I232 Information Theory
Chapter 11: Slepian-Wolf Coding

Brian Kurkoski

Japan Advanced Institute of Science and Technology

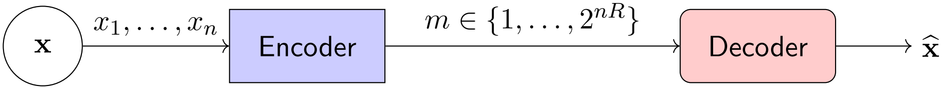
2023 May

There is no Pop Quiz for this lecture.

From Single User to Multi-User Systems

- So far, we only considered a single encoder and a single decoder:
  - Source coding: encoding of a single source
  - Channel coding: single transmitter, single receiver
- **Question:** What happens when we consider multiple sources or multiple receivers?
  - This lecture: source coding for two sources

Recall Vector Source Coding from Chapter 5



- A vector source is $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.
- Optimal encoding can compress at rate R:

$$R \geq H(\mathsf{X})$$

- Decoder produces $\widehat{\mathbf{x}} = \mathbf{x}$. Lossless source coding

11.1 Distributed Source Coding

11.1.1 Motivation for Distributed Source Coding (1/2)

Consider various types of environmental monitoring problems:

- Earthquake, fire, animal activity, acoustic submarine monitoring

Typically a large number of wireless sensors are distributed in the environment:

- Data is correlated: Physically separated but observe the same signal
- Compression is needed: Wireless transmitters need to reduce transmit power

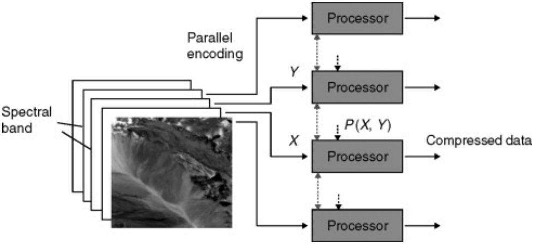
Thus, distributed source coding is suitable for such problems.

Motivation for Distributed Source Coding (2/2)

Distributed source coding suitable when encoder has limited computational power.
Consider video compression:

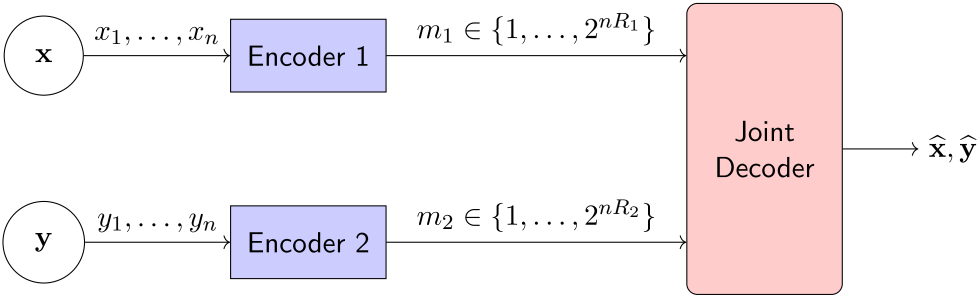
- Compression on a power-constrained device with limited computational power
- Adjacent video frames are correlated
- Each frame is encoded separately, using distributed coding

Thus, distributed source coding reduces complexity even when the encoder has access to the correlated data streams.



https://www.sciencedirect.com

11.1.2 Distributed Source Coding



We have two sources $\mathbf{x}$ and $\mathbf{y}$. Source coding is **distributed**:

- Encoder 1 sees only $\mathbf{x}$ and compresses with rate R_1
- Encoder 2 sees only $\mathbf{y}$ and compresses with rate R_2.

Assume that sources $\mathbf{x}$ and $\mathbf{y}$ are **correlated**

Correlated Sources

Correlated sources are considered $\mathbf{x}$ and $\mathbf{y}$ are considered:

$$\mathbf{x} = x_1, x_2, \dots, x_n$$
$$\mathbf{y} = y_1, y_2, \dots, y_n$$

Correlation means x_i, y_i are jointly distributed, but pairs (x_i, y_i) are i.i.d.:

$$p_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} p_{\mathsf{XY}}(x_i, y_i)$$

When $\mathbf{x}$ and $\mathbf{y}$ are binary, define $\mathbf{z}$ as:

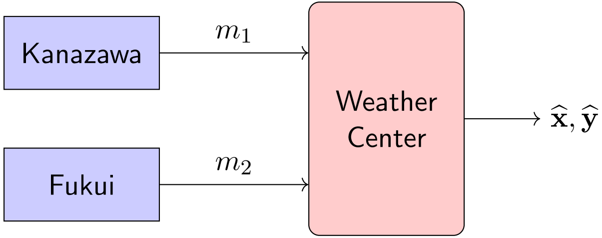
$$\mathbf{z} = \mathbf{x} \oplus \mathbf{y},$$

so that $z_i = 0$ if the two source sequences agree in position i. If they disagree, then $z_i = 1$.

Example: Transmitting the Weather

$$\mathbf{x} = 1000011101$$

$$\mathbf{y} = 1000110101$$



The weather in Kanazawa and Fukui is correlated. Transmit to the weather center:

$$0 = \text{rainy}$$
$$1 = \text{sunny}$$

x and y differ with probability p: $\Pr[x \neq y] = p$. Difference sequence:

$$\mathbf{z} = \mathbf{x} \oplus \mathbf{y} = 0000101000.$$

Take $p_X(x) = p_Y(y) = [\frac{1}{2}, \frac{1}{2}]$,

Three Encoding Strategies

1. Naive Encoders: Encoders ignore correlation. Encoders do not communicate.
2. Communicating Encoders: Encoders use correlation. Encoders communicate.
3. Slepian-Wolf encoding: Encoders use correlation. Encoders do not communicate.

Our main interest is Slepian-Wolf encoding, but introduce naive encoders and communicating encoders to show the significance of Slepian-Wolf encoding.

11.1.3 Naive Encoders

Naive encoders ignore the correlation in X, Y.

- Encoder 1 compresses $\mathbf{x}$ are rate R_1.
- Encoder 2 compresses $\mathbf{y}$ are rate R_2.

What are the achievable rate pairs (R_1, R_2)?

- Using what we already know:

$$H(\mathsf{X}) \leq R_1$$
$$H(\mathsf{Y}) \leq R_2$$

11.1.3 Naive Encoders

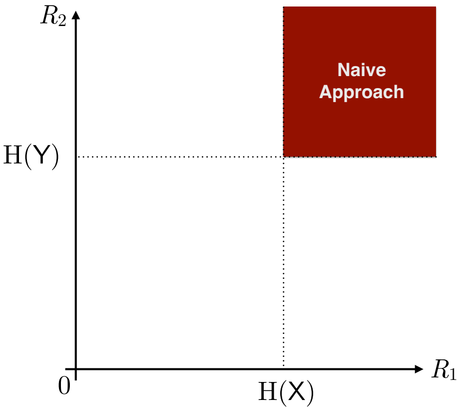
Naive encoders ignore the correlation in X, Y.

- Encoder 1 compresses $\mathbf{x}$ are rate R_1.
- Encoder 2 compresses $\mathbf{y}$ are rate R_2.

What are the achievable rate pairs (R_1, R_2)?

- Using what we already know:

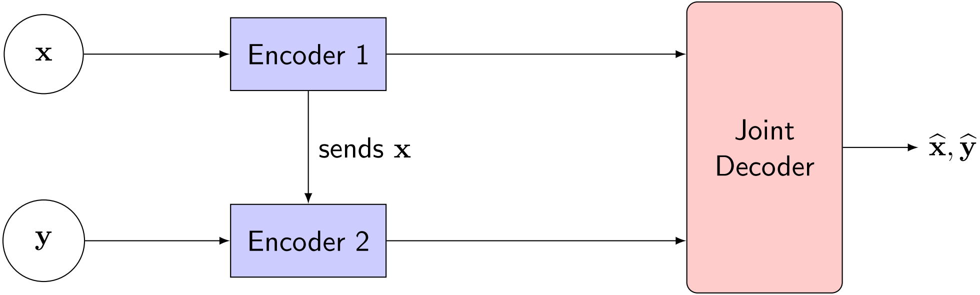
$$H(\mathsf{X}) \leq R_1$$
$$H(\mathsf{Y}) \leq R_2$$



Red region indicates achievable rate pairs (R_1, R_2) for naive encoders.

11.1.4 Communicating Encoders

With communicating encoders, temporarily ignore the distributed source coding framework, and allow Encoder 1 to send $\mathbf{x}$ to Encoder 2.



We have two sources $\mathbf{x}$ and $\mathbf{y}$. Source coding is **distributed**:

- Encoder 1 compresses $\mathbf{x}$ with rate $R_1 = H(\mathsf{X})$
- Encoder 2 uses $\mathbf{x}$ and to compress $\mathbf{y}$ with rate $R_2 = H(\mathsf{Y}|\mathsf{X})$.

Example: Transmitting the Weather

Weather Example on the whiteboard:

1. Numerical values of achievable rates for naive scheme
2. "Joint Encoding": Encoder 2 knows $\mathbf{x}$.
3. Numerical values of achievable rates for joint encoding.

Slepian-Wolf Theorem

As before, Encoder 1 and Encoder 2 do not communicate.

Slepian-Wolf Theorem: The achievable rate region for the pair of rates, (R_1, R_2) is the set of points that satisfy:

$$H(\mathsf{X}|\mathsf{Y}) \leq R_1$$
$$H(\mathsf{Y}|\mathsf{X}) \leq R_2,$$
$$H(\mathsf{X}, \mathsf{Y}) \leq R_1 + R_2$$

Surprising result: Encoders which do not communicate achieve the same rate as the joint encoder.

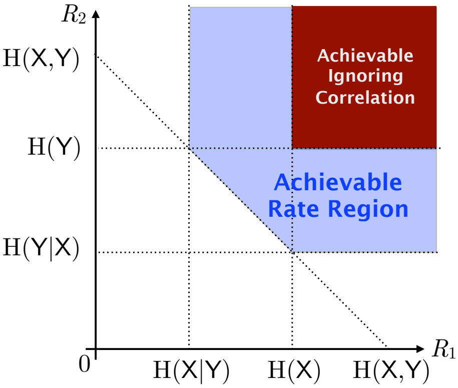
Proof technique:

- Lower bounds: Joint encoding. We cannot do better than this.
- Achievability: random binning with typical sequences.

Slepian-Wolf Theorem

Slepian-Wolf Theorem: The achievable rate region for the pair of rates, (R_1, R_2) is the set of points that satisfy:

$$H(\mathsf{X}|\mathsf{Y}) \leq R_1$$
$$H(\mathsf{Y}|\mathsf{X}) \leq R_2,$$
$$H(\mathsf{X},\mathsf{Y}) \leq R_1 + R_2$$



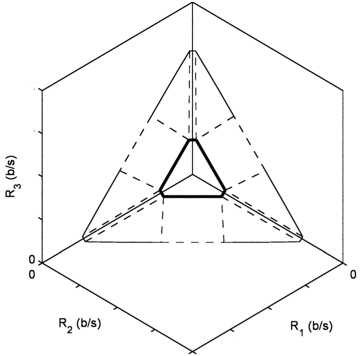
Distributed coding achieves lower rates by considering source correlation.

Rate Region of Three-User Distributed Source Coding

While Slepian-Wolf is given for two sources, it can be generalized to three or more sources.

Three-user capacity region:

$$
\begin{aligned}
R_1 &\geq H(X_1|X_2, X_3), \\
R_2 &\geq H(X_2|X_1, X_3), \\
R_3 &\geq H(X_3|X_1, X_2), \\
R_1 + R_2 &\geq H(X_1, X_2|X_3), \\
R_1 + R_3 &\geq H(X_1, X_3|X_2), \\
R_2 + R_3 &\geq H(X_2, X_3|X_1), \\
R_1 + R_2 + R_3 &\geq H(X_1, X_2, X_3).
\end{aligned}
$$

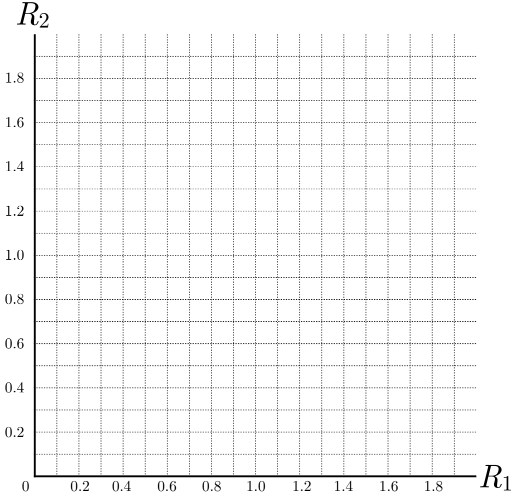


Find the Slepian-Wolf achievable rate region for this source

$P_{X,Y}$		Y = 0	Y = 1
X	0	3/8	1/8
	1	1/8	3/8

Note:

- $\log_2 3 \approx 1.585$

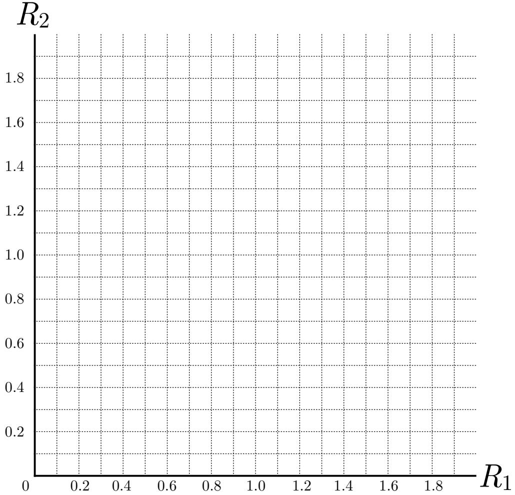


Find the Slepian-Wolf achievable rate region for this source

$P_{X,Y}$		$Y=0$	$Y=1$
X	0	3/8	1/8
	1	1/8	3/8

Note:

- $\log_2 3 \approx 1.585$



11.2 Finite-Length Slepian-Wolf Coding

Consider a concrete, finite length scheme. Assume:

- $\mathbf{x}$ and $\mathbf{y}$ are binary vectors, with finite length n.
- $p_{\mathsf{X}}(x) = p_{\mathsf{Y}}(y) = [\frac{1}{2}, \frac{1}{2}]$
- Correlation: assume $\mathbf{x}$ and $\mathbf{y}$ differ in at most t positions:

$$\mathbf{x} \oplus \mathbf{y} = \mathbf{z} \text{ has } t \text{ or fewer ones}$$

11.2 Finite-Length Slepian-Wolf Coding

Consider a concrete, finite length scheme. Assume:

- $\mathbf{x}$ and $\mathbf{y}$ are binary vectors, with finite length n.
- $p_{\mathsf{X}}(x) = p_{\mathsf{Y}}(y) = [\frac{1}{2}, \frac{1}{2}]$
- Correlation: assume $\mathbf{x}$ and $\mathbf{y}$ differ in at most t positions:

$$\mathbf{x} \oplus \mathbf{y} = \mathbf{z} \text{ has } t \text{ or fewer ones}$$

Slepian-Wolf Encoding: Partition all sequences $\mathcal{Y}^n$ into 2^m bins. Each bin has an m-bit label.

1. Encoder 1 transmits $\mathbf{x}$ uncompressed, with rate $R_1 = 1$.
2. Encoder 2 finds the bin that $\mathbf{y}$ belongs to. Encoder 2 transmits the bin label with rate $R_2 = m/n$.

11.2 Finite-Length Slepian-Wolf Coding

Consider a concrete, finite length scheme. Assume:

- $\mathbf{x}$ and $\mathbf{y}$ are binary vectors, with finite length n.
- $p_{\mathsf{X}}(x) = p_{\mathsf{Y}}(y) = [\frac{1}{2}, \frac{1}{2}]$
- Correlation: assume $\mathbf{x}$ and $\mathbf{y}$ differ in at most t positions:

$$\mathbf{x} \oplus \mathbf{y} = \mathbf{z} \text{ has } t \text{ or fewer ones}$$

Slepian-Wolf Encoding: Partition all sequences $\mathcal{Y}^n$ into 2^m bins. Each bin has an m-bit label.

1. Encoder 1 transmits $\mathbf{x}$ uncompressed, with rate $R_1 = 1$.
2. Encoder 2 finds the bin that $\mathbf{y}$ belongs to. Encoder 2 transmits the bin label with rate $R_2 = m/n$.

Decoding: 1. The decoder already has $\mathbf{x}$.

2. The decoder looks in bin m. It chooses $\widehat{\mathbf{y}}$ to be the sequence in bin m that differs from $\mathbf{x}$ in the smallest number of positions.

Finite-Length Slepian-Wolf Coding

Key point regarding the partition of sequences:

- Sequences in one bin should differ in a large number positions
- Specifically: If the source sequence differ in $\leq t$ positions, then any two sequences in a bin should differ in $\geq 2t+1$ positions.

A partition is also called a codebook.

Example 1: Binning

Consider Slepian-Wolf compression of binary sources $\mathbf{x}$ and $\mathbf{y}$ with length $n = 3$ which differ in 1 or fewer positions.

Use the codebook with four bins, given by the following table.

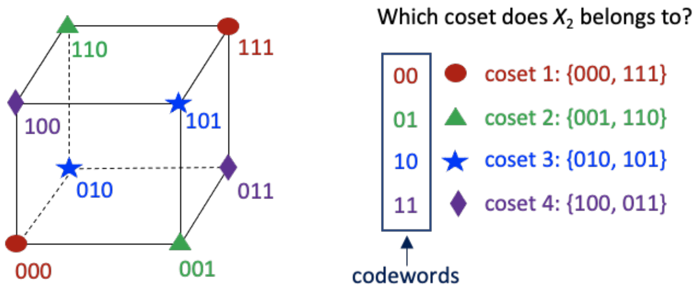
$\mathbf{s} = 00$	$\mathbf{s} = 01$	$\mathbf{s} = 10$	$\mathbf{s} = 11$
000	001	010	011
111	110	101	100

Each bin has two codewords.

Suppose $\mathbf{x} = 010$ and $\mathbf{y} = 110$, which differ in one position. ★2

Visualization of Binning

Visualize all $n = 3$ sequences using the "Hamming cube".



Sequences in each bin differ in at least 3 positions. "Coset" means "bin".

In our example, decoder receives $\mathbf{x} = 010$ and $\mathbf{s} = 01$, corresponding to coset 2.

Example 2: Linear Block Codes

Binning operations may be performed using linear block codes.

The code is specified by an r-by-n *parity-check matrix* $\mathbf{H}$.

Encoder 1 transmits $\mathbf{x}$ and Encoder 2 transmits the *syndrome* $\mathbf{s}$:

$$\mathbf{s}_2 = \mathbf{y}\mathbf{H}^{\mathrm{t}}$$

Decoder already has $\mathbf{x}$. The decoder recovers $\mathbf{y}$ from $\mathbf{s}_2$ and $\mathbf{x}$ as follows. Using the syndrome table[1] ϕ compute:

$$\begin{aligned}
\mathbf{c} &= \phi(\mathbf{s}_2) \\
\mathbf{s}_1 &= (\mathbf{x} \oplus \mathbf{c})\mathbf{H}^{\mathrm{t}} \\
\widehat{\mathbf{e}} &= \phi(\mathbf{s}_1) \\
\widehat{\mathbf{y}} &= \mathbf{x} \oplus \widehat{\mathbf{e}},
\end{aligned}$$

Above, all operations are modulo-2.

[1] gives the lowest-weight vector in bin s. Found from $\mathbf{H}$

Linear Code Example with $n = 3$

Repeat previous example using the parity check matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

which has syndrome table:

$\mathbf{s}$	$\phi(\mathbf{s})$
0 0	0 0 0
0 1	0 0 1
1 0	0 1 0
1 1	1 0 0

With $\mathbf{x} = 010$ and $\mathbf{y} = 110$, perform encoding and decoding. ★3

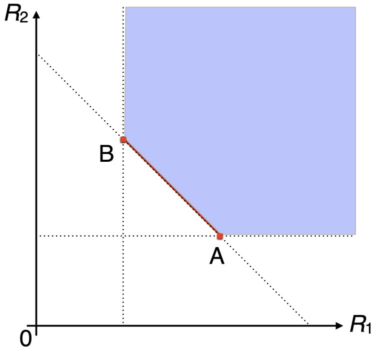
Switch Role of Encoder 1 and Encoder 2

So far, we considered achieving point A:

- Encoder 1 does not compress:
  $R_1 \geq H(\mathsf{X})$
- Encoder 2 compresses $R_2 \geq H(\mathsf{Y}|\mathsf{X})$

Achieve point B by switching roles:

- Encoder 1 compresses $R_1 \geq H(\mathsf{X}|\mathsf{Y})$
- Encoder 2 does not compress:
  $R_1 \geq H(\mathsf{Y})$



How to achieve other points between A and B?

Time Sharing (Also Called Rate Splitting)

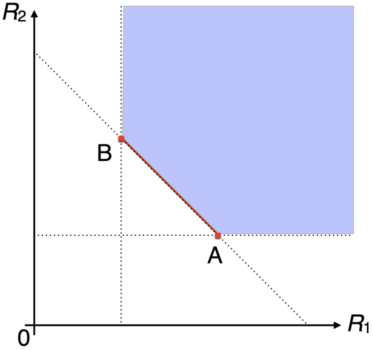
Let $\mathrm{A} = (R_1^1, R_2^1)$ and $\mathrm{B} = (R_1^2, R_2^2)$ be two achievable rates. Then, any rate pair given by:

$$(R_1, R_2) = k\mathrm{A} + (1-k)\mathrm{B}$$

for $0 \leq k \leq 1$ is also achievable.

Use code A for fraction k, and use code B for fraction $(1-k)$.

Since $n \to \infty$, we have infinite bits available for time sharing.



This also shows that the achievable rate region is convex.

Summary

- Example of network information theory: more than one transmitter/receiver
- Distributed source coding: multiple sources, each with own encoder
- Slepian-Wolf Theorem for lossless source coding:
  - Achievable rate region bounded by three inequalities
  - Surprising result: distributed encoding has the same rate as communicating encoders
  - Slepian-Wolf coding using binning

Class Info

- Tutorial Hours: Monday, May 29 at 13:30. Ask questions about homework.
- Homework 10 on LMS. Deadline: Monday, May 29 at 18:00
- Last lecture: Wednesday, May 31. Review of Information Theory. No Pop Quiz.
- Final exam: Monday, June 5 at 9:00.

Final Exam

The exam is closed book. You may use:

- **Two** pages of notes, A4-sized paper, double-sided OK.
- Blank scratch paper

You may not use anything else: No printed materials, including books, lecture notes, and slides. No notes (except as above). No internet-connected devices. No calculators. You may need to perform a 2×2 matrix inverse.

Exam Content

- Covers the whole course: Chapters 1–7, 9–10
- Emphasizes second half: channel capacity, differential entropy, rate-distortion.
- Study Homeworks.

Past exam will be provided.