Multi-use Information Theory 2 — Nov 8th, 2010

Lecture 5

Lecturer: Haim Permuter — *Scribe: Avihay Shirazi*

I. Coordination capacity

A reminder: A rate distortion function for the source $X \sim$ i.i.d. $p_0(x)$, the reconstruction $\hat{X}$ and some distortion measure $d: \ \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$, is given by

$$R(D) = \min_{p(\hat{x}|x): \ \mathbb{E}\left[d(X^n,\hat{X}^n)\right] \le D} I(X;\hat{X}), \tag{1}$$

where

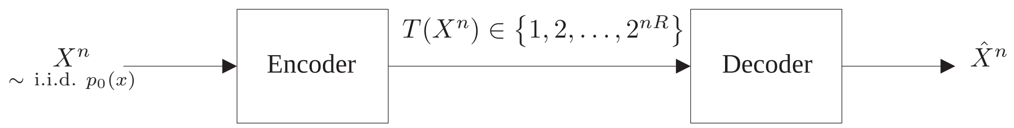

Fig. 1: The basic rate distortion problem.

$$d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i). \tag{2}$$

Our goal in the rate distortion problem is that $\mathbb{E}\left[d(X^n, \hat{X}^n)\right] \le D$.

Now, let us look at another goal; say that we do not want to obey a distortion constrain, but rather, we want to establish coordination between the nodes of the network summarized by a joint probability distribution. For example, if the encoder sends $T(X^n) \in \{1, 2, \dots, 2^{nR}\}$, then we want the decoder to do an action such that some empirical joint PMF between the encoder and the decoder wil be arbitrarily close to the goal PMF. This problem is called a coordination problem, as depicted in Figure 2.

Rate coordination

Let us phrase the problem described in Figure 2 in a mathematical way:
For $X \sim$ i.i.d. $p_0(x)$, the demand for the empirical joint PMF $P_{x^n y^n}(x, y)$ is to be arbitrarily close to the goal PMF $p_0(x)p(y|x)$. i.e., for every $\epsilon > 0$, there exists $N_0(\epsilon) \in \mathbb{N}$ such that $\forall n > N_0(\epsilon)$,

$$\|P_{x^n y^n}(x, y) - p_0(x)p(y|x)\|_{\text{TV}} < \epsilon, \tag{3}$$

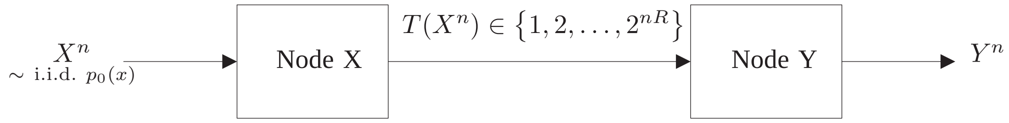

Fig. 2: *Rate coordination problem.* Node X is assigned the action X^n chosen by nature according to it's PMF, and the node Y is producing the sequence Y^n according to the index $T(X^n)$ and some decoding map $g : \{1, 2, \ldots, 2^{nR}\} \to \mathcal{Y}^n$.

where the TV norm is defined in the following definition.

Definition 1 (Total variation norm)**.**

$$\|p(x,y) - q(x,y)\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} |p(x,y) - q(x,y)|. \tag{4}$$

Example 1. Let $\mathcal{X} = \{1,2,3\}$, $\mathcal{Y} = \{1,2,3\}$ and $X \sim \mathrm{Uniform}\{1,2,3\}$. Our goal is

$$Y = \begin{cases} 1 \text{ or } 2 & \text{if } X = 1 \\ 2 \text{ or } 3 & \text{if } X = 2 \\ 3 \text{ or } 1 & \text{if } X = 3 \end{cases} \tag{5}$$

where Y distributes uniformly for each choice of X.

How many bits do we need to send? What is the minimal rate R that insures that we reach our goal with probability 1?

Example 2 (Broadcast)**.** Consider the example depicted in Figure 3. What are the rates (R_1, R_2) that can achieve the joint type (the empirical joint distribution) $P_{x^n y^n z^n}(x,y,z) = p_0(x)p(y,z|x)$, when the actions at node X are specified by nature and distributed according to $p(x^n) = \prod_{i=1}^n p_0(x_i)$?

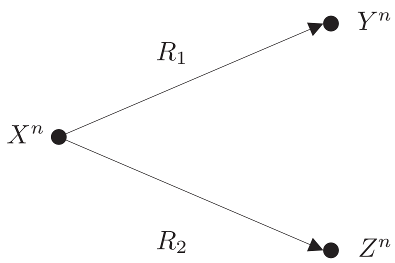

Fig. 3: *Broadcast rate coordination.*

Notice that in the rate distortion problem, every distortion D dictates a set of possible types of sequences, while the coordination problem dictates a specific type (e.g., only $p(y|x)$ for the simple 1 sender and 1 receiver case). Recall that the joint type of (x^n, y^n) is defined as follows

$$P_{x^n y^n}(x,y) = \frac{1}{n} N(x,y|x^n,y^n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i=x, y_i=y\}}. \tag{6}$$

Definition 2 (A coordination code)**.** The coordination code $(2^{nR}, n)$ is defined by the mappings

$$\text{Encoder} \quad f: \quad \mathcal{X}^n \longrightarrow \{1,2,\dots,2^{nR}\}, \tag{7}$$

$$\text{Decoder} \quad g: \quad \{1,2,\dots,2^{nR}\} \longrightarrow \mathcal{Y}^n. \tag{8}$$

Definition 3 (An achievable rate)**.** A rate R is said to be achievable for a coordination $p_0(x)p(y|x)$ if there exists a sequence of coordination codes $(2^{nR}, n)$ such that

$$\|P_{x^n y^n}(x,y) - p_0(x)p(y|x)\|_{\mathrm{TV}} \to 0 \quad \text{in probability.} \tag{9}$$

In other words, for every $\epsilon > 0$ there exists $\delta > 0$ and $N_0 \in \mathbb{N}$ such that

$$\forall n > N_0, \qquad \Pr\left\{\|P_{x^n y^n}(x,y) - p_0(x)p(y|x)\|_{\mathrm{TV}} < \epsilon\right\} \geq 1-\delta. \tag{10}$$

The rate coordination capacity $R^*\big(p_0(x)p(y|x)\big)$ is the infimum of all achievable retes.

Theorem 1 (The rate coordination capacity)**.** For the source $X \sim$ i.i.d. $p_0(x)$ and the desired joint distribution $p(y|x)$ (the desired coordination goal), the rate coordination capacity is given by

$$R^* = I(X;Y). \tag{11}$$

proof of Theorem 1:

Achievability: Look at the rate distortion achievability proof.

Before proving the converse, let us look at the following properties: For X^n and $Q \sim$ Uniform$\{1,2,\dots,n\}$, X_Q is a random variable.

Properties:

1) If all of the elements of the sequence X^n are identically distributed, then X_Q is independent of Q.

   Proof:

$$p(x_1 = a) = p(x_2 = a) = \cdots = p(x_n = a) = p(x_q = a|Q = q) = p(x_Q = a). \tag{12}$$

∎

2) If the elements of the sequence (X^n, Y^n) are identically distributed, then

$$\mathbb{E}\big[P_{X^n Y^n}(a,b)\big] = P_{X_Q Y_Q}(a,b). \tag{13}$$

Proof:

$$\mathbb{E}\big[P_{X^nY^n}(a,b)\big] = \sum_{x^n,y^n} p(x^n,y^n)P_{x^ny^n}(a,b) \tag{14}$$

$$\stackrel{(a)}{=} \frac{1}{n}\sum_{x^n,y^n} p(x^n,y^n)\sum_{q=1}^{n}\mathbb{1}_{\{x_q=a,y_q=b\}} \tag{15}$$

$$= \frac{1}{n}\sum_{q=1}^{n}\underbrace{\sum_{x^n,y^n} p(x^n,y^n)\mathbb{1}_{\{x_q=a,y_q=b\}}}_{P_{X_qY_q}(a,b)\text{ - PMF, not type!}} \tag{16}$$

$$\stackrel{(b)}{=} \frac{1}{n}\sum_{q=1}^{n}P_{X_qY_q}(a,b) \tag{17}$$

$$= \sum_{q=1}^{n}P_Q(q)P_{X_QY_Q|Q}(a,b|q) \tag{18}$$

$$= P_{X_QY_Q}(a,b), \tag{19}$$

where step (a) follows from the definition of the joint empirical distribution $P_{x^ny^n}(a,b) = \frac{1}{n}N(a,b|x^n,y^n) = \frac{1}{n}\sum_{q=1}^{n}\mathbb{1}_{\{x_q=a,y_q=b\}}$, and step (b) follows from $\sum_{x^n,y^n}p(x^n,y^n)\mathbb{1}_{\{x_q=a,y_q=b\}} = \Pr\Big\{\bigcup_{x^n,y^n}\big[(X^n = x^n, Y^n = y^n)\cap(X_q = a, Y_q = b)\big]\Big\} = \Pr\Big\{(X^n,Y^n)\in\mathcal{X}^n\times\mathcal{Y}^n|X_q=a,Y_q=b\Big\} = \Pr\big\{X_q=a,Y_q=b\big\} = P_{X_qY_q}(a,b)$. ■

Notice that

$$\|P_{X^nY^n}(a,b)-p_0(a)p(b|a)\|_{\mathrm{TV}}\to 0 \text{ in probability} \Longrightarrow \mathbb{E}\big[\|P_{X^nY^n}(x,y)-p_0(x)p(y|x)\|_{\mathrm{TV}}\big]\underset{n\to\infty}{\longrightarrow}0. \tag{20}$$

This follows from the fact that if $X_n\to 0$ in probability and X_n is bounded, then $\mathbb{E}\big[X_n\big]\to 0$ as $n\to\infty$.

Proof: If $X_n\to 0$ in probability, then $\forall\epsilon'>0\quad\exists N_0\in\mathbb{N},$ s.t. $n>N_0\Rightarrow\Pr\big\{|X_n|>\epsilon'\big\}<\epsilon'$, hence,

$$\big|\mathbb{E}\big[X_n\big]\big| = \big|\sum_{x\in\mathcal{X}}xP(X_n=x)\big| \tag{21}$$

$$\le \epsilon'\Pr\big\{|X_n|<\epsilon'\big\}+\max_{x\in\mathcal{X}}|x|\cdot\Pr\big\{|X_n|\ge\epsilon'\big\} \tag{22}$$

$$\le \epsilon'+\epsilon'\cdot\max_{x\in\mathcal{X}}|x| \tag{23}$$

$$= \epsilon'\big(1+\max_{x\in\mathcal{X}}|x|\big) \tag{24}$$

$$= \epsilon, \tag{25}$$

for $\epsilon' = \frac{\epsilon}{1+\max_{x\in\mathcal{X}}|x|}$, and since $\max_{x\in\mathcal{X}}|x|<\infty$, then, $\forall\epsilon>0,\quad\exists N_0\in\mathbb{N}$, s.t. $n>N_0\Rightarrow\big|\mathbb{E}\big[X_n\big]\big|<\epsilon$.

■

Example 3 (If X_n is not bounded:)**.** Let us consider the following RV

$$X_n = \begin{cases} 0 & \text{w.p. } 1 - \frac{1}{n} \\ n & \text{w.p. } \frac{1}{n} \end{cases}, \tag{26}$$

then, $X_n \to 0$ in probability, but $\lim_{n\to\infty} \mathbb{E}\big[X_n\big] = 1$!!!

We will now prove the converse for the coordination capacity problem. *Converse*: Fix a $(2^{nR}, n)$ coordination code, Now, consider

$$\begin{aligned}
nR &\geq H(T) && (27)\\
&\overset{(a)}{\geq} I(X^n; Y^n) && (28)\\
&= \sum_{i=1}^{n} I(X_i; Y^n | X^{i-1}) && (29)\\
&\overset{(b)}{=} \sum_{i=1}^{n} I(X_i; Y^n, X^{i-1}) && (30)\\
&\geq \sum_{i=1}^{n} I(X_i; Y_i) && (31)\\
&= nI(X_Q; Y_Q | Q) && (32)\\
&\overset{(c)}{=} nI(X_Q; Y_Q, Q) && (33)\\
&\geq nI(X_Q; Y_Q), && (34)
\end{aligned}$$

and since $(X_Q, Y_Q) \sim p_0(x)p(y|x)$, we get $I(X_Q; Y_Q) = I(X; Y)$, and this completes the proof of the converse. Step (a) follows from the data processing inequality, step (b) follows from the fact that X_i is independent of X^{i-1} and step (c) follows from the fact that X_Q is independent of Q. ■

Let us look at two examples of rate coordination capacity.

Example 4 (Broadcast)**.** Consider a rate-distortion problem as illustrated in Figure 4,

The goal is

$$\mathbb{E}\big[d(X,Y)\big] \leq D_Y, \tag{35}$$

$$\mathbb{E}\big[d(X,Z)\big] \leq D_Z. \tag{36}$$

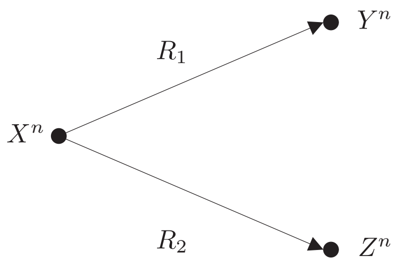

Fig. 4: *Broadcast rate coordination.*

The rate region that corresponds with this problem is

$$\begin{cases} R_1 \geq I(X;Y) \\ R_2 \geq I(X;Z) \end{cases}, \tag{37}$$

over all distributions of the form $p_0(x)p(y|x)p(z|x)$.

Let us now consider a new goal: we want the empirical distribution to be equal to $p_0(x)p(y,z|x)$. i.e., $P_{x^n y^n z^n}(x,y,z) \approx p_0(x)p(y,z|x)$. We get a rate coordination problem, and the inner bound $\mathcal{R}_{p_0,in}$ and the outer bound $\mathcal{R}_{p_0,out}$ [1, Theorem 7] are brought by

$$\mathcal{R}_{p_0,in} = \begin{cases} R_1 \geq I(X;U,Y) \\ R_2 \geq I(X;U,Z) \\ R_1 + R_2 \geq I(X;U,Y) + I(X;U,Z) + I(Y;Z|X,U) \end{cases}, \tag{38}$$

for some joint probability mass function $p_0(x)p(y,z|x)p(u|x,y,z)$, where $|U|$ is finite.

$$\mathcal{R}_{p_0,out} = \begin{cases} R_1 \geq I(X;Y) \\ R_2 \geq I(X;Z) \\ R_1 + R_2 \geq I(X;Y,Z) \end{cases}, \tag{39}$$

for some joint probability mass function $p_0(x)p(y,z|x)$.

Note that the RV U in the inner bound formulas is introduced in order to correlate the codebooks for Y^n and for Z^n. We can consider U to be a common message to both nodes, and each node gets in addition a private message.

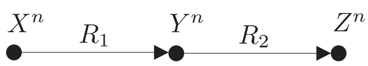

Fig. 5: *Cascade rate coordination.*

Example 5 (Cascade)**.** The rate region for the rate distortion problem illustrated in Figure 5 is

$$\begin{cases} R_1 \geq I(X;Z) + I(X;Y|Z) \\ R_2 \geq I(X;Z) \end{cases}, \tag{40}$$

Solving this problem with the coordination goal $p_0(x)p(y,z|x)$, will yield the result

$$\mathcal{R}_{p_0} = \begin{cases} R_1 \geq I(X;Y,Z) \\ R_2 \geq I(X;Z) \end{cases}, \tag{41}$$

for some joint PMF $p_0(x)p(y,z|x)$.

Achievability: First, let us define the typical set. For a given $\epsilon > 0$ and a given distribution $P_{XYZ}(x,y,z) = p_0(x)p(y,z|x)$, define $T_\epsilon^{(n)}(XYZ)$ in the following way:

$$T_\epsilon^{(n)}(XYZ) = T_\epsilon^{(n)}(P_{XYZ}) = \Big\{(x^n, y^n, z^n) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \;:\; \Big|P_{x^n y^n z^n}(x,y,z) - p_0(x)p(y,z|x)\Big| < \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}||\mathcal{Z}|} \text{for all } (x,y,z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}\Big\}. \tag{42}$$

Let X^n be a drawn i.i.d. $\sim p_0(x)$. Generate at random 2^{nR_2} sequences Z^n distributed i.i.d. according to $p(z)$ and index them $Z^n(j)$ for $j \in \{1, 2, \ldots, 2^{nR_2}\}$. Generate at random $2^{nR_1'}$ sequences Y^n distributed i.i.d. according to $p(y|z)$ and index them $Y^n(i)$ for $i \in \{1, 2, \ldots, 2^{nR_1'}\}$.
Define the mappings

$$f_n : \; \mathcal{X}^n \mapsto \{1, 2, \ldots, 2^{nR_1'}\} \times \{1, 2, \ldots, 2^{nR_2}\} \tag{43}$$

$$g_n : \; \{1, 2, \ldots, 2^{nR_1}\} \mapsto \mathcal{Y}^n \tag{44}$$

$$h_n : \; \{1, 2, \ldots, 2^{nR_2}\} \mapsto \mathcal{Z}^n. \tag{45}$$

Node X: Given x^n, first, look for an index j such that $Z^n(j) \in T_\epsilon^{(n)}(P_{XZ}|x^n)$. After obtaining j, look for an index i such that $Y^n(i) \in T_\epsilon^{(n)}(P_{XYZ}|x^n, z^n(j))$. If such (i,j) are found, send (i,j), to Node Y, otherwise send $(0,0)$.
Node Y Given (i,j), reconstruct $Y^n = g_n(i,j)$ and send j to node Z.
Node Z Given j, reconstruct $Z^n = h_n(j)$.

Analysis of an error:

$$E_1(x^n) = \Big\{\nexists j \in \{1, 2, \ldots, 2^{nR_2}\} \text{ s.t. } \quad Z^n(j) \in T_\epsilon^{(n)}(P_{XZ}|x^n)\Big\} \tag{46}$$

$$E_2(x^n, z^n(j)) = \Big\{\nexists i \in \{1, 2, \ldots, 2^{nR_1'}\} \text{ s.t. } \quad Y^n(i) \in T_\epsilon^{(n)}(P_{XYZ}|x^n, z^n(j))\Big\} \tag{47}$$

The probability for an error is bounded by $P_e^{(n)} \leq \Pr\{E_1\} + \Pr\{E_2\}$.

$$\Pr\{E_1(x^n)\} \leq \prod_j \Pr\left\{Z^n(j) \notin T_\epsilon^{(n)}(P_{XZ}|x^n)\right\} \tag{48}$$

$$\leq \exp\left\{-2^{n\left(R_2 - I(X;Z)\right)}\right\}. \tag{49}$$

The last term tends to 0 as $n \to \infty$ for $R_2 > I(X;Z)$.

As for E_2, because Y^n is generated given $z^n(j)$,

$$\Pr\{E_2(x^n, z^n(j))\} \leq \prod_i \Pr\left\{Y^n(i) \notin T_\epsilon^{(n)}(P_{XYZ}|x^n, z^n(j))\right\} \tag{50}$$

$$\leq \exp\left\{-2^{n\left(R_1' - I(X;Y|Z)\right)}\right\}. \tag{51}$$

The last term tends to 0 as $n \to \infty$ for $R_1' > I(X;Y|Z)$.

Now since we want to sent to Node Y the pair of indices (i,j), we need R_1 to be $R_1 = R_1' + R_2 = I(X;Y|Z) + I(X;Z) = I(X;Y,Z)$. This completes the proof of the achievability. ∎

Converse: For any given coordination code $(2^{R_1}, 2^{nR_2}, n)$,

$$nR_2 \geq I(X^n; Z^n) \tag{52}$$

$$= \sum_{i=1}^{n} I(X_i; Z^n | X^{i-1}) \tag{53}$$

$$\stackrel{(a)}{=} \sum_{i=1}^{n} I(X_i; Z^n, X^{i-1}) \tag{54}$$

$$\geq \sum_{i=1}^{n} I(X_; Z_i) \tag{55}$$

$$= nI(X_Q; Z_Q | Q) \tag{56}$$

$$\stackrel{(b)}{=} nI(X_Q; Z_Q, Q) \tag{57}$$

$$\geq nI(X_Q; Z_Q), \tag{58}$$

and since for all $(X_Q, Z_Q) \sim \sum_y p_0(x)p(y,z|x) = p_0(x)p(z|x)$, we get that $I(X_Q; Z_Q) = I(X;Z)$. Step (a) follows from the fact that X_i is independent of X^{i-1}, and step (b) follows from the fact that X_Q is independent of Q.

And in the same way, we get that

$$nR_1 = H(f_n(X^n)) \tag{59}$$

$$\geq I(X^n; Y^n, Z^n) \tag{60}$$

$$\geq I(X_Q; Y_Q, Z_Q), \tag{61}$$

and again, $(X_Q, Y_Q, Z_Q) \sim p_0(x)p(y, z|x)$, and therefore, $I(X_Q; Y_Q, Z_Q) = I(X; Y, Z)$. This completes the proof of the converse. ■

Note that the rate region for both the rate distortion problem and the rate coordination problem is the same (it is exactly the cut-set bound!).

REFERENCES

[1] Paul Warner Cuff, Haim H. Permuter, and Thomas M. Cover. Coordination capacity. *IEEE Trans. Inf. Theor.*, 56:4181–4206, September 2010.