Electrical Engineering 229A Lecture 24 Notes

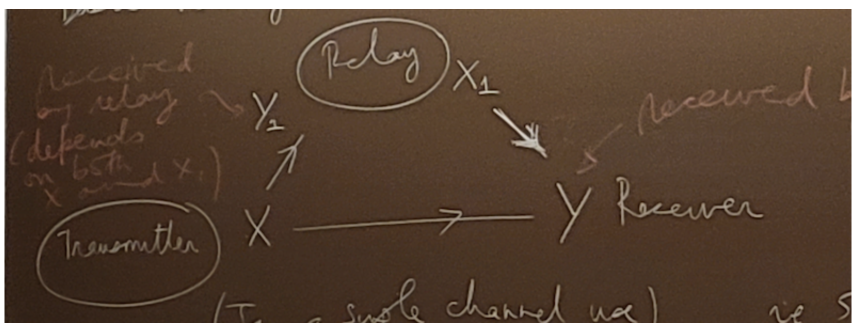
Daniel Raban

November 18, 2021

1 The Relay Channel Model, One Shot Information Theory, and Rate Distortion Theory

1.1 The relay channel model

The basic relay channel model (in the discrete memoryless case) has a transmitter, a relay, and a receiver. In a single channel use, the transmitter inputs X. The relay receives Y_1 (which depends on X and X_1) and sends X_1, and the receiver receives Y, which depends both on X and X_1.



The channel is described by $p(y, y_1 \mid x, x_1)$ with $y \in \mathcal{Y}$, $y_1 \in \mathcal{Y}_1$, $x \in \mathscr{X}$, and $x_1 \in \mathscr{X}$.

We use our Shannon persona to study the Shannon capacity asymptotically as block length goes to ∞. The new twist is that in deciding the k-th input with $1 \leq k \leq n$, the relay can use the past $k-1$ observations. The overall probability distribution is

$$p(m)p(x_{[1:n]} \mid m) \prod_{i=1}^{n} p(x_{1,i} \mid \underbrace{y_{1,1}, \dots, y_{1,i-1}}_{y_{1,[1:i-1]}}) \prod_{i=1}^{n} p(y_i, y_{1,i} \mid x_i x_{1,i})$$

in either a deterministic or random coding scheme (for proof purposes), where $m \in [M_n] = [2^{nR}]$.

In a fixed coding scheme,

$$p(x_{[1:n]} \mid m) = \mathbb{1}_{\{e_n(m)=x_{[1:n]}\}},$$

where $e_n : [M_n] \to \mathscr{X}^n$ is an encoding map, and

$$p(x_{1,i} \mid y_{1,[1:i-1]}) = \mathbb{1}_{\{f_i(y_{1,[1:i-1]})=x_{1,i}\}},$$

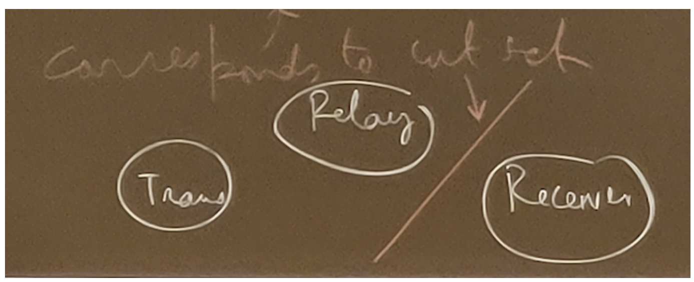
where $f_1, \ldots, f_n$ are the relay's encoding rules. We also need the decoding map $d_n : \mathscr{Y}^n \to [M_n]$.

The Shannon capacity region, defined as usual as the supremum of rates at which the error probability (asymptotically in n) goes to zero, is unknown. Here is a basic theorem in this area.

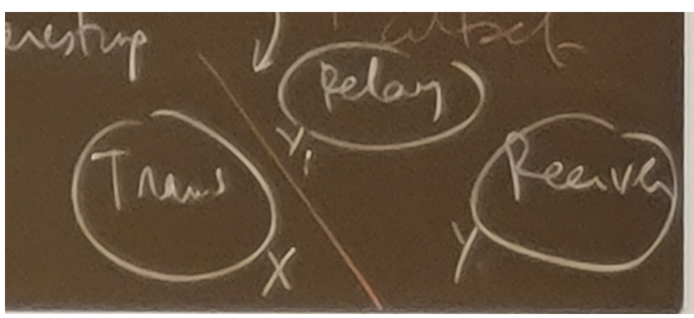
Theorem 1.1 (Cut-set bound)**.**

$$C \leq \sup_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y, Y_1 \mid X_1)\}.$$

The bound in terms of $I(X, X_1; Y)$ should be thought of as the case where the transmitter and the relay can communicate freely; here is the following cut-set in a picture:



The bound in terms of $I(X; Y, Y_1 \mid X_1)$ should be thought of as the case where the relay and the receiver can communicate freely. This cut-set looks like



The first term is more straightforward, while the second is more interesting. We'll omit the details of the proof and include them in a handout later.

1.2 One shot information theory

One shot information theory (in the discrete memoryless case) involves a single use of a DMC. A message $m \in \{1, \dots, L\}$ is encoded as $x(m) \in \mathscr{X}$, received as y via $[p(y \mid x)]$ through the channel, and decoded as $d(y) = \widehat{m}$. We want to study $p_e := \mathbb{P}(\widehat{W} \neq W)$ as a function of L, where $W \sim \mathrm{Unif}(\{1, \dots, L\})$.

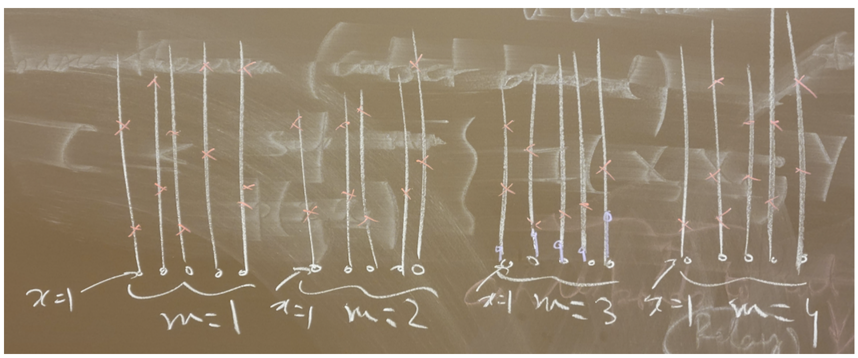
Theorem 1.2 (Poisson matching lemma)**.** *For any input distribution distribution p_X,*

$$p_e \leq \mathbb{E}\left[1 - \frac{1}{1 + L2^{-i_{X;Y}(X;Y)}}\right],$$

where $i_{X,Y}(x; y) := \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}$.

This uses a Poisson process on $\mathscr{X} \times [L] \times \mathbb{R}_+$. You may think of this as one copy of $\mathbb{R}_+$ for each $x \in X, 1 \leq m \leq 1$ and an independent rate 1 Poisson process on each line. That is, the points are placed with iid Exp(1) interarrival times.

Example 1.1. Here is what this looks like when $|\mathscr{X}| = 5$ and $L = 4$.

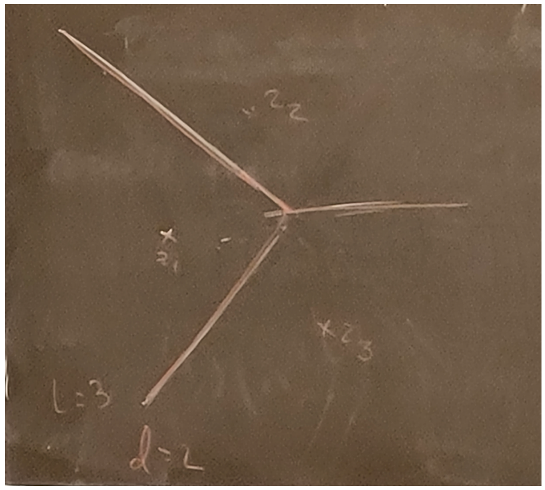


This Poisson structure is shared randomness between transmitter and receiver. The transmitter knows m and scales up the profile $(p_x(x), x \in \mathscr{X})$ in the m-th group until it hits a point in the Poisson process. If that is in line x, input x into the channel. The receiver scales up the distribution $\frac{1}{L} p_{X|Y}(x \mid Y)$ (which is computed from p_X and $p_{Y|X}$ using Bayes' rule) until it heats a point of the Poisson process. Then the receiver returns $\widehat{m}$, which is the block of lines in which the hit occurs.

1.3 Rate distortion theory

Rate distortion theory is a "Shannon mindset theory" which tries to do an asymptotic version of vector quantization. Here is the basic vector quantization problem. Say $Z \in \mathbb{R}^d$ is random, and you are allowed to place L points $z_1, \dots, z_: \in \mathbb{R}^d$. The aim is to minimize $\mathbb{E}[\min_{1 \leq \ell \leq L}(Z - z_\ell)^2]$.

Given $z_1, \ldots, z_L$, $\mathbb{R}^d$ gets decomposed into **Voronoi cells**, which are the points closest to a given z_ℓ.



But given a region D, the best choice of $z \in \mathbb{R}^d$ to map that region to will be the one that minimizes $\mathbb{E}[(Z-z)^2 \mathbb{1}_{\{Z \in D\}}]$.

In the rate distortion formulation, the block length is n, the alphabet is $\mathscr{X}$, and the finite source sequence $x_{[1:n]} \in \mathscr{X}^n$ can be represented by 2^{nR} points via $f_n : \mathscr{X}^n \to \{1, \ldots, 2^{nR}\}$. The decompressor sees $f_n(x_{[1:n]})$ and reproduces it as $\widehat{x}_{[1:n]} \in \widehat{\mathscr{X}}^n$ via $g_n : [2^{nR}] \to \widehat{\mathscr{X}}^n$. The aim is to minimize

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} d(X_i, \widehat{X}_i)\right].$$

Here, $X_1, \ldots, X_n$ are iid, $d : \mathscr{X} \times \widehat{\mathscr{X}} \to \mathbb{R}$ is some distortion measure, and $\widehat{X}_{[1:n]} = g_n(f_n(X_{[1:n]}))$.