The Role of Feedback in Communications

Solmaz Torabi

Dept. of Electrical and Computer Engineering
Drexel University
st669@drexel.edu

Advisor: Dr. John M. Walsh

References I

M. Horstein, "Sequential transmission using noiseless feedback," *Information Theory, IEEE Transactions on*, vol. 9, no. 3, pp. 136–143, 1963.

J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–i: No bandwidth constraint," *Information Theory, IEEE Transactions on*, vol. 12, no. 2, pp. 172–182, 1966.

O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1186–1222, 2011.

N. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.)," *Information Theory, IEEE Transactions on*, vol. 21, no. 1, pp. 100–102, 1975.

G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, University of Manitoba, Canada, 1998.

References II

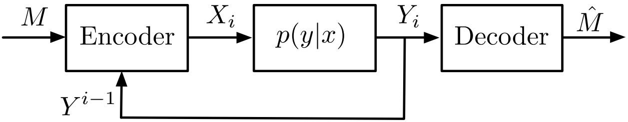
A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

R. Venkataramanan and S. S. Pradhan, "Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source," *Information Theory, IEEE Transactions on*, vol. 53, no. 6, pp. 2154–2179, 2007.

Outline

- Introduction
- Point to point communication
  - Horstein coding Scheme
  - Block Feedback Coding Scheme for BSC
  - Schalkwijk-Kailath Coding scheme
  - Posterior matching scheme
- Multiuser Channel
  - Multiple access channel
  - Two way channel
- Source coding with feedforward

DMC with feedback

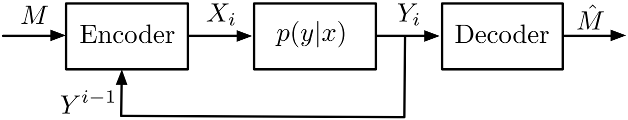


- Memoryless channel: $P_{Y_n|X^n,Y^{n-1}}(.|x^n, y^{n-1}) = P(.|x_n)$
- Messages: M
- Encoding function

$$g_i : \mathcal{M} \times \mathcal{Y}^{i-1} \rightarrow \mathcal{X}$$

- X_n is a function of $(M, Y_1, Y_2, ..., Y_{n-1})$

Point to point feedback communication system

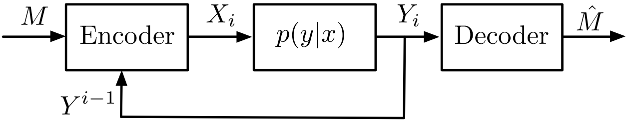


channel is memoryless if $P(y_n|x^n, y^{n-1}) = P(y_n|x_n)$

channel is used without feedback $P(x_n|x^{n-1}, y^{n-1}) = P(x_n|x^{n-1})$

DMC without feedback $P(y^N|x^N) = \prod_{n=1}^{N} P(y_n|x_n)$

Point to point feedback communication system



- If the channel is memory less, there is no information you get from the feedback that can help you increase your rate

$$C_{FB} = \max_{p(x)} I(X;Y) = C$$

Capacity of Memoryless Feedback Channel

- Shannon 56: Feedback does not increase capacity of memoryless channel
- Simplifiying schemes for attaining it:
- Horstein 63:Developed a recursive coding strategy for the BSC with noiseless feedback(sequential coding acheme, varying block length)
- Schalkwijk-Kailath 66: AWGN channel
- Shayevtitz 2008: Extends the SK and Horstein schemes to general memoryless channels
- Gaarder-Wolf 75: Feedback enlarge the capacity region of multiuser channels

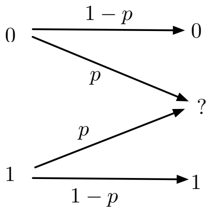
Point to point feedback communication system

Feedback can:

- Simplify coding scheme
- Improve reliability (decreases error prob. much faster)
- Increase capacity channels with memory
- Enlarge capacity region of multiuser channels (Gaarder-Wolf 1975)

Iterative refinement for BEC

- First send a message at a rate higher than the channel capacity (without coding)
- Then iteratively refine the receiver's knowledge about the message



$$n + pn + p^2 n + ... = \frac{n}{1-p}$$

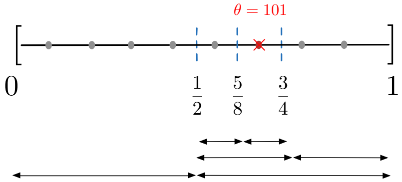
suffices to transmit n bits reliably

- We can achieve the capacity $C = 1 - p$ by simply retransmitting each bit after it is erased.
- There is no need for sophisticated error correcting codes.

noiseless binary forward channel

binary search algorithm would provide an effective procedure for transmitting the information involved in the source's choice.

- Receiver starts out with a uniform prior distribution for the message point selected
- The a priori median of the receiver distribution is $m_0 = 1/2$
- Suppose 1 was sent. Hence, the new receiver distribution is uniform over the interval $(1/2, 1)$.



Outline

- Introduction
- Point to point communication
  - Horstein coding Scheme
  - Block Feedback Coding Scheme for BSC
  - Schalkwijk-Kailath Coding scheme
  - Posterior matching scheme
- Multiuser Channel
  - Multiple access channel
  - Two way channel
- Source coding with feedforward

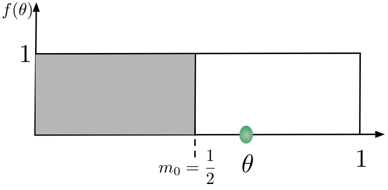
Horstein Scheme

- Horstein developed a recursive coding strategy for the BSC with noiseless feedback
- We could call Horstein's feedback scheme to binary search algorithm with lies.
- The transmitter transmits a 0 on the $(i+1)$st transmission when the true message point θ is to the left of the current median m_i, a 1 otherwise.
- However, since crossovers can occur in the channel, the transmitter does not know what the receiver's current median m_i is.
- A noiseless feedback channel is used to provide the transmitter with this information.

Horstein Scheme

- Divide the interval $[0, 1]$ into 2^{nR} equidistance subinterval.
- Represent each message by the midpoint of each interval.
- receiver has no prior knowledge of the location of θ receiver density os initially uniform. $f_0(\theta) = 1$ for $\theta \in [0, 1]$

$$x_1 = g(\theta_0) = \begin{cases} 1 & \text{if } \theta_0 \text{ is greater than } 1/2 \\ 0 & \text{o.w} \end{cases}$$



Horstein Scheme

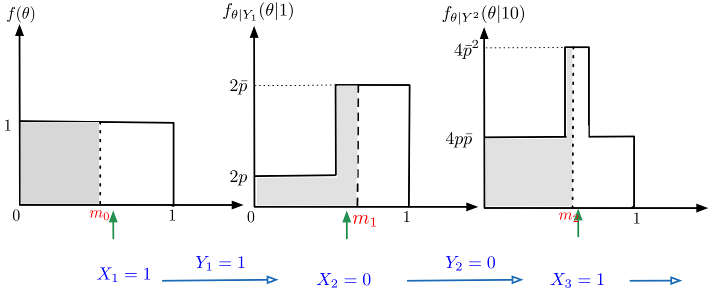
- Assume $x_1 = 1$ is sent through the channel. It gets corrupted with probability p.
- After the channel output is observed, the receiver distribution and median is updated.
- Through the noiseless feedback, the encoder learns the distribution.

$$f(\theta|y_1) = \frac{f(\theta)p(y_1|\theta)}{\int_0^1 f(\theta)p(y_1|\theta)d\theta}$$

- Assume $y_1 = 1$ is received.
  - For $0 \leq \theta \leq 1/2$, then $f(\theta|y_1 = 1) = 2p$
  - For $1/2 \leq \theta \leq 1$, then $f(\theta|y_1 = 1) = 2\bar{p}$

Horstein Scheme

- The encoder transmits 1, if $\theta >$ median of $f(\theta|y_1)$,and 0 otherwise.



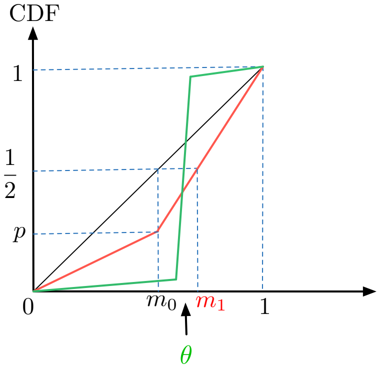
- Terminates when most of the probability mass is concentrated in the neighborhood of one of he possible message points

Horstein Scheme

The schemes can admit a simple recursive structure

$$f(\theta|y^{i-1}) = \begin{cases} 2pf(\theta|y^{i-2}) & \text{if } y_{i-1} \text{ is greater than median of } f(\theta|y^{i-1}) \\ 2\bar{p}f(\theta|y^{i-2}) & \text{o.w} \end{cases}$$

Terminates when the receiver distribution is sufficiently steep.



Horstein Scheme- Decoding

- The decoder uses maximal posterior decoding
- It finds the interval of length 2^{-nR} that maximizes

$$\int_{\beta}^{\beta+2^{-nR}} f(\theta|y^n)\, d\theta$$

Horstein Scheme- Error probability

- By analysis of the evolution of $f(\Theta|Y^i), i \in [1:n]$, based on the iterated function system, it can be shown that

$$p(\theta \notin [\beta, \beta + 2^{-nR}]) \to 0 \quad \text{as } n \to \infty \text{ if } R < C$$

- With high probability, $\theta(M)$ is the unique message point within the $[\beta, \beta + 2^{-nR})$

Outline

Block Feedback Coding Scheme for BSC

Ahlswede 1973:

- Implement the iterative refinment at the block level
- The encoder initially transmits an uncoded block of information
- It then refines the receiver's knowledge about it in subsequent blocks

Block Feedback Coding Scheme for BSC

- Tx. 1: Sends N uncoded data bits over channel.
- Ch. 1: Adds (modulo-2) N samples of Bern(p) noise
- Rx. 1: Feeds its N noisy observations back to Tx.
- Tx.2:
  - (a) Finds N samples of noise added by channel.
  - (b) Compresses noise into $NH(p)$ new data bits.
  - (c) Sends these data bits uncoded over the channel
- Ch.2: Adds (modulo-2) $NH(p)$ samples of Bern(p) noise.
- Rx.2: Feeds its $NH(p)$ noisy observations back to T_x

Block Feedback Coding Scheme for BSC

the number of channel inputs used to send the N bits would be

$$N + NH(p) + NH(p)^2 + ... = \frac{N}{1 - H(p)}$$

which corresponds to a rate of $1 - H(p)$, the capacity of BSC(p)

Outline

- Introduction
- Point to point communication
  - Horstein coding Scheme
  - Block Feedback Coding Scheme for BSC
  - Schalkwijk-Kailath Coding scheme
  - Posterior matching scheme
- Multiuser Channel
  - Multiple access channel
  - Two way channel
- Source coding with feedforward

Robbins-Monro procedure

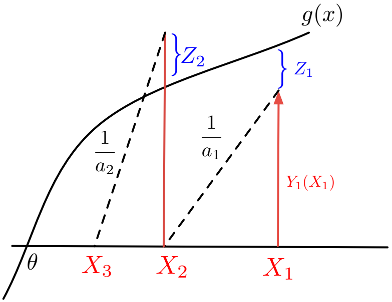
How to determine θ, a zero of a function $g(x)$, without knowing the shape of the function?

- The observations are noisy
- instead of $g(x)$, one obtains $Y(x) = g(x) + Z$

Robbins-Monro procedure

$$X_{n+1} = X_n - a_n Y_n(X_n), \quad n = 1, 2, ..$$

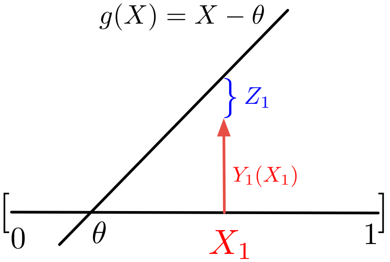
- where $\sum a_n = \infty$, and $\sum a_n^2 < \infty \Rightarrow X_n \to \theta$ Almost surely



Schalkwijk-Kailath scheme

put straight line $g(x) = x - \theta$

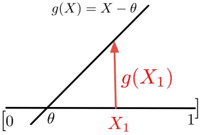
- Start with $X_1 = \frac{1}{2}$, send the receiver the number $g(X_1) = (X_1 - \theta)$
- The receiver obtains the number $Y_1(X_1) = (X_1 - \theta) + Z_1$, where $Z_1 \sim N(0, 1)$

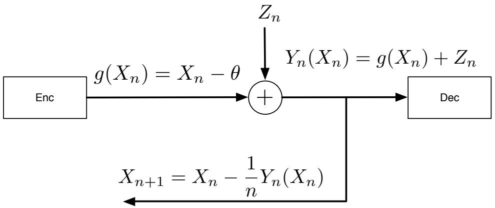


Schalkwijk-Kailath scheme

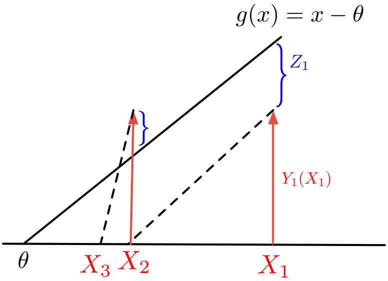
The recursion is easily solved to yield

$$X_{n+1} = \theta - \frac{1}{n}\sum_{1}^{n} Z_i \sim N(\theta, 1/n)$$





Schalkwijk-Kailath scheme



Schalkwijk-Kailath Coding

Another interpretation of SK scheme, with expected average transmitted power constraint

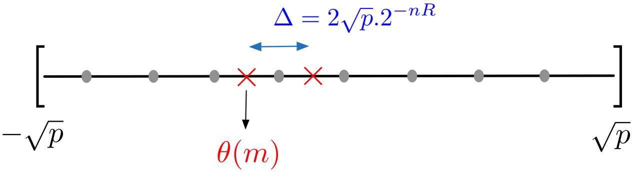
Schalkwijk-Kailath Coding

$$Y = X + Z, \quad Z \sim N(0,1)$$

- Expected average transmitted power constraint

$$\sum_{i=1}^{n} E(g_i^2(m, Y^{i-1})) \leq nP \quad m \in [1 : 2^{nR}]$$

- Divide the interval $[-\sqrt{p}, \sqrt{p}]$ into 2^{nR} message interval
- Represent each message m by the midpoint of its interval



Schalkwijk-Kailath Coding

- The transmitter first sends the message point it self:

$$X_0 = \theta(m)$$

- It is corrupted by additive Gaussian noise, so received with some bias

$$Y_0 = \theta(m) + Z_0$$

- The goal of the transmitter is to refine the receiver's knowledge of the bias
- It computes the MMSE estimate of the bias given the output sequence observed thus far, and sending the error term

Schalkwijk-Kailath Coding

- For $i = 1$, encoder learns $Z_0 = Y_0 - X_0$ and transmits

$$X_1 = \gamma_1 Z_0$$

  $\gamma_1 = \sqrt{p}$ is chosen so that $E(X_1^2) = P$

- Sends the Gaussian random variable Z_0 to the receiver, thus reducing the effect of the noise on the original transmission

- For $i \in [2 : n]$, it transmits

$$X_i = \gamma_i (Z_0 - E(Z_0 | Y^{i-1}))$$

  γ_i is chosen to meet power constraint

Schalkwijk-Kailath Decoding rule

- After the n transmissions to convey Z_0, the receiver combines its estimate of Z_0 with Y_0 to get an estimate of message point
- The receiver, uses a nearest neighbor decoding rule to recover the message point.

$$\hat{\Theta}_n = Y_0 - E(Z_0|Y^n) = \theta(m) + Z_0 - E(Z_0|Y^n)$$

Schalkwijk-Kailath Error Analysis

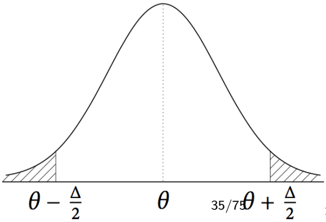
Theorem

The probability of decoding error decreases as a second-order exponent in block length for rates below capacity.

- Decoder makes an error if $\hat{\Theta}_n$ is closer to the nearest neighbors of $\theta(m)$ than to $\theta(m)$

$$|\hat{\Theta}_n - \theta(m)| > \Delta/2$$

$$p_e^n \leq 2Q(2^{nC(P)}\Delta/2), \quad Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$



Schalkwijk-Kailath Error Analysis

Distribution of $\hat{\Theta}_n$: Gaussian with mean $\theta(m)$, and variance ?

$$\hat{\Theta}_n = Y_0 - E(Z_0|Y^n) = \theta(m) + Z_0 - E(Z_0|Y^n)$$

$$I(Z_0; Y^n) = h(Z_0) - h(Z_0|Y_n) = \frac{1}{2}\log\frac{1}{Var(Z_0|Y^n)}$$

$\Rightarrow Var(Z_0|Y^n) = 2^{-nI(Z_0;Y^n)}$

Schalkwijk-Kailath Error Analysis

$$
\begin{aligned}
I(Z_0; Y^n) &= \sum_{i=1}^{n} I(Z_0; Y_i | Y^{i-1}) \\
&= \sum_{i=1}^{n} (h(Y_i | Y^{i-1}) - h(Y_i | Z_0, Y^{i-1}) \\
&= \sum_{i=1}^{n} (h(Y_i) - h(Z_i | Z_0, Y^{i-1})) \\
&= \sum_{i=1}^{n} (h(Y_i) - h(Z_i)) \\
&= \frac{n}{2} \log(1 + P) \\
&= nC(P)
\end{aligned}
$$

Schalkwijk-Kailath Error Analysis

Theorem

Channel input is independent of the previous output Y^{i-1}.

Proof.

- $Z_0 \perp Z_1$, and Gaussian
- $Y_1 = \gamma_1 Z_0 + Z_1 \Rightarrow E(Z_0|Y_1)$ is linear in Y_1
- $X_2 = \gamma_2(Z_0 - E(Z_0|Y_1)$ is Gaussian $\perp Y_1$
- Z_2 is Gaussian $\perp Y_1$
- $Y_2 = X_2 + Z_2$ is Gaussian $\perp Y_1$

...

□

Error Exponent

$$Var(Z_0|Y^n) = 2^{-2nC(P)}$$

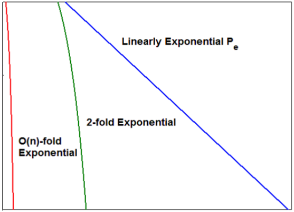
$$\hat{\Theta}_n \sim N(\theta(m), 2^{-2nC(P)})$$

[Shannon 59]: No feedback

$$p_e^{(n)} = e^{-O(n)}$$

With feedback

$$p_e^{(n)} = \exp(-\exp(O(n(C-R)))$$



Recursion rule for SK scheme

$$
\begin{aligned}
X_i &= \gamma_i(Z_0 - E(Z_0|Y^{i-1})) \\
&= \gamma_i(Z_0 - E(Z_0|Y^{i-2}) + E(Z_0|Y^{i-2}) - E(Z_0|Y^{i-1})) \\
&= \frac{\gamma_i}{\gamma_{i-1}}(X_{i-1} - E(X_{i-1}|Y^{i-1}) \\
&= \frac{\gamma_i}{\gamma_{i-1}}(X_{i-1} - E(X_{i-1}|Y_{i-1})
\end{aligned}
$$

$$
\boxed{X_i \propto X_{i-1} - E(X_{i-1}|Y_{i-1})}
$$

Important observation:

$$X_1 \propto Z_0 \sim N(0,1)$$

$$X_i \propto Z_0 - E(Z_0|Y^{i-1}) \perp Y^{i-1}$$

Schalkwijk-Kailath observation

Theorem

Channel input is independent of the previous output Y^{i-1}.

Proof.

- $Z_0 \perp Z_1$, and Gaussian
- $Y_1 = \gamma_1 Z_0 + Z_1 \Rightarrow E(Z_0|Y_1)$ is linear in Y_1
- $X_2 = \gamma_2(Z_0 - E(Z_0|Y_1)$ is Gaussian $\perp Y_1$
- Z_2 is Gaussian $\perp Y_1$
- $Y_2 = X_2 + Z_2$ is Gaussian $\perp Y_1$

...

□

Outline

- Introduction
- Point to point communication
  - Horstein coding Scheme
  - Block Feedback Coding Scheme for BSC
  - Schalkwijk-Kailath Coding scheme
  - Posterior matching scheme
- Multiuser Channel
  - Multiple access channel
  - Two way channel
- Source coding with feedforward

Posterior Matching Scheme

- At each time, the receiver calculates the a-posteriori density function of the message point. $f_n(\theta) = f_{\theta|y^n}(\theta|y^n)$
- Transmitter can track $f_n(\theta)$ as well.
- The goal is to select the transmission function g_n, for the fast concentration of $f_n(\theta)$ around θ_0

What is the best selection of the transmission functions g_i?

$$g_i(\theta, Y^{i-1}) = F_X^{-1} \circ F_{\Theta|Y^{i-1}}(\theta|Y^{i-1})$$

information regarding θ_0 still missing at the receiver is extracted by

- Encoder first extracts the information missing at the receiver from the a-posteriori by
  - Generating a random variable that is statistically independent of past observations,
  - When coupled with those observations, uniquely produces the intended message θ_0
- This information is then matched to the optimal input distribution of the channel, F_X, to achieve capacity
  - Stretching the posterior into the desired input distribution

Posterior Matching Scheme

The input to the channel is a set of random variables given by

$$X_1 = F_X^{-1}(F_\Theta(\Theta))$$
$$X_i = F_X^{-1}(F_{\Theta|Y^{i-1}}(\Theta|Y^{i-1}))$$

Note that because $F_{\Theta|Y^{i-1}}(\Theta|Y^{i-1})$ is distributed uniformly on $[0,1]$, regardless of the sequence Y^{i-1}, it follows that

- X_i is independent of Y^{i-1} and , due to the memoryless nature of the channel, Y_i is independent of Y^{i-1}
- The marginal distribution on X_i is P_X, the capacity achieving distribution, Consequently, $\{Y_i\}$ are i.i.d.

Proposition

Let X be a real-valued random variable. The random variable $Z = F_X(X)$ is uniformly distributed on $[0, 1]$.

Proof.

Let $Z = F_X(X)$

$$\begin{aligned} F_Z(x) &= p(Z \leq x) \\ &= p(F_X(X) \leq x) \\ &= p(X \leq F_X^{-1}(x)) \\ &= F_X(F_X^{-1}(x)) \\ &= x \end{aligned}$$

For any $x \in [0, 1]$, which shows that Z is a uniform random variable on $[0, 1]$ □

Proposition

Suppose that $\Theta \sim \mathcal{U}[0,1]$ *and let* X *be a real-valued random variable. Then the random variable* $Y = F_X^{-1}(\Theta)$ *has the same distribution as* X*. on* $[0,1]$*.*

Proof.

$$
\begin{aligned}
F_Y(x) &= p(Y \leq x) \\
&= p(F_X^{-1}(\Theta) \leq x) \\
&= p(\Theta \leq F_X(x)) \\
&= F_X(x)
\end{aligned}
$$

$F_X(x) \in [0,1]$ for all $x \Rightarrow X, Y$ have the same distribution

□

Posterior matching AWGN channel

- Let $p_{Y|X}$ be an AWGN channel with noise variance N
- set Gaussian input distribution $X \sim N(0, P)$ (capacity achieving for an input power constraint P
- derive posterior matching scheme in this case
- Let $SNR = P/N$

$$X_{i+1} = \sqrt{1 + SNR}\left(X_i - \frac{SNR}{1 + SNR} Y_i\right)$$

- The transmitter sends the error term pertaining to the MMSE estimate of X_i from Y_i

Posterior Matching Scheme-BSC

- Set $P_x = \text{Bern}(1/2)$ (capacity achieving)
- the PM scheme coincides with Horstein's median rule

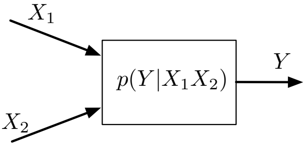
$$X_{n+1} = F_X^{-1}(F_{\Theta_0|Y^n}(\Theta_o|Y^n)) = \begin{cases} 1 & \text{if } \Theta_0 > \text{median}\{f_{\Theta_0|Y^n}(\circ|Y^n)\} \\ 0 & \text{o.w} \end{cases}$$

- F_X^{-1} quantizes above/below $\frac{1}{2}$

Outline

- Introduction
- Point to point communication
  - Horstein coding Scheme
  - Block Feedback Coding Scheme for BSC
  - Schalkwijk-Kailath Coding scheme
  - Posterior matching scheme
- Multiuser Channel
  - Multiple access channel
  - Two way channel
- Source coding with feedforward

Multiple Access channel, No feedback

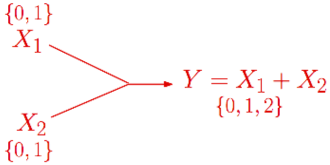


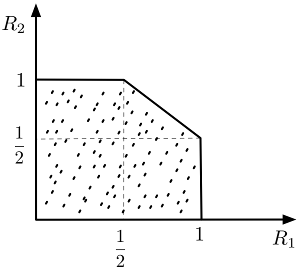
Capacity region:

$$
\begin{aligned}
R_1 &< I(X_1; Y|X_2) \\
R_2 &< I(X_2; Y|X_1) \\
R_1 + R_2 &< I(X_1X_2; Y)
\end{aligned}
$$

for all $p(X_1)p(X_2)p(Y|X_1X_2)$

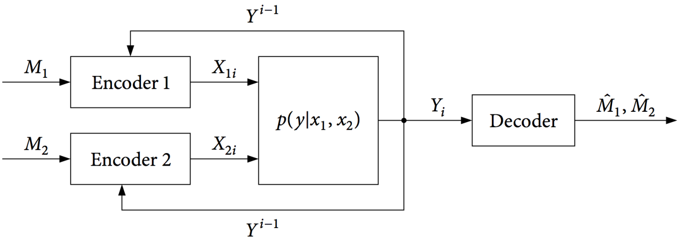
Multiple access channel, No feedback





$$R_1 < 1$$
$$R_2 < 1$$
$$R_1 + R_2 < 1.5$$

Does feedback help in MAC?



Yes! Gaarder-Wolf 1975

Erasure MAC with feedback

- $R_{sym} = 2/3$: N uncoded transmissions + $N/2$ one-sided retransmissions:

transmitter 1:	010010101011100
transmitter 2:	110100011011001
Output:	120110112022101

prob. of erasure=$1/2 \Rightarrow N/2$ bits are erased.

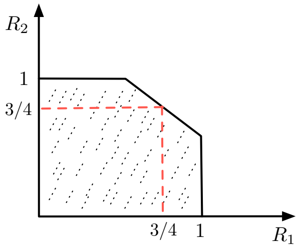
- transmitter 1 retransmits the erased bits over the next $N/2$ transmissions

N bits are sent over $N + N/2$ transmission $\Rightarrow R = 2/3$ is achievable

block feedback coding scheme:Erasure MAC with feedback

- $R_{sym} = 3/4$: N uncoded transmissions + $N/4$ two-sided retransmissions + $N/16$ + ...
- the two encoders can cooperate by each sending half of the $N/2$ erased bits over the following $N/4$ transmissions

$$R = \frac{N}{N + N/4 + N/16 + \ldots} = 3/4$$

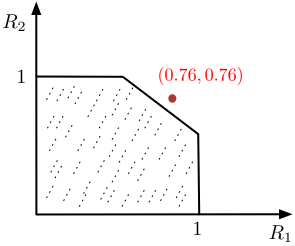


Erasure MAC with feedback (Gaarder-Wolf 1975)

- $R_{sym} = 0.7602$: N uncoded transmissions + $N/(2\log 3)$ cooperative retransmission
- Can cooperate and use three symbols: $(0,0),(1,1)$ and $(1,0)$

resolve erasure at $\log_2 3$ bit/channel use

$$R = \frac{N}{N + N/(2\log 3)} = 0.7602$$



Can we do better? Cover-Leung inner bound

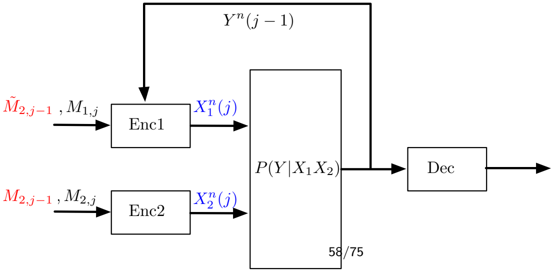
$R_{sym} = 0.7911$ (Cover-Leung 1981)

Theorem

(R_1, R_2) is achievable for MAC with feedback if

$$\begin{aligned} R_1 &< I(X_1; Y|X_2, U), \\ R_2 &< I(X_2; Y|X_1, U) \\ R_1 + R_2 &< I(X_1, X_2; Y) \end{aligned}$$

for some $p(u)p(x_1|u)p(x_2|u)$



Cover-Leung Achievability proof

- Block Markov coding:

Messages are sent over b blocks of transmission.

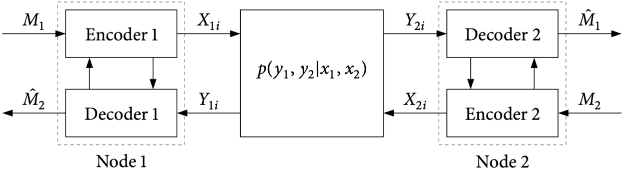
		$\overleftrightarrow{\;n\;}$			
M_1	M_2	M_3	$\cdots$	M_{b-1}	1
Block 1	Block 2	Block 3	$\cdots$	Block $b-1$	Block b

Block	1	2	$\cdots$	$b-1$	b
X_1	$x_1^n(m_{11}\|1)$	$x_1^n(m_{12}\|\tilde{m}_{21})$	$\cdots$	$x_1^n(m_{1,b-1}\|\tilde{m}_{2,b-2})$	$x_1^n(m_{1b}\|\tilde{m}_{2,b-1})$
(X_1, Y)	$\tilde{m}_{21} \to$	$\tilde{m}_{22} \to$	$\cdots$	$\tilde{m}_{2,b-1}$	$\emptyset$
X_2	$x_2^n(m_{21}\|1)$	$x_2^n(m_{22}\|m_{21})$	$\cdots$	$x_2^n(m_{2,b-1}\|m_{2,b-2})$	$x_2^n(1\|m_{2,b-1})$
Y	$\hat{m}_{11}$	$\leftarrow \hat{m}_{12}, \hat{m}_{21}$	$\cdots$	$\leftarrow \hat{m}_{1,b-1}, \hat{m}_{2,b-2}$	$\leftarrow \hat{m}_{1b}, \hat{m}_{2,b-1}$

Outline

- Introduction
- point to point communication
  - Horstein coding Scheme
  - Block Feedback Coding Scheme for BSC
  - Schalkwijk-Kailath Coding scheme
  - Posterior matching scheme
- Multiuser Channel
  - Multiple access channel
  - Two way channel
- Source coding with feedforward

Two way channel

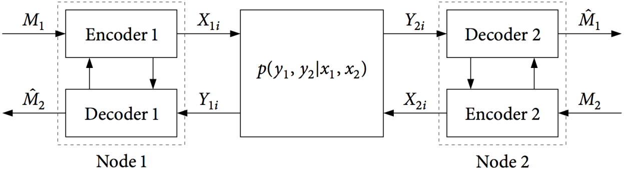


Shannoninner bound:

$$R_1 < I(X_1; Y|X_2)$$
$$R_2 < I(X_2; Y|X_1)$$

for some $p(x_1)p(x_2)$

Two way channel



Shannon outer bound:

$$R_1 < I(X_1; Y|X_2)$$
$$R_2 < I(X_2; Y|X_1)$$

for some $p(x_1, x_2)$

Directed information

1 Entropy

$$H(Y^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1})$$

2 Conditional entropy

$$H(Y^n|X^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X^n)$$

3 Causally-conditioned entropy

$$H(Y^n||X^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X^i)$$

$1 - 2 \Rightarrow I(Y^n; X^n)$ Mutual information
$1 - 3 \Rightarrow I(Y^n \rightarrow X^n)$ Directed information

Directed infromation

Directed information from a random vector A^N to another random vector B^N is:

$$I(A^N \to B^N) = \sum_{n=1}^{N} I(A^n; B_n | B^{n-1})$$

mutual information:

$$I(A^N; B^N) = \sum_{n=1}^{N} I(A^N; B_n | B^{n-1})$$

TWC- Capacity region Kramer 2003

Theorem

Let $\mathcal{R}_N$ be the set of rate pairs (R_1, R_2)

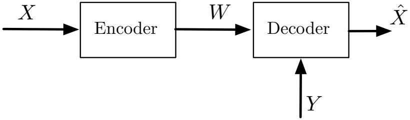
$$R_1 \leq \frac{1}{N} I(X_1^N \to Y^N || X_2^N)$$
$$R_2 \leq \frac{1}{N} I(X_2^N \to Y^N || X_1^N)$$

for some $p(x_1^N || y^{N-1}) p(x_2^N || y^{N-1})$. *Then* $\mathcal{R} = \cup_N \mathcal{R}_N$

Outline

- Introduction
- Point to point communication
  - Horstein coding Scheme
  - Block Feedback Coding Scheme for BSC
  - Schalkwijk-Kailath Coding scheme
  - Posterior matching scheme
- Multiuser Channel
  - Multiple access channel
  - Two way channel
- Source coding with feedforward

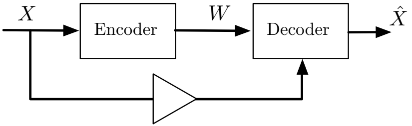
Source coding with side information



Time	1	2	3	4	5	6	7	8	9	10
Source	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Encoder	–	–	–	–	W	–	–	–	–	W
Side info	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
Decoder	–	–	–	–	–	$\hat{X}_1$	$\hat{X}_2$	$\hat{X}_3$	$\hat{X}_4$	$\hat{X}_5$

Block length $= 5$

Source coding with feedforward



Time	1	2	3	4	5	6	7	8	9	10
Source	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Encoder	–	–	–	–	W	–	–	–	–	W
Side info	–	–	–	–	–	–	X_1	X_2	X_3	X_4
Decoder	–	–	–	–	–	$\hat{X}_1$	$\hat{X}_2$	$\hat{X}_3$	$\hat{X}_4$	$\hat{X}_5$

Block length = 5 Delay = 6 Delay 1 FF

Source coding with Feedforward

$(N, 2^{NR})$ source code:
encoding function

$$f : \mathcal{X}^N \to \{1, .., 2^{NR}\}$$

Decoding function:

$$g_n : \{1, .., 2^{NR}\} \times \mathcal{X}^{n-1} \to \hat{\mathcal{X}}, \quad n = 1, .., N$$

Directed information

1 Entropy

$$H(Y^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1})$$

2 Conditional entropy

$$H(Y^n|X^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X^n)$$

3 Causally-conditioned entropy

$$H(Y^n||X^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X^i)$$

$1-2 \Rightarrow I(Y^n; X^n)$ Mutual information
$1-3 \Rightarrow I(Y^n \rightarrow X^n)$ Directed information

Directed infromation

Directed information from a random vector A^N to another random vector B^N is:

$$I(A^N \to B^N) = \sum_{n=1}^{N} I(A^n; B_n | B^{n-1})$$

mutual information:

$$I(A^N; B^N) = \sum_{n=1}^{N} I(A^N; B_n | B^{n-1})$$

Directed information from the reconstruction $\hat{X}^N$ to the source X^N:

$$I(\hat{X}^N \to X^N) = I(X^N; \hat{X}^N) - \sum_{n=2}^{N} I(X^{n-1}; \hat{X}_n | \hat{X}^{n-1})$$

- a direct coding theorem for a general source with feed-forward assuming that the joint random process $\{X_n, \hat{X}_n\}$ is discrete, stationary and ergodic

- for stationary and ergodic joint processes, the directed information rate exists and is defined by

$$I(\hat{X} \to X) = \lim_{N\to\infty} \frac{1}{N} I(\hat{X}^N \to X^N)$$

Theorem

For a discrete stationary and ergodic source X characterized by a distribution P_X, all rates R such that

$$R \geq R^*(D) = \inf_{P_{\hat{X}|X}:\lim_{N\to\infty} E[d_N(X^N,\hat{X}^N)]\leq D} I(\hat{X} \to X)$$

are achievable at expected distortion D.

Proof.

The proof uses AEP for directed qualities.

$$-\frac{1}{N}\log P(\hat{X}^N||X^N) \to H(\hat{X}||X) \;\; w.p.1$$

define a new kind of typicality that we call "directed typicality".

□

Thank you

Questions?