On Coding with Feedback in presence of Side Information

Tsachy Weissman

Stanford

Based on work with

Neri Merhav

Technion

Channel Coding

- Message: $I \in \{1, \ldots, 2^{nR}\}$
- Encoder: $X^n(I) = (X_1(I), \ldots, X_n(I))$
- Channel output: $Y^n = (Y_1, \ldots, Y_n)$
- Receiver: $\hat{I}(Y^n)$

Capacity: C = max rate for reliable communications

- $R < C$: can get $P(I \neq \hat{I}(Y^n)) \to 0$
- $R > C$: cannot get $P(I \neq \hat{I}(Y^n)) \to 0$

Capacity of a Memoryless Channel

[Shannon 48]:

$$C = \max_{P(X)} I(X;Y)$$

Channel Coding with (Perfect) Feedback

- Message: $I \in \{1, \ldots, 2^{nR}\}$
  - Encoder: $X_1(I)$
  - Channel output: Y_1
  - Encoder: $X_2(I, Y_1)$
  - Channel output: Y_2
  - $\vdots$
  -
  - Encoder: $X_n(I, Y_1, \ldots, Y_{n-1})$
  - Channel output: Y_n
- Receiver: $\hat{I}(Y^n)$

Capacity: C = max rate for reliable communications

- $R < C$: can get $P(I \neq \hat{I}(Y^n)) \to 0$
- $R > C$: cannot get $P(I \neq \hat{I}(Y^n)) \to 0$

Capacity of Memoryless Feedback Channel

[Shannon 56]:
Feedback does not increase capacity of memoryless channel

Capacity of Memoryless Feedback Channel

[Shannon 56]:
Feedback does not increase capacity of memoryless channel

[Shannon 56]...[Schalkwijk&Kailath 66]...[Ahlswede 71]...[Ooi&Wornell 98]:
Simplifies schemes for attaining it

[Shannon 56]...[Schalkwijk&Kailath 66]...[Burnashev 76]:
Improves reliability

Schalkwijk-Kailath Scheme

Gaussian channel: $Y_i = X_i + N_i$; N_i i.i.d.$\sim N(0, \sigma^2)$

Initialization:

- Given a message $m \in \{0, 1, \ldots, M-1\}$, $M = 2^{nR}$, let $\theta = (m + 1/2)/M$
- For $i = 1$, set $X_{1,1} = 0.5$ and transmit $\alpha(X_{1,1} - \theta)$
- At receiver, compute $X_{2,1} = X_{1,2} = X_{1,1} - \frac{Y_1}{\alpha}$ and send $X_{2,1}$ back to transmitter

Recursion: For $i = 2, 3, \ldots, n$

- Transmit $\alpha^{i-1} g(X_{i,1} - \theta)$
- At receiver, compute $X_{i,2} = X_{i,1} - \frac{Y_i}{\alpha^{i-1} g}$, and $X_{(i+1),1} = \frac{1}{\alpha^2} X_{i,1} + \left(1 - \frac{1}{\alpha^2}\right) X_{i,2}$
- Send $X_{(i+1),1}$ back to the transmitter

Decode m by quantizing $X_{(n+1),1}$ to its message interval.

Properties of Schalkwijk-Kailath Scheme

- Cute
- Simple
- $P_e \to 0$ double exponentially rapidly (for $R < C$)

Properties of Schalkwijk-Kailath Scheme

- Cute
- Simple
- $P_e \to 0$ double exponentially rapidly (for $R < C$)

Good also for non-white additive Gaussian noise:

[Kim 05]: $MA(1)$

The Gel'fand-Pinsker Channel

- S_i i.i.d.$\sim p(s)$
- For message index $I \in \{1, \ldots, 2^{nR}\}$, i-th channel input is $X_i(I, S^n)$
- Memoryless channel $p(y|x, s)$ generates Y_i
- Decoder gives $\hat{I}(Y^n)$

Capacity of the Gel'fand-Pinsker Channel

[Gel'fand&Pinsker 80], [Heegard&El Gamal 83]:

$$\begin{aligned} C &= \max_{P(U,X|S)} \left[I(U;Y) - I(U;S) \right] \\ &= \max_{P(U|S),f} \left[I(U;Y) - I(U;S) \right], \end{aligned}$$

where $X = f(U,S)$

Example: Costa's Channel

- Channel $Y_i = X_i + S_i + N_i$
- $\{S_i\}$ is 'dirty paper'
- Encoder 'writes on dirty paper', with power limited to P
- Decoder sees writings on dirty paper, corrupted by $\{N_i\}$ i.i.d. $\sim N(0, \sigma_Z^2)$

[Costa 83]:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_N^2} \right)$$

(the interference free capacity)

Gel'fand-Pinsker Channel with Feedback

- S_i i.i.d.$\sim p(s)$
- For message index $W \in 2^{nR}$, i-th channel input is $X_i(W, S^n, Y^{i-1})$
- Memoryless channel $p(y|x, s)$ generates Y_i
- Decoder gives $\hat{W}(Y^n)$.

Gel'fand-Pinsker Channel with Feedback (cont.)

Observation: *Feedback does not increase capacity of the Gel'fand-Pinsker channel*

Proof: Original converse proof of [Gelfand&Pinsker '80] general enough to accommodate feedback:

- $U_i - (X_i, S_i) - Y_i$, where $U_i = (W, S_{i+1}^n, Y^{i-1})$, continues to be a Markov chain since
- $P(y_i|w, s^n, y^{i-1}) = p(y_i|x_i, s_i)$ so $(W, S^n, Y^{i-1}) - (X_i, S_i) - Y_i$ is a Markov chain
- therefore, since U_i is a deterministic function of (W, S^n, Y^{i-1}), so is $U_i - (X_i, S_i) - Y_i$ $\square$

The Shannon Channel (causal state information)

[Shannon 58]

- S_i i.i.d.$\sim p(s)$
- For message index $I \in 2^{nR}$, i-th channel input is $X_i(I, S^i)$
- Memoryless channel $p(y|x,s)$ generates Y_i
- Decoder gives $\hat{I}(Y^n)$.

$$C = \max_{X(\cdot):\mathcal{S}\to\mathcal{X}} I(X(\cdot);Y)$$

Observation: *Feedback does not increase capacity of the Shannon channel.*

Proof sketch: Independence between U_i and S_i perseveres. $\square$

Gel'fand-Pinsker Channel with Feedback (cont.)

Observation: *Feedback does not increase capacity of the Gel'fand-Pinsker channel*

Questions:

- Can it simplify the schemes for achieving it ?
- Can it improve the reliability ?

Writing on Dirty Paper with Feedback

- Channel $Y_i = X_i + S_i + N_i$
- $\{S_i\}$ is interference signal
- $\{N_i\}$ i.i.d. $\sim N(0, \sigma_N^2)$
- Transmission power limited to P
- $X_i = X_i(I, S^n, Y^{i-1})$

[Costa 83] + **Observation**:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_N^2} \right)$$

Towards Schalkwijk-Kailath for general S^n

Encoder: Operate *exactly* as in S-K scheme using $Y^{i-1} - S^{i-1}$ for time i

Decoder: Operate *exactly* as in S-K scheme

Towards Schalkwijk-Kailath for general S^n

Encoder: Operate *exactly* as in S-K scheme using $Y^{i-1} - S^{i-1}$ for time i

Decoder: Operate *exactly* as in S-K scheme

Suggestion: Interference cancellation (notwithstanding Costa's advice)

Towards Schalkwijk-Kailath for general S^n

Encoder: Operate *exactly* as in S-K scheme using $Y^{i-1} - S^{i-1}$ for time i

Decoder: Operate *exactly* as in S-K scheme

Suggestion: Interference cancellation (notwithstanding Costa's advice)

Attempt I: Cancellation at the end. Instead of X_n take

$$X'_n = X_n - ?,$$

where $?$ will cancel the interference, resulting in decision identical to case $S^n \equiv 0$.

Unfortunately, we get

$$? = \alpha^n \sum_{i=1}^{n} \frac{S_i}{\alpha^{i-1}}$$

Way too much power :-(

Towards Schalkwijk-Kailath for general S^n (cont.)

But what were we expecting ?

Suggested scheme depends on SI causally so, since feedback does not improve capacity of Shannon channel, can achieve at most $C_{\text{causal SI}}$

But for Costa channel [Eggers, Su, Girod 00], [Erez, Shamai, Zamir 04]:

$$C_{\text{causal SI}} < C_{\text{non-causal SI}} = \frac{1}{2}\log\left(1+\frac{P}{\sigma_N^2}\right)$$

Towards Schalkwijk-Kailath for general S^n (cont.)

Encoder: Operate *exactly* as in S-K scheme using $Y^{i-1} - S^{i-1}$ for time i

Decoder: Operate *exactly* as in S-K scheme

Suggestion: Interference cancellation (notwithstanding Costa's advice)

Towards Schalkwijk-Kailath for general S^n (cont.)

Encoder: Operate *exactly* as in S-K scheme using $Y^{i-1} - S^{i-1}$ for time i

Decoder: Operate *exactly* as in S-K scheme

Suggestion: Interference cancellation (notwithstanding Costa's advice)

Attempt II: *Pre*-Cancellation. Instead of $X_1 = \alpha(1/2 - \theta)$ take

$$X_1' = \alpha(1/2 - \theta - ?),$$

where $?$ will cancel the interference, resulting in decision identical to case $S^n \equiv 0$.

We get

$$? = \frac{S_1}{\alpha} + \left(1 - \frac{1}{\alpha^2}\right) \frac{1}{g} \sum_{i=2}^{n} \frac{S_i}{\alpha^{i-1}} \;\Rightarrow\; E(X_1')^2 \le B < \infty$$

Resulting power $\le P + \frac{B}{n} \to P$

Comments on Derived Scheme

- Simple capacity achieving scheme. No binning.
- While in absence of feedback [Costa 83] idea is not to 'fight' the interference, here pre–cancelling approach is fruitful.
- Scheme gives exactly the same estimation and decoding as the Schalkwijk-Kailath scheme, for every realization of the noise process, and as if S^n were non–existent ($S^n \equiv 0$). In particular, same error probability.
- Law of $\{S_n\}$ is immaterial so long as

$$E\left[\left(\sum_{i=1}^{\infty} \frac{S_i}{\alpha^{i-1}}\right)^2\right] < \infty.$$

- Non–causal dependence of transmission on S^n is only via X_1.

Feedback Coding for the DMC

[Ahlswede 71], [Ooi&Wornell 98]:

- *Transmitter:* Map N message bits into $n_1 = N/H(X)$ channel inputs iid $\sim X$
- Send X^{n_1} through channel
- *Channel:* Corrupts X^{n_1}, receiver gets Y^{n_1}
- *Transmitter:*
  - Using Y^{n_1}, compresses X^{n_1} into $n_1 H(X|Y)$ new data bits
  - Precode new data bits into $n_2 = n_1 H(X|Y)/H(X)$ channel inputs iid $\sim X$
- Send $X_{n_1+1}^{n_1+n_2}$ through channel
- $\vdots$

Overall number of channel uses:

$$\frac{N}{H(X)} + \frac{N}{H(X)}\left(\frac{H(X|Y)}{H(X)}\right) + \frac{N}{H(X)}\left(\frac{H(X|Y)}{H(X)}\right)^2 + \cdots = \frac{N}{H(X) - H(X|Y)} = \frac{N}{C}$$

Feedback Coding for the Discrete Gel'fand-Pinsker Channel

Let S, U, X, Y be capacity achieving. Using similar idea as in Ahlswede's scheme, replacing conditional entropy coding with Slepian-Wolf coding gives overall number of channel uses:

$$\begin{aligned}
\frac{N}{H(U|S)} + \frac{N}{H(U|S)}\left(\frac{H(U|Y)}{H(U|S)}\right) + \frac{N}{H(U|S)}\left(\frac{H(U|Y)}{H(U|S)}\right)^2 + \cdots &= \\
\frac{N}{H(U|S) - H(U|Y)} &= \\
\frac{N}{I(U;Y) - I(U;S)} &= \frac{N}{C}
\end{aligned}$$

Source Coding with "Feedforward"

- $I = I(X^n) \in \{1, \ldots, 2^{nR}\}$
- $\hat{X}_i = \hat{X}_i(I, X^{i-1})$
- would like small $\rho(X^n, \hat{X}^n) = \frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \hat{X}_i)$

Motivation:

- Competitive Predictability [Weissman&Merhav 03]
- Source Coding with Feedforward [Pradhan 04]
- Source Coding with Fixed Lag Side Information [Martinian&Wornell 04]

Fact 1. *Feedforward does not improve the achievable Rate-Distortion performance for a memoryless source.*

Source Coding with "Feedforward"

Cute schemes for feedforward coding of memoryless sources:

- Gaussian source: [Pradhan 04]
- Discrete memoryless sources: [Martinian & Wornell 04]

Wyner-Ziv Coding

- Source and Side information are iid drawings of pairs (X_i, Y_i)
- Encoder: $I(X^n) = \{1, \ldots, 2^{nR}\}$
- Decoder: $\hat{X}^n(I, Y^n)$
- Goal: minimize $\rho(X^n, \hat{X}^n) = \frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \hat{X}_i)$

[Wyner&Ziv 76]

$$R(D) = \min_{p(u|x), f: E\rho(X, f(U,Y)) \leq D} [I(X;U) - I(Y;U)]$$

Feedforward Wyner-Ziv Coding

- Source and Side information are iid drawings of pairs (X_i, Y_i)
- Encoder: $I(X^n) = \{1, \ldots, 2^{nR}\}$
- Decoder: $\hat{X}_i(I, Y^n, X^{i-1})$
- Goal: minimize $\rho(X^n, \hat{X}^n) = \frac{1}{n}\sum_{i=1}^{n} \rho(X_i, \hat{X}_i)$

Feedforward Wyner-Ziv Coding (cont.)

Observation *Feedforward does not improve the rate distortion tradeoff in the Wyner-Ziv problem.*

Proof: Original converse proof carries over almost unchanged.

- In notation of [Cover&Thomas, Section 14.9], only need to add X^{i-1} to W_i, resulting in $W_i = (T, Y^{i-1}, Y_{i+1}^n, X^{i-1})$
- Converse carries over verbatim, since $W_i - X_i - Y_i$ continues to form a Markov chain under modified W_i $\square$

Feedforward Wyner-Ziv Coding (cont.)

Observation *Feedforward does not improve the rate distortion tradeoff in the Wyner-Ziv problem.*

But can it

- Simplify the schemes ?
- Improve the error exponent ?

A S-K-type Scheme for W-Z Coding with Feedforward

Source and SI:

1. $\{Y_i\}$ arbitrarily distributed side-information signal available only at decoder
2. Source signal $X_i = Y_i + N_i$, $\{N_i\}$ i.i.d. $N(0, \sigma^2)$, independent of $\{Y_i\}$

Schalkwijk-Kailath-type scheme leads to:

- Scheme as simple as channel coding one, with no binning required
- Achieves conditional R-D function for case where side information available at both sides, for arbitrarily distributed side information $\{Y_i\}$
- Implies that for Wyner-Ziv problem, when pairs (X_i, Y_i) are i.i.d., with Y_i arbitrarily distributed and $X_i = Y_i + N_i$ for N_i Gaussian there is no loss due to absence of side information at encoder
  (also follows by direct examination of Wyner-Ziv function [Wyner 78])
- Non-causal dependence on SI only through one number
- Significantly improves on source coding exponent of [Marton 74]

A Scheme for the Discrete Wyner-Ziv Source with Feedforward

- Can obtain scheme for the finite-alphabet Wyner-Ziv source
- Scheme builds on ideas of [Martinian&Wornell 04], replacing conditional entropy coding by Slepian-Wolf coding
- Achieves $R_{WZ}(D)$

Conclusions

Feedback:

- Does not increase capacity of Gel'fand-Pinsker channel
- Does not improve rate-distortion performance for Wyner-Ziv source

Does however simplify the schemes, and their error performance:

- Schalkwijk-Kailath scheme with pre-cancellation achieves the Costa "writing on dirty paper" capacity
- Schalkwijk-Kailath-type scheme for achieving the Wyner-Ziv R-D function
- Simple achievability schemes also for the discrete-alphabet case

Remark: Performance analysis for schemes also gives insight on fundamental limits

Question: Schemes for causal state information