precise correspondence never established
  - No generalization ever provided
- Can feedback facilitate achieving capacity without coding in general?

Main Results

[Shayevitz&Feder '07,'08,'09]

- Formalize the common underlying *posterior matching principle*
- Devise a generic feedback scheme
  - Suitable for any memoryless channel $P_{Y|X}$ and any desired input distribution P_X
  - Achieves any rate $R < I(X;Y)$ under general conditions
  - Simple, sequential, no coding
  - Horstein & SK are special cases
- Corollary: The Horstein scheme achieves the BSC capacity (verifying a longstanding conjecture)
- Error exponent and model mismatch analysis
- Applications to joint source-channel coding

Posterior Matching

The Basic Principle

- Say the receiver has observed the output sequence Y^n
- What information is it still missing?
- A reasonable answer: any r.v. which
  - is statistically independent of previous outputs Y^n
  - Together with Y^n uniquely determines Θ_0
- However...
  - Many possible distributions
  - Channel input may have constraints (e.g., power, discreteness)
- *Match* the distribution to the channel!

The Posterior Matching (PM) Scheme

- Set some input distribution P_X
- The next channel input is given by

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

- $F_X, F_{\Theta_0|Y^n}$ are c.d.f.'s
- $X_{n+1} \sim P_X$ and independent of Y^n
- Y^n is i.i.d. with the "correct" marginal
- A two step procedure
  - Zoom in on the posterior
  - Match to the channel

The Posterior Matching (PM) Scheme

- Set some input distribution P_X
- The next channel input is given by

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

- $F_X, F_{\Theta_0|Y^n}$ are c.d.f.'s
- $X_{n+1} \sim P_X$ and independent of Y^n
- Y^n is i.i.d. with the "correct" marginal
- A two step procedure
  - Zoom in on the posterior
  - Match to the channel

Recursive Representation

Lemma: If P_{XY} has a p.d.f., then the transmission scheme is given by

$$X_1 = F_X^{-1}(\Theta_0)$$

$$X_{n+1} = F_X^{-1} \circ F_{X|Y}(X_n|Y_n)$$

- $F_X^{-1} \circ F_{X|Y}(\cdot|\cdot)$ is called the *PM kernel*
- The next input is generated by applying the PM kernel to the last input/output pair only – simple!
- (X_n, Y_n) constitutes a Markov chain over $\mathbb{R}^2$
- By construction, P_{XY} is an invariant distribution

The AWGN Channel

- Let $P_{Y|X}$ be an AWGN channel with an input power constraint P
- Set $P_X = \mathcal{N}(0, \mathrm{P})$ (capacity achieving)
- PM kernel is linear, and yields

$$X_{n+1} = \sqrt{1+\mathrm{SNR}}\left(X_n - \frac{\mathrm{SNR}}{1+\mathrm{SNR}} Y_n\right)$$

- Precisely the SK scheme!
- MMSE error term – *uncorrelated* with previous outputs
- Mind the difference – In the PM scheme transmission is *independent* of previous outputs, coincides only in the Gaussian case

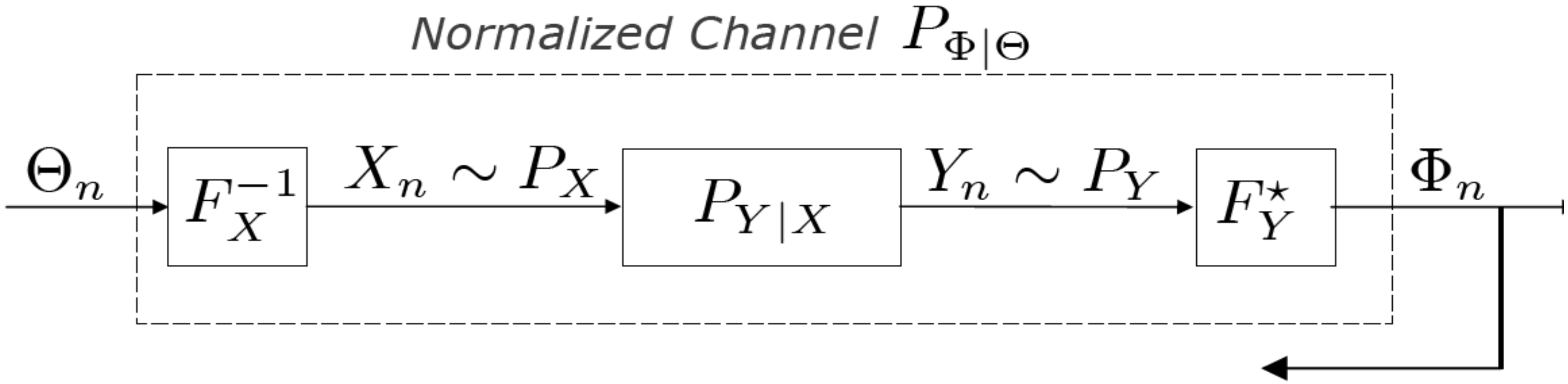
The BSC

- Set $P_X = \mathrm{Bern}\left(\frac{1}{2}\right)$ (capacity achieving)
- P_X has no proper p.d.f. – recursion rule is invalid!
- Nevertheless, the PM scheme coincides with Horstein's median rule

$$X_{n+1} = F_X^{-1} \circ F_{\Theta_0|Y^n}(\Theta_0|Y^n) = \begin{cases} 0 & \Theta_0 < \mathrm{median}\left\{f_{\Theta_0|Y^n}(\circ|Y^n)\right\} \\ 1 & \text{o.w.} \end{cases}$$

- F_X^{-1} quantizes above/below $\frac{1}{2}$
- Can we find a simple recursion rule nonetheless?

Normalized Recursive Representation



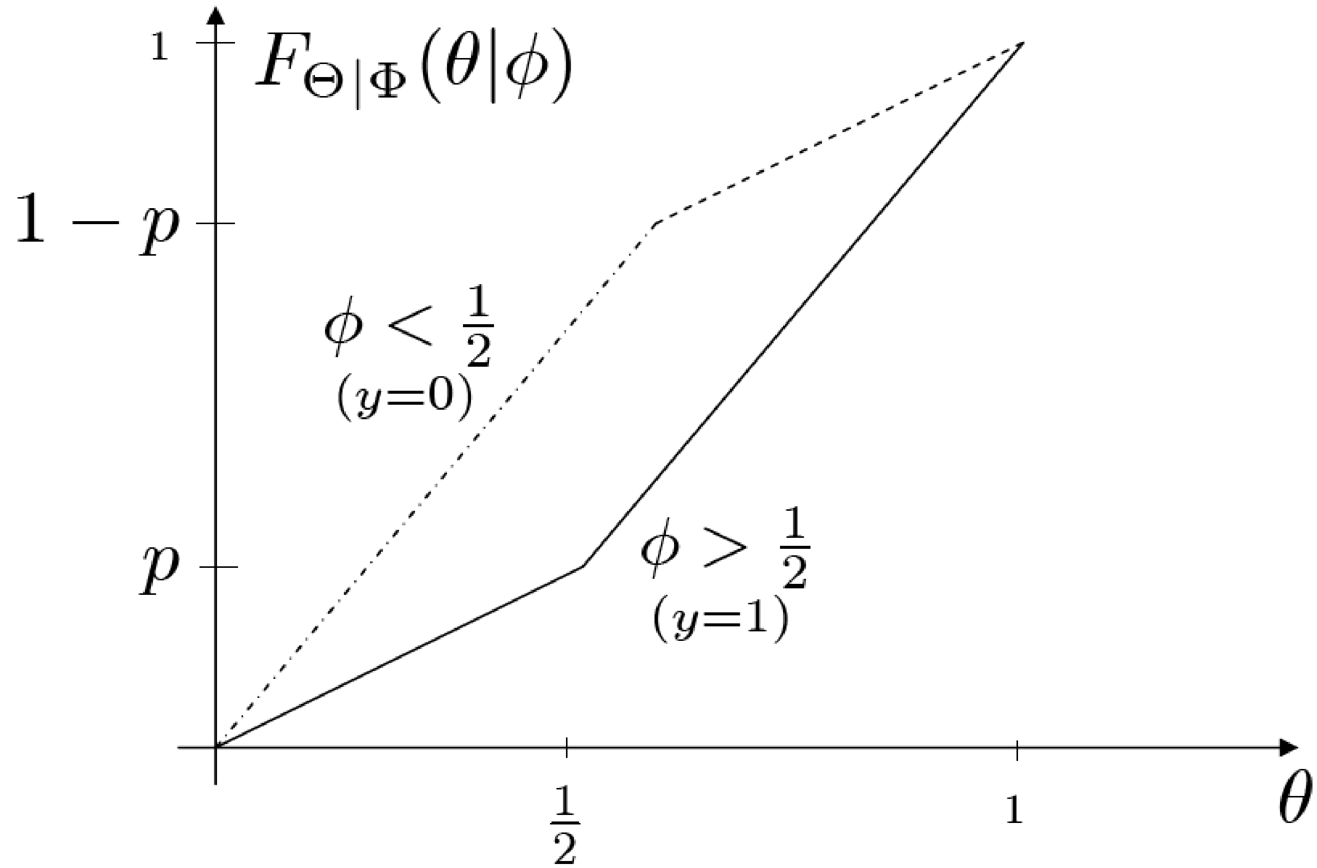
- A “message point” channel, $\Theta_n, \Phi_n \sim \mathrm{Unif}(0,1)$
- Preserves mutual information, common framework for discrete/continuous/mixed alphabet input distributions/channels
- PM scheme recursive representation

$$\Theta_1 = \Theta_0\,, \quad \Theta_{n+1} = F_{\Theta|\Phi}(\Theta_n|\Phi_n)$$

- (Θ_n, Φ_n) constitute a Markov chain over $(0,1)^2$, having $P_{\Theta\Phi}$ as an invariant distribution

The Normalized BSC

- The PM kernel for the normalized BSC is given by



- Equivalent to the original Horstein scheme (set $X_n = F_X^{-1}(\Theta_n)$)

Exponential Noise with a Mean Input Constraint

- Let $P_{Y|X}$ be an additive noise $\sim \mathrm{Exp}(b)$ channel
- Impose a mean input constraint $\mathbb{E}(X_n) = a$
- Capacity achieving distribution [Verdú'96]

$$f_X(x) = \frac{b}{a+b}\,\delta(x) + \frac{a}{(a+b)^2}\, e^{-\frac{x}{a+b}}\,, \quad C = \log\left(1 + \frac{a}{b}\right)$$

- Corresponding (normalized) PM scheme given by

$$\Theta_{n+1} = \begin{cases} \frac{a+b}{b} \cdot \Theta_n \cdot (1 - \Phi_n)^{\frac{a}{b}} & \Theta_n \leq \frac{b}{a+b} \\ \left(\frac{a}{a+b} \cdot \frac{1-\Phi_n}{1-\Theta_n}\right)^{\frac{a}{b}} & \Theta_n > \frac{b}{a+b} \end{cases}$$

- The actual input is $X_n = (a+b)\ln\left(\frac{a}{(a+b)(1-\Theta_n)}\right) \mathbb{1}_{[\frac{b}{a+b},1)}(\Theta_n)$

Achieving the Mutual Information

Theorem: Under some general conditions on the input/channel pair $(P_X, P_{Y|X})$, the corresponding PM scheme achieves any rate $R < I(X;Y)$, with either a fixed/variable optimal decoding rule, within an input constraint $(\eta, \mathbb{E}\eta(X))$

- Specifically, the Theorem holds for
  - Discrete memoryless channels (up to some small issues)
    - Corollary: The Horstein scheme achieves capacity
  - Input/channel pairs $(P_X, P_{Y|X})$ with a joint p.d.f. P_{XY} continuous over a convex support, and not "too wild"

Proof Outline

Achievability Conditions

- (REG) $I(X;Y) < \infty$, and some mild regularity conditions
- (ERG) The invariant distribution $P_{\Theta\Phi}$ for the Markov chain (Θ_n, Φ_n) is ergodic.
- (FIX) The (normalized) posterior matching kernel has no universal fixed points, in the sense that

$$\mathbb{P}(F_{\Theta|\Phi}(\theta|\Phi) = \theta) \; < \; 1$$

for any $\theta \in (0,1)$

Iterated Function System (IFS)

- An IFS $\{S_n(s_0)\}_{n=1}^{\infty}$ over a measurable space $\mathfrak{F}$ is generated by a measurable function $\omega_\phi(s)$ mapping $\mathfrak{F}$ to itself, as follows:

$$S_1 = s_0\,, \qquad S_{n+1}(s_0) = \omega_{\Phi_n} \circ \omega_{\Phi_{n-1}} \circ \cdots \circ \omega_{\Phi_1}(s)$$

  where $\{\Phi_n\}_{n=1}^{\infty}$ is an i.i.d control sequence, and $s_0 \in \mathfrak{F}$ is the initial point

- $\{S_n(s_0)\}_{n=1}^{\infty}$ is a Markov chain over $\mathfrak{F}$
- Let $\lambda : \mathfrak{F} \mapsto \mathbb{R}^+$, let $\xi : [0,1) \mapsto [0,1)$ be $\cap$-convex and $\xi(x) < x$ over $(0,1)$ (a *contraction*)
- Lemma: If $\mathbb{E}(\lambda(\omega_{\Phi_1}(s))) \leq \xi(\lambda(s))$ for any $s \in \mathfrak{F}$, then $\lambda(S_n(s_0)) \to 0$ in probability

Receiver Tracking Process as an IFS

- The receiver tracks the posterior $P_{\Theta_0|\Phi^n}$
- A recursive representation for the posterior c.d.f.

$$F_{\Theta_0|\Phi^{n+1}}(\theta|\Phi^{n+1}) = F_{\Theta|\Phi}(\cdot|\Phi_{n+1}) \circ F_{\Theta_0|\Phi^{n+1}}(\theta|\Phi^n)$$

- Recall $\{\Phi_n\}_{n=1}^{\infty}$ is i.i.d
- Thus the posterior c.d.f. is an IFS
  - Evolves over a function space $\mathfrak{F}_c$ of c.d.f.-like functions
  - Generated by the PM kernel (via function composition)
  - Initialized at $F_{\Theta_0}(\theta) = \theta$
  - Controlled by the channel output sequence $\{\Phi_n\}_{n=1}^{\infty}$

Contraction of the Posterior

- Under assumptions (REG)+(FIX), we can find a length function $\lambda : \mathfrak{F}_c \mapsto \mathbb{R}^+$ and a contraction $\xi(\cdot)$ such that
  - $\mathbb{E}\lambda(F_{\Theta|\Phi}(\cdot|\Phi) \circ s) \le \xi(\lambda(s)) \,, \quad s \in \mathfrak{F}_c$
  - Loosely speaking, $\lambda(s) \approx 0$ implies that s “close” a unit step function
- Therefore,
  - The posterior c.d.f. “tends” to a unit step function about Θ_0
  - Any fixed interval containing Θ_0 will be reliably decoded eventually
  - $R = 0$ is achievable!

Achievability of $R < I(X;Y)$

- Expand the posterior p.d.f. at the message point (Bayes rule, memoryless channel, i.i.d. output)

$$\frac{f_{\Theta_0|\Phi^n}(\Theta_0|\Phi^n)}{f_{\Theta_0|\Phi^{n-1}}(\Theta_0|\Phi^{n-1})} = \frac{f_{\Phi_n|\Theta_0,\Phi^{n-1}}(\Phi_n|\Theta_0,\Phi^{n-1})}{f_{\Phi_n|\Phi^{n-1}}(\Phi_n|\Phi^{n-1})} = f_{\Phi|\Theta}(\Phi_n|\Theta_n)$$

- Taking the logarithm, applying the recursion n times and using (ERG)+SLLN, we get

$$\frac{1}{n}\log f_{\Theta_0|\Phi^n}(\Theta_0|\Phi^n) = \frac{1}{n}\sum_{k=1}^{n}\log f_{\Phi|\Theta}(\Phi_k|\Theta_k) \to \mathbb{E}\left(\log f_{\Phi|\Theta}(\Phi|\Theta)\right)$$

$$= I(\Theta;\Phi) = I(X;Y) \qquad \text{a.s.}$$

- Roughly, $f_{\Theta_0|\Phi^n}(\Theta_0|\Phi^n) \approx 2^{nI(X;Y)}$

Achievability of $R < I(X;Y)$

- Assume that for all $k \in [n]$

$$F_{\Theta_0|\Phi^n}(\Theta_0 + 2^{-nR}|\Phi^n) - F_{\Theta_0|\Phi^n}(\Theta_0 - 2^{-nR}|\Phi^n) < \varepsilon$$

- $F_{\Theta_0|\Phi^n}(\theta|\Phi^k)$ is what the input to the normalized channel at time $k+1$ would have been, had the message point been θ
- Hence, the assumption implies that the input is insensitive to a 2^{-nR} perturbation in the message point
- Using (REG)+(ERG)+SLLN again, this can be roughly translated into

$$f_{\Theta_0|\Phi^n}(\Theta_0 \pm 2^{-nR}|\Phi^n) \approx 2^{n(I(X;Y)-\delta)}$$

where $\varepsilon \to 0$ implies $\delta \to 0$

Achievability of $R < I(X;Y)$

- If $R < I(X;Y) - \delta$ we get a contradiction

$$\int f_{\Theta_0|\Phi^n}(\theta|\Phi^n)d\theta \approx 2^{n(I(X;Y)-R-\delta)} \longrightarrow \infty$$

- Hence, with high probability there exists $k_0 \in [n]$ so that

$$F_{\Theta_0|\Phi^n}(\Theta_0 + 2^{-nR}|\Phi^n) - F_{\Theta_0|\Phi^n}(\Theta_0 - 2^{-nR}|\Phi^n) \geq \varepsilon$$

- Due to the repetitive nature of the scheme, one can imagine transmission to have started at time k_0 with the message point Θ_{k_0}
- 2^{-nR} neighborhood of $\Theta_0 \Rightarrow \varepsilon$-neighborhood of Θ_{k_0}
- Invoking zero rate result, this ε-neighborhood can be decoded in sublinear time $\rightarrow$ Achievability proved!

Examples Revisited

- BSC – Verifying the Horstein scheme achieves capacity
- AWGN channel – reconfirming the SK achieves capacity
- Exponential noise, mean constraint – The explicit PM scheme described achieves capacity!

Further Results

- Error probability analysis, providing closed form error exponent expressions for a range of rates (sometimes strictly below capacity)
- Channel model mismatch
  - Scheme designed for $(P_X, P_{Y|X})$
  - True channel is $P_{Y^*|X^*}$
  - Scheme induces some stationary input distribution P_{X^*}
  - Penalty in rate relative to $I(X^*;Y^*)$ is

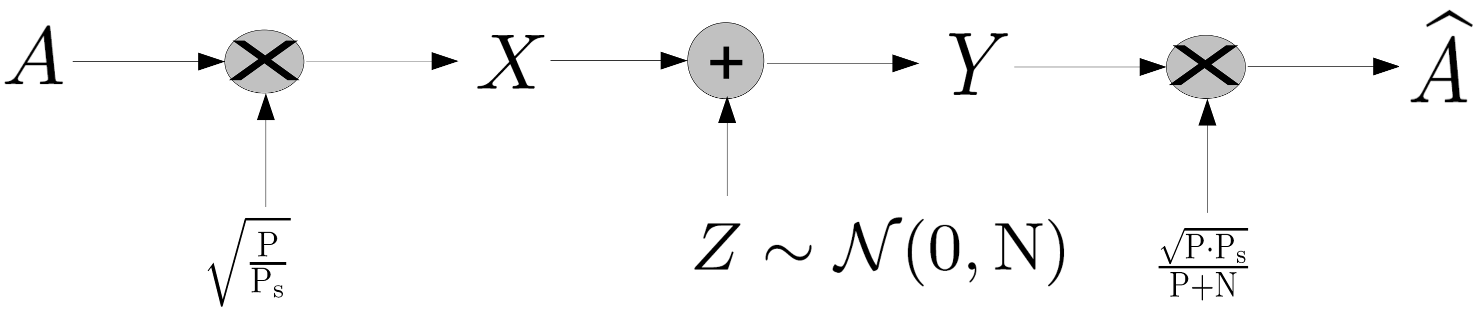
$$D(P_{Y^*|X^*} \| P_{Y|X} \mid P_{X^*}) - D(P_{Y^*} \| P_Y)$$

  - Robustness of the SK scheme

Application to Joint-Source Channel Coding (JSCC) with Feedback

A Well Known Gaussian Example

- Gaussian source $A \sim \mathcal{N}(0, \mathrm{P_s})$
- AWGN channel, input power constraint P
- Scalar linear transmission scheme ("uncoded"):



- Achieves optimal performance under quadratic distortion!

$$D = \mathbb{E}(A - \widehat{A})^2 = \mathrm{P_s} \cdot \left(1 + \frac{\mathrm{P}}{\mathrm{N}}\right)^{-1} \quad \Rightarrow \quad \mathrm{R(D) = C}$$

AWGN with Bandwidth Expansion

- Suppose m AWGN channels uses per source sample available
- Optimal distortion given by

$$R(D) = mC \quad \Rightarrow \quad D = \mathrm{P_s} \cdot \left(1 + \frac{\mathrm{P}}{\mathrm{N}}\right)^{-\mathrm{m}}$$

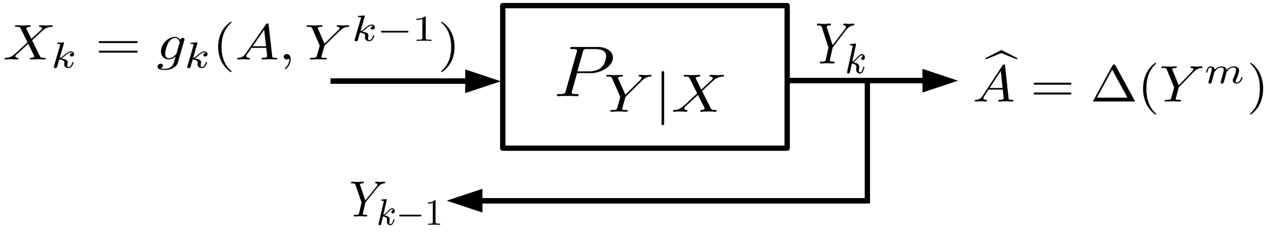
- Decays exponentially with the *bandwidth expansion factor* (BEF) m
- Optimal performance cannot generally be attained by a scalar $1:m$ *joint source-channel coding* (JSCC) scheme
- Assume an instantaneous noiseless feedback link is available
- The SK scheme achieves optimal performance!

What About Non Gaussian Settings?

- For a finite BEF, can optimal performance be achieved via feedback for other sources/channels/distortion measures?
- In general no, unless some unique relation is satisfied [Gastpar '02]
- So what can we do anyway?
  - Apply the PM principle
  - Show that the resulting PM-JSCC scheme has "good performance"

Problem Setting

- Source $A \sim P_A$ over an alphabet $\mathcal{A} \subseteq \mathbb{R}$
- Memoryless channel $P_{Y|X}$ with feedback, BEF $= m$
- A general $1 : m$ JSCC transmission scheme:



- General (η, u) input constraint: $\mathbb{E}(\eta(X_k)) \leq u$ for $k = 1, \ldots, m$
- General distortion measure $d : \mathcal{A}^2 \mapsto \mathbb{R}^+$
- Performance measured by the average distortion $D = \mathbb{E}(d(A, \widehat{A}))$

The PM-JSCC Scheme

- Set some channel input distribution P_X (design parameter)
- The transmission functions are defined for $k = 0, 1, \ldots m-1$ by

$$X_{k+1} = F_X^{-1} \circ F_{A|Y^k}(A|Y^k)$$

- Again, $X_{k+1} \sim P_X$ independent of Y^k, Y^m is i.i.d. with marginal P_Y, and we have a recursive representation:

$$X_1 = F_X^{-1} \circ F_A(A)\,, \quad X_{k+1} = F_X^{-1} \circ F_{X|Y}(X_k|Y_k)$$

- Seems to satisfy Gastpar's optimality conditions when possible
- Optimal exponential decay for a quadratic distortion measure
- What happens otherwise (which is usually the case..)?

Distortion Analysis Outline

- Suppose $\widehat{X}_m$ is some estimate of X_m
- Corresponds to a unique estimate $\widehat{X}_1$ of X_1, given by reversing the transmission scheme

$$\widehat{X}_1 = \omega_{Y_1} \circ \cdot \circ \omega_{Y_{m-2}} \circ \omega_{Y_{m-1}}(\widehat{X}_m)$$

where $\omega_y(\cdot) = F^{-1}_{X|Y}(\cdot|y) \circ F_X(\cdot)$ is the *inverse PM kernel*

- $\widehat{X}_1$ is generated by a time-reversed IFS (RIFS) with kernel $\omega_y(\cdot)$ and an i.i.d control sequence Y^m
- Now simply $\widehat{A} = F_A^{-1} \circ F_X(\widehat{X}_1)$

Distortion Analysis Outline (cont.)

- Set a fixed interval $J_m \subseteq \mathcal{X}$ so that $\mathbb{P}(X_m \in J_m) = P_X(J_m) = 1 - \delta$
  - Thus $\mathbb{P}(X_1 \in J_1) = 1 - \delta$, where the random interval J_1 is generated by running the RIFS over (the edges of) J_m
  - $\mathbb{P}(A \in J_A) = 1 - \delta$, where $J_A = F_A^{-1} \circ F_X(J_1)$
  - Now set $\widehat{A}$ to be any point within J_A (suboptimal)
- If the RIFS kernel $\omega_y(\cdot)$ is *contractive on the average*, then J_1 is exponentially smaller than J_m w.h.p.
- If $F_A^{-1} \circ F_X$ is M-Lipschitz, J_A is also exponentially small w.h.p.
- Two sources of distortion
  - $A \in J_A$ and J_A exponentially small (high prob.) $\Rightarrow$ small distortion
  - $A \notin J_A$ or J_A large (low prob.) – $\sup_{a,b \in \mathcal{A}} d(a,b) = d_{\max} < \infty$

Distortion Analysis Outline (cont.)

Theorem: Let $\omega_y(\cdot)$ be the inverse PM kernel and define

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

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

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

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

Corollary: If P_X has a polynomially (or faster) decaying tail and $d(a,b) = |a-b|^\gamma$, the distortion decays exponentially with the BEF m.

Uniform Source, Uniform Noise

- $A \sim \mathrm{Unif}(0,1)$, $P_{Y|X}$ is an additive channel with noise $\sim \mathrm{Unif}(0,1)$
- Set (for example) $P_X = \mathrm{Unif}(0,1)$
- The PM-JSCC scheme is given by

$$X_1 = A\,, \qquad X_{k+1} = \frac{X_k}{Y_k} \cdot \mathbb{1}_{(0,1]}(Y_k) + \frac{X_k - Y_k + 1}{2 - Y_k} \cdot \mathbb{1}_{(1,2)}(Y_k)$$

- Very simple interpretation:
  - Begin by transmitting the uncoded source $X_1 = A$
  - Given Y_1 find the interval of feasible inputs, and an affine transformation that stretches this interval to $(0,1)$
  - Generate X_2 by applying the transformation to X_1
  - Repeat the two steps above for X_k, Y_k

Uniform Source, Uniform Noise (cont.)

- The inverse PM kernel is

$$\omega_y(s) = s f_Y(y) + (y-1)\cdot \mathbb{1}_{(1,2)}(y)$$

- The contraction factor is given by

$$r_q = \sup_s \mathbb{E}\left(\frac{\partial}{\partial s}\omega_Y(s)\right)^q = \mathbb{E}(f_Y(Y))^q = \left(1+\frac{q}{2}\right)^{-1} < 1$$

- Assume a distortion measure $d(a,b) = |a-b|^\gamma$, bounded over $(0,1)$
- The conditions of the Theorem hold, and we get

$$D \leq \inf_{q>0}\left(\left(\frac{\gamma}{q}\right)^{\frac{q}{q+\gamma}} + \left(\frac{q}{\gamma}\right)^{\frac{\gamma}{q+\gamma}}\right)\cdot\left(\frac{1}{1+\frac{q}{2}}\right)^{\frac{m\gamma}{q+\gamma}}$$

Uniform Source, Uniform Noise (cont.)

- The distortion decays exponentially with the BEF:

$$\lim_{m\to\infty} -\frac{1}{m}\log D \;\geq\; \sup_{q>0} \frac{\gamma}{q+\gamma}\log\left(1+\frac{q}{2}\right)$$

- For a quadratic distortion measure ($\gamma = 2$) the exponent attained by the PM-JSCC scheme is lower bounded by $\frac{\log e}{e}$
- This should be contrasted with the exponent promised by the separation principle, which is $\log e$

Further Research

- JSCC setting:
  - Obtain tighter bounds by “measuring” contraction “relative” to the distortion measure
  - Finite block optimality of the PM-JSCC scheme?
  - Sensitivity to channel variations (graceful degradation)
  - relation to non-feedback JSCC schemes?
- Communication setting:
  - Channels with memory, achieve the *directed information rate* – requires a different interpretation of the PM principle
  - Multi-terminal channels
  - Noisy feedback