ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 26 Information Theory and Coding
Notes on Channel Capacity with Constraints December 11, 2007

Capacity with Constraints

In this note we prove the achievability part of the channel coding theorem for memoryless channels under input constraints.

Recall the setting: we are given a channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, described by the conditional probabilities $P(y|x)$. We are also given a *cost function* $\rho : \mathcal{X} \to [0, \infty)$, $\rho(x)$ is the cost of input letter x.

A block code with M messages and block length n is a mapping from a set of M messages $\{1, \dots, M\}$ to channel input sequences of length n. Thus, a block code is specified when we specify the M channel input sequences $\mathbf{c}_1 = (c_{1,1}, \dots, c_{1,n}), \dots, \mathbf{c}_M = (c_{M,1}, \dots, c_{M,n})$ the messages are mapped into. We will call $\mathbf{c}_m$ the codeword for message m.

To send message m with such a block code we simply give the sequence $\mathbf{c}_m$ to the channel as input.

The cost of codeword $\mathbf{c}_m = (c_{m,1}, \dots, c_{m,n})$ is defined to be $\rho(\mathbf{c}_m) = \frac{1}{n}\sum_{i=1}^n \rho(c_{m,1})$. The code is said to obey a cost constraint P if each codeword has cost less than or equal to P.

A decoder for such a block code is a mapping from channel output sequences $\mathcal{Y}^n$ to the set of M messages $\{1, \dots, M\}$. For a given decoder, let $D_m \subset \mathcal{Y}^n$ denote the set of channel outputs which are mapped to message m. Since an output sequence $\mathbf{y}$ is mapped to exactly one message, D_m's form a collection of disjoint sets whose union is $\mathcal{Y}^n$.

We define the rate of a block code with M messages and block length n as

$$\frac{\ln M}{n},$$

and given such a code and a decoder we define

$$P_{e,m} = \sum_{\mathbf{y} \notin D_m} P(\mathbf{y}|\mathbf{c}_m),$$

the probability of a decoding error when message m is sent. Further define

$$P_{e,\text{ave}} = \frac{1}{M}\sum_{m=1}^{M} P_{e,m} \quad \text{and} \quad P_{e,\max} = \max_{1 \le m \le M} P_{e,m}$$

as the average and maximal (both over the possible messages) error probability of such a code and decoder.

Given a channel and a cost function, we say that a rate R can be achieved under cost constraint P if for every $\delta > 0$ there is a block code with rate at least R, each codeword having cost at most $P + \delta$ and $P_{e,m} < \delta$ for every m. The capacity of a channel under a cost constraint P is the supremum of achievable rates.

Theorem 1. *The capacity of a channel under cost constraint P is given by*

$$C = \max I(X;Y)$$

where the maximum is taken over all input distributions p_X that satisfy $E[\rho(X)] \le P$.

In the class we proved that the capacity is at most C. In this note we will show that for any distribution p_X on the input alphabet of the channel for which $E[\rho(X)] \le P$, all rates up to $I(X;Y)$ are achievable. This then says that capacity is at least C, proving the theorem.

To this end, suppose we are given a p_X for which $E[\rho(X)] \le P$ and a rate $R < I(X;Y)$. Our task is to find, for each $\delta > 0$, a code with rate at least R, maximal error probability at most δ and whose codewords obey cost constraint $P + \delta$. Suppose then $\delta > 0$ is given, and consider constructing a block code of block length n and $M = 2 \times 2^{nR}$ codewords by randomly choosing each letter of each codeword independently, according to the distribution p_X. (Note that we are choosing twice the number of codewords needed, the reason will become clear later when we will eliminate half the chosen codewords.) Being defined as a result of a random experiment, such a code is a random variable, with codewords $\mathbf{C}_1, \dots, \mathbf{C}_M$. The probability that a particular codeword with codewords $\mathbf{c}_1 = (c_{1,1}, \dots, c_{1,n}, \dots, \mathbf{c}_M = (c_{M,1}, \dots, c_{M,n})$ is constructed is

$$\prod_{m=1}^{M} \prod_{i=1}^{n} p_X(c_{m,n}).$$

Consequently, the quantities $P_{e,m}$ and also the costs of each codewords are all random variables.

Let A_ϵ be the set of jointly typical sequences of length n with respect to the distribution $p_{XY}(x, y) = p_X(x)P(y|x)$, that is, the set

$$\Big\{(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \\ \Big|-\frac{1}{n} \log p_X(\mathbf{x}) - H(X)\Big| < \epsilon \\ \Big|-\frac{1}{n} \log p_Y(\mathbf{y}) - H(Y)\Big| < \epsilon \\ \Big|-\frac{1}{n} \log p_{XY}(\mathbf{x}, \mathbf{y}) - H(X, Y)\Big| < \epsilon\Big\}.$$

Since $R < I(X;Y)$ one can find $\epsilon > 0$ such that $R < I(X;Y) - 3\epsilon$. Fix such an ϵ and consider a decoder that operates as follows: Given a $\mathbf{y} \in \mathcal{Y}^n$, if there is exactly one m for which $(\mathbf{C}_m, \mathbf{y}) \in A_\epsilon$, and if this $\mathbf{C}_m$ has cost less than $P + \epsilon$, the decoder declares m as its decision. Otherwise the decoder is free in its decision (but we will assume that an error is made).

Let us now upper bound the expected probability of error, $E[P_{e,m}]$ (the expectation is over the random choice of the code). By symmetry, it is sufficient to consider $E[P_{e,1}]$, i.e., to assume that the transmitted codeword is $\mathbf{c}_1$. The $E[P_{e,1}]$ equals,

$$\sum_{\mathbf{c}_1} \cdots \sum_{\mathbf{c}_M} \sum_{\mathbf{y}} p_X(\mathbf{c}_1) \dots p_X(\mathbf{c}_M) P(\mathbf{y}|\mathbf{c}_1) 1\{\rho(\mathbf{c}_1) \ge P + \delta \text{ or (not } E_1) \text{ or } E_2 \text{ or} \dots \text{or } E_M\}$$

where E_m stands for "$(\mathbf{c}_m, \mathbf{y}) \in A_\epsilon$". Upper bounding

$$1\{\rho(\mathbf{c}_1) \ge P + \delta \text{ or (not } E_1) \text{ or } E_2 \text{ or} \dots \text{or } E_M\} \\ \le 1\{\rho(\mathbf{c}_1) \ge P + \delta\} + 1\{\text{not } E_1\} + \sum_{m=2}^{M} 1\{E_m\}$$

we see that the $E[P_{e,1}]$ is upper bounded by

$$\Pr\{(\mathbf{C}_1, \mathbf{Y}) \notin A_\epsilon\} + \Pr\{\rho(\mathbf{C}_1) \geq P + \delta\} + \sum_{m=2}^{M} \Pr\{(\mathbf{C}_m, \mathbf{Y}) \in A_\epsilon\}.$$

Observe now that the first two terms have probabilities that approach zero as n tends to infinity by the law of large numbers. The sum consists of $M-1$ terms each of which is upper bounded by $2^{-n[I(X;Y)-3\epsilon]}$ (from the properties of jointly typical sets). Since ϵ was chosen so that $R < I(X;Y) - 3\epsilon$, the sum also approaches zero as n tends to infinitity. Thus, we can find an n such that

$$E[P_{e,m}] \leq \delta/2$$

for each $m = 1, \ldots, M$. This means that

$$E\Big[\sum_{m=1}^{M} P_{e,m}\Big] \leq (M/2)\delta$$

and thus there must exist a particular code with codewords $\mathbf{c}_1, \ldots, \mathbf{c}_M$ such that

$$\sum_{m=1}^{M} P_{e,m} \leq (M/2)\delta.$$

Observe now that in this sum, there can't be more than $M/2$ terms whose value exceeds δ (otherwise the sum could not be upper bounded by $(M/2)\delta$. Thus if we throw away from our code the codewords $\mathbf{c}_m$ for which $P_{e,m} \geq \delta$ we will throw away at most $M/2$ codewords and be left with at least $M/2 = 2^{nR}$ codewords for each of which $P_{e,m} < \delta$.

Also note that if a codeword $\mathbf{c}_m$ had $\rho(\mathbf{c}_m) \geq P + \delta$ then $P_{e,m}$ would have equaled 1: the decoder, by construction never decodes such an m. Thus all the codewords that remain not only have error probility less than δ but also satisfy the cost constraint. Thus we see that we have constructed a code with all the required properties and all rates up to the C in the theorem are achievable.