EE276 Information Theory	Lecture 17 - 03/05/2024

Lecture 17: Joint Source-Channel Coding

Lecturer: Tsachy Weissman

1 Note on random coding for lossy compression

1. Even for small n, random coding gets close to $R(D)$ with high probability. This is in contrast to communication where very large blocklengths are required to achieve reliable communication.

2. Random coding yields practical schemes. That is, in this context, random coding has practical applications in addition to its use in our proofs.

These two remarks will be addressed as part of the next homework assignment.

2 JSCC Problem setting

In this lecture, we study the joint source-channel coding problem (also referred to as joint source-channel communication), which combines the lossy compression problem and the channel communication problem discussed in previous lectures. In particular, we consider the following setting:

$$U^N \rightarrow \boxed{\text{transmitter/encoder}} \overset{X^n}{\rightarrow} \boxed{\text{noisy channel } P_{Y^n|X^n}} \overset{Y^n}{\rightarrow} \boxed{\text{receiver/decoder}} \rightarrow V^N$$

We want to communicate an iid sequence U^N of length N through a memoryless noisy channel $P_{Y|X}$ and the decoder reconstructs U^N as V^N. We are interested in the expected distortion $E\left[d\left(U^N, V^N\right)\right]$, and the rate of communication

$$\text{rate} = \frac{N}{n} = \frac{\#\text{ source symbols}}{\#\text{ channel use}}. \tag{1}$$

Our goal is to understand what is the smallest possible average distortion given a fixed communication rate and a particular channel model. More precisely, we introduce the following definition of achievable rate-distortion pairs.

Definition 1. *A rate-distortion pair (ρ, D) is achievable if $\forall \varepsilon > 0$, there exists a scheme with $\frac{N}{n} \geq \rho - \varepsilon$ and $E\left[d\left(U^N, V^N\right)\right] \leq D + \varepsilon$.*

A scheme in the joint source-channel coding setup is defined as follows.

Definition 2. *A scheme in the JSCC setting is an (encoder$[N \rightarrow n]$, decoder$[n \rightarrow N]$) pair.*

This is a generalization of both communication (where U^N is iid Ber(0.5) and maximum allowed distortion is 0) and lossy compression (where the channel is a noiseless binary symmetric channel). This occurs in a number of practical scenarios such as transmission of audio/video over a noisy medium where some distortion is acceptable, while the rate must be sufficient high for uninterrupted streaming. Also observe that in contrast to all our previous problems, here bits do not occur anywhere in the problem description.

3 The Source-Channel Separation Theorem

3.1 Negative Result

In this section, we find an upper bound on the achievable rate of a scheme, given a distortion level. We begin with introducing the following remarks.
Remark Note that under any scheme:

1. $U^N - X^n - Y^n - V^N$ is a Markov relation.
2. Recall from the proof of the converse in the channel coding theorem (Lecture 11, Section 3):
$$I(X^n; Y^n) \leq n \cdot C. \tag{2}$$
3. Recall from the proof of the converse in the lossy compression theorem (Lecture 17, Section 2), if $E\left[d\left(U^N, V^N\right)\right] \leq D$, then
$$I\left(U^N, V^N\right) \geq N \cdot R(D). \tag{3}$$

Therefore, for any scheme satisfying $E\left[d(U^N, V^N)\right] \leq D$, we have

$$n \cdot C \geq I\left(X^n, Y^n\right) \geq I\left(U^N, V^N\right) \geq N \cdot R(D), \tag{4}$$

where the first inequality follows from remark 2, the second inequality follows from remark 1, and the third inequality follows from remark 3. Then we have

$$\frac{N \cdot R(D)}{n} = \text{rate} \cdot R(D) \leq C. \tag{5}$$

In other words, if (ρ, D) is achievable, then $\rho \cdot R(D) \leq C$.

3.2 Positive Result

Now we turn to study a positive result. Consider the following "separation" scheme:

$$U^N \to \boxed{\begin{array}{l}\text{good compressor}\\ \text{for distortion}\approx D\end{array}} \xrightarrow{N\cdot R(D)\text{ bits}} \boxed{\begin{array}{c}\text{good channel encoder for}\\ \text{communicating reliably } N\cdot R(D)\text{ bits}\end{array}} \xrightarrow{X^n} \boxed{\text{noisy channel}}$$

$$\xrightarrow{Y^n} \boxed{\text{channel decoder}} \xrightarrow{\text{original } N\cdot R(D)\text{ bits}} \boxed{\text{source decoder}} \to V^N$$

By applying this scheme, we will get $E\left[d\left(U^N, V^N\right)\right] \approx D$ provided that $C \geq \frac{N \cdot R(D)}{n} = \text{rate} \cdot R(D)$. Hence, for communication rate ρ and distortion level D, if $\rho \cdot R(D) \leq C$, then (ρ, D) is achievable.

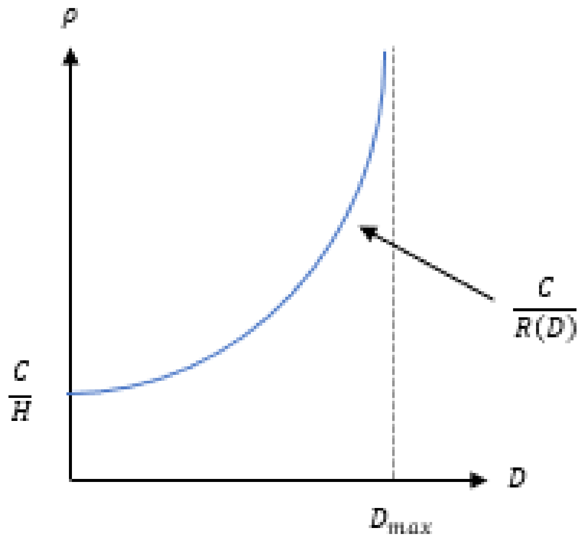
3.3 The Source-Channel Separation Theorem

We combine the positive and negative results to achieve the following theorem.

Theorem 3 (Source-Channel Separation Theorem). *The rate-distortion pair (ρ, D) is achievable iff*

$$\rho \cdot R(D) \leq C.$$

Figure 1: SCST Curve



This theorem is one of the most important takeaways from the class. It implies that the optimal architecture in JSCC is the "separation" scheme defined above, where we compress the source (which is not necessarily represented by bits) into a bit representation, and then we encode those bits for reliable transmission across the channel. This is significant since we have shown that the source and channel coding problems can be considered separately in a JSCC setting without losing performance. Also, even for transmission of analog sources over analog channels, it is optimal to first convert the source into bits.

Figure 1 allows us to visualize which (ρ, D) pairs are achievable. Any point under the curve $\frac{C}{R(D)}$ is achievable, as per the Source-Channel Separation Theorem. In fact, the separation architecture allows us to achieve points arbitrarily close to this curve.

Now we study an example with binary source and binary symmetric channel.

Example 4 (Binary source and binary symmetric channel)**.** Source: $U \sim \text{Ber}(p)$, $0 \leq p \leq 1/2$; Channel: $\text{BSC}(q)$, $0 \leq q \leq 1/2$; Distortion: Hamming.

Recall from the previous lectures that $R(D) = h_2(p) - h_2(D)$ when $0 \leq D \leq p$, and $C = 1 - h_2(q)$. See Figure 2 for the achievable region.

In particular, for $p = 1/2$, consider the following scheme: encoding: $X_i = U_i$, decoding: $V_i = Y_i$. The expected distortion of this scheme is $P(V_i \neq U_i) = P(Y_i \neq X_i) = q$. Also note that under this scheme, $\rho = 1$. This scheme achieves exactly the fundamental limit set forth in the Source-Channel Separation Theorem. In this case, information theory can save from attempting to improve on this simple scheme using some complicated method. Also, this simple scheme is the exception rather than the norm, usually highly non-trivial coding is required to achieve the optimum. See Figure 3 for the achievable region.

Figure 2: SCST Curve for Example 1. Area below the curve is achievable.

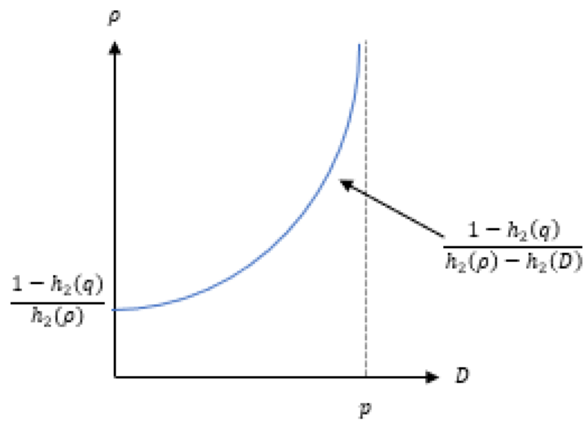


Figure 3: SCST Curve for Example 1, $p = 0.5$. Area below the curve is achievable.