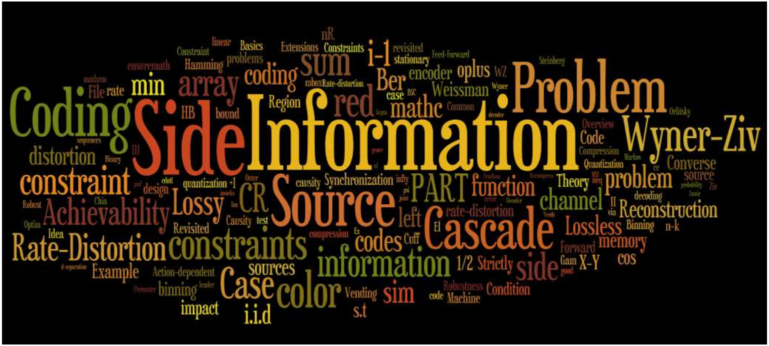
Source Coding With Side Information

O. Simeone

NJIT

National Chiao-Tung University, Jan. 2015

Overview



- Classical problem (1973-) with many applications
- This lecture: Overview covering both classical and recent results with emphasis on intuition
- Background: Basics of information theory and coding theory

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

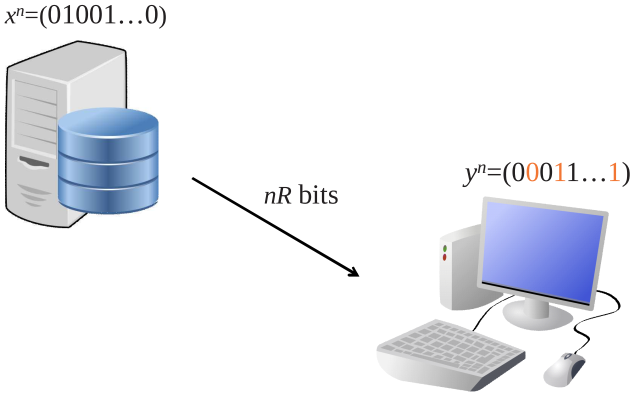
PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

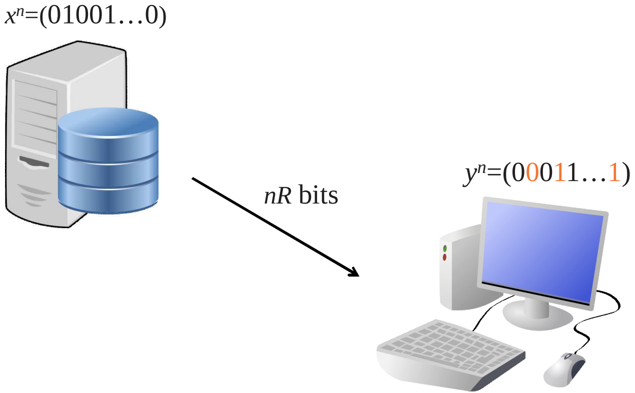
- Action-dependent side information
- Cascade problems
- Sources with memory

Example: File Synchronization



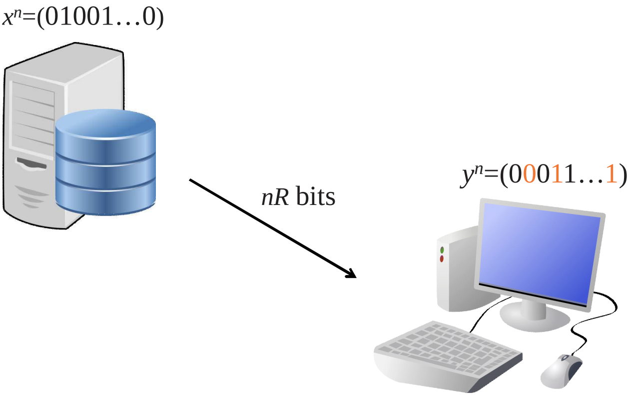
- Goal: File synchronization with minimum (bit) rate [Orlitsky '91]
- R rate (bits per source symbol)
- Key observation: leveraging the side information, $R < 1$ bits may suffice

Example: File Synchronization



- Goal: File synchronization with minimum (bit) rate [Orlitsky '91]
- R rate (bits per source symbol)
- Key observation: leveraging the side information, $R < 1$ bits may suffice

Example: File Synchronization



- Source coding (compression) with side information
- Encoder: x^n (and possibly y^n) $\rightarrow nR$ bits (message or code)
- Decoder: $(nR \text{ bits}, y^n) \rightarrow x^n$

Example: File Synchronization

- $n = 3$
- At most one substitution
- If y^3 is known at the encoder, only the error pattern $x^3 \oplus y^3$ must be communicated (delta compression)
- Ex.: $x^3 = (100)$ and $y^3 = (110)$, $x^3 \oplus y^3 = (010)$
- Required number of bits

$$nR = \log_2(\#\text{error patterns}) = \log_2\left(\binom{3}{0} + \binom{3}{1}\right) = 2 \text{ bits}$$

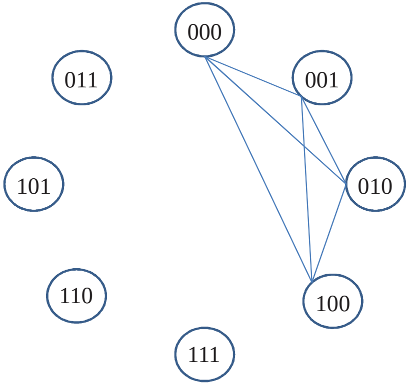
Example: File Synchronization

- $n = 3$
- At most one substitution
- If y^3 is known at the encoder, only the error pattern $x^3 \oplus y^3$ must be communicated (delta compression)
- Ex.: $x^3 = (100)$ and $y^3 = (110)$, $x^3 \oplus y^3 = (010)$
- Required number of bits

$$\begin{aligned} nR &= \log_2(\#\text{error patterns}) \\ &= \log_2\left(\binom{3}{0} + \binom{3}{1}\right) = 2 \text{ bits} \end{aligned}$$

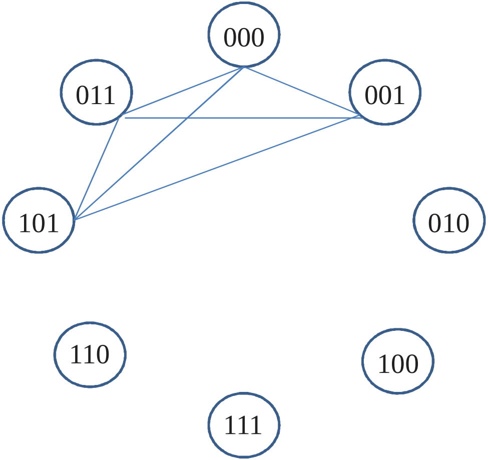
Example: File Synchronization

- When y^3 is not known at the encoder...
- Need to assign code (color) to each x^n so that the decoder can recover x^n based on code and y^n
- What does the decoder know based on y^n? Equivocation set at the receiver: for $y^n = (000)$



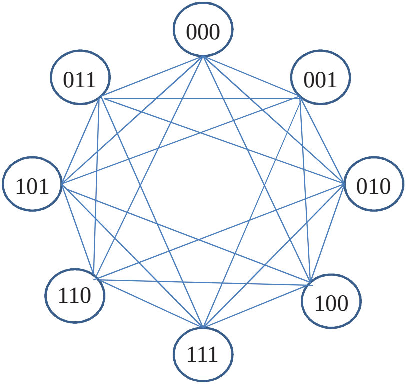
Example: File Synchronization

- Equivocation set at the receiver: for $y^n = (001)$



Example: File Synchronization

- Each source sequence is connected to all others at a Hamming distance of at most 2

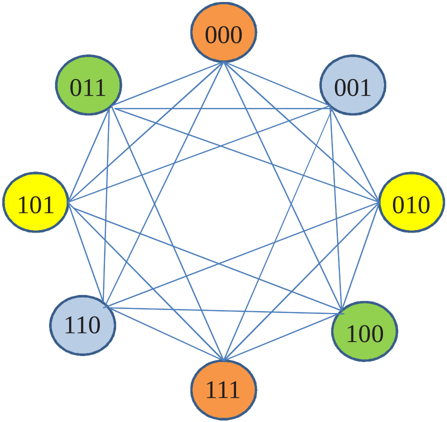


(characteristic graph [Witsenhausen '76]

- The encoder must assign different codes to sequences connected by an edge

Example: File Synchronization

- The encoder can thus encode with the same label ("color") sequences that are not connected by an edge (independent set)



- A set with the same color forms a "bin"

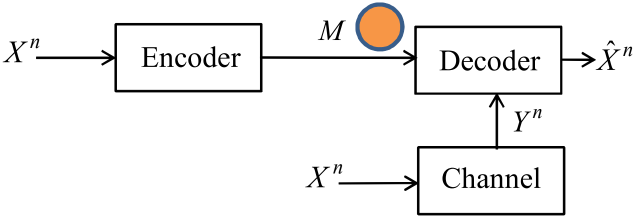
Example: File Synchronization

- Encoding:

Bin	Code
$\mathscr{C}_1 = \{000, 111\}$	00
$\mathscr{C}_2 = \{001, 110\}$	01
$\mathscr{C}_3 = \{100, 011\}$	10
$\mathscr{C}_4 = \{010, 101\}$	11

- Same rate as the case in which y^3 is known at the encoder! (But this is not always the case for zero error [Orlitsky '91])

Example: File Synchronization
Bins



- Channel decoding at the decoder...
- Bins as channel codes?

Example: File Synchronization

Bins

- $\mathscr{C}_1 = \{000, 111\}$ is a $(3, 1)$ repetition code: can correct one bit flip
- ... zero-error channel code for the channel between X and Y
- $\mathscr{C}_2 = \{001, 110\}$ is $\mathscr{C}_1 \oplus (001)$... coset of $\mathscr{C}_1$ ((001) coset leader)
- $\mathscr{C}_3 = \{100, 011\}$ is $\mathscr{C}_1 \oplus (100)$ ... coset of $\mathscr{C}_1$ ((100) coset leader)
- $\mathscr{C}_4 = \{010, 101\}$ is $\mathscr{C}_1 \oplus (010)$ ... coset of $\mathscr{C}_1$ ((010) coset leader)

Example: File Synchronization

Bins

- $\mathscr{C}_1 = \{000, 111\}$ is a $(3,1)$ repetition code: can correct one bit flip
- ... zero-error channel code for the channel between X and Y

- $\mathscr{C}_2 = \{001, 110\}$ is $\mathscr{C}_1 \oplus (001)$... coset of $\mathscr{C}_1$ ((001) coset leader)

- $\mathscr{C}_3 = \{100, 011\}$ is $\mathscr{C}_1 \oplus (100)$ ... coset of $\mathscr{C}_1$ ((100) coset leader)
- $\mathscr{C}_4 = \{010, 101\}$ is $\mathscr{C}_1 \oplus (010)$ ... coset of $\mathscr{C}_1$ ((010) coset leader)

Example: File Synchronization

Bins

- $\mathscr{C}_1 = \{000, 111\}$ is a $(3, 1)$ repetition code: can correct one bit flip
- ... zero-error channel code for the channel between X and Y

- $\mathscr{C}_2 = \{001, 110\}$ is $\mathscr{C}_1 \oplus (001)$... coset of $\mathscr{C}_1$ ((001) coset leader)

- $\mathscr{C}_3 = \{100, 011\}$ is $\mathscr{C}_1 \oplus (100)$ ... coset of $\mathscr{C}_1$ ((100) coset leader)
- $\mathscr{C}_4 = \{010, 101\}$ is $\mathscr{C}_1 \oplus (010)$ ... coset of $\mathscr{C}_1$ ((010) coset leader)

Example: File Synchronization

Code

- Parity matrix for $\mathscr{C}_1 = \{000, 111\}$

$$H = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 1 \end{array} \right]$$

- Codewords must satisfy: $H \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = H \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
- (00) is the code for the sequences in $\mathscr{C}_1$
- (00) is also known as the syndrome for $\mathscr{C}_1$

Example: File Synchronization

Code

- The code $\mathscr{C}_2 = \{001, 110\}$ is obtained in the same way

$$H\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = H\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

- (01) is referred to as the syndrome for $\mathscr{C}_2$
- Check syndromes for $\mathscr{C}_3$ (10), and for $\mathscr{C}_4$ (11)

Example: File Synchronization

Cosets and Syndromes

Bin (Coset)	Code (Syndrome)
$\mathscr{C}_1 = \{000, 111\}$	00
$\mathscr{C}_2 = \{001, 110\}$	01
$\mathscr{C}_3 = \{100, 011\}$	10
$\mathscr{C}_4 = \{010, 101\}$	11

- ... The operation at the decoder can be interpreted as channel decoding... [Orlitsky and Viswanathan '03]
- Binning is related to the concept of hashing [MacKay '03]

Example: File Synchronization

Cosets and Syndromes

Bin (Coset)	Code (Syndrome)
$\mathscr{C}_1 = \{000, 111\}$	00
$\mathscr{C}_2 = \{001, 110\}$	01
$\mathscr{C}_3 = \{100, 011\}$	10
$\mathscr{C}_4 = \{010, 101\}$	11

- ... The operation at the decoder can be interpreted as channel decoding... [Orlitsky and Viswanathan '03]
- Binning is related to the concept of hashing [MacKay '03]

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

Slepian-Wolf/ Wyner-Ziv Problem

A Theory of Source Coding with Side Information

- Tractable and relevant

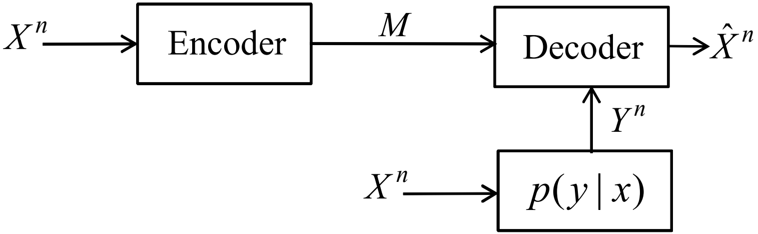
"In theory, yes, Mrs. Wilkins. But also in theory, no"

Slepian-Wolf/ Wyner-Ziv Problem

- How much compression is possible for a given "correlation" between x^n and y^n?
- How to encode optimally?

Slepian-Wolf Problem

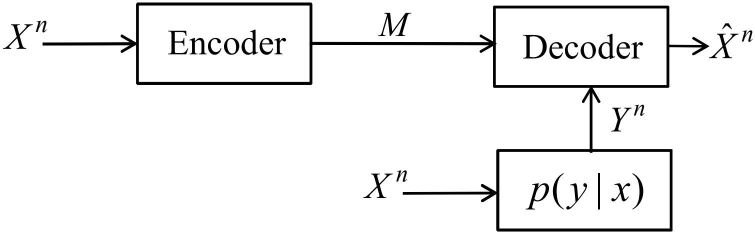
Slepian and Wolf '73



- $X^n \sim p(x)$ i.i.d., discrete finite alphabet
- Memoryless side information "channel": $Y^n|X^n = x^n \sim \prod_{i=1}^n p(y_i|x_i)$
- Fixed-length codes: $M \in \{1, 2, ..., 2^{nR}\}$ (nR bits)
- Correlation between X and Y is captured by $p(x,y) = p(x)p(y|x)$

Slepian-Wolf Problem

Doubly Symmetric Binary Source (DSBS)



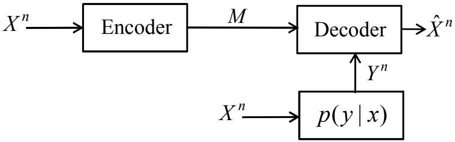
- Binary Symmetric Source (BSS): $X_i \underset{\text{i.i.d.}}{\sim} Ber(1/2)$
- Binary Symmetric Channel (BSC(q)):

$$Y_i = X_i \oplus Q_i \text{ with } Q_i \underset{\text{i.i.d.}}{\sim} Ber(q) \text{ and indep. of } X^n$$

- $q = \Pr[X_i \neq Y_i]$ ($q \leq 1/2$)

Slepian-Wolf Problem

Slepian and Wolf '73

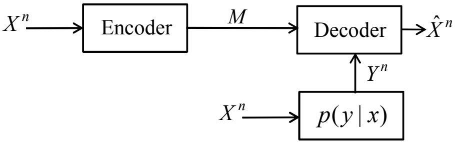


- Lossless compression: R is achievable if for any $\varepsilon > 0$, there exists a sufficiently large n_0 such that for $n \geq n_0$

$$\Pr[\hat{X}^n \neq X^n] \leq \varepsilon$$

- What is the infimum R^{SW} of all achievable rates? How to achieve it?

Slepian-Wolf Problem



- File synchronization [Orlitsky '91], distributed file backup and file sharing systems [Suel and Memon '02], reference based genome compression [Chern et al '12], ...

Shannon Theory

- Entropy $H(A) = -E[\log_2 p(A)]$: measure of "randomness" or "uncertainty"
  - For $A \sim Ber(p)$, $H(A) = H(p) = -p\log_2 p - (1-p)\log_2(1-p)$
- Mutual information $I(A;B) = E\left[\log_2 \frac{p(A,B)}{p(A)p(B)}\right]$: measure of "correlation"
- Conditional entropy $H(A|B) = H(A) - I(A;B)$: measure of "residual uncertainty"
- Conditional mutual information $I(A;B|C) = H(A|C) - H(A|B,C)$: measure of "conditional correlation"

Lossless Compression

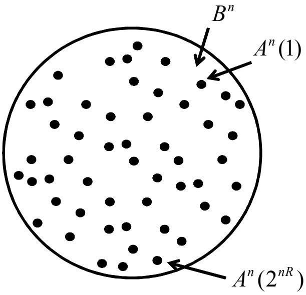
El Gamal and Kim '11

- $A^n \sim p(a)$ i.i.d. (ex.: $Ber(q)$)
- $H(A)$ bits/sample are necessary and sufficient for lossless compression
- ... Consequence of the fact that, with high probability, A^n is one of the $2^{nH(A)}$ typical sequences (ex.: for $Ber(q)$, sequence with approximately nq ones)
- Achievable in practice with Huffman coding, arithmetic coding, LZ techniques,... (e.g., [Sayood '12])

Packing Lemma (Channel Coding)

El Gamal and Kim '11

- Memoryless channel $p(b|a)$ (ex.: BSC(q))

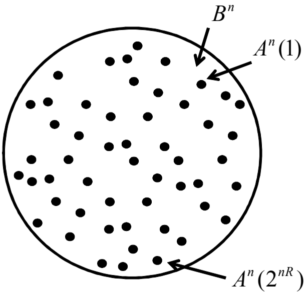


- Random codebook of 2^{nR} codewords $A^n(m)$, $m = 1, ..., 2^{nR}$
- $A^n(m) \sim p(a)$ i.i.d. and independent
- If $R < I(A;B)$, then $\Pr[\hat{M} \neq M] \to 0$ as $n \to \infty$

Packing Lemma (Channel Coding)

El Gamal and Kim '11

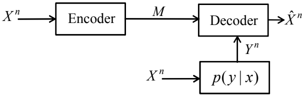
- Memoryless channel $p(b|a)$ (ex.: BSC(q))



- Random codebook of 2^{nR} codewords $A^n(m)$, $m = 1, ..., 2^{nR}$
- $A^n(m) \sim p(a)$ i.i.d. and independent
- If $R < I(A;B)$, then $\Pr[\hat{M} \neq M] \to 0$ as $n \to \infty$

Slepian-Wolf Problem

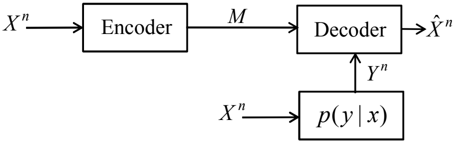
Doubly Symmetric Binary Source (DSBS)



- Encoder:
  - Divide the set of all binary sequences randomly and uniformly into codebooks (bins) of sizes $2^{nI(X;Y)} = 2^{n(1-H(q))} = 2^n/2^{nH(q)}$
  - (There are $2^{nH(q)}$ codebooks or bins)
  - The code associated to any binary sequence 2^n is the index of the bin $\rightarrow R > H(q)$
- Decoder:
  - Channel decoding within the bin (successful with high probability by the packing lemma)

Slepian-Wolf Problem

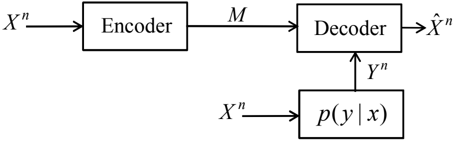
Slepian and Wolf '73



- If the encoder knew Y^n, it could calculate the error pattern $Q^n = X^n \oplus Y^n \sim Ber(q)$ i.i.d., which requires $H(q)$ bits/sample for lossless compression
- "It follows" that $R^{SW} = H(q)$

Slepian-Wolf Problem

Slepian and Wolf '73



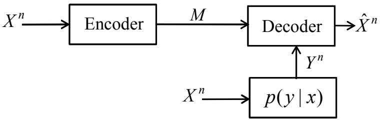
- The results can be generalized (see discussion around the more general Wyner-Ziv problem)

$$
\begin{aligned}
R^{SW} &= H(X) - I(X;Y) \\
&= H(X|Y)
\end{aligned}
$$

- This is also the rate required if the encoder knows Y^n (as in the example)!

Wyner-Ziv Problem

Wyner and Ziv '76



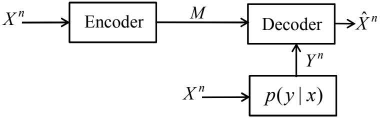
- Lossy compression: Distortion measure $0 \leq d(x, \hat{x}) < \infty$,
- Ex.: Hamming distortion $d(x, \hat{x}) = 1(x \neq \hat{x})$; MSE $d(x, \hat{x}) = (x - \hat{x})^2$
- Average per-block distortion

$$\frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] \leq D$$

- Ex.: With Hamming distortion $\frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] = \frac{1}{n} \sum_{i=1}^{n} \Pr[X_i \neq \hat{X}_i]$

Wyner-Ziv Problem

Wyner and Ziv '76



- Lossy compression: Distortion measure $0 \leq d(x,\hat{x}) < \infty$,
- Ex.: Hamming distortion $d(x,\hat{x}) = 1(x \neq \hat{x})$; MSE $d(x,\hat{x}) = (x-\hat{x})^2$

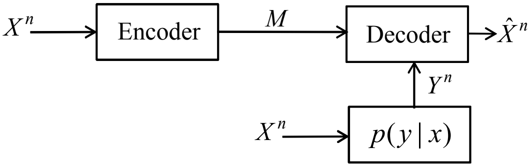
- Average per-block distortion

$$\frac{1}{n}\sum_{i=1}^{n} E[d(X_i,\hat{X}_i)] \leq D$$

- Ex.: With Hamming distortion
$\frac{1}{n}\sum_{i=1}^{n} E[d(X_i,\hat{X}_i)] = \frac{1}{n}\sum_{i=1}^{n} \Pr[X_i \neq \hat{X}_i]$

Wyner-Ziv Problem

Wyner and Ziv '76

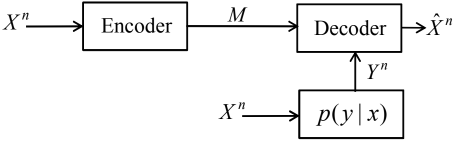


- Lossy functional reconstruction $d(x, y, \hat{x}) = d(f(x, y), \hat{x})$ [Yamamoto '82] [Feng et al '04]

$$\frac{1}{n}\sum_{i=1}^{n} E[d(X_i, Y_i, \hat{X}_i)] \leq D$$

Wyner-Ziv Problem

Wyner and Ziv '76



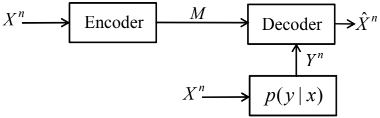
- (R, D) is achievable if for any $\varepsilon > 0$, there exists a sufficiently large n_0 such that for $n \geq n_0$

$$\frac{1}{n}\sum_{i=1}^{n} E[d(X_i, Y_i, \hat{X}_i)] \leq D + \varepsilon$$

- The rate-distortion function $R_{X|Y}^{WZ}(D)$ is the infimum of all R such that (R, D) is achievable

Wyner-Ziv Problem

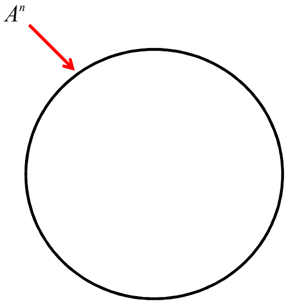
Wyner and Ziv '76



- Sensor networks [Draper and Wornell '04], video coding [Aaron et al '02],...
- Systematic source-channel coding: analog+digital audio/video broadcasting [Shamai et al '98]
- Denoising: fidelity-boosting sequence constrained to rate R [Jalali et al '10]
- Networks: Relay channel [Cover and El Gamal '79], distributed uplink reception [Sanderovich et al '09],...

Covering Lemma
El Gamal and Kim '11

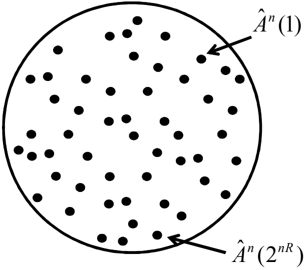
- Source $A^n \sim p(a)$ i.i.d. (ex.: $A^n \underset{\text{i.i.d.}}{\sim} Ber(1/2)$)



Covering Lemma

El Gamal and Kim '11

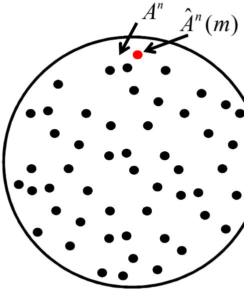
- Cover (quantize) source A^n using a codebook of 2^{nR} codewords $\hat{A}^n(m)$ for $m = 1, ..., 2^{nR}$



Covering Lemma

El Gamal and Kim '11

- For "most" sequences A^n, we want to find a sequence $\hat{A}^n(m)$ such that $(A^n,\hat{A}^n(m))$ are "close"

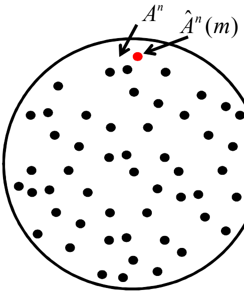


- Ex.: $\frac{1}{n}\sum_{i=1}^{n} 1(a_i \neq \hat{a}_i) \leq D = 0.2$
- How large should R be?

Covering Lemma

El Gamal and Kim '11

- "Closeness": $(A^n, \hat{A}^n(m))$ are jointly typical according to a desired joint distribution $p(a, \hat{a}) = p(a)p(\hat{a}|a)$



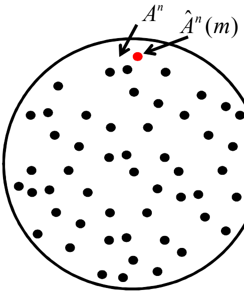
- Ex.: $p(a, \hat{a})$:

$A \backslash \hat{A}$	0	1
0	0.4	0.1
1	0.1	0.4

Covering Lemma

El Gamal and Kim '11

- "Closeness": $(A^n, \hat{A}^n(m))$ are jointly typical according to a desired joint distribution $p(a,\hat{a}) = p(a)p(\hat{a}|a)$

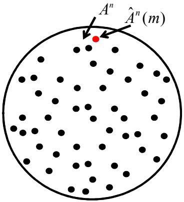


- Ex.: $p(a,\hat{a})$:

$A \backslash \hat{A}$	0	1
0	0.4	0.1
1	0.1	0.4

Covering Lemma

El Gamal and Kim '11



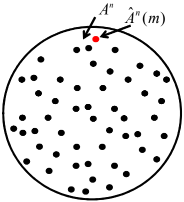
- Ex.: $p(a, \hat{a})$: $\rightarrow$ Hamming distortion $D = 0.2$

$A \backslash \hat{A}$	0	1
0	0.4	0.1
1	0.1	0.4

- $p(a)$ given by the problem
- $p(\hat{a}|a)$ is referred to as the inverse test channel
- Intuitively, we need a larger R if we desire A and $\hat{A}$ to be more "correlated"

Covering Lemma

El Gamal and Kim '11



- Ex.: $p(a, \hat{a})$:

$A \backslash \hat{A}$	0	1
0	0.4	0.1
1	0.1	0.4

$\rightarrow$ Hamming distortion $D = 0.2$

- $p(a)$ given by the problem
- $p(\hat{a}|a)$ is referred to as the inverse test channel
- Intuitively, we need a larger R if we desire A and $\hat{A}$ to be more "correlated"

Covering Lemma

El Gamal and Kim '11

- If the codewords $\hat{A}^n(m)$ are selected i.i.d. $\sim p(\hat{a})$ and independently (random codebook), then if

$$R > I(A; \hat{A}),$$

we have

$$\Pr[\nexists\ \hat{A}^n(m) \text{ jointly typical with } A^n] \to 0 \text{ as } n \to \infty$$

- Quantization defined by the inverse test channel $p(\hat{a}|a)$

Covering Lemma

El Gamal and Kim '11

- If the codewords $\hat{A}^n(m)$ are selected i.i.d. $\sim p(\hat{a})$ and independently (random codebook), then if

$$R > I(A;\hat{A}),$$

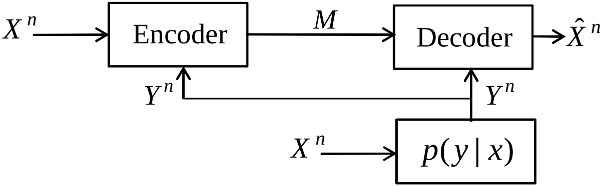
we have

$$\Pr[\nexists\ \hat{A}^n(m) \text{ jointly typical with } A^n] \to 0 \text{ as } n \to \infty$$

- Quantization defined by the inverse test channel $p(\hat{a}|a)$

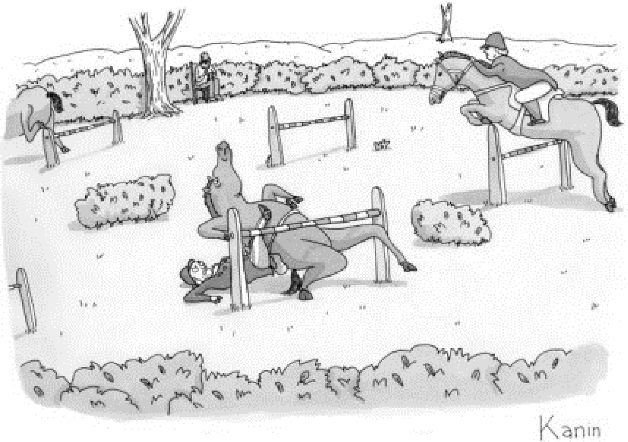
Conditional Rate-Distortion Theory

Gray '72



Conditional Rate-Distortion Theory

Converse



"Over, damn you, over!"

Conditional Rate-Distortion Theory

Converse

- Using data processing inequality, memorylessness and convexity properties of the rate-distortion function, we get the lower bound:

$$R_{X|Y}(D) \geq \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y)$$
$$\text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

- Proof: See Appendix-A
- Note that $p(x,y,\hat{x}) = p(x,y)p(\hat{x}|x,y)$

Conditional Rate-Distortion Theory

Achievability

- DSBS(q) ($Y = X \oplus Q$ where $Q \sim Ber(q)$) with Hamming distortion
- The lower bound

$$R_{X|Y}(D) \geq \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y)$$
$$\text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

leads to

$$R_{X|Y}(D) \geq H(q) - H(D)$$
$$= I(Q;\hat{Q}),$$

where $p(q|\hat{q})$ is BSC(D)

- ... lower bound achievable by compressing the error Q^n with inverse test channel $p(\hat{q}|q)$ (delta compression)

Conditional Rate-Distortion Theory

Achievability

- DSBS(q) ($Y = X \oplus Q$ where $Q \sim Ber(q)$) with Hamming distortion
- The lower bound

$$R_{X|Y}(D) \geq \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y)$$
$$\text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

leads to

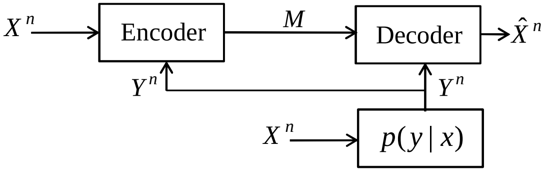
$$R_{X|Y}(D) \geq H(q) - H(D)$$
$$= I(Q;\hat{Q}),$$

where $p(q|\hat{q})$ is BSC(D)

- ... lower bound achievable by compressing the error Q^n with inverse test channel $p(\hat{q}|q)$ (delta compression)

Conditional Rate-Distortion Theory

Achievability



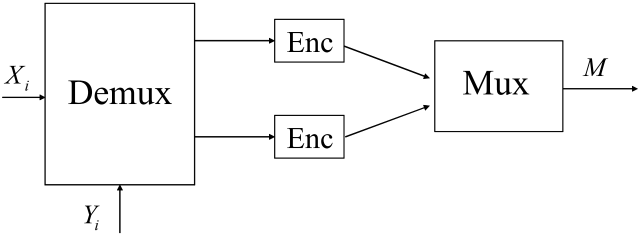
- In order to obtain a more generally applicable scheme, note that:

$$R_{X|Y}(D) \geq \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y) = \min_{p(\hat{x}|x,y)} \sum_{y\in\mathcal{Y}} p(y) I(X;\hat{X}|Y=y)$$

$$\text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

Conditional Rate-Distortion Theory

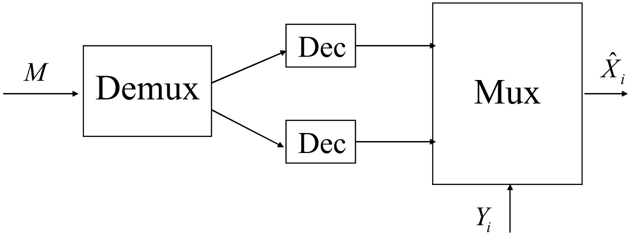
Achievability



- Context-adaptive encoding
- Each subsequence is of length $np(y)$ and is encoded using the inverse test channel $p(\hat{x}|x,y)$
- Number of bits produced for each subsequence $np(y)I(X;\hat{X}|Y=y)$

Conditional Rate-Distortion Theory

Achievability



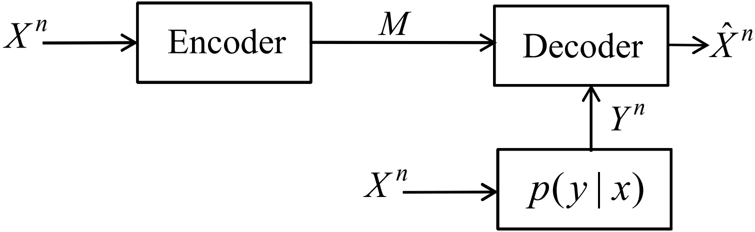
Conditional Rate-Distortion Theory

- The achievable rate matches the lower bound: rate-distortion function

$$R_{X|Y}(D) = \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y)$$
$$\text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

Wyner-Ziv Problem

Wyner and Ziv '76



- $R_{X|Y}(D) = \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y)$ s.t. $E[d(X,Y,\hat{X})] \leq D$ not achievable!
- Problem: the encoder cannot implement the inverse test channel $p(\hat{x}|x,y)$

Wyner-Ziv Problem

Converse

- Lower bound:

$$R^{WZ}_{X|Y}(D) \geq \min_{p(u|x),f(u,y)} I(X;U|Y)$$
$$\text{s.t. } E[d(X,Y,f(U,Y))] \leq D$$

- Proof: See Appendix-B

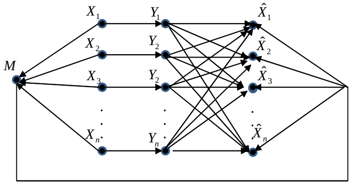
Wyner-Ziv Problem

Converse

- Lower bound:

$$R^{WZ}_{X|Y}(D) \geq \min_{p(u|x), f(u,y)} I(X;U|Y)$$
$$\text{s.t. } E[d(X,Y,f(U,Y))] \leq D$$

- Proof: See Appendix-B
- Key point: We have the Markov chain $U - X - Y$ so that $p(x,y,u) = p(x,y)p(u|x)$



Wyner-Ziv Problem

Converse

- Lower bound:

$$R_{X|Y}^{WZ}(D) \geq \min_{p(u|x), f(u,y)} I(X;U|Y)$$

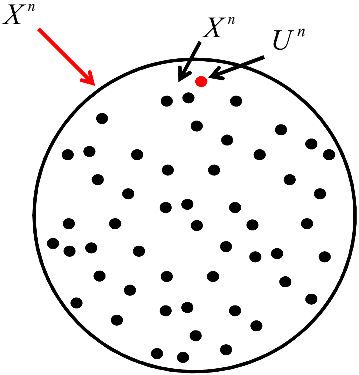
$$= \min_{p(u|x), f(u,y)} I(X;U) - I(U;Y)$$

$$\text{s.t. } E[d(X,Y,f(U,Y))] \leq D$$

Wyner-Ziv Problem

Achievability: 1. Quantization

$$I(X;U) - I(U;Y)$$

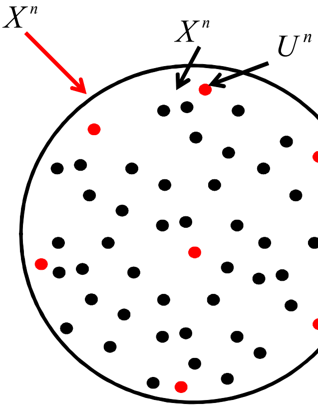


- Quantization with inverse test channel $p(u|x)$

Wyner-Ziv Problem

Achievability: 2. Binning

$$I(X;U) - I(U;Y)$$



- $2^{nI(U;Y)}$ codewords per bin: the decoder can detect the codeword inside the bin based on the side information Y^n

Wyner-Ziv Problem

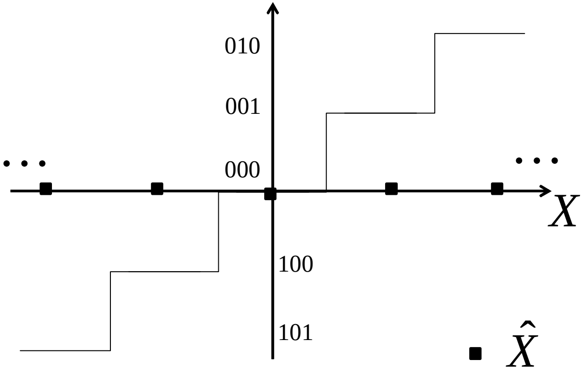
Achievability: 3. Estimate

- Final estimate: symbol-by-symbol function $\hat{X}_i = f(U_i, Y_i)$
- Rate-distortion function:

$$R_{X|Y}^{WZ}(D) = \min_{p(u|x), f(u,y)} I(X; U|Y)$$
$$\text{s.t. } E[d(X, Y, f(U, Y))] \leq D$$

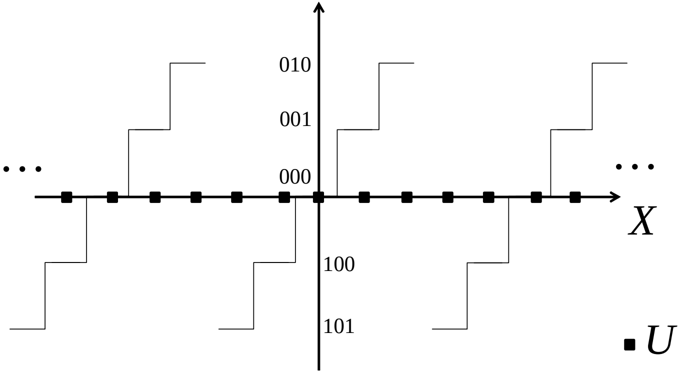
Wyner-Ziv Problem

A One-Dimensional View



Wyner-Ziv Problem

A One-Dimensional View



Wyner-Ziv Problem

Summary

- If Y^n is known at the encoder

$$R_{X|Y}(D) = \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y) \quad \text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

- Achievability: Delta compression or mux/demux

- If Y^n not known at the encoder

$$R^{WZ}_{X|Y}(D) = \min_{p(u|x),f(u,y)} I(X;U|Y) \quad \text{s.t. } E[d(X,Y,f(U,Y))] \leq D$$

- Achievability: Quantization, binning and estimate

Wyner-Ziv Problem

Summary

- If Y^n is known at the encoder

$$R_{X|Y}(D) = \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y) \quad \text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

- Achievability: Delta compression or mux/demux

- If Y^n not known at the encoder

$$R^{WZ}_{X|Y}(D) = \min_{p(u|x),f(u,y)} I(X;U|Y) \quad \text{s.t. } E[d(X,Y,f(U,Y))] \leq D$$

- Achievability: Quantization, binning and estimate

Wyner-Ziv Problem

- As seen, we have $R_{X|Y}(0) = R^{WZ}_{X|Y}(0)$
- The property extends to Gaussian sources with MSE distortion [Wyner '78]; and binary sources with erased side information and Hamming/erasure distortion [Diggavi et al '09] [Weissman and Verdú '08]

Wyner-Ziv Problem

Computation of the Rate-Distortion Function

- Introducing Shannon strategies: $T : \mathscr{Y} \to \hat{\mathscr{X}}$, we can write

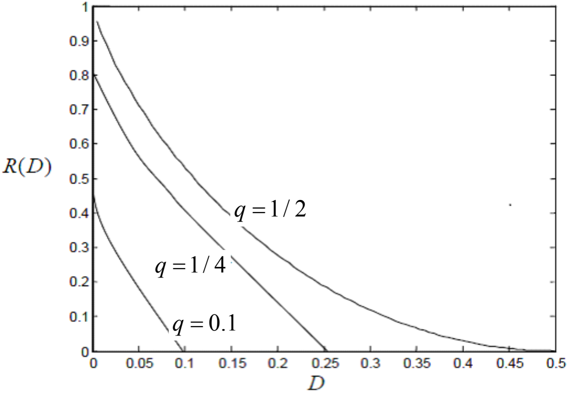
$$R_{X|Y}^{WZ}(D) = \min_{p(t|x)} I(X;T|Y)$$
$$\text{s.t. } E[d(X,Y,T(Y))] \leq D$$

- Convex problem in $p(t|x)$
- Can be solved using alternating optimization *à la* Blahut-Arimoto [Dupuis et al '04]

Wyner-Ziv Problem

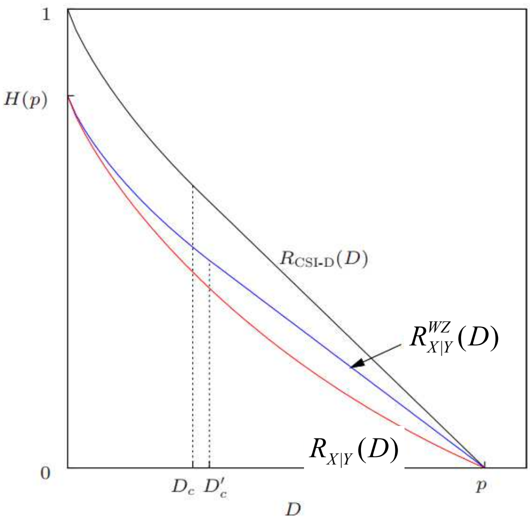
Example

- DSBS(q) with Hamming distortion
- $R^{WZ}_{X|Y}(D) = l.c.e.\{H(q*D) - H(D), (q,0)\}$ for $D \leq q$



Wyner-Ziv Problem

Example



$(q = 1/4)$

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

Code Design

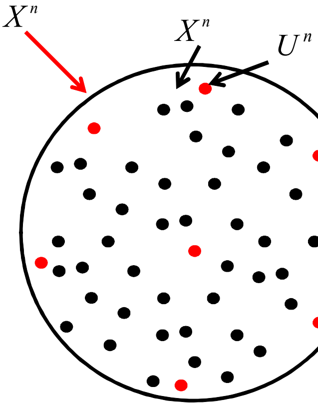
From Theory To Implementation

"In theory, theory and practice are the same. In practice, they are not."

(A. Einstein)

Code Design

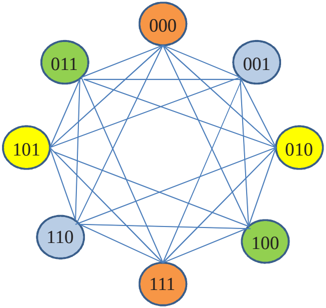
- Quantization, Binning, Estimation



- Random codes lead to exponential quantization and decoding complexities
- Goal: Design structured codes with feasible quantization/decoding

Code Design

- 1. Lossless case: Binning via syndromes and coset decoding

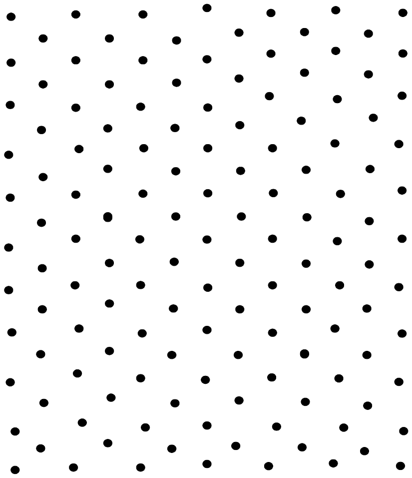


- 2. Lossy case, no side information: Quantization as decoding
- 3. Lossy case, side information: Quantization, binning and estimation

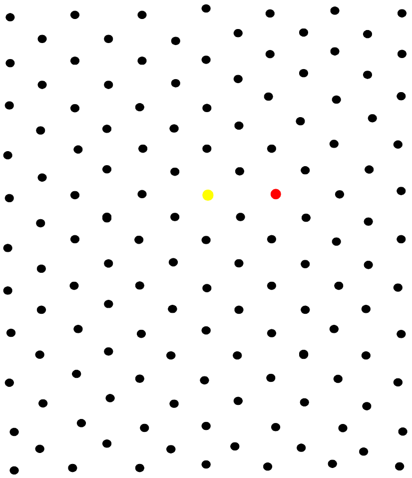
Lossless Case

- DSBS(q) with Hamming distortion
- With $D = 0$, no need for quantization ($U = X$)
- $R^{WZ}_{X|Y}(0) = H(X|Y) = H(q)$

Lossless Case



Lossless Case



- X^n in red, Y^n in yellow

Lossless Case

From the Theory

- As seen, from the packing lemma, the decoder can decode within a bin of $2^{nI(X;Y)} = 2^{n(1-H(q))}$ randomly selected sequences X^n
- The number of bins is $2^n/2^{n(1-H(q))} = 2^{nH(q)}$
- Rate of the message $R = H(q) = R^{WZ}_{X|Y}(0)$
- Linear codes are optimal for communication over a BSC [Gallager '68]
- ... therefore, bins can be constructed from linear codes

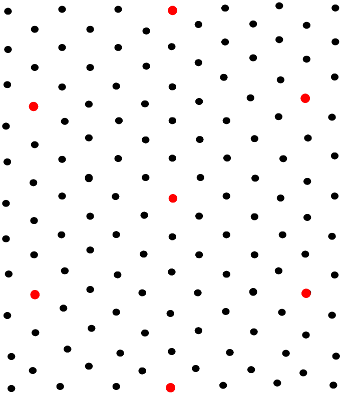
Lossless Case

From the Theory

- As seen, from the packing lemma, the decoder can decode within a bin of $2^{nI(X;Y)} = 2^{n(1-H(q))}$ randomly selected sequences X^n
- The number of bins is $2^n / 2^{n(1-H(q))} = 2^{nH(q)}$
- Rate of the message $R = H(q) = R^{WZ}_{X|Y}(0)$

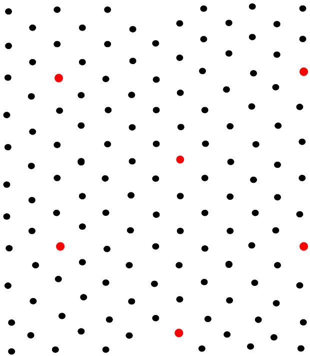
- Linear codes are optimal for communication over a BSC [Gallager '68]
- ... therefore, bins can be constructed from linear codes

Lossless Case



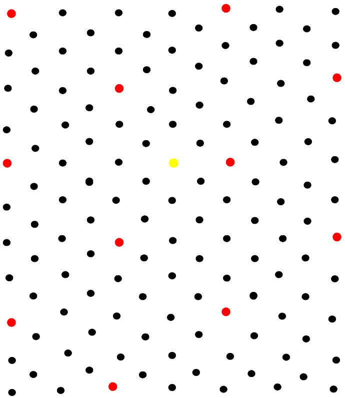
- Fix a linear code ($(n-k) \times n$ parity matrix H) with $k/n = 1 - H(q)$

Lossless Case



- ... and cosets
- Described by syndrome $s^{n-k} = Hx^n$ and by the corresponding coset leader $f(s^{n-k})$
- The coset leader $f(s^{n-k})$ is the offset with the minimal number of ones

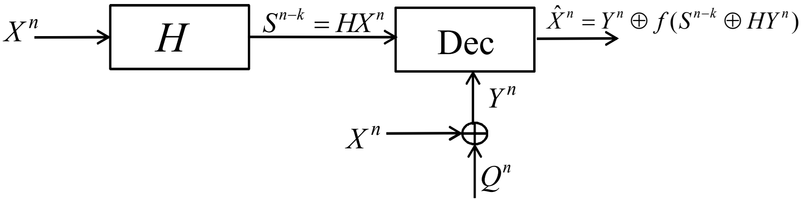
Lossless Case



- The encoder informs the decoder about the syndrome $s^{n-k} = Hx^n$, and hence about the bin
- The decoder decodes inside the bin

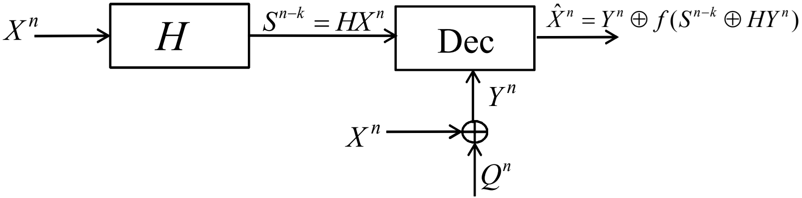
Lossless Case

- To decode, calculate $HY^n \oplus S^{n-k} = H(Y^n \oplus X^n) = HQ^n$ estimate Q^n and then calculate $\hat{X}^n = Y^n \oplus \hat{Q}^n$
- If $\frac{n-k}{n} > H(q)$, then from HQ^n, we can recover Q^n as $f(HQ^n)$ with high probability [Ancheta '76]



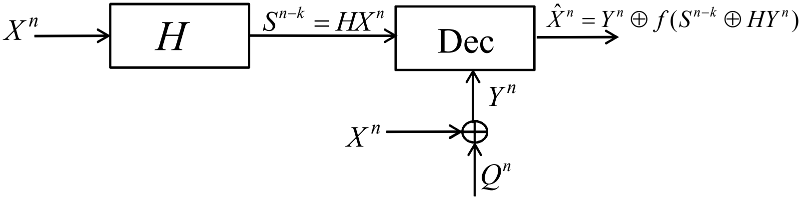
- Linear encoding
- Decoder still has an exponential complexity

Lossless Case, Side Information



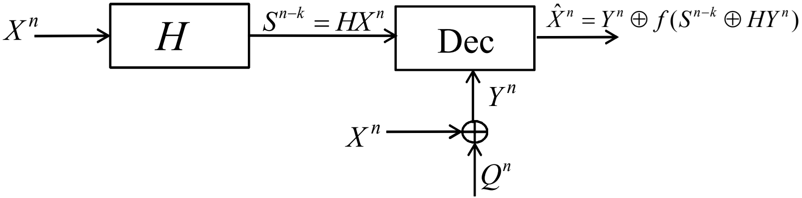
- The complexity at the decoder can be drastically reduced by adopting specific classes of linear codes
- LDPC codes are capacity-achieving for BSC and are known to be efficiently decoded using message passing [MacKay '99]
- Polar codes are capacity-achieving for BSC and can be efficiently decoded via successive decoding [Arıkan '09]

Lossless Case, Side Information



- The complexity at the decoder can be drastically reduced by adopting specific classes of linear codes
- LDPC codes are capacity-achieving for BSC and are known to be efficiently decoded using message passing [MacKay '99]
- Polar codes are capacity-achieving for BSC and can be efficiently decoded via successive decoding [Arıkan '09]

Lossless Case, Side Information

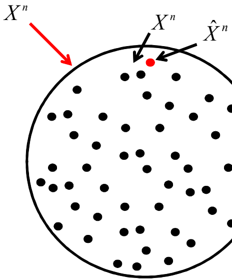


- The complexity at the decoder can be drastically reduced by adopting specific classes of linear codes
- LDPC codes are capacity-achieving for BSC and are known to be efficiently decoded using message passing [MacKay '99]
- Polar codes are capacity-achieving for BSC and can be efficiently decoded via successive decoding [Arıkan '09]

Lossy Case, No Side Information

From the Theory

- No need for binning ($U = \hat{X}$)
- Random coding



- For BSS: $R_X(D) = 1 - H(D)$ for $0 \leq D \leq 1/2$ and $R_X(D) = 0$ otherwise

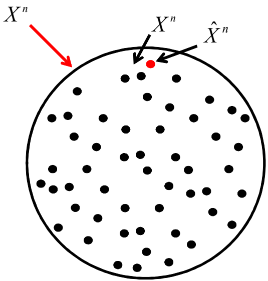
Lossy Case, No Side Information

- Can linear encoders be optimal for $D > 0$ (as in the lossless case)?
- No, non-linear operations are generally necessary [Ancheta '76] [Massey '78]

Lossy Case, No Side Information

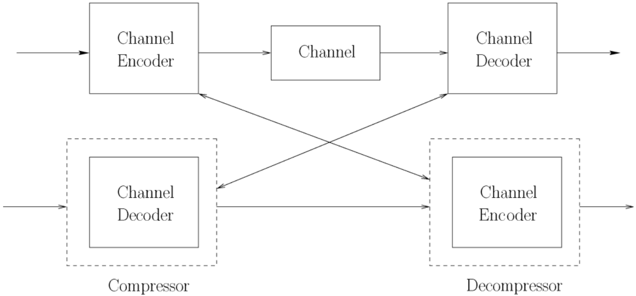
- Can linear encoders be optimal for $D > 0$ (as in the lossless case)?
- No, non-linear operations are generally necessary [Ancheta '76] [Massey '78]

Lossy Case, No Side Information



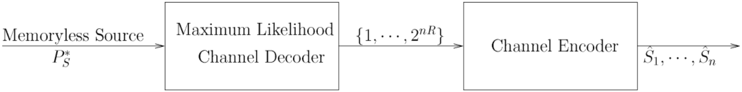
- ... Quantization is akin to channel decoding on the test channel $p(x|\hat{x})$!
- (cf. binning where channel decoding is performed at the decoder's side)

Lossy Case, No Side Information



[Gupta and Verdú '11]

Lossy Case, No Side Information



[Gupta and Verdú '11]

- Source and channel coding problems related by functional duality [Pradhan et al '03]

Lossy Case, No Side Information

- The dual channel is the test channel $p(x|\hat{x})$
- Example 1: Binary source $Ber(1/2)$ with Hamming distortion $D \leq p$
- The dual channel coding problem is: BSC $X = \hat{X} \oplus Z$ with $Z \sim Ber(D)$
- Example 2: $X_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, MSE distortion $D \leq \sigma^2$
- The dual channel coding problem is: $X = \hat{X} + Z$ with $Z \sim \mathcal{N}(0, D)$ and cost constraint $E[\hat{X}^2] \leq \sigma^2 - D$

Lossy Case, No Side Information

- The dual channel is the test channel $p(x|\hat{x})$
- Example 1: Binary source $Ber(1/2)$ with Hamming distortion $D \leq p$
- The dual channel coding problem is: BSC $X = \hat{X} \oplus Z$ with $Z \sim Ber(D)$
- Example 2: $X_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, MSE distortion $D \leq \sigma^2$
- The dual channel coding problem is: $X = \hat{X} + Z$ with $Z \sim \mathcal{N}(0, D)$ and cost constraint $E[\hat{X}^2] \leq \sigma^2 - D$

Lossy Case, No Side Information

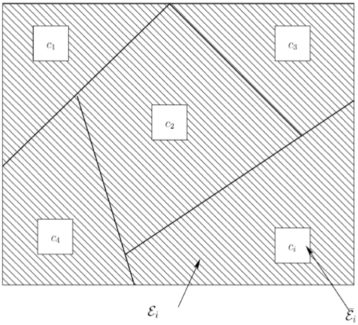
- Capacity-achieving codes exist whose ML channel decoder for the dual channel is an optimal quantizer [Gupta and Verdú '11]
- LDPC codes achieve the rate-distortion bound with ML decoder as encoder [Matsunaga and Yamamoto '03]

Lossy Case, No Side Information

- Capacity-achieving codes exist whose ML channel decoder for the dual channel is an optimal quantizer [Gupta and Verdú '11]
- LDPC codes achieve the rate-distortion bound with ML decoder as encoder [Matsunaga and Yamamoto '03]

Lossy Case, No Side Information

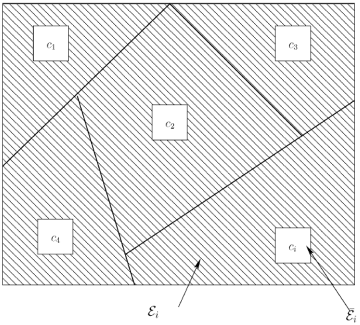
- Not *every* optimal channel decoder is an optimal quantizer [Gupta and Verdú '11]
- (see also [Csiszár and Körner '11] proof of rate-distortion theorem)



- Message passing decoders fail as quantizers with LDPC codes [Martinian and Yedidia '03] (see Appendix-C)

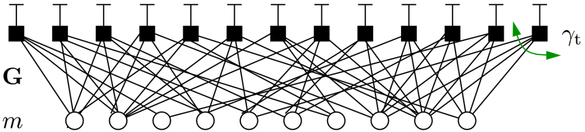
Lossy Case, No Side Information

- Not *every* optimal channel decoder is an optimal quantizer [Gupta and Verdú '11]
- (see also [Csiszár and Körner '11] proof of rate-distortion theorem)



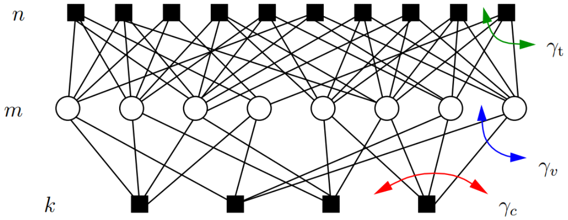
- Message passing decoders fail as quantizers with LDPC codes [Martinian and Yedidia '03] (see Appendix-C)

Lossy Case, No Side Information



- Low Density Generator Matrix (LDGM) codes achieve the rate-distortion function for BSS with Hamming distortion [Sun et al '10]
- LDGM codes allow low-complexity encoding via message-passing-type algorithm [Wainwright et al '10] [Sun et al '10]
- Unlike channel coding, many codewords "close" to to the source signal $\rightarrow$ need modifications of the message passing strategy

Lossy Case, No Side Information



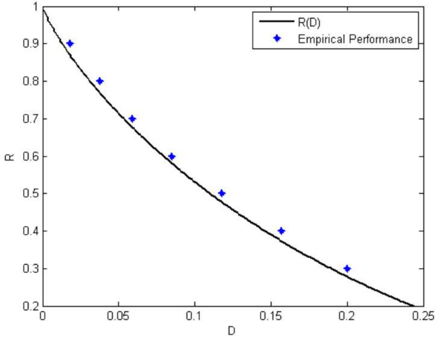
- Performance of LDGM codes with finite check degree bounded as [Dimakis et al '07] [Kudekar and Urbanke '08]: $R \geq R(D)/(1 - e^{-(1-D)\gamma/R})$
- Compound LDGM-LDPC codes achieve the rate-distortion function with bounded degrees [Martinian and Wainwright '06]

Lossy Case, No Side Information

- Polar codes achieve the rate-distortion bound with successive encoding [Korada and Urbanke '10]
- For more general alphabet and non-uniform distributions, non-linear codes obtained via multilevel mappings [Sun et al '10] or by limiting the set of w's [Gupta and Verdú '09]

Lossy Case, No Side Information

- BSS with Hamming distortion

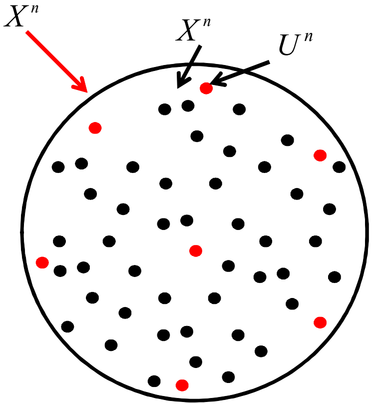


[Wainwright and Maneva '05]

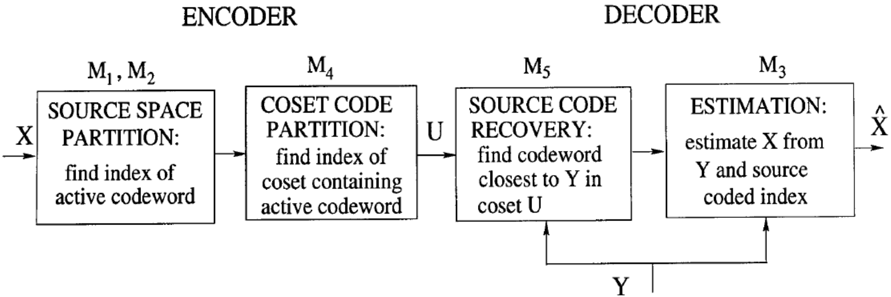
Lossy Case, Side Information

From the Theory

- Quantization, Binning, Estimation



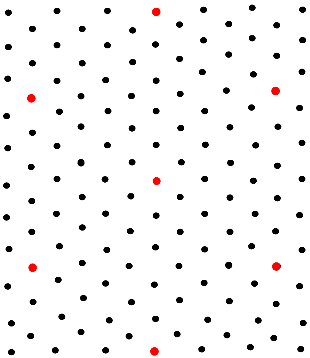
Lossy Case, Side Information



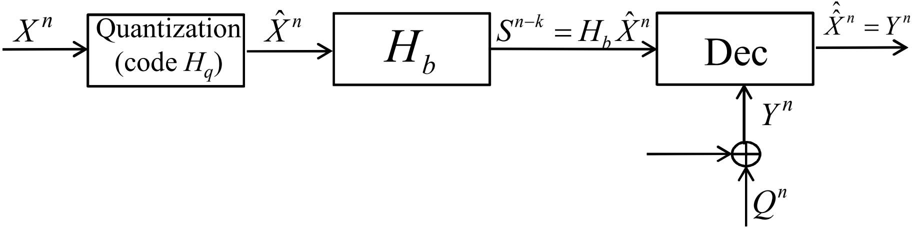
[Pradhan and Ramchandran '03]

Lossy Case, Side Information

- Quantization: linear code – $(n-k_q)\times n$ parity matrix H_q $(n-k_q)\times n$ with $\frac{k_q}{n}=I(X;U)=1-H(D)$
- Binning: nested linear code [Zamir et al '02] – $(n-k_b)\times n$ parity matrix $H_b=\begin{bmatrix} H_q \\ \Delta H_b \end{bmatrix}$ with $\frac{k_b}{n}=I(U;Y)=1-H(D*q)$



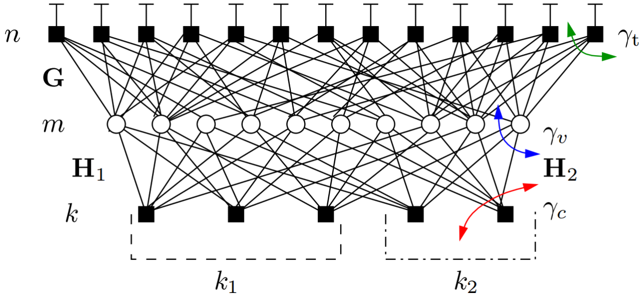
Lossy Case, Side Information



(estimation step dealt with via time-sharing)

Lossy Case, Side Information

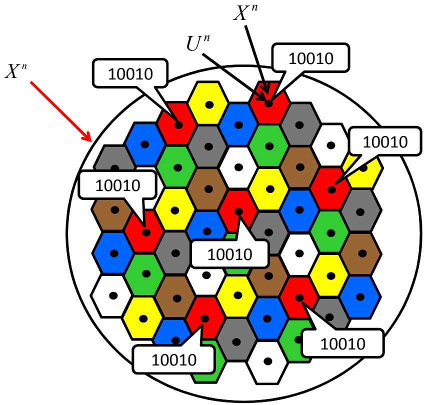
- LDGM do not provide good channel codes
- Compound LDGM-LDPC code achieves the rate-distortion function with side information with bounded degrees [Martinian and Wainwright '06]



- Polar codes [Korada and Urbanke '10]

Lossy Case, Side Information

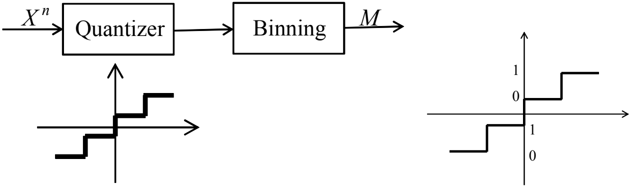
- Vector quantization and vector binning: nested lattice codes [Zamir et al '02] (see also [Fleming et al '04] [Saxena et al '10])



- Fine code good for quantization and coarse code good for channel coding [Zamir et al '02]

Lossy Case, Side Information

- Gaussian sources with MSE distortion
- Scalar quantizer followed by binning



- Scalar quantization followed by vector binning is optimal at high rate [Liu et al '06] (cf. [Ziv '85])

Lossy Case, Side Information

- DISCUS based on trellis codes for quantization and binning [Pradhan and Ramchandran '03]
- TCQ for quantization and LDPC for binning [Yang et al '09]

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

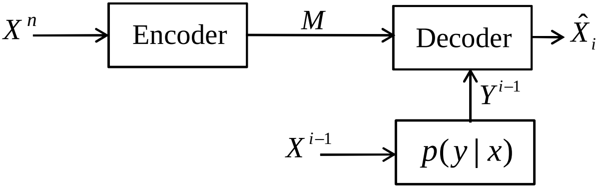
PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

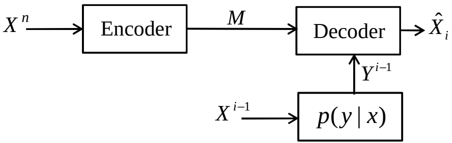
PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

Strictly Causal Side Information



Strictly Causal Side Information



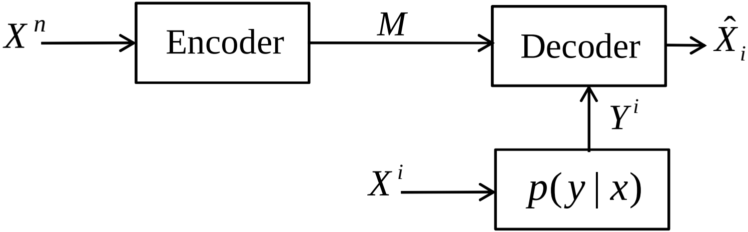
- Lower bound

$$R_{X|Y}^{WZ-C}(D) \geq \min_{p(\hat{x}|x)} I(X;\hat{X})$$
$$\text{s.t. } E[d(X,Y,\hat{X})] \leq D$$

- Proof: See Appendix-D
- ... delayed side information is not useful for memoryless sources!

Causal Side Information

Weissman and El Gamal '06



Causal Side Information

Weissman and El Gamal '06

- Lower bound:

$$R_{X|Y}^{WZ-C}(D) \geq \min_{p(u|x)} I(X;U)$$
$$\text{s.t. } E[d(X,Y,f(U,Y))] \leq D$$

- Proof: See Appendix-E

Causal Side Information

Weissman and El Gamal '06

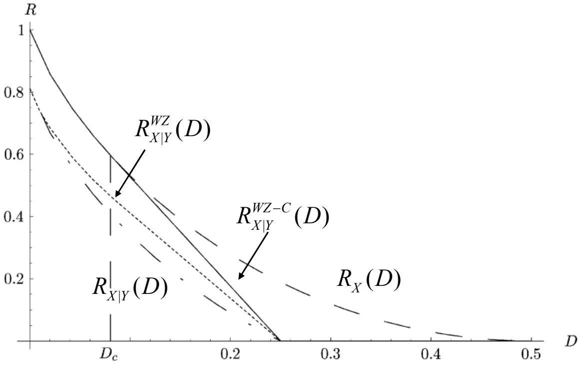
- Achievability via standard compression with inverse test channel $p(u|x)$ (... no need for binning!)
- Rate-distortion function

$$R_{X|Y}^{WZ-C}(D) = \min_{p(u|x)} I(X;U)$$
$$\text{s.t. } E[d(X,Y,f(U,Y))] \leq D$$

Causal Side Information

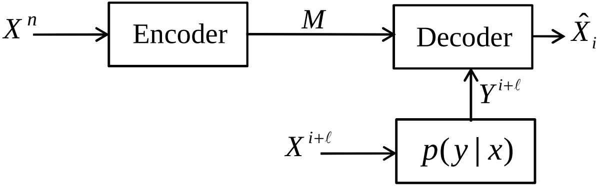
Weissman and El Gamal '06

- DSBS(1/4) with Hamming distortion



Limited Look-Ahead Side Information

Weissman and El Gamal '06



Limited Look-Ahead Side Information

Weissman and El Gamal '06

- Treat the source as a super-source with n/k symbols made of successive k-blocks of X^n
- Achievable rate:

$$R_k(D) = \frac{1}{k} \min_{p(u|x^k)} I(X^k; U)$$

s.t.

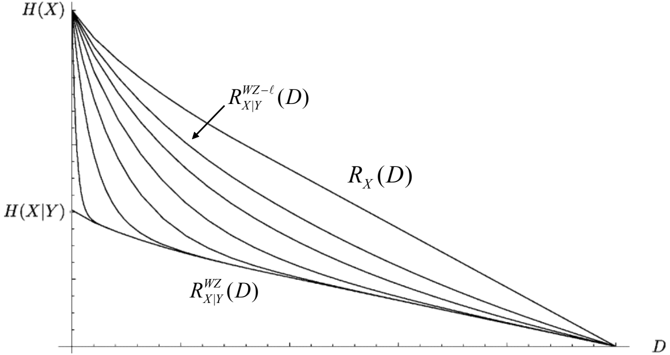
$$\frac{1}{k-\ell} \sum_{i=1}^{k-\ell} E[d(X_i, Y_i, f_i(U, Y^{i+\ell}))] \leq D$$

- Rate-distortion function

$$R_{X|Y}^{WZ-\ell}(D) = \lim_{k\to\infty} R_k(D)$$

Limited Look-Ahead Side Information

Weissman and El Gamal '06



Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

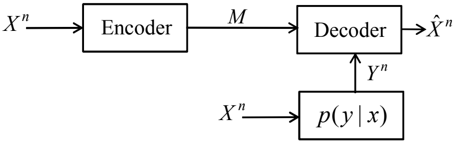
Common Reconstruction Constraint

- In the Wyner-Ziv problem, two roles for the side information sequence:

1. Reduce the rate required for "digital" communication between encoder and decoders via binning
2. Improve the source estimate

Common Reconstruction Constraint

Steinberg '09



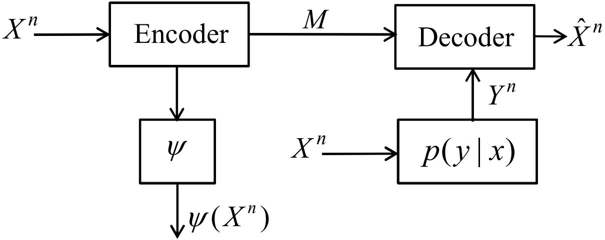
- Wyner-Ziv rate-distortion function

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

- $f(U,Y)$ cannot be reproduced at the encoder
- Might be unacceptable, e.g., for transmission of sensitive medical, financial or military information

Common Reconstruction Constraint

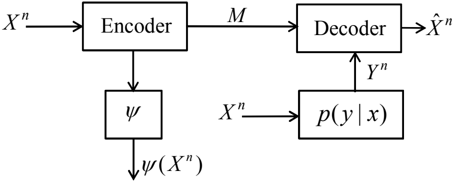
Steinberg '09



- Common Reconstruction (CR) constraint: $\hat{X}^n$ should be reproducible by the encoder ($\psi(X^n) = \hat{X}^n$ w.h.p.)

Common Reconstruction Constraint

Steinberg '09

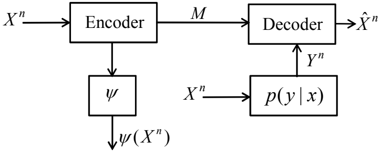


- Rate-distortion function:

$$R_{X|Y}^{WZ-CR}(D) = \min_{p(\hat{x}|x):E[d(X,\hat{X})]\leq D} I(X;\hat{X}|Y)$$

Common Reconstruction Constraint

Steinberg '09



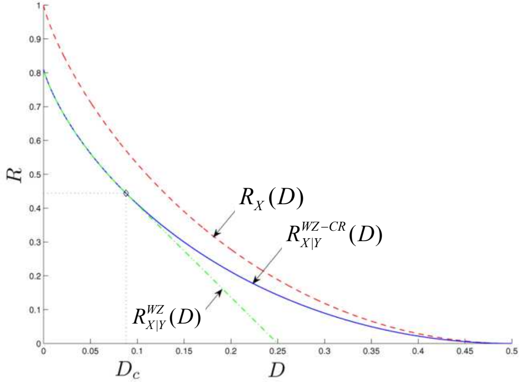
$$R_{X|Y}^{WZ-CR}(D) = \min_{p(\hat{x}|x):E[d(X,\hat{X})]\leq D} I(X;\hat{X}|Y)$$

- Achievability: 1. compression + 2. binning, but no estimate at the receiver
- Converse: Identify ψ_i with $\hat{X}_i$
- Dual to causal constraint
- Extension to distortion-based CR constraints [Lapidoth et al '11]

Common Reconstruction Constraint

Steinberg '09

- DSBS(1/4) with Hamming distortion



Overview

PART I: Basics

- The Wyner-Ziv problem
- Code design

PART II: Constraints

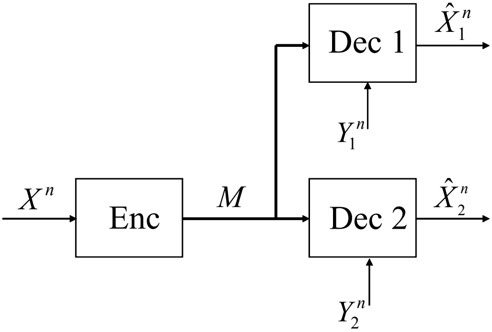
- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

Robust Source Coding

[Heegard and Berger '85] [Kaspi '94]



- Distortion constraints: $\frac{1}{n}\sum_{i=1}^{n} E[d_1(X_i, \hat{X}_{1i})] \leq D_1$ and $\frac{1}{n}\sum_{i=1}^{n} E[d_2(X_i, \hat{X}_{2i})] \leq D_2$
- Sensor network with unreliable remote measurements, file sharing with unreliable remote file transfers

Robust Source Coding

[Heegard and Berger '85] [Kaspi '94]

- Assume that Y_1 is (physically/stochastically) degraded side information with respect to Y_2: $p(x, y_2)p(y_1|y_2)$
- Rate-distortion function

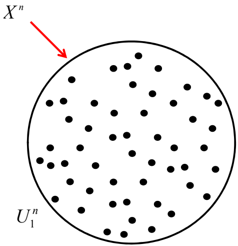
$$R^{HB}_{X|Y_1Y_2}(D_1, D_2) = \min_{p(u_1,u_2|x),\mathrm{f}_j(u_j,y_j)} I(X;U_1|Y_1) + I(X;U_2|Y_2,U_1)$$

$$\text{s.t. } \mathsf{E}[d_1(X,\mathrm{f}_j(U_j,Y_j))] \le D_j, \text{ for } j = 1,2$$

Robust Source Coding

Achievability: 1. Quantization

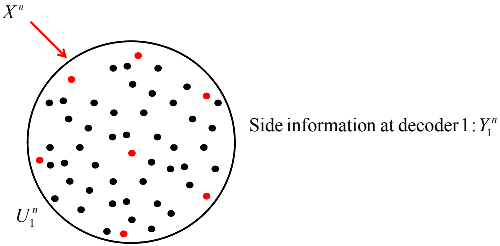
$$I(X;U_1|Y_1) + I(X;U_2|Y_2,U_1)$$



Robust Source Coding

Achievability: 2. Binning

$$\color{red}{I(X;U_1|Y_1)} + I(X;U_2|Y_2,U_1)$$

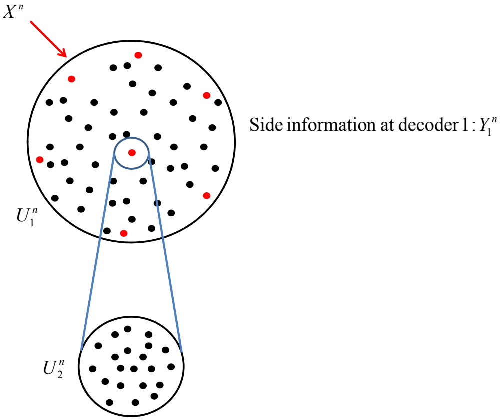


- Since $\color{red}{I(X;U_1|Y_1)} \geq I(X;U_1|Y_2)$, decoder 2 can also decode

Robust Source Coding

Achievability: 3. Successive refinement

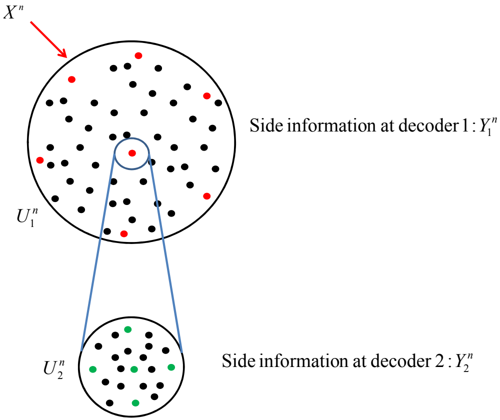
$$I(X;U_1|Y_1)+I(X;U_2|Y_2,U_1)$$



- $I(X;U_2|Y_2,U_1)=I(X;U_2|U_1)-I(U_2;Y_2|U_1)$

Robust Source Coding

Achievability: 4. Binning



- $I(X;U_2|Y_2,U_1) = I(X;U_2|U_1) - I(U_2;Y_2|U_1)$

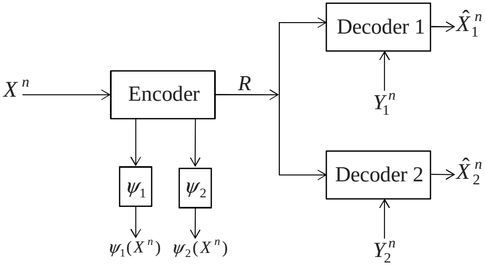
Robust Source Coding

Achievability: 5. Estimate

- Final estimates: symbol-by-symbol function $\hat{X}_1 = f_1(U_1, Y_1)$ and $\hat{X}_2 = f_2(U_2, Y_2)$

HB Problem with CR constraint

Ahmadi et al '13



- CR constraints

$$\psi_j(X^n) = \hat{X}_j^n \qquad , j = 1, 2 \text{ w.h.p.}$$

- If the side information Y_1 is stochastically degraded with respect to Y_2, the rate-distortion function for the HB problem with CR is given by

$$R^{HB}_{X|Y_1Y_2}(D_1, D_2) = \min_{p(\hat{x}_1,\hat{x}_2|x)} I(X;\hat{X}_1|Y_1) + I(X;\hat{X}_2|Y_2,\hat{X}_1),$$

s.t.

$$\mathrm{E}[d_j(X,\hat{X}_j)] \leq D_j, \text{ for } j = 1,2$$

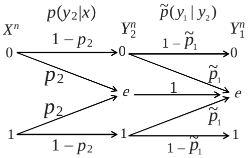
HB Problem with CR constraint

Sketch of Proof of Achievability

- As for conventional HB, successive refinement and binning
- ... but $\hat{X}_1^n$ and $\hat{X}_2^n$ generated by the encoder (no estimation) to satisfy the CR constraint

HB Problem with CR constraint

Binary Source with Erased Side Information



- Binary source $X \sim \mathrm{Ber}(\frac{1}{2})$
- For point-to-point set-up (erasure probability p) under Hamming distortion

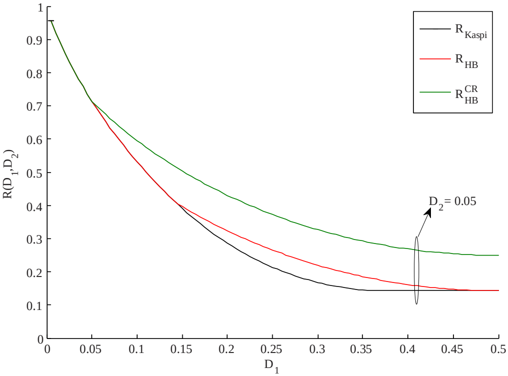
$$R_{X|Y}^{WZ-CR}(D) = p(1 - H(D)) \qquad \text{for } D \leq 1/2$$

and zero otherwise

- With no CR constraint:
$R_{X|Y}^{WZ}(D) = R_{X|Y}(D) = p(1 - H(D/p))$ for $D \leq p/2$ and zero otherwise [Diggavi et al '09]

HB Problem with CR constraint

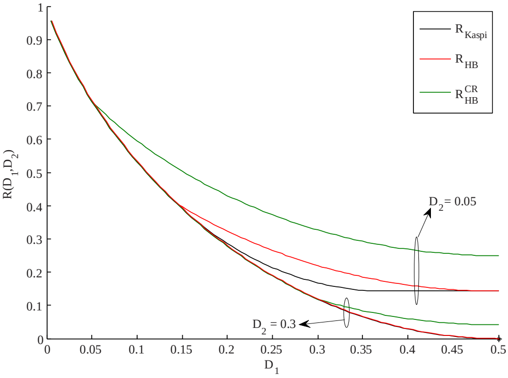
Binary Source with Erased Side Information



- $p_1 = 1$ (Decoder 1 has no side information) and $p_2 = 0.35$, and two values of the distortion $D_2 = 0.05, 0.3$
- $R_{X|Y_1,Y_2}(D_1, D_2) \leq R^{HB}_{X|Y_1 Y_2}(D_1, D_2) \leq R^{HB-CR}_{X|Y_1,Y_2}(D_1, D_2)$

HB Problem with CR constraint

Binary Source with Erased Side Information



- $p_1 = 1$ (Decoder 1 has no side information) and $p_2 = 0.35$, and two values of the distortion $D_2 = 0.05, 0.3$
- $R_{X|Y_1,Y_2}(D_1, D_2) \leq R^{HB}_{X|Y_1Y_2}(D_1, D_2) \leq R^{HB-CR}_{X|Y_1,Y_2}(D_1, D_2)$

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

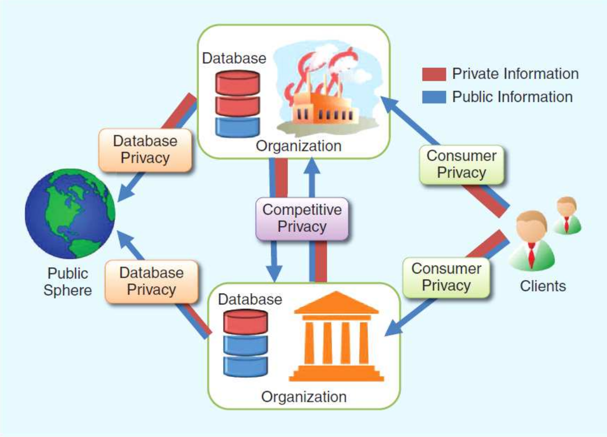
PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

Privacy


GET ALL THE INFORMATION YOU CAN, WE'LL THINK OF A USE FOR IT LATER.

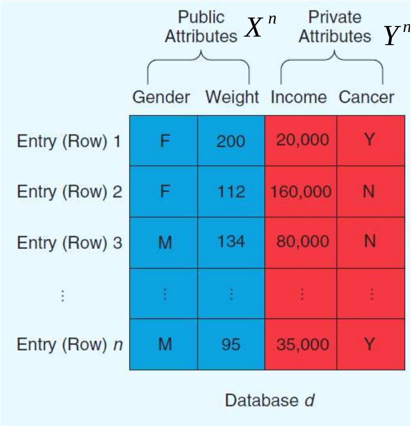
Utility vs. Privacy



[Sankar et al '13]

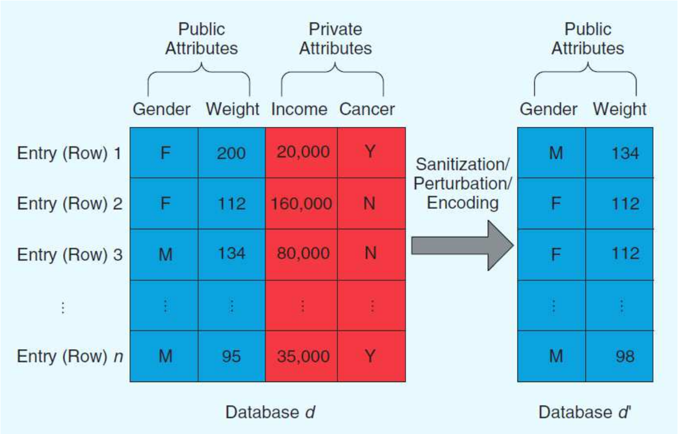
- E.g., de-anonymizing the Netflix dataset: movie ratings can be used to infer a user's political affiliation, gender, etc.

Utility vs. Privacy



[Sankar et al '13]

Utility vs. Privacy

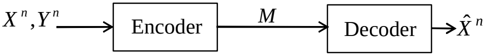


[Sankar et al '13]

- Privacy-preserving transformation

Source Coding with Privacy Constraints

Yamamoto '83



- Average per-block distortion:

$$\frac{1}{n}\sum_{i=1}^{n} E[d(X_i, \hat{X}_i)] \leq D$$

- Equivocation (privacy) constraint:

$$\frac{1}{n}H(Y^n|M) \geq E$$

Source Coding with Privacy Constraints

Yamamoto '83

- Rate-distortion-equivocation function:

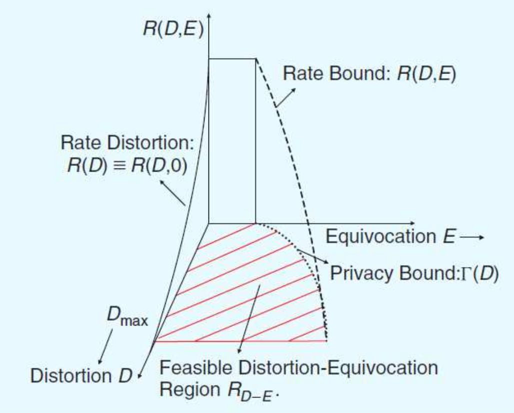
$$R(D,E) = \min_{p(\hat{x}|x,y)} I(X,Y;\hat{X})$$

s.t.

$$E[d(X,X)] \leq D$$
$$H(Y|\hat{X}) \geq E$$

- The optimal privacy-preserving transformation is quantization

Source Coding with Privacy Constraints



Source Coding with Privacy Constraints

- A number of extensions and variations on the theme: [Gunduz et al '08] [Villard and Piantanida '10] [Tandon et al '13],...

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

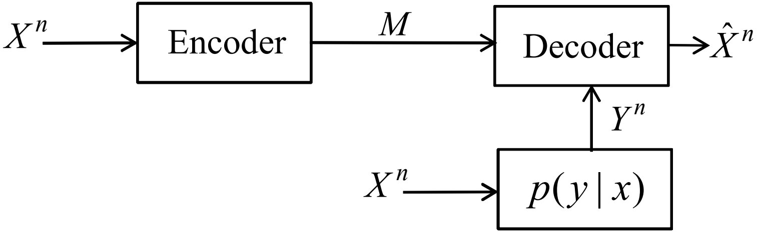
PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

Side Information "Vending Machine"

Permuter and Weissman '11

- Wyner-Ziv problem



Side Information "Vending Machine"

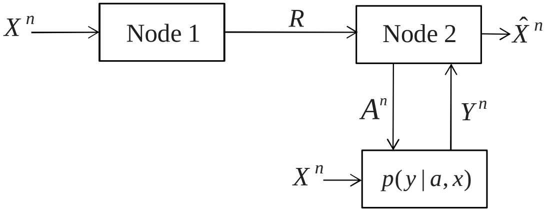
Permuter and Weissman '11

- Generalization of the Wyner-Ziv problem where the side information can be controlled via cost-constrained actions

Side Information "Vending Machine"

Permuter and Weissman '11

- Generalization of the Wyner-Ziv problem where the side information can be controlled via cost-constrained actions



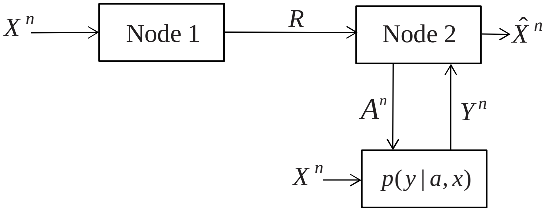
- Action $A^n(M)$ is cost constrained as $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\Lambda(A_i)] \leq \Gamma$

- Ex.: If $A = 1$, $Y = X + Z$; and if $A = 0$, $Y = \phi$. With $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[A_i] \leq \Gamma$

Side Information "Vending Machine"

Permuter and Weissman '11

- Generalization of the Wyner-Ziv problem where the side information can be controlled via cost-constrained actions

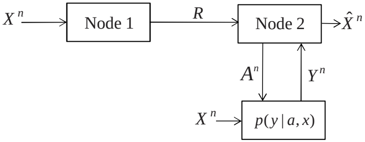


- Action $A^n(M)$ is cost constrained as $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\Lambda(A_i)] \leq \Gamma$

- Ex.: If $A = 1$, $Y = X + Z$; and if $A = 0$, $Y = \phi$. With $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[A_i] \leq \Gamma$

Side Information "Vending Machine"

Permuter and Weissman '11



- Key issue: Adaptive vs. non-adaptive data acquisition
- Rate-distortion-cost function

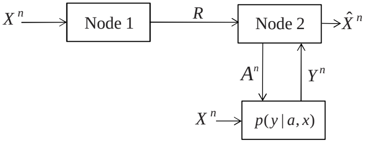
$$R(D,\Gamma) = \min_{p(a,u|x),\, f(U,Y)} I(X;A) + I(X;U|A,Y)$$

subject to $\mathrm{E}[\Lambda(A)] \leq \Gamma$ and $\mathrm{E}[d(X,f(U,Y))] \leq D$

- Result extends to actions $A_i(M, Y^{i-1})$ [Choudhuri and Mitra '12]

Side Information "Vending Machine"

Permuter and Weissman '11



- Key issue: Adaptive vs. non-adaptive data acquisition
- Rate-distortion-cost function

$$R(D,\Gamma) = \min_{p(a,u|x),\, f(U,Y)} I(X;A) + I(X;U|A,Y)$$

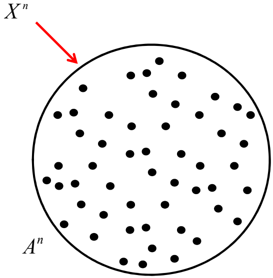
subject to $\mathrm{E}[\Lambda(A)] \leq \Gamma$ and $\mathrm{E}[d(X,f(U,Y))] \leq D$

- Result extends to actions $A_i(M,Y^{i-1})$ [Choudhuri and Mitra '12]

Side Information "Vending Machine"

Achievability: 1. Quantization

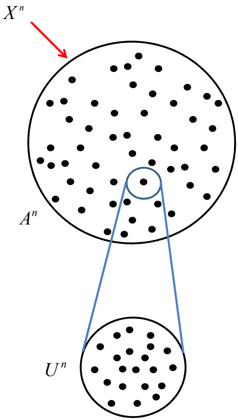
$$I(X;A) + I(X;U|A,Y)$$



Side Information "Vending Machine"

Achievability: 2. Successive refinement

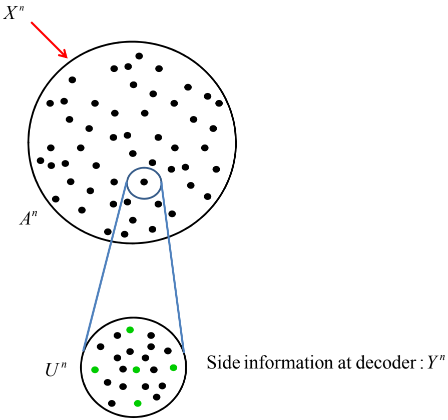
$$I(X;A) + I(X;U|A,Y)$$



- $I(X;U|A,Y) = I(X;U|A) - I(U;Y|A)$

Side Information "Vending Machine"

Achievability: 3. Binning



- $I(X;U|A,Y) = I(X;U|A) - I(U;Y|A)$

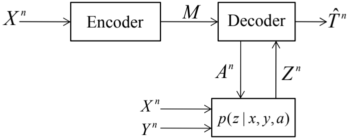
Side Information "Vending Machine"

Achievability: 4. Estimate

- Final estimate $\hat{X} = f(U, Y)$.
- ... Adaptive data acquisition

Side Information "Vending Machine"

Comparison with Non-Adaptive Approach



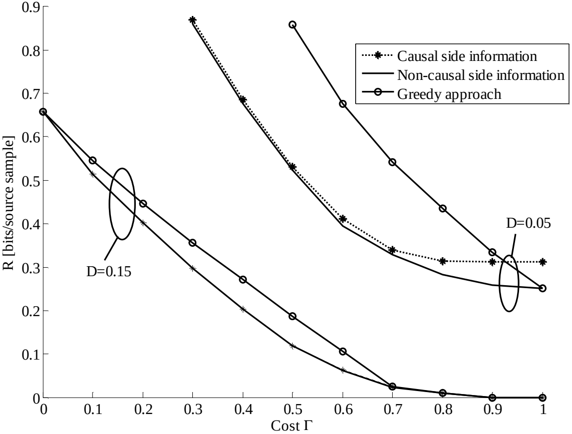
- (X, Y) DSBS(q)
- $T = \mathrm{f}(X, Y) = X \otimes Y$ (binary product)
- Measurement of Z_i as $p(z|x, y, a)$ (to measure or not to measure)

$$Z_i = \begin{cases} Y_i & \text{if} \quad A_i = 1 \\ 1 & \text{if} \quad A_i = 0 \end{cases}$$

- Cost constraint: $\Lambda(A_i) = 1$ if $A_i = 1$ and $\Lambda(A_i) = 0$ if $A_i = 0$

Side Information "Vending Machine"

Comparison with Non-Adaptive Approach



- "Greedy" approach: choose A_i independently of M
- For $\Gamma = 1$, the greedy approach is clearly optimal
- Joint design of control and compression

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

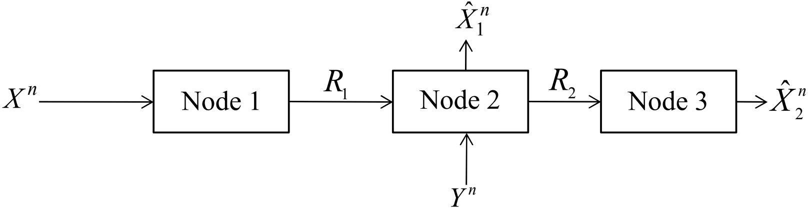
PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

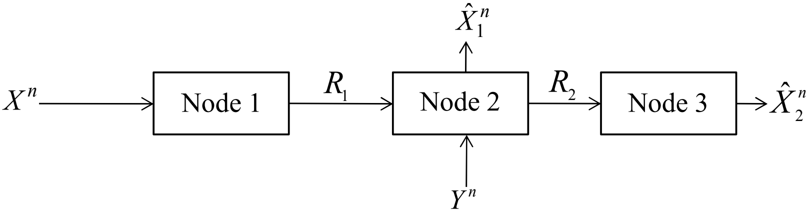
- Action-dependent side information
- Cascade problems
- Sources with memory

Cascade Source Coding



- Fixed-length codes: $M_1 \in \{1, 2, ..., 2^{nR_1}\}$ (nR_1 bits), $M_2 \in \{1, 2, ..., 2^{nR_2}\}$ (nR_2 bits),
- $\frac{1}{n}\sum_{i=1}^{n} E[d_j(X_i, Y_i, \hat{X}_{ji})] \leq D_j$ for $j = 1, 2$.
- Sensor networks, computer networks

Cascade Source Coding

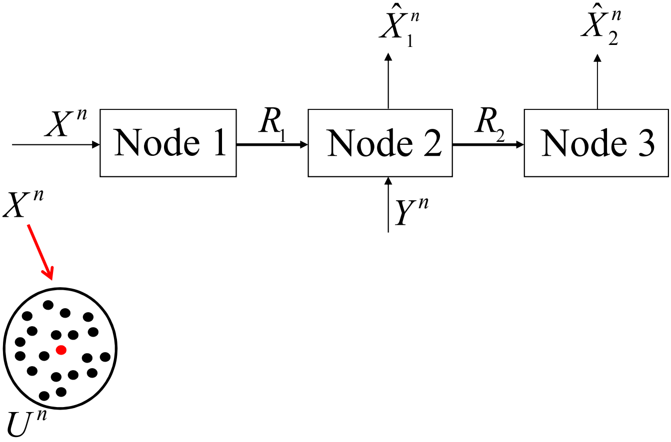


- The problem of determining the rate-distortion region is still open
- The main issue is the operation at Node 2 [Vasudevan et al '06] [Gu and Effros '06] [Bakshi et al '07] [Cuff et al '09]

Cascade Source Coding

Idea 1: Recompress [Cuff et al '09]

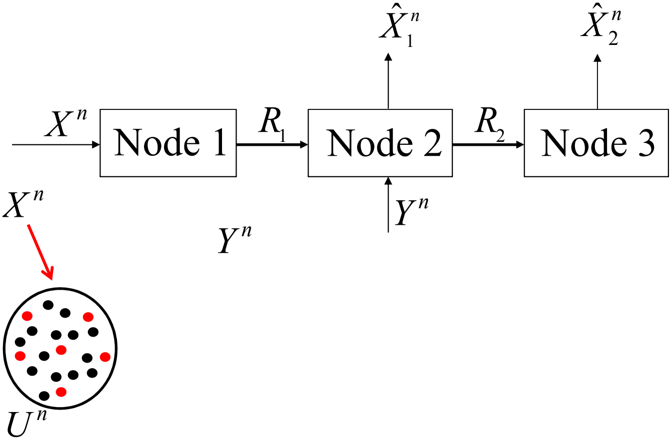
- $R_1 \geq I(X;U|Y)$
- $R_2 \geq I(Y,U;\hat{X}_2)$



Cascade Source Coding

Idea 1: Recompress [Cuff et al '09]

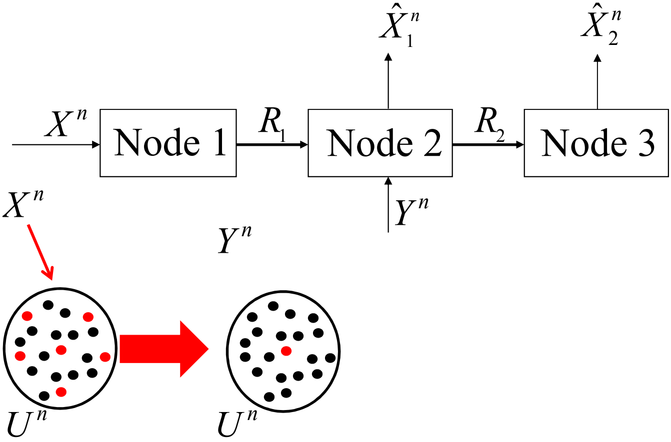
- $R_1 \geq I(X; U|Y)$
- $R_2 \geq I(Y, U; \hat{X}_2)$



Cascade Source Coding

Idea 1: Recompress [Cuff et al '09]

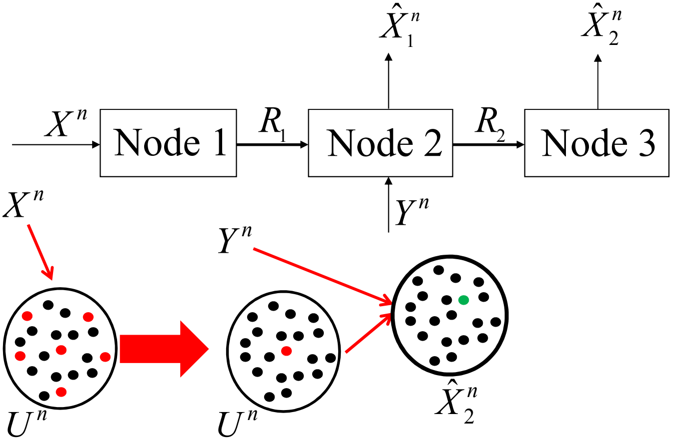
- $R_1 \geq I(X;U|Y)$
- $R_2 \geq I(Y,U;\hat{X}_2)$



Cascade Source Coding

Idea 1: Recompress [Cuff et al '09]

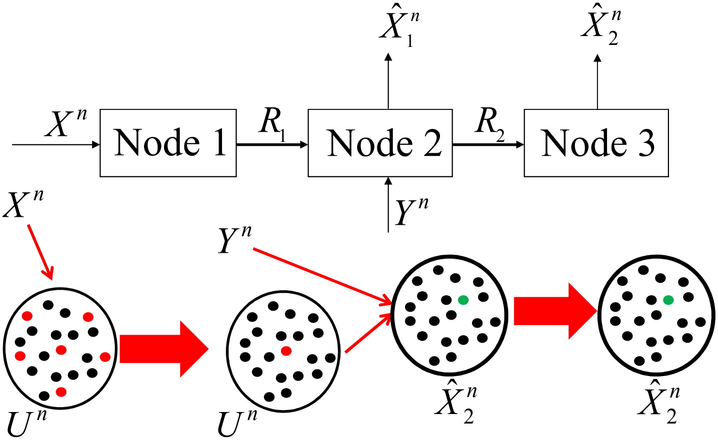
- $R_1 \geq I(X;U|Y)$
- $R_2 \geq I(Y,U;\hat{X}_2)$



Cascade Source Coding

Idea 1: Recompress [Cuff et al '09]

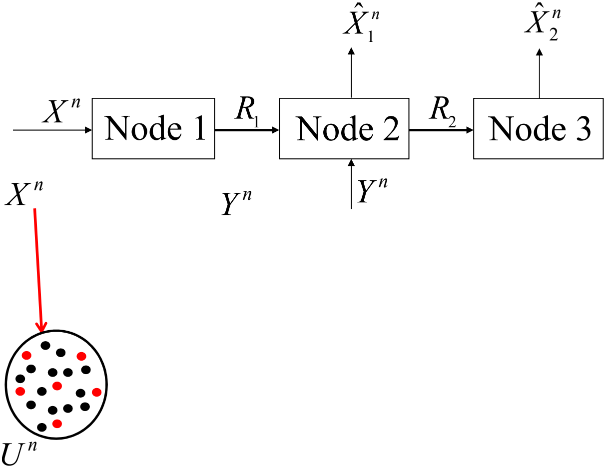
- $R_1 \geq I(X; U|Y)$
- $R_2 \geq I(Y, U; \hat{X}_2)$



Cascade Source Coding

Idea 2: Forward [Cuff et al '09]

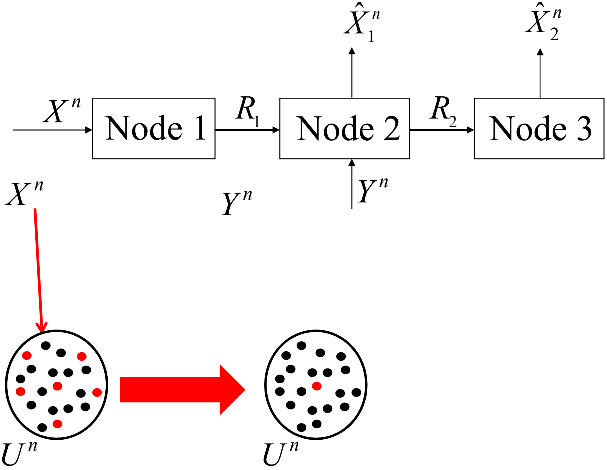
- $R_1 \geq I(X;U|Y)$
- $R_2 \geq I(X;U) + I(Y;\hat{X}_2|U)$



Cascade Source Coding

Idea 2: Forward [Cuff et al '09]

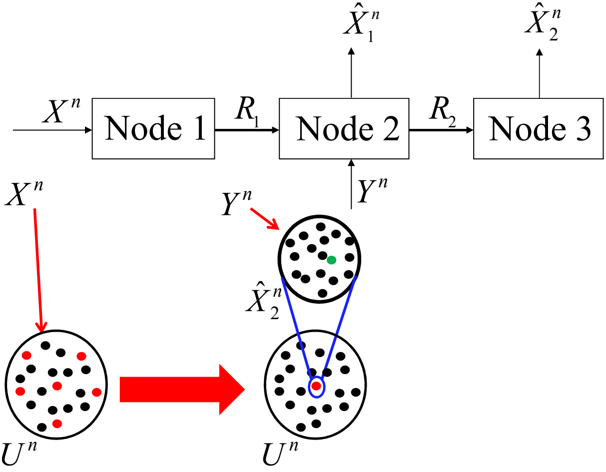
- $R_1 \geq I(X;U|Y)$
- $R_2 \geq I(X;U) + I(Y;\hat{X}_2|U)$



Cascade Source Coding

Idea 2: Forward [Cuff et al '09]

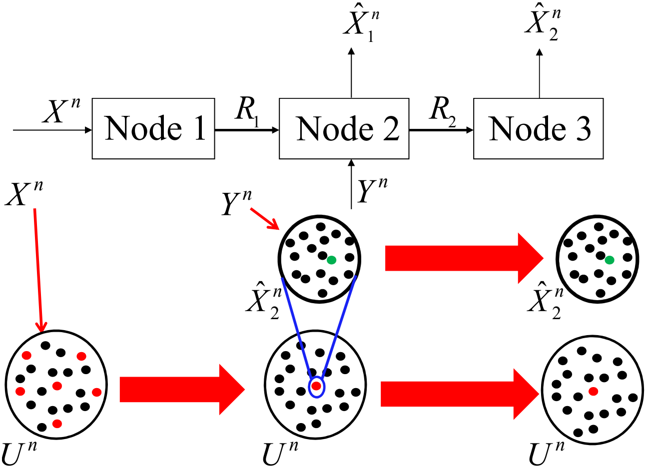
- $R_1 \geq I(X;U|Y)$
- $R_2 \geq I(X;U) + I(Y;\hat{X}_2|U)$



Cascade Source Coding

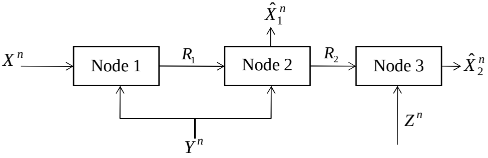
Idea 2: Forward [Cuff et al '09]

- $R_1 \geq I(X;U|Y)$
- $R_2 \geq I(X;U) + I(Y;\hat{X}_2|U)$



Cascade Source Coding

When Forward is Optimal [Chia et al '11]



- If $X - Y - Z$, the rate-cost region is given by union of the set of all of rate tuples (R_1, R_2) that satisfy the inequalities

$$R_1 \geq I(X; \hat{X}_1, U|Y)$$
$$R_2 \geq I(X, Y; U|Z),$$

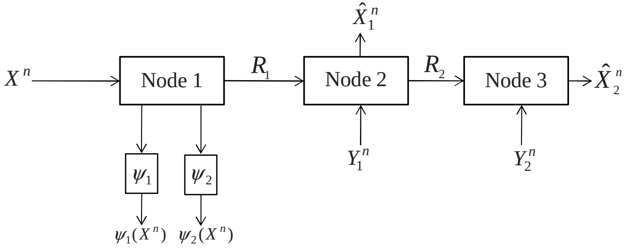
  where the mutual informations are evaluated with respect to the joint pmf

$$p(x, y, z, \hat{x}_1, u) = p(x, y)p(z|y)p(\hat{x}_1, u|x, y)$$

  for some pmf $p(\hat{x}_1, u|x, y)$ that satisfies $\mathrm{E}[d_j(X, \hat{X}_j)] \leq D_j$ for $j = 1, 2$.
- See Appendix-F

Cascade Source Coding Problem with CR constraint

Ahmadi et al '13



- Distortion constraints:

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[d_j(X,\hat{X}_j)\right] \leq D_j \text{ for } j = 1,2$$

- CR constraints:

$$\psi_1(X^n) \neq \hat{X}_1 \quad \text{w.h.p.}$$
$$\psi_2(X^n) \neq \hat{X}_2 \quad \text{w.h.p.}$$

Common Reconstruction Constraint

When the CR Constraint Simplifies the Problem

- Simultaneous transmission of data and state over state-dependent channels [Steinberg '09]
- Joint source-channel coding for the degraded broadcast channel [Steinberg '09]
- Multiple description problem [Tandon et al '12]

Cascade Source Coding Problem with CR constraint

Rate-Distortion Function ($X - Y_1 - Y_2$) [Ahmadi et al '13]

- If $X - Y_1 - Y_2$, the rate-distortion cost region is given by union of the set of all of rate tuples (R_1, R_2) that satisfy the inequalities

$$R_1 \geq I(X; \hat{X}_1 \hat{X}_2 | Y_1)$$
$$R_2 \geq I(X; \hat{X}_2 | Y_2),$$

  where the mutual informations are evaluated with respect to the joint pmf

$$p(x, y_1, y_2, \hat{x}_1, \hat{x}_2) = p(x, y_1) p(y_2 | y_1) p(\hat{x}_1, \hat{x}_2 | x)$$

  for some $p(\hat{x}_1, \hat{x}_2 | x)$ that satisfies $\mathrm{E}[d_j(X, \hat{X}_j)] \leq D_j$, for $j = 1, 2$.
- See Appendix-G

Overview

PART I: Basics

- Example: File Synchronization
- The Wyner-Ziv problem
- Code design

PART II: Constraints

- Causality constraints
- Reconstruction constraints
- Robustness constraints
- Privacy constraints

PART III: Extensions

- Action-dependent side information
- Cascade problems
- Sources with memory

The impact of memory

Lossless Compression, No Side Information

- Consider stationary and ergodic source X^n
- Lossless compression requires a rate equal to the entropy rate [Cover and Thomas '06]

$$\begin{aligned} H(\mathscr{X}) &= \lim_{n\to\infty} \frac{1}{n} H(X^n) \\ &= \lim_{n\to\infty} H(X_n|X^{n-1}) \\ &\leq H(X) \end{aligned}$$

The impact of memory

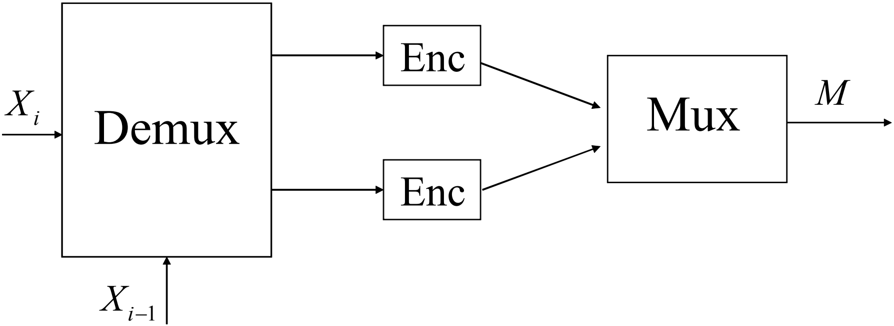
Lossless Compression, No Side Information

- For stationary Markov chains: $H(\mathscr{X}) = H(X_2|X_1)$

The impact of memory

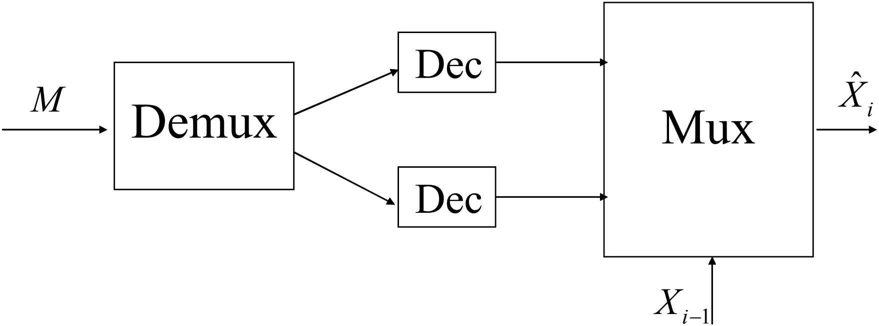
Lossless Compression, No Side Information

- Achieving rate $R = H(X_2|X_1)$ via context-adaptive encoding (e.g., JBIG)



The Impact of Memory

Lossless Compression, No Side Information



The impact of memory

Lossy Compression, No Side Information

- Rate-distortion function for stationary ergodic sources [Gallager '68]

$$R(D) = \lim_{n\to\infty} \frac{1}{n} \min_{p(\hat{x}^n|x^n) \text{ s.t. } \frac{1}{n}\sum_{i=1}^{n} E[d(X_i,\hat{X}_i)] \leq D} I(X^n;\hat{X}^n)$$

- Can be evaluated for some sources such as stationary Gaussian processes

The impact of memory

Lossy Compression, No Side Information

- Hard to evaluate in general
- For stationary Markov sources

$$\begin{aligned} I(X^n; \hat{X}^n) &= \sum_{i=1}^{n} H(X_i|X_{i-1}) - H(X_i|X^{i-1}, \hat{X}^n) \\ &\geq \sum_{i=1}^{n} H(X_i|X_{i-1}) - H(X_i|X_{i-1}, \hat{X}_i) \\ &= \sum_{i=1}^{n} I(X_i; \hat{X}_i|X_{i-1}) \end{aligned}$$

- This leads to the lower bound [Gray '73]
  $R(D) \geq R_{X_2|X_1}(D) = \min_{p(\hat{x}|x_1,x_2)} I(X_2; \hat{X}|X_1)$ s.t. $E[d(X_2, \hat{X})] \leq D$

Strictly Causal Side Information Revisited

Lossy Compression, No Side Information

- In the lossless case, $R(0) = H(X_2|X_1)$
- However, in the lossy case ($D > 0$), we can have $R(D) > R_{X_2|X_1}(D)$

Strictly Causal Side Information Revisited

Lossy Compression, No Side Information

- X_i binary Markov chain with symmetric transition probabilities $p(1|0) = p(0|1) = \varepsilon$ ($\varepsilon \leq 1/2$)
- Note that $X_i = X_{i-1} \oplus Z_i$, where $Z_i \sim Ber(\varepsilon)$ i.i.d. (innovation process)
- From [Weissman and Merhav, 03] for $0 \leq D \leq \varepsilon$

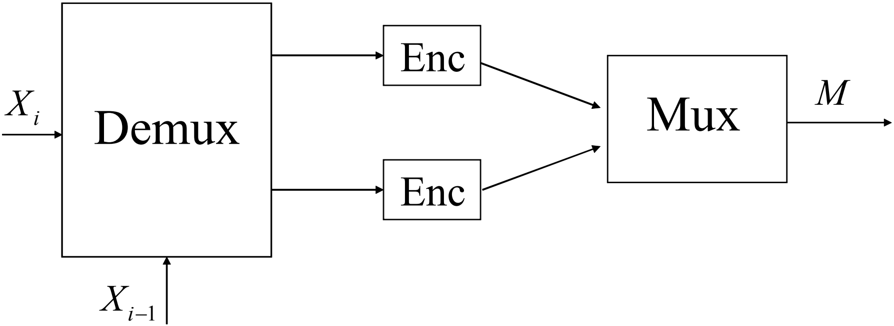
$$R_{X_2|X_1}(D) = H(\varepsilon) - H(D) \underset{\text{[Gray '70]}}{<} R(D)$$

for D larger than a critical value

The impact of memory

Lossy Compression, No Side Information

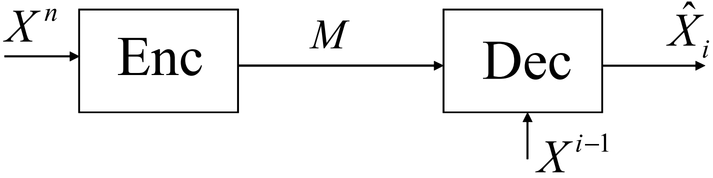
- Unlike the lossless case $R_{X_2|X_1}(D) = \min_{p(\hat{x}|x_1,x_2)} I(X_2;\hat{X}|X_1)$ generally not achievable... no known general single-letter expression



Strictly Causal Side Information Revisited

Sources Coding with Feed-Forward

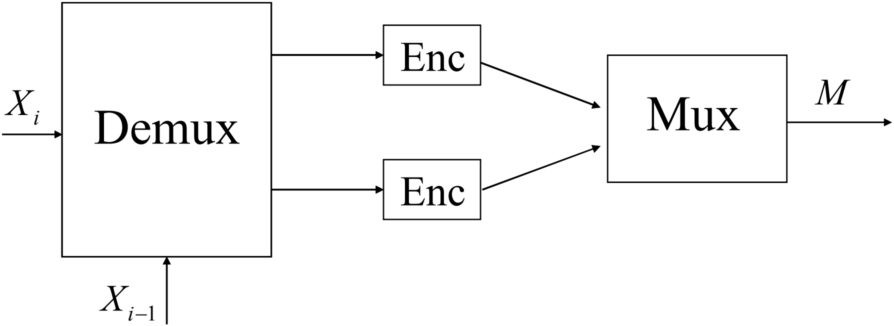
- Achievable if the receiver has strictly casual side information X^{i-1}



- Studied for general sources [Weissman and Merhav '03] [Venkataramanan and Pradhan '07] [Naiss and Permuter '11]
- For stationary Markov sources: $R_{FF}(D) = R_{X_2|X_1}(D)$

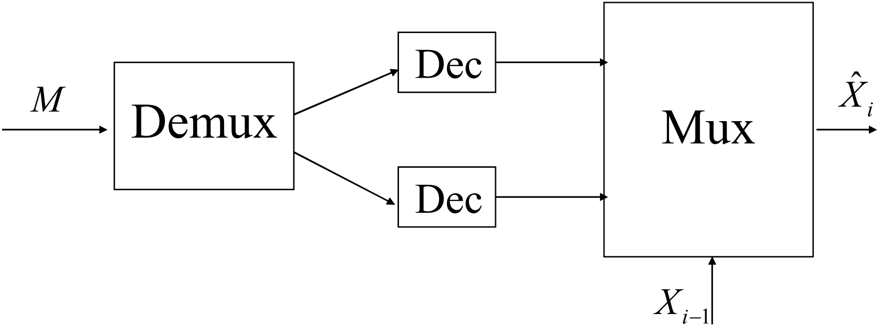
Strictly Causal Side Information Revisited

Sources Coding with Feed-Forward: Achievability



Strictly Causal Side Information Revisited

Sources Coding with Feed-Forward: Achievability



Strictly Causal Side Information Revisited

Generalizing Conditional Lower Bound

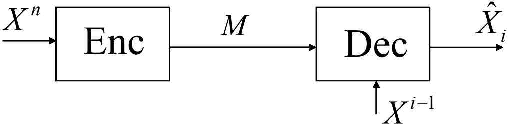
- Generalizing to stationary and ergodic sources [Gray '73]

$$R(D) \geq R_{X_m|X^{m-1}}(D) - (H(X_m|X^{m-1}) - H(\mathscr{X}))$$

- Achievable for order-m stationary Markov sources with feed-forward

Strictly Causal Side Information Revisited

Source Coding with Feed-Forward



- Rate-distortion-cost function for stationary ergodic sources [Venkataramanan and Pradhan '07] [Naiss and Permuter '11]

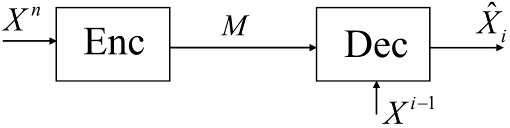
$$R(D) = \lim_{n\to\infty} \frac{1}{n} \min_{p(\hat{x}^n|x^n) \text{ s.t. } \frac{1}{n}\sum_{i=1}^n E[d(X_i,\hat{X}_i)] \leq D} I(X^n \to \hat{X}^n)$$

- Directed information [Massey '90]

$$I(\hat{X}^n \to X^n) = \sum_{i=1}^n I(X_i^n; \hat{X}_i | \hat{X}^{i-1}, X^{i-1})$$
$$= \sum_{i=1}^n I(X_i; \hat{X}^i | X^{i-1})$$

Strictly Causal Side Information Revisited

Source Coding with Feed-Forward



- Rate-distortion-cost function for stationary ergodic sources [Venkataramanan and Pradhan '07] [Naiss and Permuter '11]

$$R(D) = \lim_{n\to\infty} \frac{1}{n} \min_{p(\hat{x}^n|x^n) \text{ s.t. } \frac{1}{n}\sum_{i=1}^n E[d(X_i,\hat{X}_i)] \leq D} I(X^n \to \hat{X}^n)$$

- Directed information [Massey '90]

$$I(\hat{X}^n \to X^n) = \sum_{i=1}^n I(X_i^n; \hat{X}_i | \hat{X}^{i-1}, X^{i-1})$$
$$= \sum_{i=1}^n I(X_i; \hat{X}^i | X^{i-1})$$

Strictly Causal Side Information Revisited

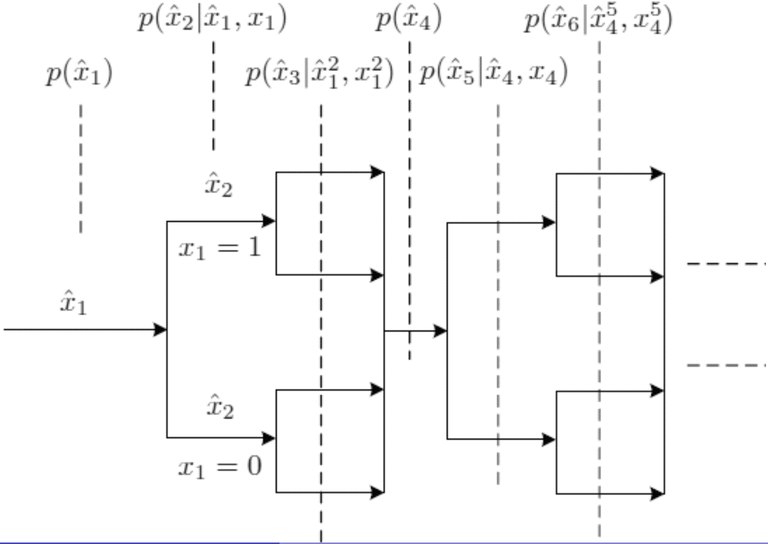
Source Coding with Feed-Forward: Converse

- Converse: easy (try!)

Strictly Causal Side Information Revisited

Source Coding with Feed-Forward: Achievability

- Codetrees



Strictly Causal Side Information Revisited

Source Coding with Feed-Forward: Achievability

- Compare with codetrees with general feed-forward (the encoder does not know $\hat{X}^n$) [Simeone '13]
- Related results for Hidden Markov Models [Simeone and Permuter '13]

Conclusions

Further issues:

- Vector sources [Gastpar et al '06] [Yu and Sharma '11]
- Fixed-to-variable codes [Alon and Orlitsky '96] [Koulgi et al '03] [Zhao and Effros '03] [Grangetto et al '09]
- Rateless codes [Draper '04] [Eckford and Yu '05] [Caire et al '05]
- Universal coding [Merhav and Ziv '06] [Jalali et al '10]
- Computational complexity [Gupta et al '08]
- Real-time operation [Teneketzis '06] [Kaspi and Merhav '12]
- Source-channel coding [Wernersson et al '09] [Akyol et al '10]
- Interactive coding [Orlitsky '92] [Ma and Ishwar '11]
- Quantum source coding [Yard and Devetak '09]
- Distributed source coding [El Gamal and Kim '11]

References

[Aaron et al '02] A. Aaron, S. Rane, E. Setton, B. Girod, Transformdomain Wyner-Ziv codec for video, in: SPIE Visual Communications and Image Processing Conf., San Jose, CA, 2004

[Ahmadi et al '13] B. Ahmadi, R. Tandon, O. Simeone and H. V. Poor, "Heegard-Berger and Cascade Source Coding Problems with Common Reconstruction Constraints," IEEE Trans. Inform. Theory, March 2013.

[Akyol et al '10] Akyol, Emrah, Kenneth Rose, and Tor Ramstad. "Optimal mappings for joint source channel coding." Information Theory Workshop (ITW), 2010 IEEE. IEEE, 2010.

[Alon and Orlitsky '96] Alon, Noga, and Alon Orlitsky. "Source coding and graph entropies." Information Theory, IEEE Transactions on 42, no. 5 (1996): 1329-1339.

[Ancheta '76] Ancheta Jr, T. "Syndrome-source-coding and its universal generalization." IEEE Trans. Information Theory, 1976.

[Arıkan '09] E. Arıkan, "Channel polarization: A method for constructing capacity achieving codes for symmetric binary-input memoryless channels," IEEE Trans. Inform. Theory, Jul. 2009.

References

[Bakshi et al '07] Bakshi, Mayank, Michelle Effros, WeiHsin Gu, and Ralf Koetter. "On network coding of independent and dependent sources in line networks." In Information Theory, 2007. ISIT 2007. IEEE International Symposium on, pp. 1096-1100. IEEE, 2007.

[Berger '71] Berger, Toby. Rate-Distortion Theory. John Wiley & Sons, Inc., 1971.

[Caire et al '03] aire, Giuseppe, Shlomo Shamai, and Sergio Verdu. "Lossless data compression with error correcting codes." In Information Theory, 2003. Proceedings. IEEE International Symposium on, p. 22. IEEE, 2003.

[Caire et al '05] Caire, Giuseppe, Shlomo Shamai, Amin Shokrollahi, and Sergio Verdú. "Fountain codes for lossless data compression." DIMACS Series in Discrete Mathematics and Theoretical Computer Science 68 (2005): 1-18.

[Chern et al '12] Chern, Bobbie, Idoia Ochoa, Alexandros Manolakos, Albert No, Kartik Venkat, and Tsachy Weissman. "Reference Based Genome Compression." arXiv preprint arXiv:1204.1912 (2012).

References

[Choudhuri and Mitra '12] Choudhuri, Chiranjib, and Urbashi Mitra. "How useful is adaptive action?." in Proc. Globecom2012 (2012).
[Chia et al '11] Chia, Yeow-Khiang, Haim H. Permuter, and Tsachy Weissman. "Cascade, Triangular, and Two-Way Source Coding With Degraded Side Information at the Second User." I nformation Theory, IEEE Transactions on 58, no. 1 (2012): 189-206.
[Cover and El Gamal '79] Cover, T., and A. EL Gamal. "Capacity theorems for the relay channel." Information Theory, IEEE Transactions on 25, no. 5 (1979): 572-584.
[Cover and Thomas '06] Cover, Thomas M., and Joy A. Thomas. Elements of information theory. Wiley-interscience, 2006.
[Csiszár and Körner '11] Csiszar, Imre, and János Körner. Information theory: coding theorems for discrete memoryless systems. Cambridge University Press, 2011.
[Cuff et al '09] Cuff, Paul, Han-I. Su, and Abbas El Gamal. "Cascade multiterminal source coding." In Information Theory, 2009. ISIT 2009. IEEE International Symposium on, pp. 1199-1203. IEEE, 2009.

References

[Diggavi et al '09] Diggavi, S., E. Perron, and Emre Telatar. "Lossy source coding with Gaussian or erased side-information." In Information Theory, 2009. ISIT 2009. IEEE International Symposium on, pp. 1035-1039.

[Dimakis et al '07] Dimakis, A. G., M. J. Wainwright, and K. Ramchandran. "Lower bounds on the rate-distortion function of LDGM codes." In Information Theory Workshop, 2007. ITW'07. IEEE, pp. 650-655. IEEE, 2007.

[Dimakis et al '09] Dimakis, Alexandros G., and Pascal O. Vontobel. "LP decoding meets LP decoding: a connection between channel coding and compressed sensing." In Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on, pp. 8-15. IEEE, 2009.

[Doshi et al '10] Doshi, Vishal, Devavrat Shah, Muriel Médard, and Michelle Effros. "Functional compression through graph coloring." Information Theory, IEEE Transactions on 56, no. 8 (2010): 3901-3917.

[Draper '04] Draper, Stark C. "Universal incremental slepian-wolf coding." In Proc. 42nd Allerton Conf. on Communication, Control and Computing, pp. 1332-1341. 2004.

References

[Draper and Wornell '04] Draper, Stark C., and Gregory W. Wornell. "Side information aware coding strategies for sensor networks." Selected Areas in Communications, IEEE Journal on 22, no. 6 (2004): 966-976.

[Dupuis et al '04] Dupuis, Frédéric, Wei Yu, and Frans MJ Willems. "Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information." In Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on, p. 179. IEEE, 2004.

[Eckford and Yu '05] Eckford, Andrew W., and Wei Yu. "Rateless slepian-wolf codes." In Proc. Asilomar conference on signals, systems and computers, pp. 1757-1761. 2005.

[El Gamal and Kim '11] A. El Gamal and Y. Kim, Network Information Theory, Cambridge University Press, Dec 2011.

[Feng et al '04] Feng, Hanying, Michelle Effros, and Serap Savari. "Functional source coding for networks with receiver side information." In Proceedings of the Allerton Conference on Communication, Control, and Computing, pp. 1419-1427. 2004.

References

[Fleming et al '04] Fleming, Michael, Qian Zhao, and Michelle Effros. "Network vector quantization." Information Theory, IEEE Transactions on 50, no. 8 (2004): 1584-1604.
[Gallager '68] Gallager, Robert G. "Information theory and reliable communication." (1968).
[Gastpar et al '06] Gastpar, Michael, Pier Luigi Dragotti, and Martin Vetterli. "The distributed karhunen–loeve transform." Information Theory, IEEE Transactions on 52, no. 12 (2006): 5177-5196.
[Gersho and Gray '92] Gersho, Allen, and Robert M. Gray. Vector quantization and signal compression. Vol. 159. Springer, 1992.
[Grangetto et al '09] Grangetto, Marco, Enrico Magli, and Gabriella Olmo. "Distributed arithmetic coding for the Slepian–Wolf problem." Signal Processing, IEEE Transactions on 57, no. 6 (2009): 2245-2257.
[Gray '72] R. M. Gray, "Conditional rate-distortion theory," Stanford Univ., Stanford, CA, Electronics Laboratories Tech. Rep. 6502-2, Oct. 1972.

References

[Gray '73] Gray, R. "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions." Information Theory, IEEE Transactions on 19, no. 4 (1973): 480-489.

[Gu and Effros '06] Gu, Wei-Hsin, and Michelle Effros. "Source coding for a simple multi-hop network." In Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on, pp. 2335-2339. IEEE, 2005.

[Gupta et al '08] Gupta, Ankit, Sergio Verdu, and Tsachy Weissman. "Rate-distortion in near-linear time." In Information Theory, 2008. ISIT 2008. IEEE International Symposium on, pp. 847-851. IEEE, 2008.

[Gupta and Verdú '09] Gupta, Ankit, and Sergio Verdú. "Nonlinear sparse-graph codes for lossy compression." Information Theory, IEEE Transactions on 55, no. 5 (2009): 1961-1975.

[Gupta and Verdú '11] Gupta, Ankit, and Sergio Verdú. "Operational duality between lossy compression and channel coding." Information Theory, IEEE Transactions on 57, no. 6 (2011): 3171-3179.

References

[Heegard and Berger '85] C. Heegard and T. Berger, "Rate distortion when side information may be absent," IEEE Trans. Inf. Theory, vol. 31, no. 6, pp. 727-734, Nov. 1985.

[Jalali et al '10] Jalali, Shirin, Sergio Verdú, and Tsachy Weissman. "A Universal Scheme for Wyner–Ziv Coding of Discrete Sources." Information Theory, IEEE Transactions on 56, no. 4 (2010): 1737-1750.

[Kanlis et al '96] A. Kanlis, S.Khudanpur, and P.Narayan. Typicality of a good rate-distortion code. Problems of Information Transmission (Problemy Peredachi Informatsii), January 1996. Special issue in honor of M. S. Pinsker.

[Kaspi '94] A. Kaspi, "Rate-distortion when side-information may be present at the decoder," IEEE Trans. Inf. Theory, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.

[Kaspi and Merhav '12] Kaspi, Yonatan, and Neri Merhav. "On real-time and causal secure source coding." In Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, pp. 353-357. IEEE, 2012.

References

[Koller and Friedman '09] Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.

[Korada and Urbanke '10] Korada, Satish Babu, and Rüdiger L. Urbanke. "Polar codes are optimal for lossy source coding." Information Theory, IEEE Transactions on 56, no. 4 (2010): 1751-1768.

[Koulgi et al '03] Koulgi, Prashant, Ertem Tuncel, Shankar L. Regunathan, and Kenneth Rose. "On zero-error source coding with decoder side information." Information Theory, IEEE Transactions on 49, no. 1 (2003): 99-111.

[Kramer '08] Kramer, Gerhard. Topics in multi-user information theory. Vol. 4. Now Pub, 2008.

[Kudekar and Urbanke '08] Kudekar, Shrinivas, and Ruediger Urbanke. "Lower bounds on the rate-distortion function of individual LDGM codes." In Turbo Codes and Related Topics, 2008 5th International Symposium on, pp. 379-384. IEEE, 2008.

References

[Liu et al '06] Liu, Zhixin, Samuel Cheng, Angelos D. Liveris, and Zixiang Xiong. "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design." Information Theory, IEEE Transactions on 52, no. 10 (2006): 4358-4379.

[Ma and Ishwar '11] Ma, Nan, and Prakash Ishwar. "Some results on distributed source coding for interactive function computation." Information Theory, IEEE Transactions on 57, no. 9 (2011): 6180-6195.

[MacKay '99] MacKay, David J. C. "Good error-correcting codes based on very sparse matrices." Information Theory, IEEE Transactions on 45, no. 2 (1999): 399-431.

[MacKay '03] MacKay, David JC. Information theory, inference and learning algorithms. Cambridge university press, 2003.

[Martinian and Wainwright '06] Martinian, Emin, and Martin Wainwright. "Low density codes achieve the rate-distortion bound." In Data Compression Conference, 2006. DCC 2006. Proceedings, pp. 153-162. IEEE, 2006.

References

[Martinian and Yedidia '03] Martinian, Emin, and Jonathan S. Yedidia. "Iterative quantization using codes on graphs." arXiv preprint cs/0408008 (2004).

[Matsunaga and Yamamoto '03] Y. Matsunaga and H. Yamamoto, "Coding Theorem for Lossy Data Compression by LDPC Codes," IEEE Trans. Inform. Theory, Sept. 2003.

[Massey '78] J. L. Massey, "Joint source channel coding," in Communication Systems and Random Process Theory, 1978.

[Massey '90] Massey, J. "Causality, feedback and directed information." In Proc. int. symp. information theory application, pp. 303-305. 1990.

[Merhav and Ziv '06] Merhav, Neri, and Jacob Ziv. "On the Wyner-Ziv problem for individual sequences." IEEE Trans. Inf. Theory 2006.

[Naiss and Permuter '11] I. Naiss and H. Permuter, "Computable bounds for rate distortion with feed-forward for stationary and ergodic sources," arXiv:1106.0895v1.

References

[Orlitsky '91] Orlitsky, Alon. "Interactive communication: Balanced distributions, correlated files, and average-case complexity." In Foundations of Computer Science, 1991. Proceedings., 32nd Annual Symposium on, pp. 228-238. IEEE, 1991.

[Orlitsky '92] Orlitsky, Alon. "Average-case interactive communication." Information Theory, IEEE Transactions on 38, no. 5 (1992): 1534-1547.

[Orlitsky and Roche '01] Orlitsky, Alon, and James R. Roche. "Coding for computing." Information Theory, IEEE Transactions on 47, no. 3 (2001): 903-917.

[Orlitsky and Wiswanathan '03] Orlitsky, Alon, and Krishnamurthy Viswanathan. "One-way communication and error-correcting codes." Information Theory, IEEE Transactions on 49, no. 7 (2003): 1781-1788.

[Permuter and Weissman '11] H. Permuter and T. Weissman, "Source coding with a side information "vending machine"," IEEE Trans. Inf. Theory, vol. 57, pp. 4530–4544, Jul 2011.

References

[Pradhan et al '03] Pradhan, S. Sandeep, Jim Chou, and Kannan Ramchandran. "Duality between source coding and channel coding and its extension to the side information case." Information Theory, IEEE Transactions on 49, no. 5 (2003): 1181-1203.
[Pradhan and Ramchandran '03] Pradhan, S. Sandeep, and Kannan Ramchandran. "Distributed source coding using syndromes (DISCUS): Design and construction." Information Theory, IEEE Transactions on 49, no. 3 (2003): 626-643.
[Sanderovich et al '09] Sanderovich, Amichai, Oren Somekh, H. Vincent Poor, and Shlomo Shamai. "Uplink macro diversity of limited backhaul cellular network." Information Theory, IEEE Transactions on 55, no. 8 (2009): 3457-3478.
[Saxena et al '10] Saxena, Ankur, Jayanth Nayak, and Kenneth Rose. "Robust distributed source coder design by deterministic annealing." Signal Processing, IEEE Transactions on 58, no. 2 (2010): 859-868.

References

[Shamai et al '98] Shamai, Shlomo, Sergio Verdú, and Ram Zamir. "Systematic lossy source/channel coding." Information Theory, IEEE Transactions on 44, no. 2 (1998): 564-579.
[Simeone '13] Simeone, Osvaldo. "Source Coding with in-Block Memory and Causally Controllable Side Information." arXiv preprint arXiv:1301.0926 (2013).
[Simeone and Permuter '13] Simeone, Osvaldo, and Haim H. Permuter. "Source coding when the side information may be delayed." to appear in IEEE Trans. Inform. Theory, 2013.
[Slepian and Wolf '73] Slepian, David, and Jack Wolf. "Noiseless coding of correlated information sources." Information Theory, IEEE Transactions on 19, no. 4 (1973): 471-480.
[Steinberg '09] Y. Steinberg, "Coding and common reconstruction," IEEE Trans. Inform. Theory, vol. 55, no. 11, 2009.

References

[Suel and Memon '02] T. Suel and N.Memon. Algorithms for delta compression and remote file synchronization. In K. Sayood, editor, Lossless Compression Handbook. Academic Press, 2002.

[Sun et al '10] Sun, Zhibin, Mingkai Shao, Jun Chen, Kon Wong, and Xiaolin Wu. "Achieving the rate-distortion bound with low-density generator matrix codes." Communications, IEEE Transactions on 58, no. 6 (2010): 1643-1653.

[Tandon et al '11] R. Tandon, S. Mohajer, and H. V. Poor, "Cascade source coding with erased side information," in Proc. IEEE Symp. Inform. Theory, St. Petersburg, Russia, Aug. 2011.

[Tandon et al '12] R. Tandon, B. Ahmadi, O. Simeone and H. V. Poor, "Gaussian Multiple Descriptions with Common and Constrained Reconstruction Constraints," in Proc. IEEE International Symposium on Information Theory (ISIT 2012), Cambridge, MA, USA, July 1-6, 2012.

[Teneketzis '06] Teneketzis, Demosthenis. "On the structure of optimal real-time encoders and decoders in noisy communication." Information Theory, IEEE Transactions on 52, no. 9 (2006): 4017-4035.

References

[Vaezi and Labeau '12] Vaezi, Mojtaba, and Fabrice Labeau. "Distributed lossy source coding using real-number codes." In Vehicular Technology Conference (VTC Fall), 2012 IEEE, pp. 1-5. IEEE, 2012.
[Vasudevan et al '06] D. Vasudevan, C. Tian, and S. N. Diggavi, "Lossy source coding for a cascade communication system with sideinformations," In Proc. Allerton, Sept. 2006.
[Venkataramanan and Pradhan '07] R. Venkataramanan and S. S. Pradhan, "Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source," IEEE Trans. Inform. Theory, vol. 53, no. 6, pp. 2154-2179, Jun. 2007.
[Wainwright et al '10] Wainwright, Martin J., Elitza Maneva, and Emin Martinian. "Lossy source compression using low-density generator matrix codes: Analysis and algorithms." Information Theory, IEEE Transactions on 56, no. 3 (2010): 1351-1368.

References

[Wainwright and E. Maneva '05] M. J. Wainwright and E. Maneva, “Lossy source coding via messagepassing and decimation over generalized codewords of LDGM codes," IEEE International Symp. Inf. Theory, Adelaide, Australia, Sept. 2005.

[Wainwright and Martinian '09] Wainwright, Martin J., and Emin Martinian. "Low-density graph codes that are optimal for binning and coding with side information." Information Theory, IEEE Transactions on 55, no. 3 (2009): 1061-1079.

[Weissman and El Gamal '06] T. Weissman and A. El Gamal, “Source coding with limited-look-ahead side information at the decoder,” IEEE Trans. Inf. Theory, vol. 52, no. 12, pp. 5218–5239, Dec. 2006.

[Weissman and Merhav '03] T. Weissman and N. Merhav, “On competitive prediction and its relation to rate-distortion theory,” IEEE Trans. Inform. Theory, vol. 49, no. 12, pp. 3185- 3194, Dec. 2003.

[Weissman and Merhav '04] Weissman, Tsachy, and Neri Merhav. "On competitive prediction and its relation to rate-distortion theory." Information Theory, IEEE Transactions on 49, no. 12 (2003): 3185-3194.

References

[Weissman and Ordentlich '05] Weissman, Tsachy, and Erik Ordentlich. "The empirical distribution of rate-constrained source codes." Information Theory, IEEE Transactions on 51, no. 11 (2005): 3718-3733.

[Weissman and Verdú '08] Verdú, Sergio, and Tsachy Weissman. "The information lost in erasures." Information Theory, IEEE Transactions on 54, no. 11 (2008): 5030-5058.

[Wernersson and Skoglund '09] Wernersson, Niklas, Johannes Karlsson, and Mikael Skoglund. "Distributed quantization over noisy channels." Communications, IEEE Transactions on 57, no. 6 (2009): 1693-1700.

[Witsenhausen '76] Witsenhausen, H. "The zero-error side information problem and chromatic numbers (corresp.)." Information Theory, IEEE Transactions on 22, no. 5 (1976): 592-593.

[Wyner '74] Wyner, Aaron. "Recent results in the Shannon theory." Information Theory, IEEE Transactions on 20, no. 1 (1974): 2-10.

References

[Wyner '78] Wyner, Aaron D. "The rate-distortion function for source coding with side information at the decoder\ 3-II: General sources." Information and Control 38, no. 1 (1978): 60-80.

[Wyner and Ziv '76] Wyner, Aaron, and Jacob Ziv. "The rate-distortion function for source coding with side information at the decoder." Information Theory, IEEE Transactions on 22, no. 1 (1976): 1-10.

[Yamamoto '82] Yamamoto, Hirosuke. "Wyner-Ziv theory for a general function of the correlated sources (Corresp.)." Information Theory, IEEE Transactions on 28, no. 5 (1982): 803-807.

[Yang et al '09] Yang, Yang, Samuel Cheng, Zixiang Xiong, and Wei Zhao. "Wyner-Ziv coding based on TCQ and LDPC codes." Communications, IEEE Transactions on 57, no. 2 (2009): 376-387.

[Yard and Devetak '09] Yard, Jon T., and Igor Devetak. "Optimal quantum source coding with quantum side information at the encoder and decoder." Information Theory, IEEE Transactions on 55, no. 11 (2009): 5339-5351.

References

[Yu and Sharma '11] Yu, Chao, and Gaurav Sharma. "Distributed estimation and coding: A sequential framework based on a side-informed decomposition." Signal Processing, IEEE Transactions on 59, no. 2 (2011): 759-773.

[Zamir et al '02] Zamir, Ram, Shlomo Shamai, and Uri Erez. "Nested linear/lattice codes for structured multiterminal binning." Information Theory, IEEE Transactions on 48, no. 6 (2002): 1250-1276.

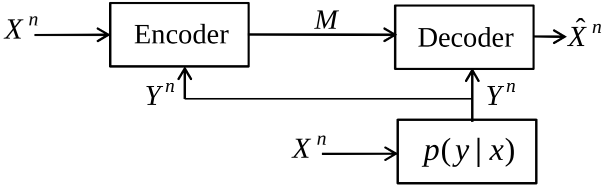
[Zhao and Effros '03] Zhao, Qian, and Michelle Effros. "Lossless and near-lossless source coding for multiple access networks." Information Theory, IEEE Transactions on 49, no. 1 (2003): 112-128.

[Ziv '85] Ziv, Jacob. "On universal quantization." Information Theory, IEEE Transactions on 31, no. 3 (1985): 344-347.

Appendix-A

Conditional Rate-Distortion Theory

Converse



- Factorization of the joint pmf:
  $p(x^n, y^n, m, \hat{x}^n) = \prod_{i=1}^{n} p(x_i)p(y_i|x_i) \cdot 1(m|x^n, y^n) \cdot 1(\hat{x}^n|m, y^n)$
- From data processing inequality considerations:

$$\begin{aligned} nR &\geq H(M) \\ &\geq H(M|Y^n) \\ &= H(M|Y^n) - H(M|X^n, Y^n) \\ &= I(X^n; M, \hat{X}^n|Y^n) \end{aligned}$$

Conditional Rate-Distortion Theory

Converse

$$\begin{aligned}
nR &\geq I(X^n;\hat{X}^n|Y^n) \\
&= H(X^n|Y^n) - H(X^n|Y^n,\hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|X^{i-1},Y^n,\hat{X}^n) \\
&\geq \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|Y_i,\hat{X}_i) \\
&= \sum_{i=1}^{n} I(X_i;\hat{X}_i|Y_i)
\end{aligned}$$

- Using convexity properties of the rate-distortion function (see, e.g., [El Gamal and Kim '11]), we get the lower bound:

$$R_{X|Y}(D) \geq \min_{p(\hat{x}|x,y)} I(X;\hat{X}|Y)$$

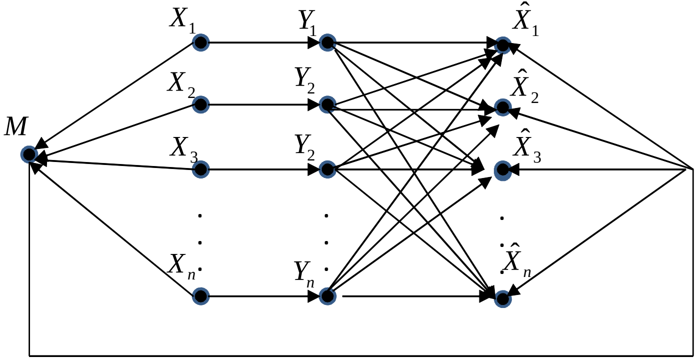
Appendix-B

Wyner-Ziv Problem

Converse

- Factorization of the joint pmf:

$$p(x^n, y^n, m, \hat{x}^n) = \prod_{i=1}^{n} p(x_i)p(y_i|x_i) \cdot 1(m|x^n) \cdot 1(\hat{x}^n|m, y^n)$$



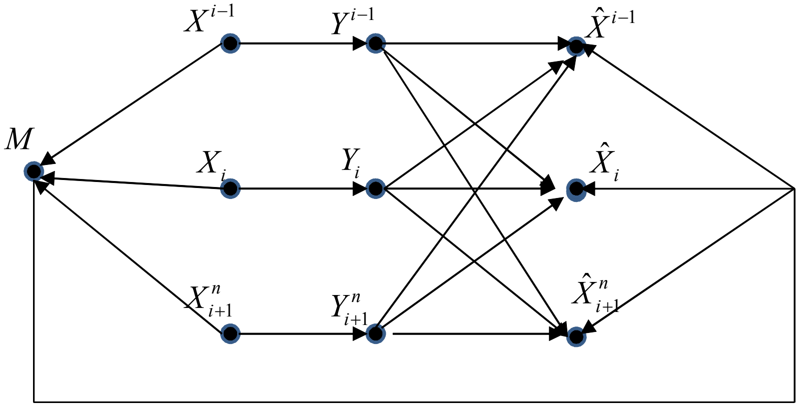
- See [Kramer '08] and [Koller and Friedman '09]

Wyner-Ziv Problem

Converse

- Factorization of the joint pmf:

$$p(x^n, y^n, m, \hat{x}^n) = \prod_{i=1}^{n} p(x_i)p(y_i|x_i) \cdot 1(m|x^n) \cdot 1(\hat{x}^n|m, y^n)$$



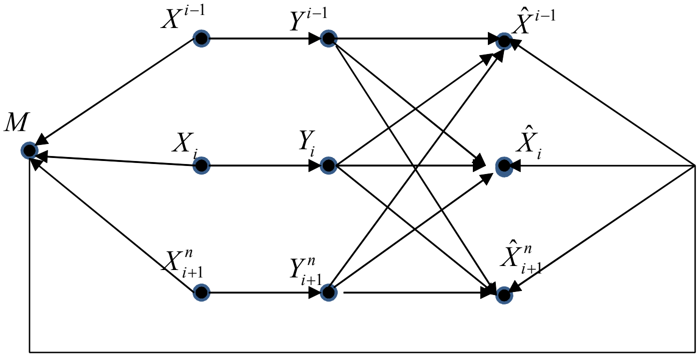
Wyner-Ziv Problem

Converse

$$
\begin{aligned}
nR &\geq H(M) \\
&\geq H(M|Y^n) \\
&= H(M|Y^n) - H(M|X^n, Y^n) \\
&= I(X^n; M|Y^n) \\
&= H(X^n|Y^n) - H(X^n|Y^n, M) \\
&= \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|X^{i-1}, Y^n, M) \\
&= \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|\underbrace{X^{i-1}, Y^{n\setminus i}, M}_{\triangleq U_i}, Y_i) \\
&= \sum_{i=1}^{n} I(X_i; U_i|Y_i)
\end{aligned}
$$

Wyner-Ziv Problem

Converse

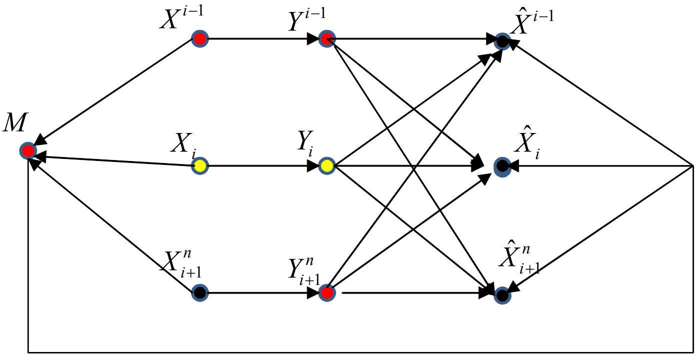


- $U_i = \{X^{i-1}, Y^{n\backslash i}, M\}$ satisfies $U_i - X_i - Y_i$, and hence $p(u_i, x_i, y_i) = p(x_i, y_i)p(u_i|x_i)$
- $\hat{X}_i = f(M, Y^n) = f(U_i, Y_i)$

Wyner-Ziv Problem

Converse

- d-separation test for conditional independence [Kramer '08]

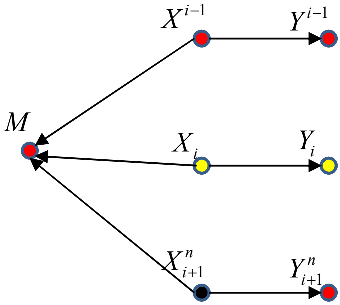


- $U_i = \{X^{i-1}, Y^{n\backslash i}, M\}$ satisfies $U_i - X_i - Y_i$

Wyner-Ziv Problem

Converse

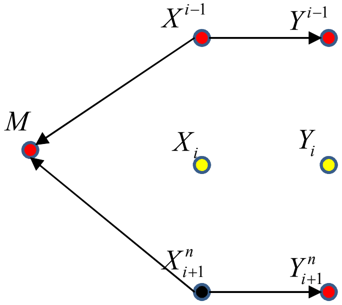
- d-separation test for conditional independence [Kramer '08]:
- 1. Include only edges and vertices moving backward from the involved vertices



Wyner-Ziv Problem

Converse

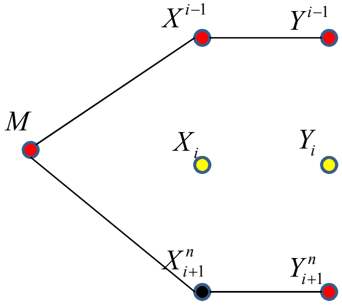
- d-separation test for conditional independence [Kramer '08]:
- 2. Remove all edges coming out of the conditioning variable X_i



Wyner-Ziv Problem

Converse

- d-separation test for conditional independence [Kramer '08]:
- 3. Make edges undirected



- If no path between U_i and Y_i in the resulting *undirected* graph, then we have $U_i - X_i - Y_i$

Appendix C

LDPC Codes and Compression

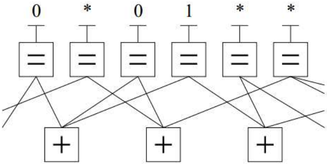


Figure 1: Using an LDPC code for binary erasure quantization. The boxes with = signs are repetition nodes: all edges connected to an = box must have the same value. The boxes with + signs are check nodes: the modulo 2 sum of the values on edges connected to a + box must be 0. The source consists of 0's, 1's, and erasures represented by $*$'s. Erasures may be quantized to 0 or 1 while incurring no distortion. A non-zero distortion must be incurred for the source shown above since the left-most check cannot be satisfied.

[Martinian and Yedidia '03]

LDPC Codes and Compression

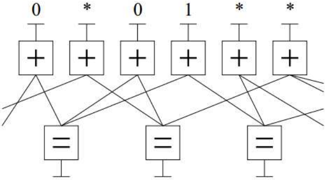


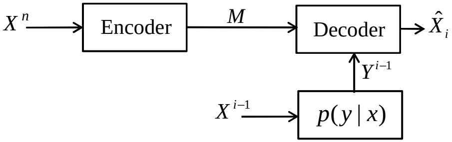
Figure 2: Using the dual of an LDPC code for binary erasure quantization. Choosing values for the variables at the bottom produces a codeword. The values for each sample of the resulting codeword are obtained by taking the sum modulo 2 of the connected variables. In contrast to Fig. 1, this structure can successfully match the source with no distortion if the bottom 3 variables are set to 0, 0, 1.

[Martinian and Yedidia '03]

Appendix D

Strictly Causal Side Information

Converse



$$nR \geq \sum_{i=1}^{n} H(X_i) - H(X_i|X^{i-1}, M)$$
$$\geq \sum_{i=1}^{n} H(X_i) - H(X_i|Y^{i-1}, M)$$

by data processing inequality since $X_i - (X^{i-1}, M) - (Y^{i-1}, M)$ (check with d-separation!)

Strictly Causal Side Information

Converse

$$
\begin{aligned}
nR &\geq \sum_{i=1}^{n} H(X_i) - H(X_i|X^{i-1}, M) \\
&\geq \sum_{i=1}^{n} H(X_i) - H(X_i|Y^{i-1}, M) \\
&= \sum_{i=1}^{n} H(X_i) - H(X_i|Y^{i-1}, M, \hat{X}_i) \\
&= \sum_{i=1}^{n} I(X_i; \hat{X}_i)
\end{aligned}
$$

- ... delayed side information not useful (but it can be for sources with memory)

Appendix E

Causal Side Information

Converse

$$nR \geq \sum_{i=1}^{n} H(X_i) - H(X_i|X^{i-1}, M)$$

$$\geq \sum_{i=1}^{n} H(X_i) - H(X_i|\underbrace{Y^{i-1}, M}_{\triangleq U_i})$$

$$= \sum_{i=1}^{n} I(X_i; U_i)$$

- We also have $\hat{X}_i = f(U_i, Y_i)$ and $U_i - X_i - Y_i$
- It follows that

$$R_{X|Y}^{WZ-C}(D) \geq \min_{p(u|x)} I(X;U)$$

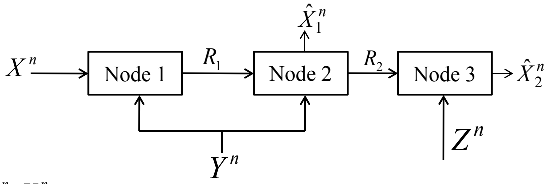
$$\text{s.t. } E[d(X, Y, f(U, Y))] \leq D$$

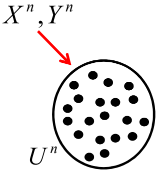
Appendix F

Cascade Source Coding Problem

When Forward is Optimal [Chia et al '11]

- $R_1 \geq I(X; \hat{X}_1, U|Y)$
- $R_2 \geq I(X, Y; U|Z)$

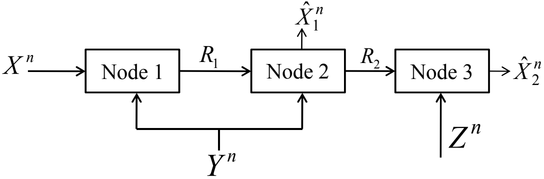


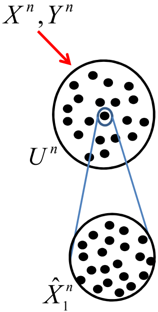


Cascade Source Coding Problem

When Forward is Optimal [Chia et al '11]

- $R_1 \geq I(X; \hat{X}_1, U|Y)$
- $R_2 \geq I(X, Y; U|Z)$

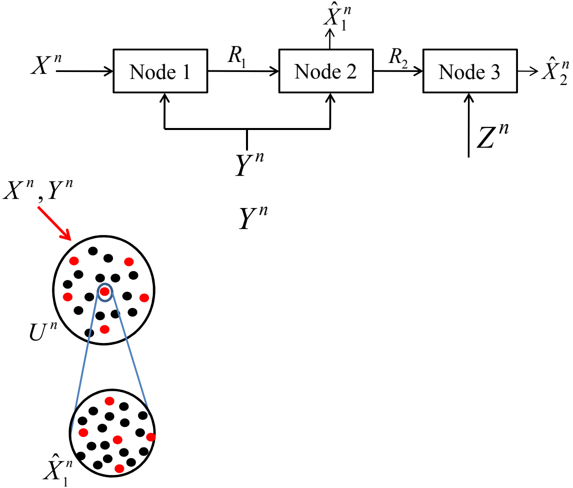




Cascade Source Coding Problem

When Forward is Optimal [Chia et al '11]

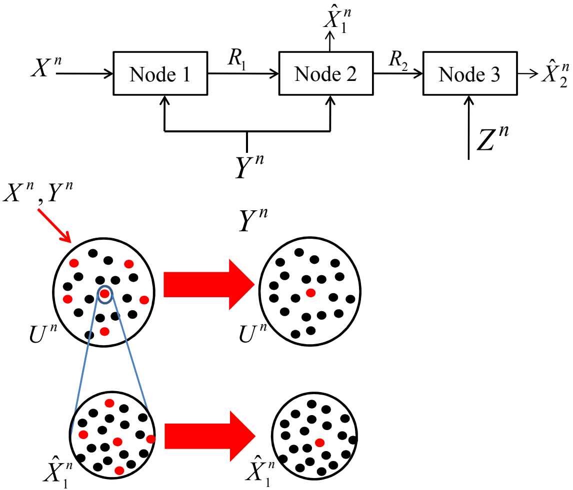
- $R_1 \geq I(X;\hat{X}_1, U|Y)$
- $R_2 \geq I(X, Y; U|Z)$



Cascade Source Coding Problem

When Forward is Optimal [Chia et al '11]

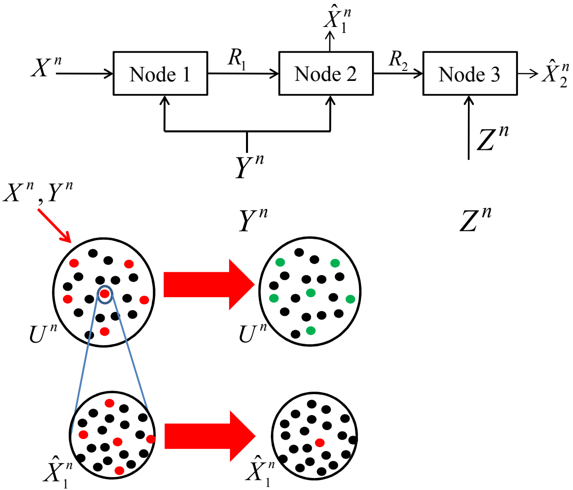
- $R_1 \geq I(X; \hat{X}_1, U|Y)$
- $R_2 \geq I(X, Y; U|Z)$



Cascade Source Coding Problem

When Forward is Optimal [Chia et al '11]

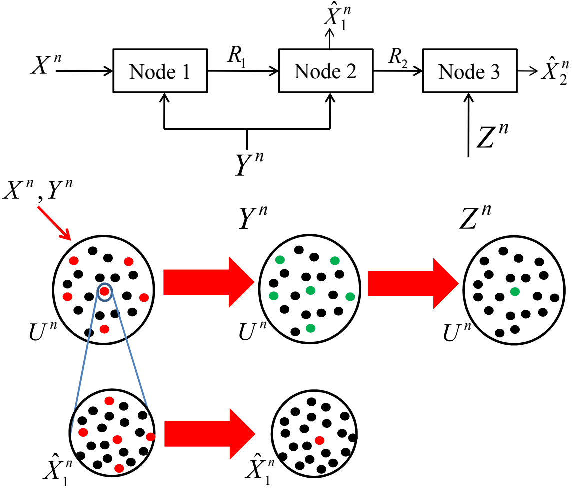
- $R_1 \geq I(X;\hat{X}_1, U|Y)$
- $R_2 \geq I(X, Y; U|Z)$



Cascade Source Coding Problem

When Forward is Optimal [Chia et al '11]

- $R_1 \geq I(X; \hat{X}_1, U|Y)$
- $R_2 \geq I(X, Y; U|Z)$

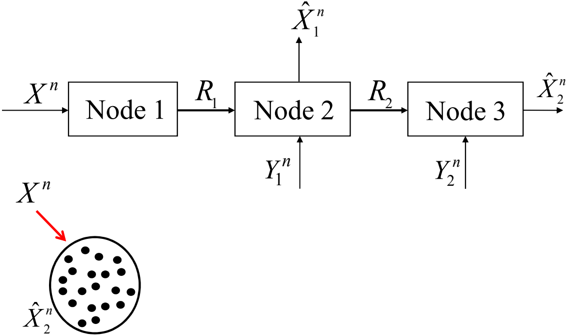


Appendix G

Cascade Source Coding Problem with CR constraint

Rate-Distortion Region ($X - Y_1 - Y_2$)

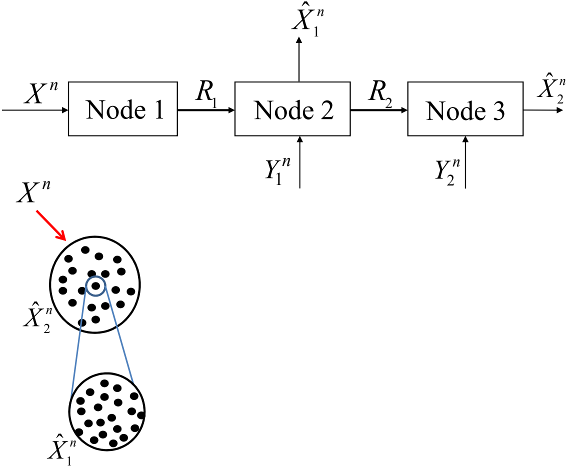
- $R_1 \geq I(X; \hat{X}_1 \hat{X}_2 | Y_1)$
- $R_2 \geq I(X; \hat{X}_2 | Y_2)$



Cascade Source Coding Problem with CR constraint

Rate-Distortion Region ($X - Y_1 - Y_2$)

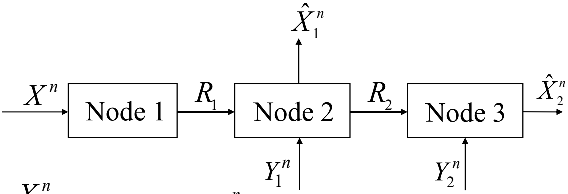
- $R_1 \geq I(X; \hat{X}_1 \hat{X}_2 | Y_1)$
- $R_2 \geq I(X; \hat{X}_2 | Y_2)$

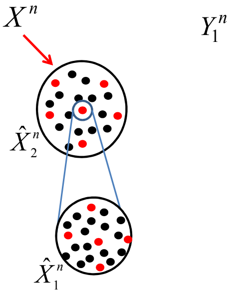


Cascade Source Coding Problem with CR constraint

Rate-Distortion Region ($X - Y_1 - Y_2$)

- $R_1 \geq I(X; \hat{X}_1 \hat{X}_2 | Y_1)$
- $R_2 \geq I(X; \hat{X}_2 | Y_2)$

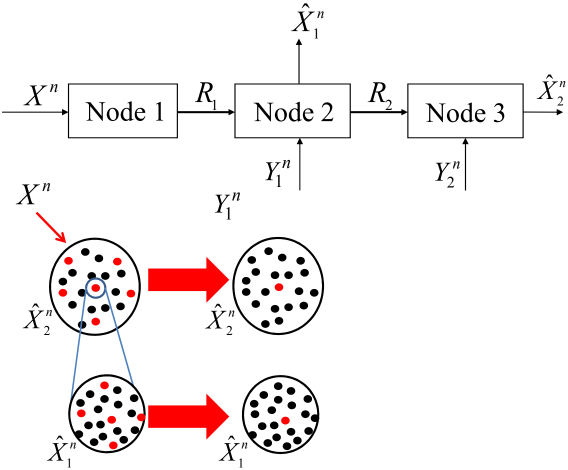




Cascade Source Coding Problem with CR constraint

Rate-Distortion Region ($X - Y_1 - Y_2$)

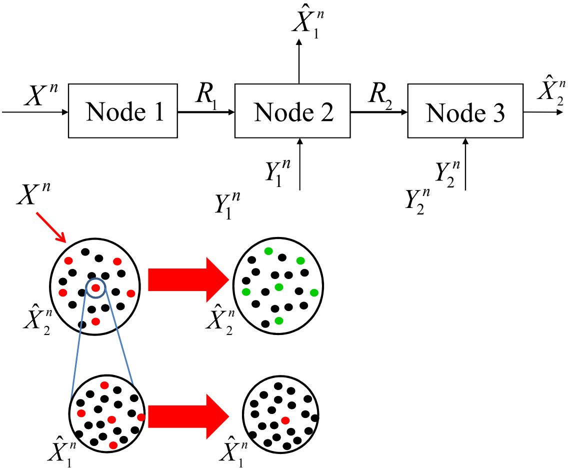
- $R_1 \geq I(X; \hat{X}_1 \hat{X}_2 | Y_1)$
- $R_2 \geq I(X; \hat{X}_2 | Y_2)$



Cascade Source Coding Problem with CR constraint

Rate-Distortion Region ($X - Y_1 - Y_2$)

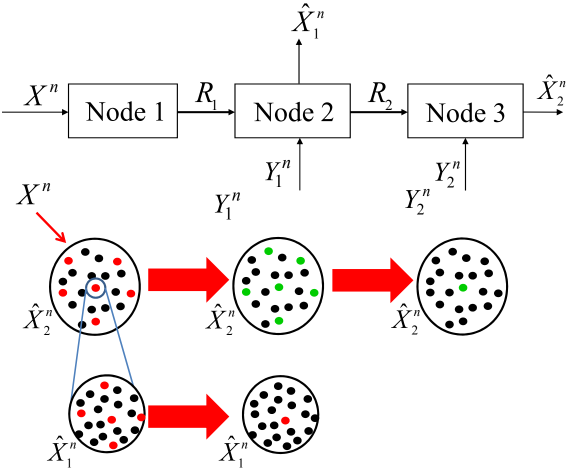
- $R_1 \geq I(X; \hat{X}_1 \hat{X}_2 | Y_1)$
- $R_2 \geq I(X; \hat{X}_2 | Y_2)$



Cascade Source Coding Problem with CR constraint

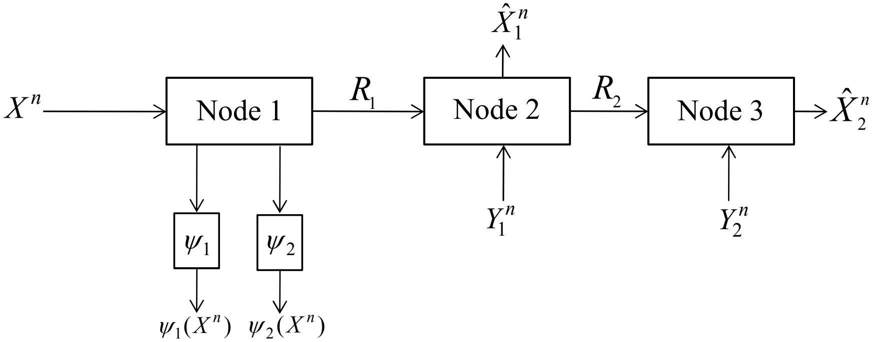
Rate-Distortion Region ($X - Y_1 - Y_2$)

- $R_1 \geq I(X; \hat{X}_1 \hat{X}_2 | Y_1)$
- $R_2 \geq I(X; \hat{X}_2 | Y_2)$



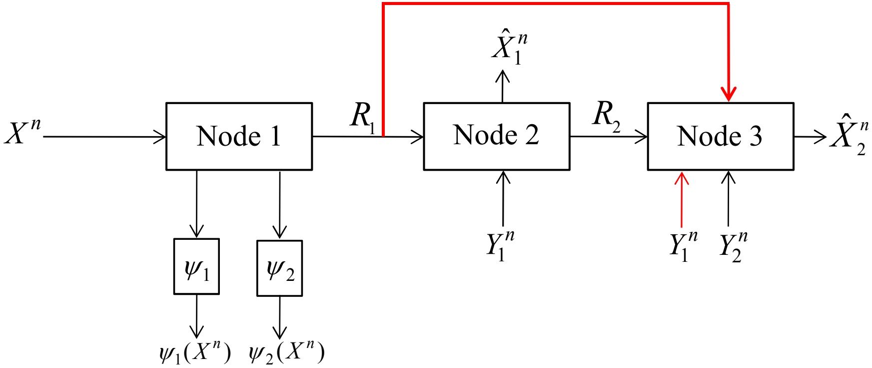
Cascade Source Coding Problem with CR constraint

Outer bound to the Rate-Distortion Region ($X - Y_1 - Y_2$)



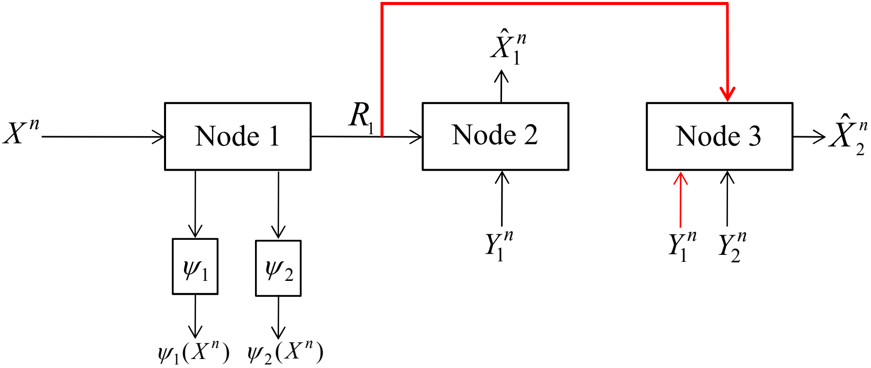
Cascade Source Coding Problem with CR constraint

Outer bound to the Rate-Distortion Region ($X - Y_1 - Y_2$)



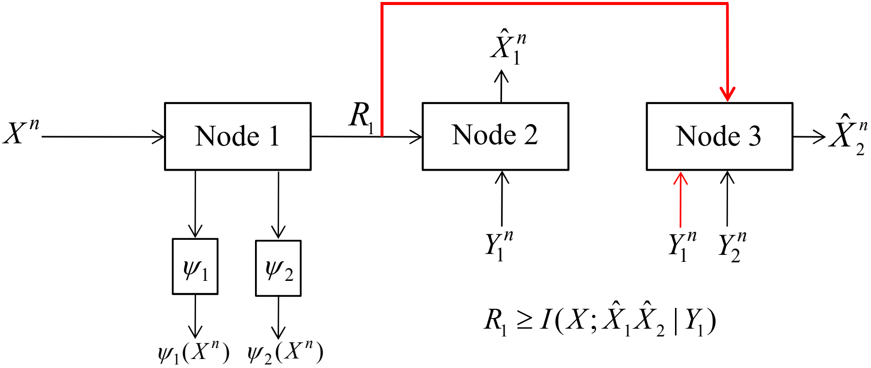
Cascade Source Coding Problem with CR constraint

Outer bound to the Rate-Distortion Region ($X - Y_1 - Y_2$)



Cascade Source Coding Problem with CR constraint

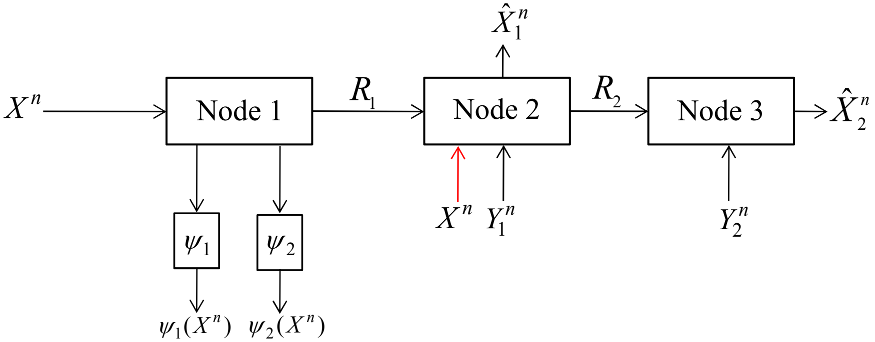
Outer bound to the Rate-Distortion Region ($X - Y_1 - Y_2$)



$$R_1 \geq I(X; \hat{X}_1 \hat{X}_2 \mid Y_1)$$

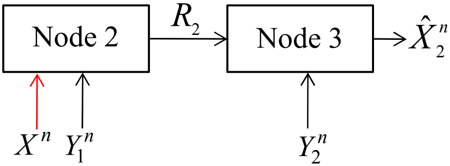
Cascade Source Coding Problem with CR constraint

Outer bound to the Rate-Distortion Region ($X - Y_1 - Y_2$)



Cascade Source Coding Problem with CR constraint

Outer bound to the Rate-Distortion Region ($X - Y_1 - Y_2$)



$$R_2 \geq I(X; \hat{X}_2 | Y_2)$$