Electrical Engineering 229A Lecture 19 Notes

Daniel Raban

October 28, 2021

1 Capacity of an Additive White Gaussian Noise Channel

1.1 Shannon capacity of a additive white Gaussian noise channel

In the additive white Gaussian noise (AWGN) model, we send inputs real-valued X_i and receive real-valued outputs Y_i, where $Y_i = X_i + Z_i$, and $Z_1, Z_2, \dots \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. At block length n, we have an encoding map $e_n : [M_n] \to \mathbb{R}^n$ and a decoding map $d_n : \mathbb{R}^n \to [M_n]$. We assume an **input power constraint**, which, in Shannon's formulation, says that each codeword is required to have power at most P: If $X^n(m)$ denotes $e_n(m)$, then

$$\frac{1}{n}\sum_{i=1}^n X_i^n(m) \leq P.$$

We want to find for which rates R we have

$$\liminf_n \frac{1}{n}\log M_n \geq R \quad \text{with} \quad \mathbb{P}(d_n(e_n(W_n)) \neq W_n) \to 0$$

for some sequence $((e_n, d_n), n \geq 1)$ with $W_n \sim \mathrm{Unif}([M_n])$.

Theorem 1.1. *The supremum over rates for which communication is possible is*

$$\sup_{X \sim f : \int_{-\infty}^{\infty} x^2 f(x)\, dx \leq P} I(X; X+Z),$$

which equals $\frac{1}{2}\log(1 + \frac{P}{\sigma^2})$ *and is achieved by* $X \sim N(0, P)$.

This quantity is called the **Shannon capacity**. The achievability part of the proof will use a random coding argument and requires the concept of ε-weakly typical sequences. The converse part of the proof involves Fano's inequality. Let's first see why the last claim is true:

Lemma 1.1. *If* $\mathbb{E}[X^2] \leq P$*, then* $I(X; X+Z) \leq \frac{1}{2}\log(1+\frac{P}{\sigma^2})$*, with equality if and only if* $X \sim N(0, P)$.

Proof.

$$\begin{aligned} I(X; X+Z) &= h(X+Z) - h(X+Z \mid X) \\ &= h(X+Z) - h(Z) \\ &= h(X+Z) - \frac{1}{2}\log(2\pi e\sigma^2). \end{aligned}$$

Since $X \amalg Z$ and $\mathbb{E}[Z_1] = 0$, we also have

$$\begin{aligned} \mathbb{E}[(X+Z)^2] &= \mathbb{E}[X^2] + \mathbb{E}[Z^2] \\ &\leq P + \sigma^2. \end{aligned}$$

So

$$h(X+Z) \leq \frac{1}{2}\log(2\pi e(P+\sigma^2))$$

with equality iff $X \sim N(0, P)$. So

$$\begin{aligned} I(X, X+Z) &\leq \frac{1}{2}\log\left(\frac{P+\sigma^2}{\sigma^2}\right) \\ &= \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right). \end{aligned}$$

□

1.2 Weak-typicality for differential entropy

Definition 1.1. For $X \sim f$ with differential entropy $h(X)$ and $\varepsilon > 0$, the set of ε**-weakly typical sequences** for the density f sis

$$A_\varepsilon^n := \left\{ x^n \in \mathbb{R}^n : \left| -\frac{1}{n}\log\prod_{i=1}^n f(x_i) - h(X) \right| < \varepsilon \right\} \subseteq \mathbb{R}^n$$

By the weak law of large numbers,

$$\mathbb{P}(X^n \in A_\varepsilon^n) = 1$$

if $X_i \overset{\text{iid}}{\sim} f$. This is because $\mathbb{E}[\log \frac{1}{f(X)}] = h(X)$ when $X \sim f$.

Proposition 1.1. *For all n,*

$$\mathrm{Vol}(A_\varepsilon^n) \leq 2^{nh(X)} 2^{n\varepsilon}.$$

Proof.

$$\begin{aligned} 1 &\geq \int_{A_\varepsilon^n} \prod_{i=1}^n f(x_i) dx^n \\ &\geq \int_{A_\varepsilon^n} 2^{-nh(X)} 2^{-n\varepsilon}\, dx^n \\ &= \mathrm{Vol}(A_\varepsilon^n) 2^{-nh(X)} 2^{-n\varepsilon}. \end{aligned}$$

□

Proposition 1.2. *Given $\delta > 0$, for all sufficiently large n,*

$$\mathrm{Vol}(A_\varepsilon^n) \geq (1-\delta) 2^{nh(X)} 2^{-n\varepsilon}.$$

Proof. For sufficiently large n,

$$\begin{aligned}
(1-\delta) &\leq \int_{A_\varepsilon^n} \prod_{i=1}^n f(x_i)\, dx^n \\
&\leq \int_{A_\varepsilon^n} 2^{-nh(X)} 2^{n\varepsilon}\, dx^n \\
&= \mathrm{Vol}(A_\varepsilon^n) 2^{-nh(X)} 2^{n\varepsilon}.
\end{aligned}$$

□

Definition 1.2. Let $(X_1, Y_1), (X_2, Y_2), \dots$ be iid with $(X_i, Y_i) \sim f(x,y)$. The set of **ε-jointly weakly typical sequences** for f is

$$\begin{aligned}
A_\varepsilon^n := \Bigg\{ (x^n, y^n) : &\left| -\frac{1}{n} \log \prod_{i=1}^n f(x_i) - h(X) \right| \leq \varepsilon, \\
&\left| -\frac{1}{n} \log \prod_{i=1}^n f(y_i) - h(Y) \right| \leq \varepsilon, \\
&\left| -\frac{1}{n} \log \prod_{i=1}^n f(x_i, y_i) - h(X,Y) \right| \leq \varepsilon, \Bigg\}.
\end{aligned}$$

With this definition in mind, we can show the following.

Lemma 1.2. *If $\widetilde{X}^n \overset{d}{=} X^n$, $\widetilde{Y}^n \overset{d}{=} Y^n$, and $\widetilde{X}^n \amalg \widetilde{Y}^n$, then*

$$(1-\delta) 2^{-nI(X;Y)} 2^{-3n\varepsilon} \leq \mathbb{P}((\widetilde{X}^n, \widetilde{Y}^n) \in A_\varepsilon^n) \leq 2^{-nI(X;Y)} 2^{3n\varepsilon}.$$

The upper bound holds for all n, and the lower bound holds for all sufficiently large n.

1.3 Proof of Shannon's channel coding theorem for an AWGN channel

Now we can prove the theorem.

Proof. Achievability: Generate a random codebook

$$\begin{bmatrix} X_1(1) & \cdots & X_n(1) \\ X_1(2) & \cdots & X_n(2) \\ \vdots & & \vdots \\ X_1(M_n) & \cdots & X_n(M_n) \end{bmatrix},$$

where each $X_n(i) \sim \mathcal{N}(0, P-\eta)$ is iid over i and n. Let $W_n \sim \text{Unif}([M_n])$. The decoding rule is

$$d_n(Y^n) = \begin{cases} m & (X^n(m), Y^n) \text{ are } \varepsilon\text{-jointly weakly typical and for all } m' \neq m, \\ & \quad (X^n(m), Y^n) \text{ are not } \varepsilon\text{-jointly weakly typical} \\ \text{arbitrary} & \text{either no or } \geq 2\ X^n(m) \text{ are } \varepsilon\text{-jointly typical with } Y^n. \end{cases}$$

By symmetry,

$$\begin{aligned} \mathbb{P}(d_n(e_n(W_n)) \neq W_n) &= \mathbb{P}(d_n(e_n(1) \neq 1)) \\ &\leq P(E_{0,n}) + \sum_{m \neq 2} P(E_{m,n}), \end{aligned}$$

where $E_{0,n}$ is the event that $(X^n(1), Y^n)$ is not ε-jointly weakly typical and $E_{m,n}$ for $m \geq 2$ is the event that $(X^n(1), Y^n)$ is ε-jointly weakly typical. Then $\mathbb{P}(E_{0,n}) \to 0$ as $n \to \infty$, and for each $2 \leq m \leq M_n$, $\mathbb{P}(E_{m,n}) \leq 2^{-nI(X;Y)} 2^{3n\varepsilon}$. So if $M_n = 2^{nR}$ with $R < I(X;Y) - 3\varepsilon$, then $\mathbb{P}(d_n(e_n(W_n)) \neq W_n) \to 0$ as $n \to \infty$.

Converse: Consider any $((e_n, d_n), n \geq 1)$. We have $W_n \sim \text{Unif}([M_n])$ and the Markov chain $W_n - X^n - Y^n - \widehat{W}_n$ with $X^n = e_n(W_n)$, $Y = X + Z$, and $\widehat{W}_n = d_n(Y^n)$. Suppose $M_n = \lceil 2^{nR} \rceil$. The data-processing inequality gives

$$H(W_n \mid Y^n) \leq H(W_n \mid \widehat{W}_n).$$

Note that W_n is a discrete random variable, and Y^n is a continuous random variable. Here, we mean $H(W_n \mid Y^n) = \int_{-\infty}^{\infty} H(W_n \mid Y_n = y) f(y)\, dy$. If $p_e(n) := \mathbb{P}(\widehat{W}_n \neq W_n)$, then Fano's inequality gives

$$H(W_n \mid Y^n) \leq 1 + nRp_e(n).$$

Also, the data processing inequality gives

$$\begin{aligned} H(W_n) &= I(W_n; Y^n) + H(W_n \mid Y^n) \\ &\leq I(X^n; Y^n) + H(W_n \mid Y^n) \\ &= h(Y^n) - \sum_{i=1}^{n} h(Y_i \mid X^n, Y^{i-1}) + H(W_n \mid Y^n) \end{aligned}$$

Use $0 \leq D(f(y^n) \mid\mid \prod_{i=1}^n f(y_i)) = \int_{\mathbb{R}^n} f(y^n) \log \frac{f(y^n)}{\prod_{i=1}^n f(y_i)}\, dy^n = -h(Y^n) + \sum_{i=1}^n h(Y_i)$.

$$\leq \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Y_i \mid X^n, Y^{i-1}) + H(W_n \mid Y^n)$$

Use the Markov chain $Y_i - X_i - (X^{i-1}, X_{i+1}^n, Y^{i-1})$

$$\leq \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Y_i \mid X_i) + H(W_n \mid Y^n)$$

$$= \sum_{i=1}^{n} I(X_i; Y_i) + H(W_n \mid Y^n)$$

Let $P_i := \mathbb{E}[X_i^2]$, and recall that $Y_i = X_i + Z_i$, where $Z \sim \mathcal{N}(0, \sigma^2)$ and $X_i \amalg Z_i$.

$$\begin{aligned}
&\le \sum_{i=1}^{n} \frac{1}{2} \log\left(1 + \frac{P_i}{\sigma^2}\right) + H(W_n \mid Y^n) \\
&\le \frac{n}{2} \log\left(1 + \frac{P}{\sigma^2}\right) + H(W_n \mid Y^n) \\
&\le \frac{n}{2} \log\left(1 + \frac{P}{\sigma^2}\right) + 1 + (\log M_n) p_e(n).
\end{aligned}$$

Since $\frac{1}{n} \log M_n \to R$ if $p_e(n) \to 0$, this gives

$$\limsup_n \frac{1}{n} \log M_n \le \frac{1}{2} \log\left(1 + \frac{P}{\sigma^2}\right). \qquad \square$$

Why is this result interesting? Suppose the FCC assigns you a bandwidth of W Hertz, and you communicate over this channel for some time T at power constraint P (with units energy per unit time). One can show that if the noise that corrupts your waveform is additive white noise with power spectral density $\frac{N_0}{2}$, then the theoretical limit of the rate at which you can communicate is

$$W \log(1 + \frac{P}{N_0 W}) \quad \text{bits/unit time.}$$

Studying the $W \to \infty$ limit and the $T \to \infty$ limit is interesting.