EE 231A Information Theory
Lecture 14
Achievability and Converse for $R(D)$

A. Statement of Rate-Distortion Theorem and the Distortion AEP
B. A Pesky Inequality
C. Random Coding Proof of Achievability of $R(D)$
D. Converse to the Channel Coding Theorem

Part 14 A: Statement of Rate-Distortion Theorem and the Distortion AEP.

Theorem 10.2.1

- The rate distortion function for an i.i.d. source X with distribution $p(x)$ and bounded distortion $d(x,\hat{x})$ is

$$R(D) = \min_{p(\hat{x}|x):\sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x})\le D} I(X;\hat{X}).$$

- This is the minimum achievable rate at distortion D.

What we need to prove for achievability

- Suppose that $(X,\hat{X}) \sim p(x,\hat{x})$ such that $E(d) < D$ and $I(X;\hat{X}) < R$. Then distortion D is achievable at rate R.

- i.e. a sequence of $(2^{nR}, n)$ codes have

$$d(\hat{x}^n, x^n) \to D \quad \text{as } n \to \infty.$$

Things we need for proof

- Distortion AEP
- 2^{-nI} inequality
- Another inequality

Distortion Typical Set

- $\mathcal{X}$ is a discrete alphabet.
- $(x,\hat{x}) \sim p(x,\hat{x})$
- $$A_{d,\epsilon}^{(n)} = \Big\{ (x^n,\hat{x}^n) : \left| -\frac{1}{n}\log p(x^n) - H(X) \right| < \epsilon$$
$$\left| -\frac{1}{n}\log p(\hat{x}^n) - H(X) \right| < \epsilon$$
$$\left| -\frac{1}{n}\log p(x^n,\hat{x}^n) - H(X,\hat{X}) \right| < \epsilon$$
$$\left| d(x^n,\hat{x}^n) - E[d(X,\hat{X})] \right| < \epsilon \Big\}$$
- $A_{d,\epsilon}^{(n)} \subset A_{\epsilon}^{(n)}$ where $A_{\epsilon}^{(n)}$ is the jointly typical set.

Distortion AEP

$$P\left(A_{d,\epsilon}^{(n)}\right) \to 1$$

- Weak law of large numbers for $(x_i, \hat{x}_i) \sim \text{i.i.d. } P(x, \hat{x})$.

A 2^{-nI} inequality

- For all $(x_i, \hat{x}_i) \in A_{d,\epsilon}^{(n)}$
  $p(\hat{x}^n) \geq p(\hat{x}^n \mid x^n) 2^{-n(I(X;\hat{X})+3\epsilon)}$

$$\begin{aligned} p(\hat{x}^n \mid x^n) &= \frac{p(\hat{x}^n, x^n)}{p(x^n)} \\ &= p(\hat{x}^n) \frac{p(\hat{x}^n, x^n)}{p(x^n) p(\hat{x}^n)} \\ &= p(\hat{x}^n) 2^{nS} \\ &\leq p(\hat{x}^n) 2^{n(I(X;\hat{X})+3\epsilon)} \end{aligned}$$

$$\begin{aligned} S &= \frac{1}{n} \sum_{i=1}^{n} \log p(x_i, \hat{x}_i) - \log p(x_i) - \log p(\hat{x}_i) \\ &\leq -H(X, \hat{X}) + \epsilon + H(X) + \epsilon + H(\hat{X}) + \epsilon \\ &\qquad \text{for } (x_i, \hat{x}_i) \in A_{d,\epsilon}^{(n)} \\ &= I(X; \hat{X}) + 3\epsilon \end{aligned}$$

Part 14B: A Pesky Inequality

Things we need for proof

- Distortion AEP
- 2^{-nI} inequality
- Another inequality

A pesky inequality

- For $0 \le x, y \le 1, n > 0$

$$(1-xy)^n \le 1 - x + e^{-yn}$$

Takes several steps….

Convexity of $g_y(x)$ $\quad 0 \le x, y \le 1, n > 0$

$$g_y\left(x\right) = \left(1-xy\right)^n$$

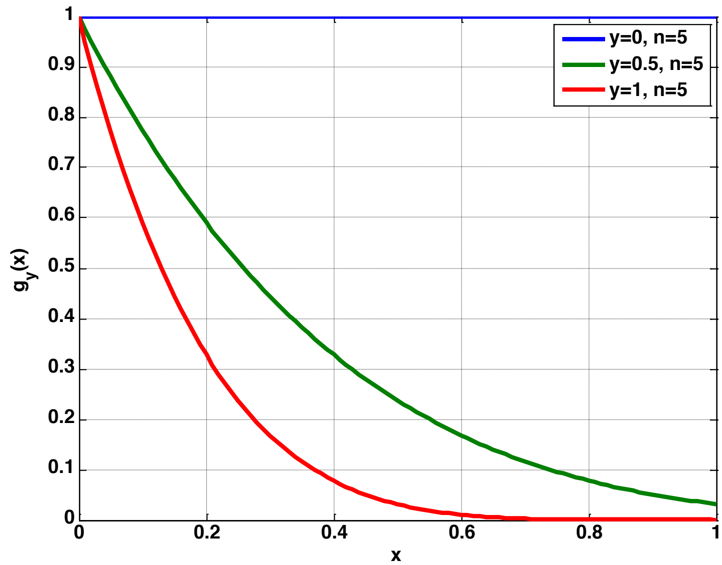
$$\frac{d}{x} g_y\left(x\right) = -yn\left(1-xy\right)^{n-1}$$

$$\frac{d^2}{x^2} g_y\left(x\right) = y^2 n\left(n-1\right)\left(1-xy\right)^{n-2}$$

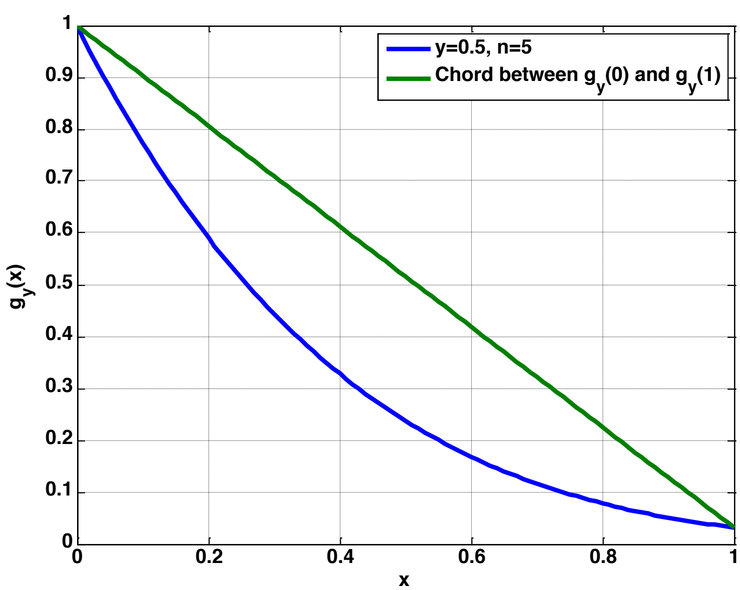
$$\ge 0 \quad \text{for } 0 \le x, y \le 1, \quad n > 0$$

Since the second derivative is positive, $g_y(x)$ is convex.

Plot of $g_y(x)$



Convexity means below any chord.



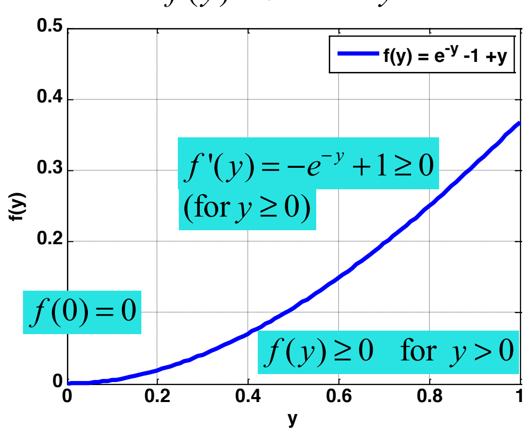
Application of Convexity

$$
\begin{aligned}
(1-xy)^n &= g_y(x) \\
&= g_y\big((1-x)\times 0 + x\times 1\big) \\
&\le (1-x)g_y(0) + x g_y(1) \qquad \text{Lies below the chord…} \\
&= (1-x)1 + x(1-y)^n
\end{aligned}
$$

$$\boxed{(1-xy)^n \le (1-x)1 + x(1-y)^n}$$

$e^{-y} \ge 1-y$

$f(y) = e^{-y} - 1 + y$



$$
\begin{aligned}
f(y) \ge 0 &\Rightarrow e^{-y} - 1 + y \ge 0 \\
&\Rightarrow e^{-y} \ge 1-y \\
&\Rightarrow (e^{-y})^n \ge (1-y)^n
\end{aligned}
$$

$$\boxed{(e^{-y})^n \ge (1-y)^n}$$

Finally, the object of our argument

$$\boxed{\left(1-xy\right)^n \le (1-x)1 + x\left(1-y\right)^n} \qquad \boxed{\left(e^{-y}\right)^n \ge \left(1-y\right)^n}$$

$$\begin{aligned}\left(1-xy\right)^n &\le (1-x)1 + x\left(1-y\right)^n \\ &= (1-x) + x\left(1-y\right)^n \\ &\le (1-x) + xe^{-yn} \\ &\le (1-x) + e^{-yn}\end{aligned}$$

$$\boxed{(1-xy)^n \le 1 - x + e^{-yn}}$$

Part 14C: Random Coding Proof of Achievability of $R(D)$

Proof of achievability

- Let $X_1, X_2, \ldots, X_n \sim$ i.i.d. $p(x)$ and $d(x, \hat{x})$ is a bounded distortion. For any D, and any rate $R > R(D)$, there is a sequence of $(2^{nR}, n)$ codes such that

$$E(d) \to D \quad \text{as } n \to \infty$$

Random Coding

- Generate 2^{nR} reproduction sequences $\hat{X}_i^n,\ i = 1, \ldots, 2^{nR}$ by drawing each $\hat{X}_i^n(k),\ k = 1, \ldots, n,$ $i = 1, \ldots, 2^{nR}$ i.i.d. $\sim p(\hat{x})$

Typical-Set *Encoder*

- For each X^n , select i such that $(X^n, \hat{X}^n(i)) \in A_{d,\epsilon}^{(n)}$ if possible.
- If two are found, choose the least i.
- If none are found, send $i = 1$.

Probability that
this happens is P_e

Reproduction Sequence

- Reproduction sequence is $\hat{X}^n(i)$

Computation of Distortion

- Compute $E[d]$ over the random selection of X^n and of our code.

$$E[d] \le P_e d_{\max} + (1-P_e)(D+\epsilon)$$

$$\boxed{\left| d(x^n,\hat{x}^n) - E[d(X,\hat{X})] \right| < \epsilon}$$

– So we need to compute P_e.

Probability of a "match"

- Fix x^n and select a single $\hat{x}^n$ randomly by choosing n values i.i.d.$\sim p(\hat{x})$.

Indicator Function

$$\Pr\{(x^n,\hat{x}^n) \notin A_{d,\epsilon}^{(n)}\} = 1 - \sum_{\hat{x}^n} p(\hat{x}^n)\mathrm{I}\left((x^n,\hat{x}^n) \in A_{d,\epsilon}^{(n)}\right)$$

- Since we choose 2^{nR} independent codewords, our randomly selected code will fail for a fixed x^n only if we fail in all 2^{nR} attempts, i.e. with probability

$$\left[1 - \sum_{\hat{x}^n} p(\hat{x}^n)\mathrm{I}\left((x^n,\hat{x}^n) \in A_{d,\epsilon}^{(n)}\right)\right]^{2^{nR}}$$

Application of 2^{-nI} inequality

$$(x_i, \hat{x}_i) \in A_{d,\epsilon}^{(n)}$$
$$p(\hat{x}^n) \geq p(\hat{x}^n \mid x^n) 2^{-n(I(X;\hat{X})+3\epsilon)}$$

$$P_e = \sum_{x^n} p(x^n) \left[1 - \sum_{\hat{x}^n} p(\hat{x}^n) \mathrm{I}\left((x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \right) \right]^{2^{nR}}$$
$$\leq \sum_{x^n} p(x^n) \left[1 - \sum_{\hat{x}^n} p(\hat{x}^n \mid x^n) 2^{-n\left(I\left(X;\hat{X}\right)+3\epsilon\right)} \mathrm{I}\left((x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \right) \right]^{2^{nR}}$$

Application of the other inequality

$$\boxed{(1 - xy)^n \leq 1 - x + e^{-yn}}$$

$$P_e \leq \sum_{x^n} p(x^n) \Big[1 - \underbrace{2^{-n\left(I\left(X;\hat{X}\right)+3\epsilon\right)}}_{y} \underbrace{\sum_{\hat{x}^n} p(\hat{x}^n \mid x^n) \mathrm{I}\left((x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \right)}_{x} \Big]^{\overbrace{2^{nR}}^{n}}$$
$$\leq \sum_{x^n} p(x^n) \left[1 - \sum_{\hat{x}^n} p(\hat{x}^n \mid x^n) \mathrm{I}\left((x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \right) + e^{-2^{-n(I+3\epsilon)} 2^{nR}} \right]$$
$$\leq \sum_{x^n} p(x^n) - \sum_{x^n} p(x^n) \sum_{\hat{x}^n} p(\hat{x}^n \mid x^n) \mathrm{I}\left((x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \right) + \sum_{x^n} p(x^n) e^{-2^{-n(I+3\epsilon)} 2^{nR}}$$
$$= 1 - \sum_{x^n} \sum_{\hat{x}^n} p(\hat{x}^n, x^n) \mathrm{I}\left((x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \right) + e^{-2^{n(R - I(X;\hat{X}) - 3\epsilon)}}$$

Probability of error converges to zero.
Distortion converges to D.

$$P_e \le \underbrace{1 - \sum_{x^n} \sum_{\hat{x}^n} p(\hat{x}^n, x^n) \mathrm{I}\left((x^n, \hat{x}^n) \in A_{d,\epsilon}^{(n)} \right)}_{\to 0 \text{ as } n \to \infty} + \underbrace{e^{-2^{n(R - I(X;\hat{X}) - 3\epsilon)}}}_{\to 0 \text{ as } n \to \infty}$$

$$E[d] \le P_e d_{\max} + (1 - P_e)(D + \epsilon)$$

$$\to D$$

Part 14D: Converse to the Rate-Distortion Theorem

Converse to rate-distortion theorem

- For any source X drawn i.i.d. $\sim p(x)$ with bounded additive distortion $d(x, \hat{x})$ and any $(2^{nR}, n)$ lossy compression code, if distortion is $\leq D$, then the rate satisfies $R \geq R(D)$.

This will take several slides to prove…

Assume a lossy compression scheme with good distortion.

- Consider a $(2^{nR}, n)$ lossy compression with encoder f_n, decoder g_n that achieves $E(d) \leq D$.

$$\hat{x}^n = g_n\left(f_n(x^n)\right)$$

A series of bounds on nR

$$
\begin{aligned}
nR &\ge H(\hat{X}^n) && \text{equality if codewords are equally likely} \\
&\ge H(\hat{X}^n) - H(\hat{X}^n \mid X^n) && \text{equality if } \hat{X}^n \text{ is a} \\
&= I(\hat{X}^n; X^n) && \text{deterministic function of } X^n \\
&= H(X^n) - H(X^n \mid \hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i) - H(X^n \mid \hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i \mid \hat{X}^n, X_{i-1}, \ldots, X_1)
\end{aligned}
$$

Continuing the bounds on nR

$$
\begin{aligned}
nR &\ge \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i \mid \hat{X}^n, X_{i-1}, \ldots, X_1) \\
&\ge \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i \mid \hat{X}_i) && \text{equality if only } \hat{X}_i \text{ is useful for reconstructing } X_i \\
&= \sum_{i=1}^{n} I(X_i; \hat{X}_i) \\
&\ge \sum_{i=1}^{n} R(Ed(X_i, \hat{X}_i)) && \text{equality if } R(D) \text{ achieved} \\
&= n \sum_{i=1}^{n} \frac{1}{n} R\left(Ed(X_i, \hat{X}_i) \right)
\end{aligned}
$$

Concluding the bounds on nR

$$nR \geq n\sum_{i=1}^{n}\frac{1}{n}R\left(Ed(X_i,\hat{X}_i)\right)$$

$$\geq nR\left(\frac{1}{n}\sum_{i=1}^{n}Ed(X_i,\hat{X}_i)\right)$$

by convexity of $R(D)$
equality if $Ed(X_i,\hat{X}_i)$
is same for all i

$$= nR\left(Ed(X^n,\hat{X}^n)\right)$$

$$\geq nR(D)$$

since $E(d) \leq D$ and $R(D)$
decreases monotonically

Thus $R > R(D)$ as required.