EE376A Information Theory Lecture 18 - 03/10/2015

Lecture 18

Lecturer: Tsachy Weissman *Scribe: Andrey Sushko, Evan Cheshire*

1 Recap

From previous lectures we've described a lossy compression scheme as follows:

$$\xrightarrow{U_1 ... U_N} \boxed{\text{Encoder}} \xrightarrow{J \in \{1,...,\text{M}\}} \boxed{\text{Decoder}} \xrightarrow{V_1 ... V_N}$$

A scheme is characterized by:

- N, M
- An encoder mapping $\mathcal{U}^N$ to a J $\in \{1, 2, ...M\}$ ($\log M$ bits are used to encode a symbol sequence, where a symbol sequence in U^N and a symbol is U_i)
- A decoder, mapping a $J \in$ {1,2,...M} to $\mathcal{V}^N$

In working with lossy compression, we examine two things:

1. rate $=\frac{\log M}{N} = \frac{\text{bits}}{\text{sourcesymbol}}$
2. expected distortion (figure of merit) $= d\left(U^N, V^N\right) = E\left[\frac{1}{N}\sum_{i=1}^{N} d\left(U_i, V_i\right)\right]$

Definition 1. *(R,D) is considered achievable if* $\exists$ *a scheme* $\forall\ \epsilon > 0$ *such that the rate,* $\frac{logM}{N} \leq R+\epsilon$ *and the expected distortion* $E\left[d\left(U^N, V^N\right)\right] \leq D+\epsilon$

Definition 2. *R(D) = inf{R':(R',D) is achievable}*

Definition 3. $R(D) = \min_{Ed(U,V)\leq D} I(U;V) = R^{(I)}D$

Theorem 4. *R(D) =* $R^{(I)}$*(D)*

$$\longleftrightarrow \left(\begin{array}{c} \text{Direct Part} : R[D] \leq R^I D \\ \text{Converse Part} : R[D] \geq R^I D \end{array} \right.$$

The direct part of this theorem was proven in the previous lecture. The converse will now be given. However, take note that, though these codes exist, the process of looking up a codeword in a table that grows exponentially is highly impractical.

2 Proof of converse

Proof

Fix a scheme satisfying $E\left[d\left(U^N, V^N\right)\right] \leq D$

Then $H(V^N)$ for V^N taking M different values is $\leq \log M$

$\log M \geq H(V^N) \geq H(V^N) - H(V^N|U^N)$
$= I\left(U^N; V^N\right)$
$= H\left(U^N\right) - H\left(U^N|V^N\right)$
$= \sum\limits_{i=1}^{N} H\left(U_i\right) - H\left(U_i|U^{i-1}, V^N\right)$ (by chain rule)
$\geq \sum\limits_{i=1}^{N} H\left(U_i\right) - H\left(U_i|U^{i-1}, V_i\right)$ (conditioning reduces entropy)
$= \sum\limits_{i=1}^{N} I\left(U_i; V_i\right)$
$\geq \sum\limits_{i=1}^{N} R^{(I)}\left(E\left[d\left(U_i, V_i\right)\right]\right)$ $\left(\text{by definition of } R^{(I)}(D)\right)$
$= N\sum\limits_{i=1}^{N} \frac{1}{N} R^{(I)}(E\left[d\left(U_i, V_i\right)\right])$ $\left(\text{average of } R^{(I)}(D) \text{ at all points}\right)$
$\geq NR^I\left(\frac{1}{N}\sum\limits_{i=1}^{N} E\left[d\left(U_i, V_i\right)\right]\right)$ (By convexity)
$\geq NR^{(I)}(D)(R^{(I)}(D)$ is non-increasing)

Rate $= \frac{\log M}{N} \geq R^{(I)}D$

□

3 Geometric Interpretation

I(U;V) the is expected distortion if both in jointly typical set, as we just proved. These figures will give a geometric interpretation to these results.

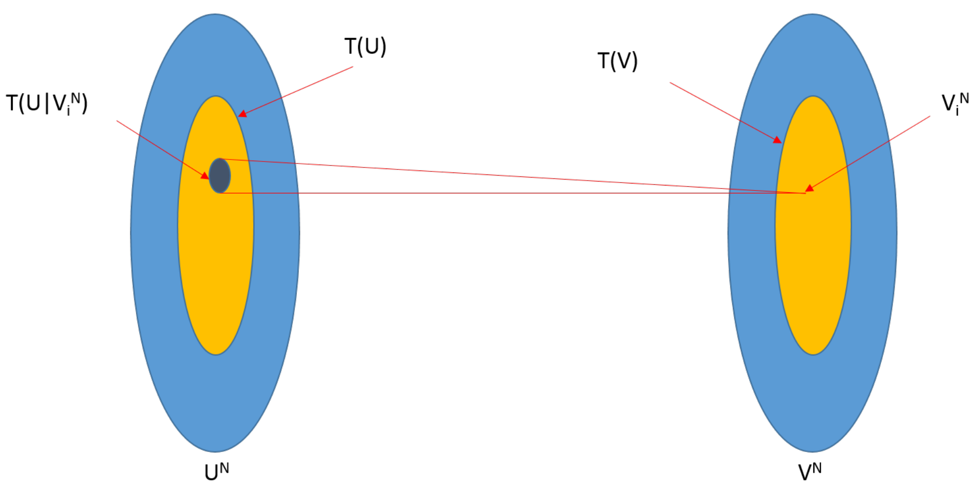

Figure 1: Distortion Function

How big does a codebook have to be so that every source sequence in the typical set has a reconstruction with which it is jointly typical. Let $T\left(U\left|V^N(i)\right.\right)$ be the set of source sequences jointly typical with the reconstruction $V^N(i)$. Therefore, to cover every source sequence, you need a codebook of at least the size of typical set of the input divided by the number of source sequences one reconstruction can cover.

The size of the codebook $=\frac{|T(U)|}{|T(U|V^N(i))|} \approx \frac{2^{\mathrm{NH}(U)}}{2^{\mathrm{NH}(U|V)}} = 2^{\mathrm{NI}(U;V)}$. This is displayed in Figure 1 on the distortion function.

Achievability: generate $V^N(i)$ i.i.d. $\sim V$, $P\left(\left(U^N, V^N(i)\right) \in \mathrm{T}(U,V)\right) \approx 2^{-nI(U;V)}$. Therefore, in order for all typical sequences to be described, $P\left(\left(U^N, V^N(i)\right) \in T(U,V) \text{ for some i}\right) \approx 1$ if codebook is sufficiently large, i.e., if R> I(U;V), as the codebook is of size $\lfloor 2^{NR} \rfloor$

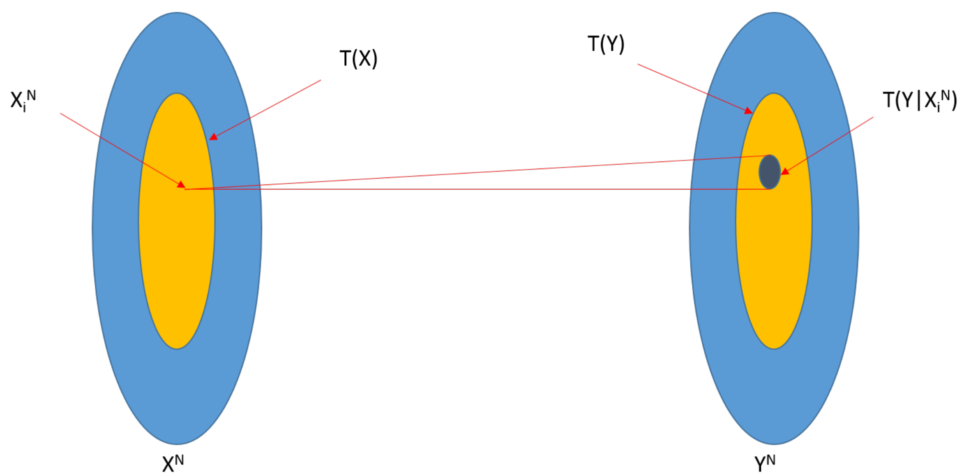

Figure 2: Communication Channel

The communication problem has a similar setup. In order to achieve reliable communication, # of messages $\leq \frac{|T(Y)|}{|T(Y|X^n(i))|} = \frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{-nI(X;Y)}$. This communication channel is diagrammed within figure 2, as the distortion is around the second portion.

Achievability: $\forall i$ s.t. $P\left(Y^n \in T\left(Y^n|X^n(i)\right)|J \neq i\right) \approx 2^{-nI(X;Y)}$. Therefore, because the number of messages was $\lfloor 2^{NR} \rfloor$, in order to guarantee that $P\left(Y^n \in T\left(Y^n|X^n(i)\right) \text{for any i} \neq J\right) \approx 0$, R < I(X;Y).

4 Joint Source Channel Coding

Now that we understand lossy compression as well as the communication problem, we can combine them into a joint source-channel coding theorem. The channel in total is represented by the following: $\xrightarrow[U_i \text{iid} \sim U]{U_1 \ldots U_N}$ $\boxed{\text{Transmitter}} \xrightarrow{X^n} \boxed{\text{Memoryless Channel}} \xrightarrow{Y^n} \boxed{\text{Receiver}} \xrightarrow{V^N=(V_1 \ldots V_N)}$

With this channel description, the goal is to find the best possible distortion given some rate and some noise during transmission

Recall Distortion $= d\left(U^N, V^N\right) = \frac{1}{N}\sum_{i=1}^{N} d\left(U_i, V_i\right)$

Rate $= \frac{N}{n} = \frac{\text{source symbols}}{\text{channel use}}$

Definition 5. *(ρ, D) is achievable if $\forall\epsilon > 0$, $\exists$ scheme with $\frac{N}{n} \geq \rho - \epsilon$ and $E\left[d\left(U^N, V^N\right)\right] \leq D + \epsilon$*

Note: under any scheme, $E\left[d\left(U^N, V^N\right)\right] \leq D$

$U^N \to X^n \to Y^n \to V^N$ is a Markov chain

So:
$nC \geq I\left(X^n; Y^n\right)$ (proven in channel coding converse thm)
$I\left(X^n; Y^n\right) \geq I\left(U^N; V^N\right))$ (Data processing inequality)
$I\left(U^N; V^N\right) \geq NR(D)$ (proven in converse of rate distortion thm)

$\frac{N \cdot R(D)}{n}$ = Rate ·R(D)≤ C

⇒If (ρ,D) achievable ⇒ ρ·R(D)≤ C

Consider the following "Separation Scheme"

$\xrightarrow{U^N}$ $\boxed{\text{Good Distortion Compressor}}$ $\xrightarrow{N \cdot R(D)\text{ bits}}$ $\boxed{\text{Reliable Channel Encoder}}$ $\xrightarrow{X^n}$ $\boxed{\text{Memoryless Channel}}$ $\xrightarrow{Y^n}$

Continued...

$\xrightarrow{Y^N}$ $\boxed{\text{Reliable Channel Decoder}}$ $\xrightarrow{N \cdot R(D)\text{ bits}}$ $\boxed{\text{Good Distortion Compressor}}$ $\xrightarrow{V^N}$

All of these pieces of hardware work correctly to insure that distortion and channel noise are handled properly.

It is guaranteed that $E\left[d\left(U^N, V^N\right)\right] \approx D$ provided that n·C≥N·R(D)
$C \geq \frac{N}{n} \cdot R(D) = \text{rate} \cdot R(D)$
⇒If R(D)≤ C then (ρ, D) is achievable

5 Source Channel Separation Theorem

As we proved in the previous section, (ρ, D) is achievable if and only if ρ·R(D) ≤C. Essentially, we have separated the problem compression from the problem of transmission and have proven that a separated solution is optimal. There is no need nor advantage to address both problems simultaneously.