Source Coding with Limited Horizon Side Information at the Decoder

Abbas El Gamal

Tsachy Weissman

EE Dept.

Stanford

Wyner Ziv Problem

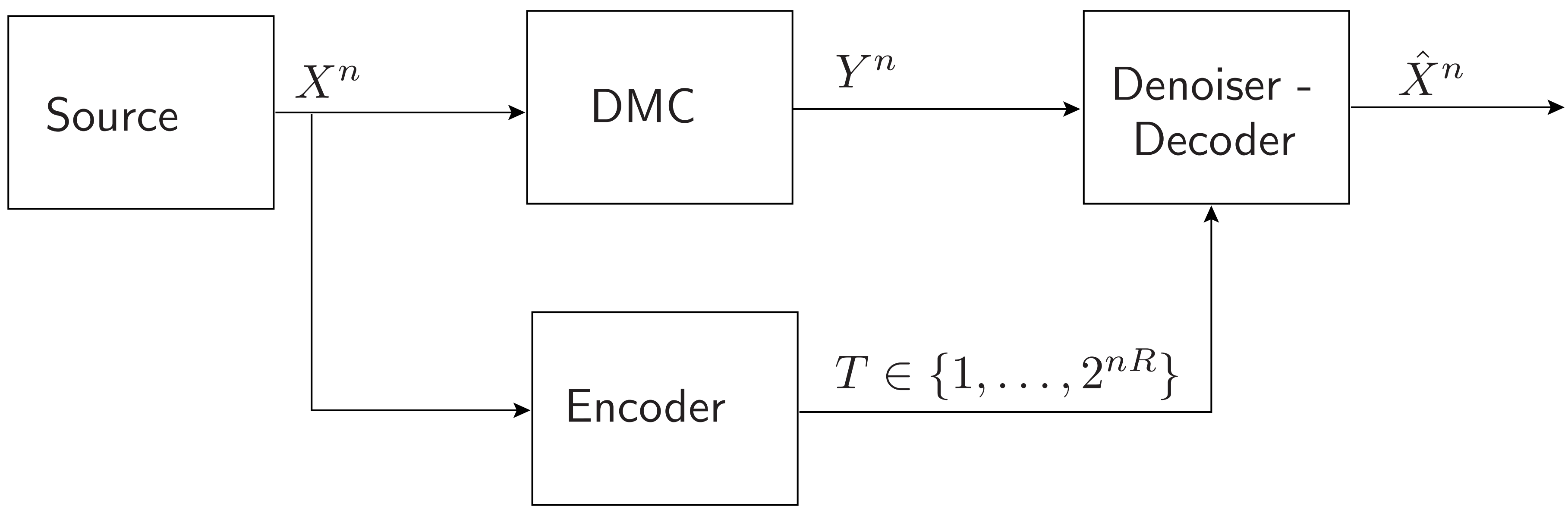


Figure 1: The Wyner-Ziv setup

Wyner Ziv Rate Distortion Function

$$R_{WZ}(D) = \min[I(X;W) - I(Y;W)] = \min I(X;W|Y)$$

$\min$ over $f : \mathcal{W} \times \mathcal{Y} \to \hat{\mathcal{X}}$, $|\mathcal{W}| \leq |\mathcal{X}| + 1$, and $P(w|x)$ such that $E\rho(X, f(W,Y)) \leq D$.

Wyner Ziv Coding with Limited Lookahead in the S.I.

- (X_i, Y_i) i.i.d. drawings of (X, Y)
- Encoding: $T(X^n) \in \{1 \ldots, 2^{nR}\}$
- Reconstruction: $\hat{X}_i(T, Y^{i+d})$, i.e., depends on the S.I. but with *limited lookahead* d

Motivation I: Sequential Zero-Delay and Delay-Constrained Source Coding

Source code specified by:

1. Encoder: sequence $\{E_i\}$, where E_i produces a code symbol $U_i \in \mathcal{U}_i$ based on observation of the source with some lookahead l, $U_i = E_i(X^{i+l})$

2. Decoder: sequence $\{D_i\}$, where D_i produces ith reconstruction symbol based on lookahead m in the code symbols and d in the side information symbols, i.e., $\hat{X}_i = D_i(U^{i+m}, Y^{i+d})$

Instantaneous rate is $\log |\mathcal{U}_i|$. Overall rate in encoding first n source symbols is $R = \frac{1}{n}\sum_{i=1}^{n} \log |\mathcal{U}_i|$.

Any source code with this structure is a member of family of schemes we allow

Motivation II: The Denoising/Filtering/Smoothing View

Given index from the encoder, decoder is a *denoiser*

$d = 0$ corresponds to filtering (sequential denoising), while $d > 0$ to fixed-lag smoothing

Example: tracking moving target whose trajectory can be described to the tracker via a rate-constrained link.

Motivation III: Broadening duality between source and channel coding

- $d = 0$

  Wyner-Ziv $\Leftrightarrow$ Gel'fand-Pinsker WZCSI $\Leftrightarrow$ Shannon Channel

- $d > 0$

  will apply our approach to characterize channel capacity when state information is available to sender with limited lookahead

Motivation IV: Connecting to Neuhoff & Gilbert Causality

- Case $d = 0$ close in spirit to causality a la Neuhoff and Gilbert. Constraint is imposed on the reconstruction, rather than on the delay introduced by the code.

- Complements [W. and Merhav, '05]

Wyner Ziv Coding with Causal S.I.

We begin by considering

- (X_i, Y_i) i.i.d. drawings of (X, Y)
- Encoding: $T(X^n) \in \{1 \ldots, 2^{nR}\}$
- Reconstruction: $\hat{X}_i(T, Y^i)$, i.e., depends on the S.I. but only *causally*

R-D Function for Wyner-Ziv Coding with *Causal* S.I.

Theorem 1. *The rate distortion function for the Wyner Ziv problem with causal side information is given by*

$$R(D) = \min I(X;W)$$

where the minimum is over all functions $f : \mathcal{W} \times \mathcal{Y} \to \hat{\mathcal{X}}$, $|\mathcal{W}| \leq |\mathcal{X}| + 1$, *and* $P(w|x)$ *such that*

$$E\rho(X, f(W,Y)) \leq D$$

[Compare with $R_{WZ}(D)$]

Proof of Theorem 1: Achievability

Encoder: Need no more than $\approx nI(X;W)$ bits to describe W^n to decoder

Decoder: Knowing W^n, reconstruct according to $\hat{X}_i = f(W_i, Y_i)$

In words: "if not allowed to look at future, past is useless"
[Reminiscent of situation in zero-delay as well as in causal source coding]

Proof of Theorem 1: Converse

$$
\begin{aligned}
nR &\geq H(T) \geq I(X^n;T) \\
&= H(X^n) - H(X^n|T) = \sum_{i=1}^{n} H(X_i) - H(X_i|T, X^{i-1}) \\
&\overset{(a)}{\geq} \sum_{i=1}^{n} H(X_i) - H(X_i|T, Y^{i-1}) \overset{(b)}{=} \sum_{i=1}^{n} I(X_i; W_i) \\
&\overset{(c)}{\geq} \sum_{i=1}^{n} R(E\rho(X_i, \hat{X}_i)) \overset{(d)}{\geq} nR\left(\frac{1}{n}\sum_{i=1}^{n} E\rho(X_i, \hat{X}_i)\right) \geq nR(D) \quad \square
\end{aligned}
$$

On Duality

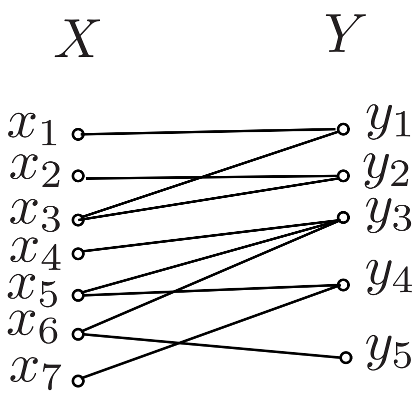
-

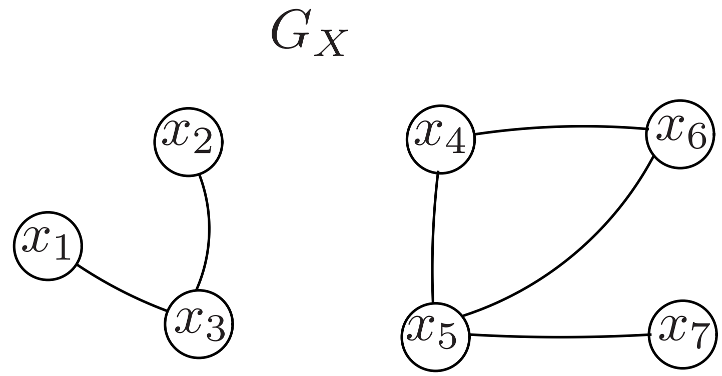
$$R_{WZ}(D) = \min[I(X;W) - I(Y;W)] \quad \Leftrightarrow \quad R(D) = \min I(X;W)$$

$$C_{GP} = \max_{p(u|s),p(x|u,s)} [I(U;Y) - I(U;S)] \quad \Leftrightarrow \quad C_{Shannon} = \max_{p(u),p(x|u,s)} I(U;Y)$$

- $R(D)$ not improved with feedforward, as observed for the other cases in [Merhav Weissman '05]

The Common Information Random Variable





- Bipartite graph has edge between x_i and y_j if and only if $P(X = x_i, Y = y_j) > 0$

- $(x_i, x_j) \in G_X$ if and only if there is y with (x_i, y) and (x_j, y) in bipartite graph
- G_X in example has two maximally connected components, so Z is binary, assuming one value on $\{x_1, x_2, x_3\}$ and another on $\{x_4, x_5, x_6, x_7\}$
- Generally, Z takes different value on each component of G_X
- Z is referred to as the 'common information random variable'

Equivalent Characterization of Rate Distortion Function

Proposition 1. *The rate distortion function of Theorem 1 is equivalently given by*

$$R(D) = \min I(X;W|Z) \ ,$$

where minimum is over exactly same set as before.

Note:

- $I(X;W) = H(W) - H(W|X) \geq H(W|Z) - H(W|X,Z) = I(X;W|Z)$ with equality if and only if W is independent of Z

- Note that for the Wyner-Ziv function $R(D) = \min I(W;X|Y)$ so conditioning on Z has no effect

Example: Doubly Symmetric Binary Source

- X unbiased input to BSC(δ). Y is output (or vice versa)
- Distortion is Hamming
- $$R(d) = \begin{cases} 1 - h(d) & 0 \leq d \leq d_c \\ -h'(d_c)d + h'(d_c)\delta & d_c < d \leq \delta, \end{cases}$$
  where d_c is solution to $(1 - h(d_c))/(d_c - \delta) = -h'(d_c)$, and $h(\cdot)$ is binary entropy
- In words: optimum performance attained by time sharing between rate distortion coding with no SI, and zero-rate decoding that uses only the SI
- In particular: side information is useless for small distortion

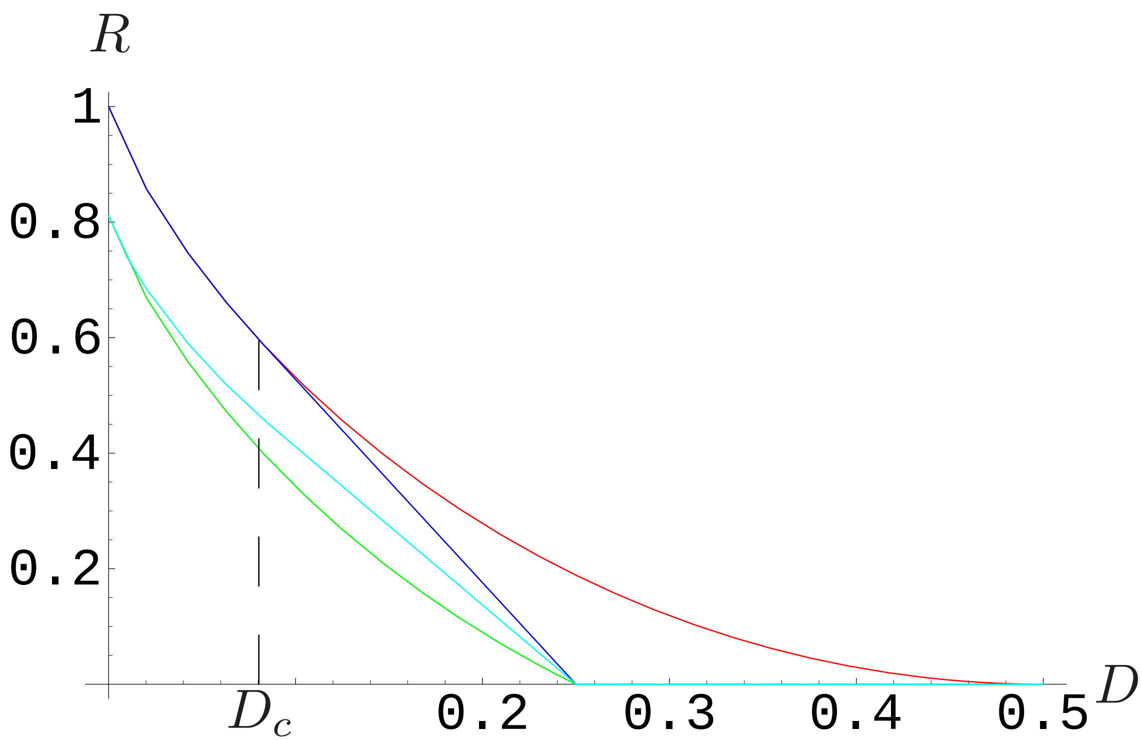


Figure 2: Rate distortion curves for doubly symmetric binary source with $\delta = 1/4$. Figure shows $R_X(D)$, $R(D)$, $R_{WZ}(D)$, $R_{X|Y}(D)$. D_c in this case can be explicitly computed and is given by $D_c = (5 - 4(19 - 3\sqrt{33})^{-1/3} - (19 - 3\sqrt{33})^{1/3})/6 \approx 0.0803566$.

Example: X, Y Jointly Gaussian

- X, Y jointly Gaussian
- Squared error distortion
- Upper bound $R(D)$ by taking $W = \alpha X + Z$ where $Z \sim \mathcal{N}(0, \sigma_Z^2) \perp X, Y$
- This gives following upper bound on rate distortion function:

$$R_{ub}(D) = \begin{cases} \frac{1}{2}\log\left[\sigma_X^2\left(\frac{1}{D} - \frac{1}{\sigma_N^2}\right)\right] & 0 < D \le \frac{\sigma_X^2\sigma_N^2}{\sigma_X^2+\sigma_N^2} \\ 0 & D \ge \frac{\sigma_X^2\sigma_N^2}{\sigma_X^2+\sigma_N^2} \end{cases}$$

Example: X, Y Jointly Gaussian (cont.)

$R_{ub}(D)$ is not necessarily convex:

Lemma 1. *1. The case $\sigma_N^2 \geq \sigma_X^2$: $R_{ub}(d)$ is convex*

2. The case $\sigma_N^2 < \sigma_X^2$: $R_{ub}(d)$ has an inflection point at $D = \sigma_N^2/2$: It is convex for $D < \sigma_N^2/2$ and concave for $\sigma_N^2/2 < D \leq \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2 + \sigma_N^2}$.

- In case $\sigma_N^2 < \sigma_X^2$ can improve by taking lower convex envelope $\underline{R}_{ub}(D)$
- Whether or not $R(D) = \underline{R}_{ub}(D)$ remains to be determined

Example: X, Y Jointly Gaussian (cont.)

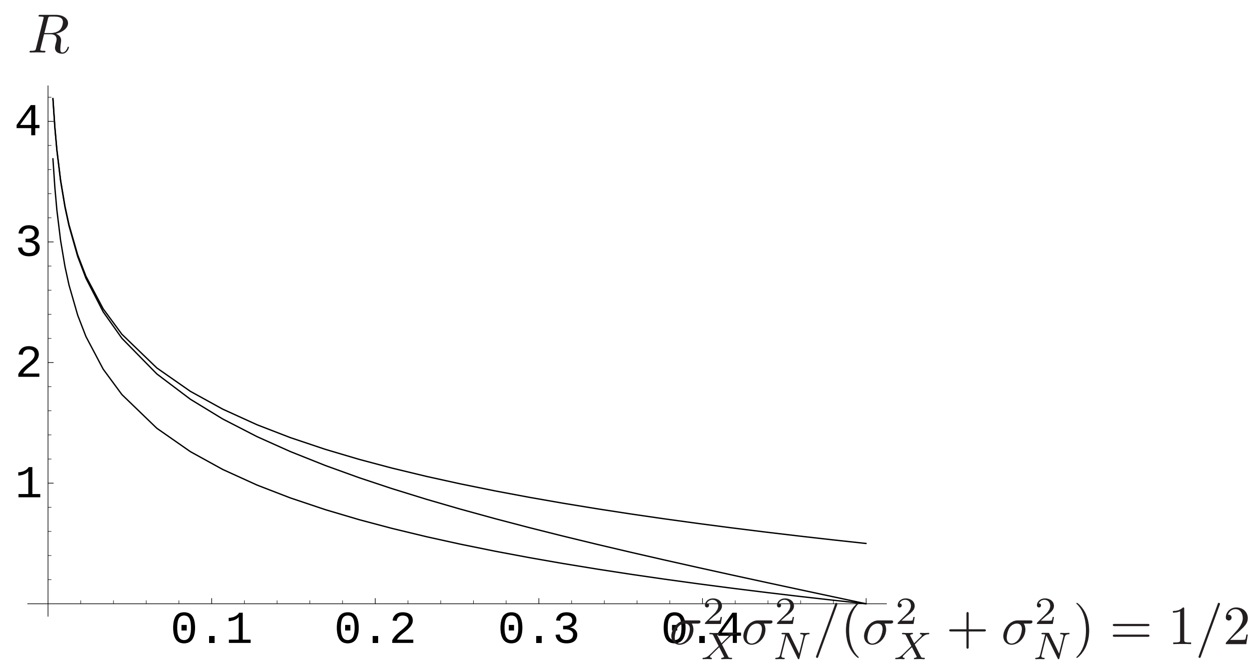


Figure 3: Typical form of the curves $R_X(D)$, $R_{ub}(D)$, and $R_{X|Y}(D)$ when $\sigma_N^2 \geq \sigma_X^2$. In this case, Lemma 1 implies that $R_{ub}(D)$ is convex. Figure shows actual curves for the case $\sigma_X = \sigma_N = 1$.

Example: X, Y Jointly Gaussian (cont.)

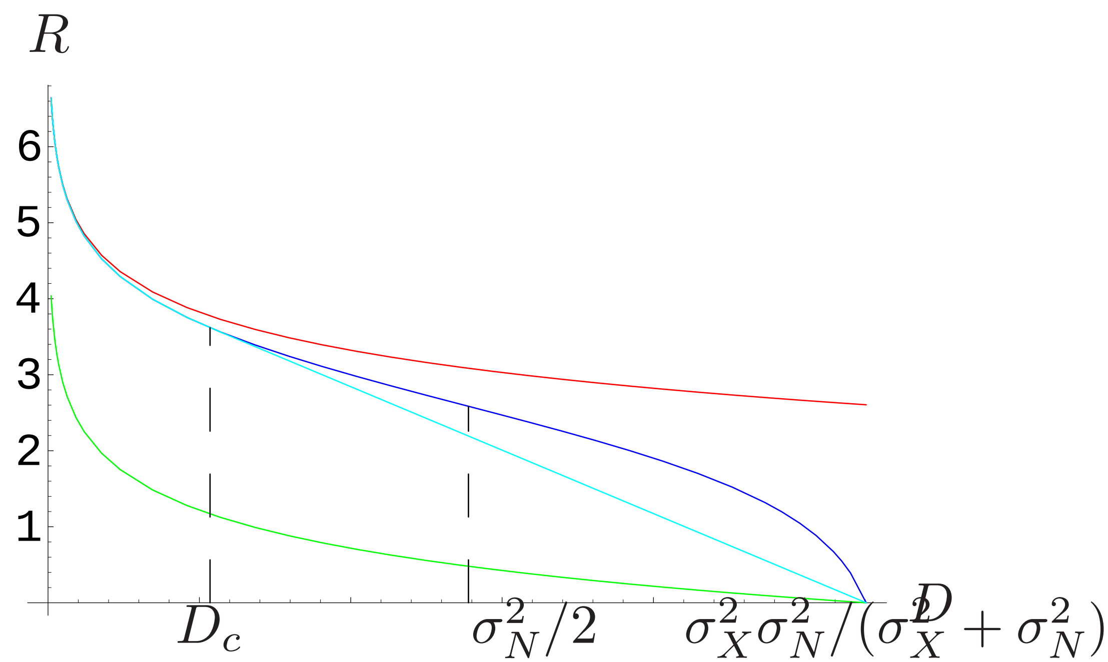


Figure 4: Curves for $\sigma_N^2 < \sigma_X^2$. Lemma implies that $R_{ub}(D)$ is not convex and can therefore be improved by its lower convex envelope $\underline{R}_{ub}(D)$. Curves shown: $R_X(D)$, $R_{ub}(D)$, $\underline{R}_{ub}(D)$, $R_{X|Y}(D)$. Figure shows curves for $\sigma_X = 1$, $\sigma_N = 1/6$. $D_c = 5.352215 \times 10^{-3}$. $D = \sigma_N^2/2$ is the inflection point, as asserted in Lemma.

Slepian-Wolf Coding with Causal S.I.

- Encoder maps X^n into $T \in \{1, \ldots, 2^{nR}\}$
- Reconstruction is of form $\hat{X}_i(T, Y^i)$
- R is achievable if exists sequence of schemes with $P(X^n \neq \hat{X}^n) \longrightarrow 0$
- Let R_{LSCSI} denote infimum over achievable rates

Slepian-Wolf Coding with Causal S.I. (cont.)

Clearly $H(X|Y) \leq R_{\text{LSCSI}} \leq H(X)$

Where in $[H(X|Y), H(X)]$ is R_{LSCSI} situated ?

Consider first three trivial cases:

$X = Y$ a.s.: $R_{LSCSI} = H(X|Y)(= 0)$.

X and Y independent: $R_{LSCSI} = H(X|Y)(= H(X))$.

U and Y independent, and $X = (U, Y)$: $R_{LSCSI} = H(X|Y)(= H(U))$.

In these cases, $R_{\text{LSCSI}} = H(X|Y)$

We will see that this is the exception rather than the rule

Slepian-Wolf Coding with Causal S.I. (cont.)

Theorem 2.

$$R_{LSCSI} = \min I(X;W) = \min I(X;W|Z)$$

where Z *is the common information r.v. and (in both minima) the minimization is over all* $P(w|x)$, $|\mathcal{W}| \leq |\mathcal{X}| + 1$, *such that* $H(X|W,Y) = 0$.

Note: writing $R_{\mathrm{LSCSI}}(P_{X,Y})$, Theorem 2 implies

$$R_{\mathrm{LSCSI}}(P_{X,Y}) = \sum_z p(z) R_{\mathrm{LSCSI}}(P_{X,Y|Z=z})$$

That is, when G_X has more than one maximally connected component, finding $R_{\mathrm{LSCSI}}(P_{X,Y})$ reduces to computing R_{LSCSI} for each component.

Slepian-Wolf Coding with Causal S.I. (cont.)

Lemma 2. *Let* W, X, Y *be discrete random variables with the Markov relation* $W \to X \to Y$*. For each* x *define* $N_W(x) = \{w : p(w|x) > 0\}$*. Then* $H(X|W, Y) = 0$ *if and only if* $N_W(x) \cap N_W(x') = \emptyset$ *whenever* $(x, x') \in G_X$

Note: When combined with Theorem 2 this implies that R_{LSCSI} depends only on the distribution of X and on G_X (on $P(y|x)$ only through its effect on G_X)

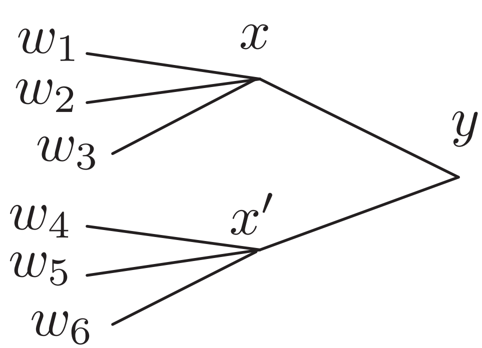


Figure 5: Illustration of condition in Lemma 2

Slepian-Wolf Coding with Causal S.I. (cont.)

Corollary 1. $R_{\mathrm{LSCSI}} = H(X)$ *whenever* G_X *is complete.*

Proof: When G_X is complete, condition for $H(X|W,Y) = 0$ is, by lemma,

$$N_W(x) \cap N_W(x') = \emptyset \quad \forall x \neq x'$$

implying

$$H(X|W) = 0$$

completing proof by an appeal to Theorem 2. □

A note on (dis)continuity

- Corollary 1 gives
$$R_{\mathrm{LSCSI}}(P_{X,Y}) = H(X)$$
in the interior of the simplex of distributions on $\mathcal{X} \times \mathcal{Y}$. This implies a discontinuity at the boundary of the simplex

- While discontinuities of this type are well-known to arise in problems such as zero-error channel coding [Shannon 1956] and the zero-error Slepian-Wolf problem [Witsenhausen 1976], it is interesting to see it arising in our setting, which assumes the standard 'near-lossless' formulation.

Evaluation of R_{LSCSI} for $|\mathcal{X}| = 1, 2, 3, 4$

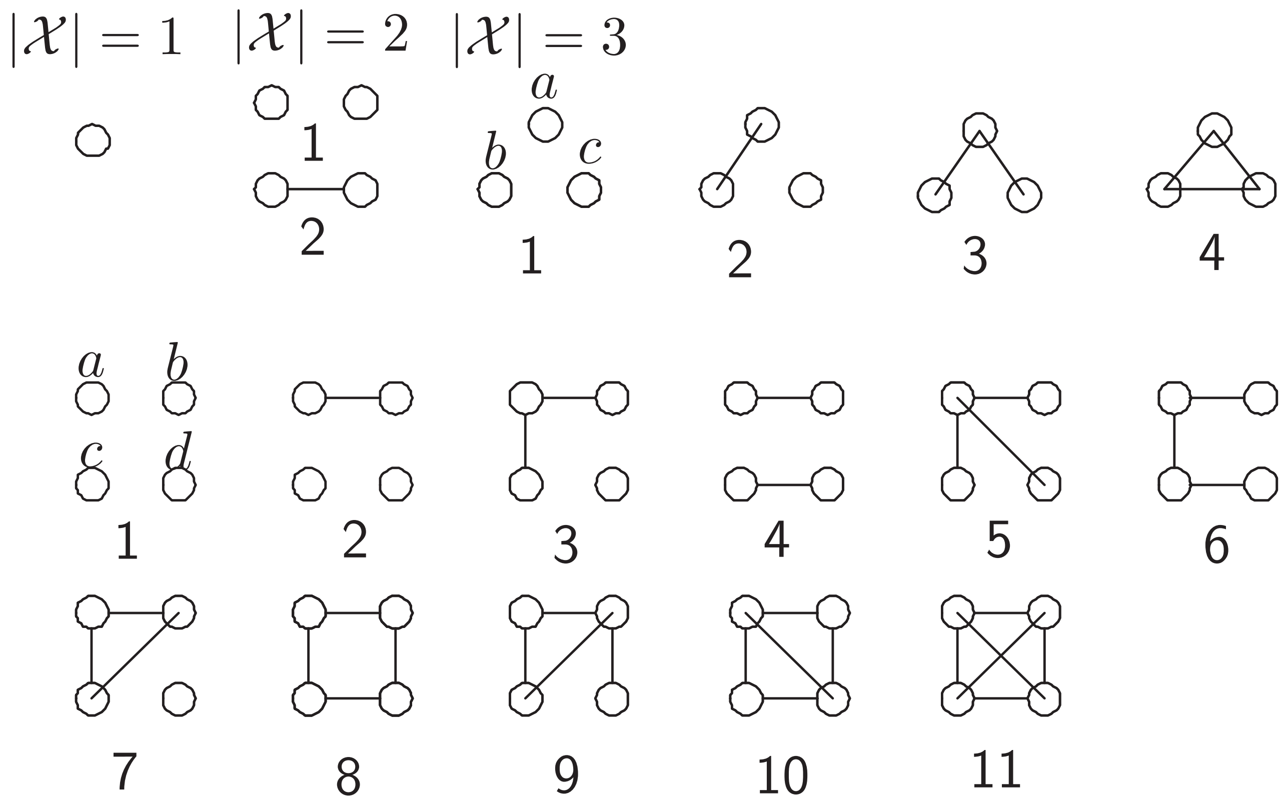


Figure 6: Possible forms of G_X for $|\mathcal{X}| = 1, 2, 3, 4$

$\lvert\mathcal{X}\rvert$	G_X	R_{LSCSI}
1	1	$0 = H(X)$
2	1	0
2	2	$H(X)$
3	1	0
3	2	$[p(a)+p(b)]h\left(\frac{p(a)}{p(a)+p(b)}\right)$
3	3	$h\left(p(a)\right)$
3	4	$H(X)$
4	1	0
4	2	$[p(a)+p(b)]\cdot h\left(\frac{p(a}{p(a)+p(b)}\right)$
4	3	$[1-p(d)]h\left(\frac{p(a)}{1-p(d)}\right)$
4	4	$[p(a)+p(b)]h\left(\frac{p(a)}{p(a)+p(b)}\right)+[p(c)+p(d)]h\left(\frac{p(c)}{p(c)+p(d)}\right)$
4	5	$h\left(p(a)\right)$
4	6	$[1-p(d)]h\left(\frac{p(a)}{1-p(d)}\right) \leq R_{\mathrm{LSCSI}} \leq h\left(p(a)+p(d)\right)$
4	7	$p(a)\log\frac{1-p(d)}{p(a)}+p(b)\log\frac{1-p(d)}{p(b)}+p(c)\log\frac{1-p(d)}{p(c)}$
4	8	$h\left(p(a)+p(d)\right)$
4	9	$I_{\alpha^*}(X;W')$, where $\alpha^* = p(a)/[p(a)+p(c)]$
4	10	$-p(a)\log p(a)-p(d)\log p(d)-[p(b)+p(c)]\log[p(b)+p(c)]$
4	11	$H(X)$

R_{LSCSI} for the Uniform Quaternary Source

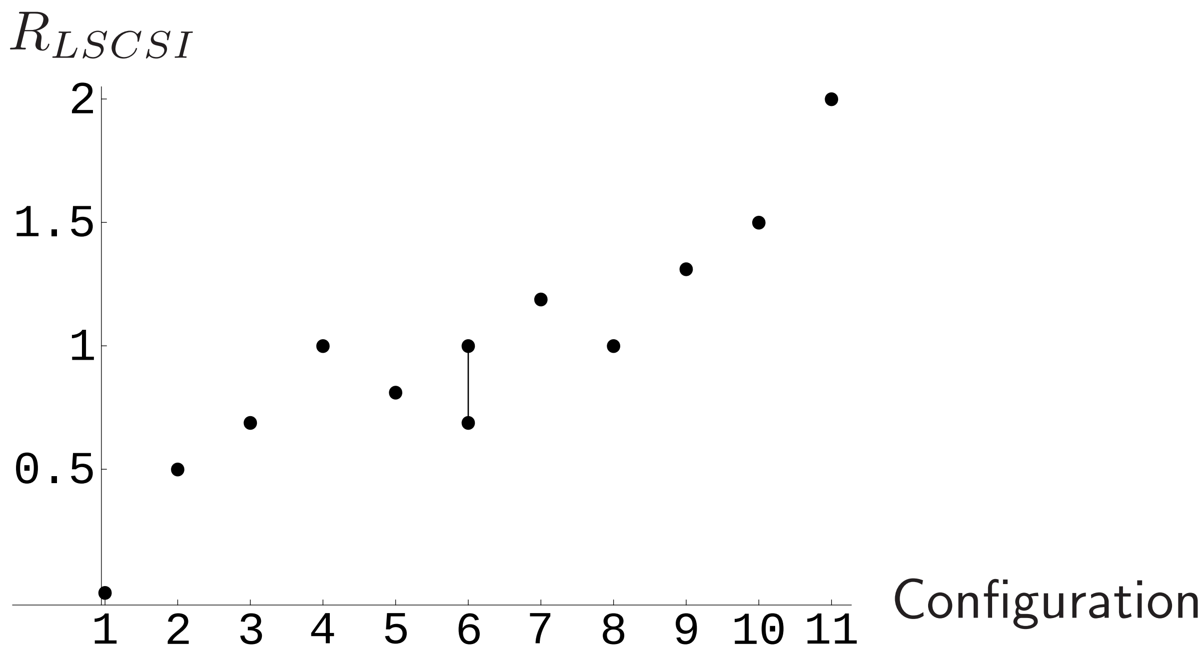


Figure 7: R_{LSCSI} for the uniform quaternary source. The x-axis corresponds to the category that G_X belongs to, as enumerated in Figure 6.

Rate Distortion with Positive S.I. Lookahead

- (X_i, Y_i) i.i.d. drawings of (X, Y)
- Encoding: $T(X^n) \in \{1 \ldots, 2^{nR}\}$
- Reconstruction: $\hat{X}_i(T, Y^{i+d})$, for $d > 0$

Rate Distortion with S.I. Lookahead (cont.)

For integer $k \geq 1$ define

$$R_{k,d}(D) = \frac{1}{k} \min I(X^k; W)$$

where the minimum is over all functions $f_i : \mathcal{W} \times \mathcal{Y}^{i+d} \to \hat{\mathcal{X}}$, $1 \leq i \leq k-d$, $|\mathcal{W}| \leq |\mathcal{X}|^k + 1$, and $P(w|x^k)$ such that

$$\frac{1}{k-d} \sum_{i=1}^{k-d} E\rho(X_i, f_i(W, Y^{i+d})) \leq D$$

Rate Distortion Function for S.I. Lookahead (cont.)

Theorem 3. *The rate distortion function for* d *lookahead,* $R_d(D)$*, is bounded, for any* $k \geq 1$*,* $0 < d < \infty$ *and* $D \geq D_{min}$*, as*

$$R_{k,d}(D) \leq R_d(D) \leq R_{k,d}(D) + \frac{d}{k} H(X)$$

and, consequently,

$$R_d(D) = \lim_{k \to \infty} R_{k,d}(D)$$

Remarks:

- Upper bound can be refined
- *Computability:* Given $\varepsilon > 0$, can obtain $R_d(D)$ to within ε
- In contrast with usual characterizations in source and in channel coding that do not give a computable approximation

Proof of Converse

$$
\begin{aligned}
knR &\geq kH(T) \\
&\geq kI(X^n; T) \\
&= kH(X^n) - k\sum_{i=1}^{n} H(X_i|T, X^{i-1}) \\
&= kH(X^n) - \sum_{j=0}^{k-1} \sum_{i=1-j}^{n-j} H(X_{i+j}|T, X^{i+j-1}) \\
&= kH(X^n) - \sum_{j=0}^{k-1} \sum_{i=1}^{n} H(X_{i+j}|T, X^{i+j-1}) \\
&\quad + \sum_{j=0}^{k-1} \left[\sum_{i=n-j+1}^{n} H(X_{i+j}|T, X^{i+j-1}) - \sum_{i=1-j}^{0} H(X_{i+j}|T, X^{i+j-1}) \right]
\end{aligned}
$$

$$
\begin{aligned}
&\stackrel{(a)}{=} \quad kH(X^n) - \sum_{j=0}^{k-1} \sum_{i=1}^{n} H(X_{i+j}|T, X^{i+j-1}) \\
&\quad + \sum_{j=0}^{k-1} \left[\sum_{i=n-j+1}^{n} H(X_{i+j}) - \sum_{i=1-j}^{0} H(X_{i+j}|T, X^{i+j-1}) \right] \\
&= \quad kH(X^n) - \sum_{j=0}^{k-1} \sum_{i=1}^{n} H(X_{i+j}|T, X^{i+j-1}) \\
&\quad + \sum_{j=0}^{k-1} \sum_{i=1-j}^{0} \left[H(X_{i+j}) - H(X_{i+j}|T, X^{i+j-1}) \right] \\
&= \quad kH(X^n) - \sum_{j=0}^{k-1} \sum_{i=1}^{n} H(X_{i+j}|T, X^{i+j-1}) \\
&\quad + \sum_{j=0}^{k-1} \sum_{i=1-j}^{0} I(X_{i+j}; T, X^{i+j-1})
\end{aligned}
$$

$$
\begin{aligned}
&\geq \sum_{i=1}^{n} H(X_i^{i+k-1}) - \sum_{i=1}^{n}\sum_{j=0}^{k-1} H(X_{i+j}|T, X^{i+j-1}) \\
&= \sum_{i=1}^{n} H(X_i^{i+k-1}) - \sum_{i=1}^{n} H(X_i^{i+k-1}|T, X^{i-1}) \\
&= \sum_{i=1}^{n} I(X_i^{i+k-1}; T, X^{i-1}) \\
&\stackrel{(b)}{\geq} \sum_{i=1}^{n} I(X_i^{i+k-1}; W_i) \\
&\stackrel{(c)}{\geq} \sum_{i=1}^{n} kR_{k,1}\left(\frac{1}{k-1}\sum_{j=0}^{k-2} E\rho(X_{i+j}, \hat{X}_{i+j})\right) \\
&\stackrel{(d)}{\geq} knR_{k,1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{k-1}\sum_{j=0}^{k-2} E\rho(X_{i+j}, \hat{X}_{i+j})\right)
\end{aligned}
$$

$$\stackrel{(e)}{\geq} \quad knR_{k,1}\left(\frac{kD_{max}}{n}+\frac{1}{n}\sum_{i=1}^{n}E\rho(X_i,\hat{X}_i)\right)$$

$$\stackrel{(f)}{\geq} \quad knR_{k,1}\left(D+\frac{kD_{max}}{n}\right)$$

Process Characterization of $R_d(D)$

For jointly stationary processes $\mathbf{X} = \{X_i\}$ and $\mathbf{W} = \{W_i\}$ let $\overline{I}(\mathbf{X}; \mathbf{W})$ denote the mutual information rate defined by

$$\overline{I}(\mathbf{X}; \mathbf{W}) = \lim_{n \to \infty} \frac{1}{n} I(X^n; W^n)$$

Theorem 4.

$$R_d(D) = \inf\{\overline{I}(\mathbf{X}; \mathbf{W}) : E\rho(X_0, \hat{X}_0^{opt}(\mathbf{W}, Y_{-\infty}^d)) \leq D\}$$

where the inf is over jointly stationary $\mathbf{W}, \mathbf{X}, \mathbf{Y}$ *with* $\mathbf{W} - \mathbf{X} - \mathbf{Y}$*, and* $\hat{X}_0^{opt}(\mathbf{W}, Y_{-\infty}^d)$ *is the optimum estimate of* X_0 *based on* $\mathbf{W}, Y_{-\infty}^d$

Gaussian $\mathbf{W}$ is Ineffectual for Gaussian X, Y

Motivated by Theorem 4, consider upper bound to $R_d(D)$

$$R_d^G(D) = \inf\{\bar{I}(\mathbf{X};\mathbf{W}) : E(X_0 - \hat{X}_0^{opt}(\mathbf{W}, Y_{-\infty}^d))^2 \leq D\},$$

where the inf is over jointly stationary and *Gaussian* $\mathbf{W}, \mathbf{X}, \mathbf{Y}$ with the Markov relation $\mathbf{W} - \mathbf{X} - \mathbf{Y}$.

If $R_d^G(D)$ is not convex, can be further improved to its convex envelope $\underline{R}_d^G(D)$

Unfortunately, $\underline{R}_d^G(D)$ turns out to be trivial in the following sense:

Theorem 5. *For every* $d \geq 0$

$$\underline{R}_d^G(D) = \underline{R}_{ub}(D),$$

where $\underline{R}_{ub}(D)$ *is the lower convex envelope of* $R_{ub}(D)$*, the upper bound on* $R(D)$ *of the causal case.*

Lossless Source Coding with Side Information Lookahead

- Encoder maps X^n into $T \in \{1, \ldots, 2^{nR}\}$
- Reconstruction is of form $\hat{X}_i(T, Y^{i+d})$
- R is achievable if exists sequence of schemes with $P(X^n \neq \hat{X}^n) \longrightarrow 0$
- Let R^d_{LSCSI} denote infimum over achievable rates

Lossless Source Coding with S.I. Lookahead (cont.)

We have seen that, under say positivity condition, $R^0_{LSCSI} = H(X)$, in contrast with $H(X|Y)$ which is achievable with no delay constraint.

It is perhaps then natural to expect $\lim_{d\to\infty} R^d_{LSCSI} = H(X|Y)$

This, as it turns out, is *not* the case

In fact, we will see that, not only is $\lim_{d\to\infty} R^d_{LSCSI} > H(X|Y)$ but

$R^d_{LSCSI} = H(X)$ for all $0 \le d < \infty$

That is, the side information is useless

Lossless Source Coding with S.I. Lookahead (cont.)

$$R^{k,d}_{LSCSI} = \frac{1}{k} \min I(X^k; W)$$

where the minimum is over all $P(w|x^k)$, $|\mathcal{W}| \leq |\mathcal{X}|^k + 1$, such that

$$H(X_i|W, Y^{i+d}) = 0 \qquad \text{for all } 1 \leq i \leq k-d.$$

Theorem 6. *For every* k, d

$$R^{k,d}_{LSCSI} \leq R^{d}_{LSCSI} \leq R^{k,d}_{LSCSI} + \frac{d}{k} R_{LSCSI}.$$

So, in particular,

$$R^{d}_{LSCSI} = \lim_{k \to \infty} R^{k,d}_{LSCSI}.$$

Lossless Source Coding with S.I. Lookahead (cont.)

Proposition 2. R^d_{LSCSI} *depends on the distribution of the pair* X, Y *only through the distribution of* X *and the bipartite graph whose edges are the pairs* (x, y) *for which* $P(x, y) > 0$.

Corollary 2. *Let* X, Y *satisfy the positivity condition* $P(x, y) > 0$. *Then, for any* $0 \leq d < \infty$, $R^d_{\mathrm{LSCSI}} = H(X)$.

Block-Length-Dependent Lookahead

Let $R_{LSCSI}^{\{d_n\}}$ denote infimum of achievable rates when $d = d_n$

Evidently, under positivity condition, $R_{LSCSI}^{\{d_n\}} = H(X)$ whenever $d_n \equiv d$

On the other hand, by Slepian-Wolf, $R_{LSCSI}^{\{d_n\}} = H(X|Y)$ when $d_n = n$

As it turns out:

1. $R_{LSCSI}^{\{d_n\}} = H(X|Y)$ provided increase of d_n with n is more than logarithmic

2. any $R > H(X|Y)$ is achievable if $d_n = C(R) \log n$ for appropriate $C(R)$

More concretely:

Block-Length-Dependent Lookahead (cont.)

Define

$$E(R) = \min_{Q_{X,Y}} \left[D(Q_{X,Y} \| P_{X,Y}) + \max\{0, R - H_Q(X|Y)\} \right]$$

This is a "random coding error exponent" for the Slepian-Wolf problem.

Theorem 7.
For every $R > H(X|Y)$, $R_{LSCSI}^{\{d_n\}} \leq R$ *provided* $d_n = \frac{1}{E(R)} \cdot \log n$.
In particular, $R_{LSCSI}^{\{d_n\}} = H(X|Y)$ *if increase of* d_n *is more than logarithmic.*

Consolidation between the Lossless and Lossy Settings

Consider rate distortion functions when ρ is Hamming.

We have seen that

$$R_d(0) = H(X) \quad \forall 0 \leq d < \infty$$

whereas

$$R_{WZ}(0) = H(X|Y)$$

On the other hand, by considering Wyner-Ziv codes for d-blocks, can show

$$\lim_{d \to \infty} R_d(D) = R_{WZ}(D) \quad \forall D > 0$$

This implies a sensitivity to the order of the limits:

$$H(X|Y) = \lim_{D \downarrow 0} \lim_{d \to \infty} R_d(D) < \lim_{d \to \infty} \lim_{D \downarrow 0} R_d(D) = H(X)$$

Consolidation between the Lossless and Lossy Settings (cont.)

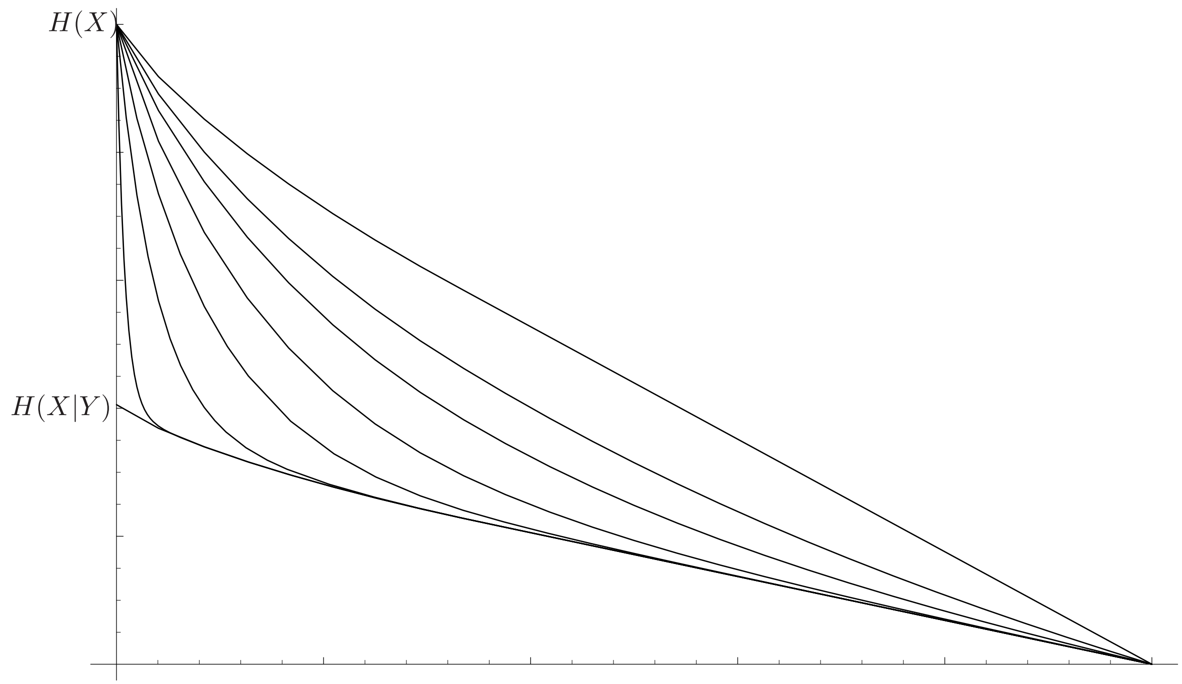


Figure 8: $H(X|Y) = \lim_{D\downarrow 0} \lim_{d\to\infty} R_d(D) < \lim_{d\to\infty} \lim_{D\downarrow 0} R_d(D) = H(X)$.

Channel Coding with Limited-Delay S.I. at the Transmitter

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

Let C_d denote the capacity
Note this bridges between Shannon and Gel'fand-Pinkser channels:

- $X_i(W, S^i)$ is Shannon's channel: $C_0 = \max_{p(u), p(x|u,s)} I(U; Y)$
- $X_i(W, S^n)$ is the Gel'fand-Pinsker: $C_\infty = \max_{p(u|s), p(x|u,s)} [I(U; Y) - I(U; S)]$

Channel Coding with Limited-Delay Transmitter S.I. (cont.)

Theorem 8. *Let*

$$C_{d,k} = \frac{1}{k} \max I(U; Y^k),$$

where the $\max$ *is over all* $P(u|s^d)$, $|\mathcal{U}| \le \min\{|\mathcal{X}|^k, |\mathcal{Y}|^k\} + |\mathcal{S}|^k - 1$, *and* f_i, $1 \le i \le k$, *such that* $X_i = f_i(U, S^{i+d})$. *Then*

$$\frac{k}{k-d} C_{d,k} - \frac{d \log |\mathcal{Y}|}{k-d} \le C_d \le C_{d,k}$$

In particular,

$$C_d = \lim_{k \to \infty} C_{d,k}$$

Note:

- Characterization in spirit of one for the source coding problem
- Lower bound can be refined
- Computability

Open Directions

- Does Gaussian W attain $\min$ for X, Y Gaussian in case $d = 0$? More precisely, is rate distortion function given by convexified version of $R_{ub}(D)$?

- Recall that $\lim_{d\to\infty} R_d(D) = R_{WZ}(D)$ for $D > D_{min}$. What is convergence rate ? Techniques for computation of the redundancy of rate distortion codes would extend to give upper bound $R_d(D) - R_{WZ}(D) = O\left(\frac{1}{d}\log d\right)$. Lower bound may prove more challenging.

- In lossless case, have seen that logarithmic growth of d_n suffices to achieve any point $R \in (H(X|Y), H(X))$. Is this also necessary ? Can we characterize $\alpha(R)$ such that $R_{LSCSI}^{\{d_n\}} = R$ when $d_n = \alpha(R) \cdot \log n$?

- Does feedback/feedforward improve on the fundamental limits for $d > 0$?