Lecture 18: Joint source-channel coding 2

Lecturer: Tsachy Weissman

In this lecture, we will review the concepts of joint source-channel coding and give an example of Gaussian source and Gaussian channel. We will also discuss an application of information theory to machine learning.

1 Review of Joint Source-Channel Coding (JSCC)

A quick summary of the concepts

1. **The model:**

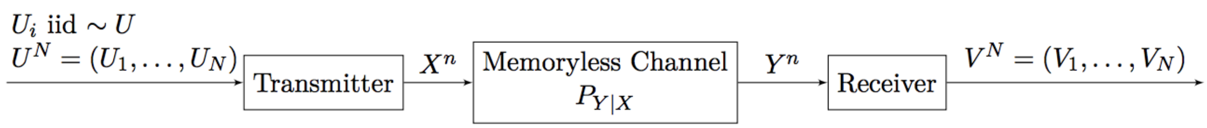

Figure 1: JSCC Problem Schematic

2. **Rate:** rate $= \frac{N}{n} \frac{\text{source symbols}}{\text{channel use}}$
3. **Distortion:** $\mathbb{E}[d(U^N, V^N)]$
4. **Achievability:** (ρ, D) is achievable if $\forall \epsilon > 0$, $\exists$schemes with $\frac{N}{n} \geq \rho - \epsilon$ and $\mathbb{E}[d(U^N, V^N)] \leq D + \epsilon$
5. **"Source-channel Separation" theorem:** (ρ, D) is achievable if and only if $\rho R(D) \leq C$.

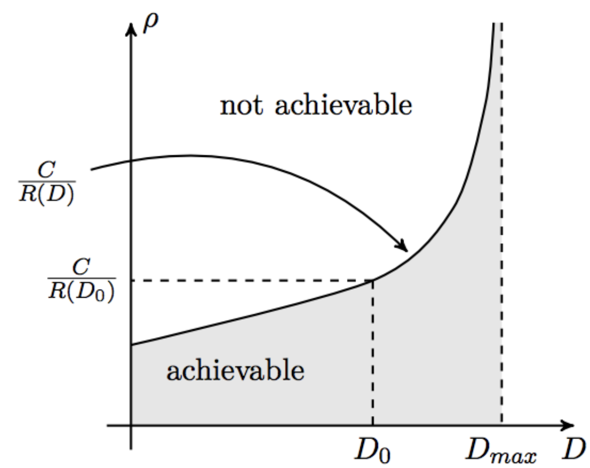

Figure 2: Example Rate Distortion Curve

2 Example: Gaussian source & Gaussian channel

Last class we gave an example of binary source & binary channel, this class we will introduce an example of Gaussian source: $U \sim \mathcal{N}(0, \sigma^2)$ and AWGN channel with transmission power constraint P, the distortion of which is defined as squared error.

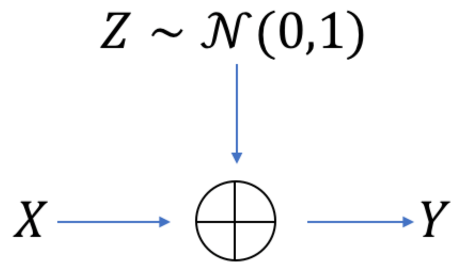

Figure 3: AWGN Channel

Recall: $R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$, $0 < D \leq \sigma^2$. $C = \frac{1}{2}\log(1+P)$.
Then, by JSCC we get

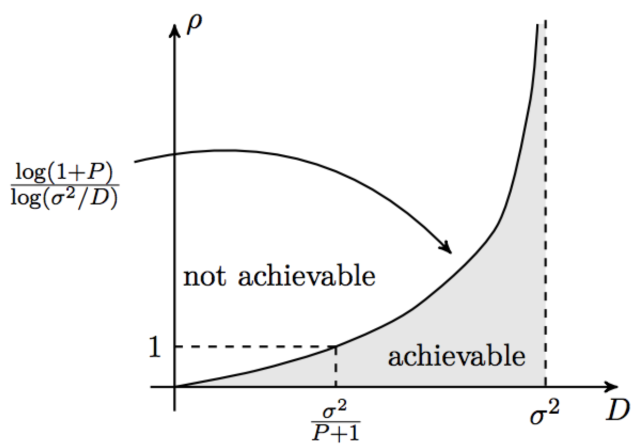

Figure 4: Rate-Distortion Curve for AWGN Channel

Observe that zero distortion is not possible for any positive rate since the source is continuous valued.

Consider the following scheme:
Rate: $\rho = 1$;

Transmit: $X_i = \sqrt{\frac{P}{\sigma^2}}U_i$ (here we rescale U_i because the power of X_i is constrained by P; however, $\text{var}(U) = \sigma^2$. In order to satisfy the power constraint, we rescale U_i to get X_i);

Receive: $Y_i = X_i + Z_i = \sqrt{\frac{P}{\sigma^2}}U_i + Z_i$;

Reconstruction: $V_i = \mathbb{E}[U_i|Y_i] = \frac{\sqrt{P/\sigma^2}\sigma^2}{(\sqrt{P/\sigma^2})^2\sigma^2+1}Y_i$.

Expected distortion achieved: $\mathbb{E}[(U_i - V_i)^2] = \frac{\sigma^2}{(\sqrt{P/\sigma^2})^2\sigma^2+1} = \frac{\sigma^2}{1+P}$

As can be seen, this simple scheme gives the optimum solution. This 'simple scheme' is rather an exception and typically we need non-trivial coding effort when $\rho \neq 1$ in this case and even for $\rho = 1$ with general sources and channels. This is illustrated by the following exercise:

Consider the following "symbol-by-symbol" scheme for $\rho = 1$:
Transmit: $X_i = f(U_i)$;
Reconstruction: $V_i = g(Y_i)$.
Exercise: for a memoryless source & channel, the symbol-by-symbol scheme is optimal **if** $g(\cdot)$ is one-to-one (injective) and $I(U;V)$ achieves $\min\limits_{E[d(U,V)]\leq D} I(U;V)$ and $I(X;Y)$ achieves $\max\limits_{P(X\mid Y)} I(X;Y)$ under the joint distribution of (U,X,Y,V) when $X = f(U)$ and $V = g(Y)$.
Proof Sketch:
Here we have a Markov Chain: $U - X - Y - Z$
Then, by the properties of Markov Chain $I(U;Y) \leq I(X;Y)$, but since $X = f(U)$, we also have by data-processing inequality that $I(X;Y) = I(f(U);Y) \leq I(U;Y)$ and hence we have $I(X;Y) = I(U;Y)$. Now, by using that $g(\cdot)$ is a one-to-one function, we also have $I(U;V) = I(U;g(Y)) = I(U;Y)$. Thus, we have shown that given conditions imply that $I(U;V) = I(X;Y)$. But, based on the conditions in the optimization problem given above, and by our formulations of Channel Coding Theorem and Rate Distortion, we can identify $I(X;Y) = C$ and $I(U;V) = R(D)$. Thus, for these set of conditions, we know by JSSC that there exist a scheme with $\rho = \frac{C}{R(D)} = \frac{I(X;Y)}{I(U;V)} = 1$ which is optimal. Thus, these conditions are sufficient for an optimal "symbol-by-symbol" scheme.

You can apply this exercise and see for yourself that the these conditions exist for both (Binary Source, Binary Channel) example as well as (Gaussian Source, Gaussian Channel) example. For instance, in the second case, we saw above that the "simple scheme" is optimal. For this simple scheme, $X = f(U) = \sqrt{\frac{P}{\sigma^2}}U$ and $X \sim \mathcal{N}(0,P)$, but remember that we have already shown that for Gaussian channel C is achieved when $X \sim \mathcal{N}(0,P)$, that is, $X = f(U)$ in our scheme indeed maximizes $I(X;Y)$. Similarly, all other relations can be established in the two examples.