Coding into a source: an inverse rate-distortion theorem

Anant Sahai
joint work with:
Mukul Agarwal    Sanjoy K. Mitter

Wireless Foundations
Department of Electrical Engineering and Computer Sciences
University of California at Berkeley

LIDS
Department of Electrical Engineering and Computer Sciences
Massachusetts Institute of Technology

2006 Allerton Conference on Communication, Control, and Computation
Suppose the aliens landed... 

- Your mission: reverse-engineer their communications technology
Suppose the aliens landed…

- Your mission: reverse-engineer their communications technology
- What assumptions should you make?
Suppose the aliens landed...

- Your mission: reverse-engineer their communications technology
- What assumptions should you make?
- They are already here!
Does information theory determine architecture?

Channel coding is an “optimal” interface.
Does information theory determine architecture?

Channel coding is an “optimal” interface.

Is it the unique interface?
Outline

1. Motivation and introduction
2. The equivalence perspective
   - Communication problems
   - Separation as equivalence
3. Main result: a direct “converse” for rate-distortion
   - Basic: finite-alphabet sources
   - Extension: real-alphabet with difference distortion
   - Conditional rate-distortion: steganography with a public cover-story
4. Application: understanding information within unstable processes
5. Conclusions and open problems
Abstract model of communication problems

- Problem: Source $S_t$, Information pattern $I_t$, and Objective $V_t$.
- Constrained resource: Noisy channel $f_c$
- Designed solution: “Encoder” $E_t$, “Decoder” $D_t$
Focus: what channels are “good enough”

- Channel $f_c$ solves the problem if $\exists \mathcal{E}, \mathcal{D}$ so system satisfies $\mathcal{V}$
- Problem $A$ is harder than problem $B$ if any $f_c$ that solves $A$ solves $B$. 
Focus: what channels are “good enough”

- Channel $f_c$ solves the problem if $\exists \mathcal{E}, \mathcal{D}$ so system satisfies $\mathcal{V}$
- Problem $A$ is harder than problem $B$ if any $f_c$ that solves $A$ solves $B$.
- Information theory is an asymptotic theory
  - Pick $\mathcal{V}$ family with an appropriate “slack” parameter
  - Consider the set of channels that solve the problem.
  - Take union over slack parameter choices.
Shannon bit-transfer problems $A_{R,\epsilon,d}$

- Source: noninteractive $X_i$ ($R$ bits): fair coin tosses
- Information pattern:
  - $D_i$ has access to $Z_1^i$
  - $E_i$ gets access to $X_1^i$
- Performance objective: $\mathcal{V}(\epsilon, d)$ is satisfied if $\mathcal{P}(X_i \neq U_{i+d}) \leq \epsilon$ for every $i \geq 0$. 
Shannon bit-transfer problems $A_{R,\epsilon,d}$

- Source: noninteractive $X_i$ ($R$ bits): fair coin tosses
- Information pattern:
  - $D_i$ has access to $Z_i^i$
  - $E_i$ gets access to $X_i^i$
- Performance objective: $\mathcal{V}(\epsilon, d)$ is satisfied if $\mathcal{P}(X_i \neq U_{i+d}) \leq \epsilon$ for every $i \geq 0$.
  - Slack parameter: permitted delay $d$
Shannon bit-transfer problems $A_{R,\epsilon,d}$

- Source: noninteractive $X_i$ ($R$ bits): fair coin tosses
- Information pattern:
  - $D_i$ has access to $Z^i_1$
  - $E_i$ gets access to $X^i_1$
- Performance objective: $V(\epsilon, d)$ is satisfied if $P(X_i \neq U_{i+d}) \leq \epsilon$ for every $i \geq 0$.
  - Slack parameter: permitted delay $d$
  - Natural orderings: larger $\epsilon, d$ is easier but larger $R$ is harder.
**Shannon bit-transfer problems** $A_{R,\epsilon,d}$

- **Source:** noninteractive $X_i$ ($R$ bits): fair coin tosses
- **Information pattern:**
  - $D_i$ has access to $Z^i_1$
  - $E_i$ gets access to $X^i_1$
- **Performance objective:** $\mathcal{V}(\epsilon, d)$ is satisfied if $\mathcal{P}(X_i \neq U_{i+d}) \leq \epsilon$ for every $i \geq 0$.
  - Slack parameter: permitted delay $d$
  - Natural orderings: larger $\epsilon, d$ is easier but larger $R$ is harder.
- **Classical capacity**

\[
C_R = \bigcap_{\epsilon > 0} \bigcap_{R' < R} \bigcup_{d > 0} \{ f_c \mid f_c \text{ solves } A_{R',\epsilon,d} \}
\]

\[
C_{\text{Shannon}}(f_c) = \sup \{ R > 0 \mid f_c \in C_R \} 
\]
Estimation to a distortion constraint: $A(F_X, \rho, D, \epsilon, d)$

- Source: *noninteractive* $X_i$ drawn iid from $F_X$
- Same information pattern
- Performance objective: $V(\rho, D, \epsilon, d)$ is satisfied if for all $\tau$,$$
\lim_{n \to \infty} P\left(\frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, U_{i+d}) > D\right) \leq \epsilon.$$

Anant Sahai (UC Berkeley)
Estimation to a distortion constraint: $A(F_X, \rho, D, \epsilon, d)$

- Source: *noninteractive* $X_i$ drawn iid from $F_X$
- Same information pattern
- Performance objective: $V(\rho, D, \epsilon, d)$ is satisfied if for all $\tau$,
  $$\lim_{n \to \infty} P\left( \frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, U_{i+d}) > D \right) \leq \epsilon.$$
  - Slack parameter: permitted delay $d$
  - Natural orderings: larger $D, d, \epsilon$ is easier
Estimation to a distortion constraint: $A(F_X, \rho, D, \epsilon, d)$

- Source: noninteractive $X_i$ drawn iid from $F_X$
- Same information pattern
- Performance objective: $V(\rho, D, \epsilon, d)$ is satisfied if for all $\tau$,
  $$\lim_{n \to \infty} P\left(\frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, U_{i+d}) > D\right) \leq \epsilon.$$
  - Slack parameter: permitted delay $d$
  - Natural orderings: larger $D, d, \epsilon$ is easier
- Channels that are good enough
  $$C_{\epsilon, (F_X, \rho, D)} = \bigcap_{\epsilon > D} \bigcap_{D' > D} \bigcup_{d > 0} \{f_c | f_c \text{ solves } A(F_X, \rho, D', \epsilon, d)\}$$
Estimation to a distortion constraint: $A(F_X, \rho, D, \epsilon, d)$

- **Source:** noninteractive $X_i$ drawn iid from $F_X$
- **Same information pattern**
- **Performance objective:** $V(\rho, D, \epsilon, d)$ is satisfied if for all $\tau$,
  \[
  \lim_{n \to \infty} P\left(\frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, U_{i+d}) > D\right) \leq \epsilon.
  \]
  - Slack parameter: permitted delay $d$
  - Natural orderings: larger $D, d, \epsilon$ is easier
- **Channels that are good enough**
  \[
  C_e(F_X, \rho, D) = \bigcap_{\epsilon > D} \bigcap_{D' > D} \bigcup_{d > 0} \{f_c | f_c \text{ solves } A(F_X, \rho, D', \epsilon, d)\}
  \]
- "Separation Theorem" if $\rho$ is finite.
  \[
  (C_R(D) \cap C^m) = (C_e(F_X, \rho, D) \cap C^m)
  \]
Classical separation revisited

Source codes

Shannon
Bit-Transfer
Equivalence proved using mutual information characterizations of $R(D)$ and $C_{\text{Shannon}}$. 
Classical separation revisited

- Equivalence proved using mutual information characterizations of $R(D)$ and $C_{\text{Shannon}}$.

- Bidirectional reductions at the problem level.
Outline

1. Motivation and introduction
2. The equivalence perspective
   - Communication problems
   - Separation as equivalence
3. Main result: a direct “converse” for rate-distortion
   - Basic: finite-alphabet sources
   - Extension: real-alphabet with difference distortion
   - Conditional rate-distortion: steganography with a public cover-story
4. Application: understanding information within unstable processes
5. Conclusions and open problems
Main Result

Suppose we have a family of black-box systems indexed by $\epsilon$ that can communicate streams from input alphabet $\{\mathcal{X}\}$ satisfying $\forall \tau$:

$$\lim_{n \to \infty} P\left(\frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, \hat{X}_i) > D\right) \leq \epsilon$$
Main Result

Suppose we have a family of black-box systems indexed by $\epsilon$ that can communicate streams from input alphabet $\{X\}$ satisfying $\forall \tau$:

$$
\lim_{n \to \infty} P\left(\frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, \hat{X}_i) > D\right) \leq \epsilon
$$

- The distortion measure $\rho$ is non-negative and additive.
Main Result

Suppose we have a family of black-box systems indexed by $\epsilon$ that can communicate streams from input alphabet $\{\mathcal{X}\}$ satisfying $\forall \tau$:

$$\lim_{n \to \infty} P\left(\frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, \hat{X}_i) > D\right) \leq \epsilon$$

- The distortion measure $\rho$ is non-negative and additive.
- The probability measure $P$ is for $\{X_i\}$ iid according to $P(x)$
Main Result

Suppose we have a family of black-box systems indexed by $\epsilon$ that can communicate streams from input alphabet $\{\mathcal{X}\}$ satisfying $\forall \tau$:

$$\lim_{n \to \infty} P\left( \frac{1}{n} \sum_{i=\tau}^{\tau+n-1} \rho(X_i, \hat{X}_i) > D \right) \leq \epsilon$$

- The distortion measure $\rho$ is non-negative and additive.
- The probability measure $P$ is for $\{X_i\}$ iid according to $P(x)$

Then, assuming access to common-randomness at both the encoder and decoder, it is possible to reliably communicate bits at all rates $R < R(D)$ bits per source symbol over the black-box system so that the probability of bit error is as small as desired.
Relationships to existing perspectives

Coding theory

- Faith in “distance” rather than probabilistic model for channel.
- Generalizes Hamming distance to arbitrary additive distortion measures.
Relationships to existing perspectives

Coding theory
- Faith in “distance” rather than probabilistic model for channel.
- Generalizes Hamming distance to arbitrary additive distortion measures.

Arbitrarily varying channels
- Attacker is not limited to a set of attack channels.
- Attacker is not forced to be causal.
- Attacker only constrained on long blocks, not individual letters.
- Attacker only promises low distortion for inputs that look like iid $P(x)$. 
Relationships to existing perspectives

- Coding theory
  - Faith in “distance” rather than probabilistic model for channel.
  - Generalizes Hamming distance to arbitrary additive distortion measures.

- Arbitrarily varying channels
  - Attacker is not limited to a set of attack channels.
  - Attacker is not forced to be causal.
  - Attacker only constrained on long blocks, not individual letters.
  - Attacker only promises low distortion for inputs that look like iid $P(x)$.

- Steganography/Watermarking
  - No cover-text
  - In conditional case, a “cover-story” instead
  - Otherwise, Merhav and Somekh-Baruch closest to this work.
Finite alphabet case

- Random encoder

- “Nearest typical neighbor” decoder

- Error events:
Finite alphabet case

- Random encoder
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.

- “Nearest typical neighbor” decoder

- Error events:
Finite alphabet case

- Random encoder
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.
  - Transmit the one corresponding to the message.
- “Nearest typical neighbor” decoder

- Error events:
Finite alphabet case

- Random encoder
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.
  - Transmit the one corresponding to the message.

- “Nearest typical neighbor” decoder
  - Define the typical codeword set $C_R$ to be codewords with type $P_X \pm \epsilon'$ for small $\epsilon'$.

- Error events:
Finite alphabet case

- Random encoder
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.
  - Transmit the one corresponding to the message.

- “Nearest typical neighbor” decoder
  - Define the typical codeword set $C_R$ to be codewords with type $P_X \pm \epsilon'$ for small $\epsilon'$.
  - Let $y_1^n$ be the received string.

- Error events:
Finite alphabet case

- **Random encoder**
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.
  - Transmit the one corresponding to the message.

- **“Nearest typical neighbor” decoder**
  - Define the typical codeword set $C_R$ to be codewords with type $P_X \pm \epsilon'$ for small $\epsilon'$.
  - Let $y_1^n$ be the received string.
  - Decode to $\hat{X}_1^n \in C_R$ closest to $y_1^n$ in a $\rho$-distortion sense.

- **Error events:**
Finite alphabet case

- Random encoder
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.
  - Transmit the one corresponding to the message.

- “Nearest typical neighbor” decoder
  - Define the typical codeword set $C_R$ to be codewords with type $P_X \pm \epsilon'$ for small $\epsilon'$.
  - Let $y_1^n$ be the received string.
  - Decode to $\hat{X}_1^n \in C_R$ closest to $y_1^n$ in a $\rho$-distortion sense.

- Error events:
  - Atypical codeword $X_1^n$. Probability tends to 0.
Finite alphabet case

- **Random encoder**
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.
  - Transmit the one corresponding to the message.

- **“Nearest typical neighbor” decoder**
  - Define the typical codeword set $C_R$ to be codewords with type $P_X \pm \epsilon'$ for small $\epsilon'$.
  - Let $y_1^n$ be the received string.
  - Decode to $\hat{X}_1^n \in C_R$ closest to $y_1^n$ in a $\rho$-distortion sense.

- **Error events:**
  - Atypical codeword $X_1^n$. Probability tends to 0.
  - Excess average distortion ($> D$) on true codeword. Probability tends to 0 by assumption.
Finite alphabet case

- Random encoder
  - Randomly draw $2^{nR}$ length $n$ codewords iid according to $P_X$ using common randomness.
  - Transmit the one corresponding to the message.

- “Nearest typical neighbor” decoder
  - Define the typical codeword set $C_R$ to be codewords with type $P_X \pm \epsilon'$ for small $\epsilon'$.
  - Let $y_1^n$ be the received string.
  - Decode to $\hat{X}_1^n \in C_R$ closest to $y_1^n$ in a $\rho$-distortion sense.

- Error events:
  - Atypical codeword $X_1^n$. Probability tends to 0.
  - Excess average distortion (> $D$) on true codeword. Probability tends to 0 by assumption.
  - There exists a false codeword has average distortion < $D$
Suppose $y^n_1$ has type $q_Y$. Consider $z^n_1 \in C_R$ that is false. Let $q_{XY}$ be the resulting joint type.
Probability of false codeword being close

- Suppose $y^n$ has type $q_Y$. Consider $z^n_1 \in C_R$ that is false. Let $q_{XY}$ be the resulting joint type.
- $q_X \in p_X \pm \epsilon'$
Suppose $y^n_1$ has type $q_Y$. Consider $z^n_1 \in C_R$ that is false. Let $q_{XY}$ be the resulting joint type.

$q_X \in p_X \pm \epsilon'$

$E_{q_{XY}}[\rho(X, Y)] \leq D$ if an error.
Probability of false codeword being close

- Suppose $y_1^n$ has type $q_Y$. Consider $z_1^n \in C_R$ that is false. Let $q_{XY}$ be the resulting joint type.
- $q_X \in p_X \pm \epsilon'$
- $E_{q_{XY}}[\rho(X, Y)] \leq D$ if an error.

- Sort $y_1^n$ and correspondingly shuffle codeword $z_1^n$
Probability of false codeword being close

- Suppose $y_1^n$ has type $q_Y$. Consider $z_1^n \in C_R$ that is false. Let $q_{XY}$ be the resulting joint type.
- $q_X \in p_X \pm \epsilon'$
- $E_{q_{XY}}[\rho(X, Y)] \leq D$ if an error.

Sort $y_1^n$ and correspondingly shuffle codeword $z_1^n$

$P(Z_1^n \text{ collides}) \leq \prod_j 2^{-nq_Y(j)D(q_{X|Y=j}||p_X)} = 2^{-nD(q_{XY}||p_Xq_Y)}$
Bounding probability of collision

- Union bound over $2^{nR}$ codewords and $(n + 1)|\mathcal{X}||\mathcal{Y}|$ joint types $q_{XY}$.
- Minimize divergence $D(q_{XY}||p_X q_Y)$ over set of joint types $q_{XY}$ satisfying:
  - $E_{q_{XY}}[\rho(X, Y)] \leq D$
  - $q_X \in p_X \pm \epsilon'$
Bounding probability of collision

- Union bound over $2^{nR}$ codewords and $(n + 1)|\mathcal{X}||\mathcal{Y}|$ joint types $q_{XY}$.
- Minimize divergence $D(q_{XY}||p_X q_Y)$ over set of joint types $q_{XY}$ satisfying:
  - $E_{q_{XY}}[\rho(X,Y)] \leq D$
  - $q_X \in p_X \pm \epsilon'$

Key step:

$$D(q_{XY}||p_X q_Y) = D(q_X||p_X) + D(q_{XY}||q_X q_Y) \leq D(q_{XY}||q_X q_Y) = I_q(X;Y)$$
Bounding probability of collision

- Union bound over $2^{nR}$ codewords and $(n + 1)|\mathcal{X}||\mathcal{Y}|$ joint types $q_{XY}$.
- Minimize divergence $D(q_{XY}||p_Xq_Y)$ over set of joint types $q_{XY}$ satisfying:
  - $E_{q_{XY}}[\rho(X,Y)] \leq D$
  - $q_X \in p_X \pm \epsilon'$

Key step:
\[ D(q_{XY}||p_Xq_Y) = D(q_X||p_X) + D(q_{XY}||q_Xq_Y) \leq D(q_{XY}||q_Xq_Y) = I_q(X;Y) \]
- Minimizing $I_q(X;Y)$ gives $R(D)$ when $\epsilon' \to 0$
Bounding probability of collision

- Union bound over $2^{nR}$ codewords and $(n + 1)|\mathcal{X}||\mathcal{Y}|$ joint types $q_{XY}$.
- Minimize divergence $D(q_{XY}||p_X q_Y)$ over set of joint types $q_{XY}$ satisfying:
  - $E_{q_{XY}}[\rho(X, Y)] \leq D$
  - $q_X \in p_X \pm \epsilon'$

  Key step:
  $D(q_{XY}||p_X q_Y) = D(q_X||p_X) + D(q_{XY}||q_X q_Y) \leq D(q_{XY}||q_X q_Y) = I_q(X; Y)$
  Minimizing $I_q(X; Y)$ gives $R(D)$ when $\epsilon' \to 0$
  If $R < R(D)$, collision probability $\to 0$ with $n$. 
Outline

1. Motivation and introduction
2. The equivalence perspective
   - Communication problems
   - Separation as equivalence
3. Main result: a direct “converse” for rate-distortion
   - Basic: finite-alphabet sources
   - Extension: real-alphabet with difference distortion
   - Conditional rate-distortion: steganography with a public cover-story
4. Application: understanding information within unstable processes
5. Conclusions and open problems
Extending to real vector alphabets

- Compact support and difference-distortion
  - Same codebook construction
Extending to real vector alphabets

- Compact support and difference-distortion
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
Extending to real vector alphabets

- Compact support and difference-distortion
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.
Extending to real vector alphabets

- Compact support and difference-distortion
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.

- Unbounded support
  - New codebook construction:
    - View $F_X(x) = (1 - \delta)F_{\tilde{X}}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
Extending to real vector alphabets

- **Compact support and difference-distortion**
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.

- **Unbounded support**
  - New codebook construction:
    - View $F_X(x) = (1 - \delta)F_{\tilde{X}}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
    - First flip $n$ iid unfair coins with $P(\text{head}) = \delta$. 

Anant Sahai (UC Berkeley)
Compact support and difference-distortion

- Same codebook construction
- Pick a fine enough quantization $\Delta$
- Reduces to finite alphabet case at decoder with a small factor increase in distortion.

Unbounded support

- New codebook construction:
  - View $F_X(x) = (1 - \delta)F_{\tilde{X}}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
  - First flip $n$ iid unfair coins with $P(\text{head}) = \delta$.
  - Mark positions as dirty or clean.
Extending to real vector alphabets

- Compact support and difference-distortion
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.

- Unbounded support
  - New codebook construction:
    - View $F_X(x) = (1 - \delta)F_{\tilde{X}}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
    - First flip $n$ iid unfair coins with $P(\text{head}) = \delta$.
    - Mark positions as dirty or clean.
    - Draw codewords from $\tilde{X}$ in clean positions and $\bar{X}$ in dirty ones.
Extending to real vector alphabets

- Compact support and difference-distortion
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.

- Unbounded support
  - New codebook construction:
    - View $F_X(x) = (1 - \delta)F_{\tilde{X}}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
    - First flip $n$ iid unfair coins with $P(\text{head}) = \delta$.
    - Mark positions as dirty or clean.
    - Draw codewords from $\tilde{X}$ in clean positions and $\bar{X}$ in dirty ones.
  - Modified decoding rule:
    - Declare error if fewer than $(1 - 2\delta)n$ clean positions.
Extending to real vector alphabets

- Compact support and difference-distortion
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.

- Unbounded support
  - New codebook construction:
    - View $F_X(x) = (1 - \delta)F_{\tilde{X}}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
    - First flip $n$ iid unfair coins with $P(\text{head}) = \delta$.
    - Mark positions as dirty or clean.
    - Draw codewords from $\tilde{X}$ in clean positions and $\bar{X}$ in dirty ones.
  - Modified decoding rule:
    - Declare error if fewer than $(1 - 2\delta)n$ clean positions.
    - Restrict codebook to clean positions only for decoding purposes.
Extension to real vector alphabets

- **Compact support and difference-distortion**
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.

- **Unbounded support**
  - New codebook construction:
    - View $F_X(x) = (1 - \delta)F_\tilde{X}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
    - First flip $n$ iid unfair coins with $P(\text{head}) = \delta$.
    - Mark positions as dirty or clean.
    - Draw codewords from $\tilde{X}$ in clean positions and $\bar{X}$ in dirty ones.
  - Modified decoding rule:
    - Declare error if fewer than $(1 - 2\delta)n$ clean positions.
    - Restrict codebook to clean positions only for decoding purposes.
  - Increases distortion by a factor of $\frac{1 + 2\delta}{1 - 2\delta}$.
Extending to real vector alphabets

- **Compact support and difference-distortion**
  - Same codebook construction
  - Pick a fine enough quantization $\Delta$
  - Reduces to finite alphabet case at decoder with a small factor increase in distortion.

- **Unbounded support**
  - New codebook construction:
    - View $F_X(x) = (1 - \delta) F_{\tilde{X}}(x) + \delta F_{\bar{X}}(x)$ where $\tilde{X}$ has compact support.
    - First flip $n$ iid unfair coins with $P(\text{head}) = \delta$.
    - Mark positions as dirty or clean.
    - Draw codewords from $\tilde{X}$ in clean positions and $\bar{X}$ in dirty ones.
  - Modified decoding rule:
    - Declare error if fewer than $(1 - 2\delta)n$ clean positions.
    - Restrict codebook to clean positions only for decoding purposes.
  - Increases distortion by a factor of $\frac{1 + 2\delta}{1 - 2\delta}$.
  - $\lim_{\Delta \to 0} \lim_{\delta \to 0} R_{\Delta, \delta}(D) = R(D)$
Conditional Rate-Distortion

- Assume “coverstory” $V^n_1$ drawn according to $P(V)$ is known to all parties: encoder, decoder, *and attacker*.
- All $R < R_{X|V}(D)$ are achievable.
Conditional Rate-Distortion

- Assume “coverstory” $V_1^n$ drawn according to $P(V)$ is known to all parties: encoder, decoder, and attacker.
- All $R < R_{X|V}(D)$ are achievable.
  - Encoder draws conditionally using $P(X|V)$. 
Conditional Rate-Distortion

- Assume “coverstory” $V_1^n$ drawn according to $P(V)$ is known to all parties: encoder, decoder, \textit{and attacker}.
- All $R < R_{X|V}(D)$ are achievable.
  - Encoder draws conditionally using $P(X|V)$.
  - Codeword typicality defined similarly (include $V_1^n$)
Conditional Rate-Distortion

- Assume “coverstory” $V_1^n$ drawn according to $P(V)$ is known to all parties: encoder, decoder, and attacker.
- All $R < R_{X|V}(D)$ are achievable.
  - Encoder draws conditionally using $P(X|V)$.
  - Codeword typicality defined similarly (include $V_1^n$)
  - Nearest-conditionally-typical codeword decoding
Conditional Rate-Distortion

- Assume “coverstory” $V^n_1$ drawn according to $P(V)$ is known to all parties: encoder, decoder, and attacker.
- All $R < R_{X|V}(D)$ are achievable.
  - Encoder draws conditionally using $P(X|V)$.
  - Codeword typicality defined similarly (include $V^n_1$)
  - Nearest-conditionally-typical codeword decoding
  - Parallel proof
Outline

1. Motivation and introduction
2. The equivalence perspective
   ▶ Communication problems
   ▶ Separation as equivalence
3. Main result: a direct “converse” for rate-distortion
   ▶ Basic: finite-alphabet sources
   ▶ Extension: real-alphabet with difference distortion
   ▶ Conditional rate-distortion: steganography with a public cover-story
4. Application: understanding information within unstable processes
5. Conclusions and open problems
Unstable Markov Processes: $R(D)$

$$X_{t+1} = AX_t + W_t \text{ where } A > 1$$
Accumulation: Look at $\{X_{kn}\}$

- Can embed $R_1 < n \log_2 A$ bits per symbol
- These bits are recovered with anytime reliability if black-box has finite error moments.
Unstable Markov Processes: two kinds of information

- Accumulation: Look at \( \{X_{kn}\} \)
  - Can embed \( R_1 < n \log_2 A \) bits per symbol
  - These bits are recovered with anytime reliability if black-box has finite error moments.
- Dissipation: Look at \( \{X_{kn-1}^{k(n-1)+1} | X_{kn}\} \)
Unstable Markov Processes: two kinds of information

- Accumulation: Look at $\{X_{kn}\}$
  - Can embed $R_1 < n \log_2 A$ bits per symbol
  - These bits are recovered with anytime reliability if black-box has finite error moments.

- Dissipation: Look at $\left\{X_{k(n-1)+1}^{kn-1} | X_{kn}\right\}$
  - Can be transformed to look iid
  - Fall under our results
Unstable Markov Processes: two kinds of information

- **Accumulation**: Look at $\{X_{kn}\}$
  - Can embed $R_1 < n \log_2 A$ bits per symbol
  - These bits are recovered with anytime reliability if black-box has finite error moments.

- **Dissipation**: Look at $\{X_{k(n-1)+1}^{kn-1} | X_{kn}\}$
  - Can be transformed to look iid
  - Fall under our results
  - Can embed $R_2 < R(D) - \log_2 A$ bits per symbol
Unstable Markov Processes: two kinds of information

- **Accumulation**: Look at \( \{X_{kn}\} \)
  - Can embed \( R_1 < n \log_2 A \) bits per symbol
  - These bits are recovered with anytime reliability if black-box has finite error moments.

- **Dissipation**: Look at \( \{X_{k(n-1)+1}^{kn-1}|X_{kn}\}\)
  - Can be transformed to look iid
  - Fall under our results
  - Can embed \( R_2 < R(D) - \log_2 A \) bits per symbol

- **Two-tiered nature of information-flow** proved by direct reduction.
Conclusions and open problems

- Already extend to stationary-ergodic processes that mix.

Traditional point-to-point source-channel separation is a consequence of a problem-level equivalence that can be proved using direct reductions in both directions.
Conclusions and open problems

Traditional point-to-point source-channel separation is a consequence of a problem-level equivalence that can be proved using direct reductions in both directions.

- Already extend to stationary-ergodic processes that mix.
- Can the gap between $R_{seq}(D)$ and $R(D)$ be used to carry information?
Conclusions and open problems

- Already extend to stationary-ergodic processes that mix.
- Can the gap between \( R_{\text{seq}}(D) \) and \( R(D) \) be used to carry information?
- Apply equivalence perspective to better understand multiterminal problems.

Traditional point-to-point source-channel separation is a consequence of a problem-level equivalence that can be proved using direct reductions in both directions.
Conclusions and open problems

Traditional point-to-point source-channel separation is a consequence of a problem-level equivalence that can be proved using direct reductions in both directions.

- Already extend to stationary-ergodic processes that mix.
- Can the gap between $R_{\text{seq}}(D)$ and $R(D)$ be used to carry information?
- Apply equivalence perspective to better understand multiterminal problems.
- Is there another reason to suspect that “channel coding” is fundamental to communication?