Lecture 9: AWGN channel and the Joint AEP

Lecturer: Tsachy Weissman Scribe: Tadafumi Ikezu, Alex Kern, Jack O'Reilly, Daniel Wright

1 Midterm Announcement

The midterm will cover through lossless compression, and will be open notes and open book. Electronics are permitted only for viewing PDFs.

2 Communication Setting

- Encoder which takes as input m bits and outputs X^n.
- Memoryless channell defined by $P_{Y|X}$ which takes as input X^n and outputs Y^n.
- Decoder which takes as input Y^n and outputs m bits.

Note that we can think of the m bits as equivalent to a message, J, uniformly distributed on $\{1, 2, \ldots, M\}$ where $M = 2^m$. The decoder, in this interpretation, outputs a message $\hat{J}(Y^n)$.

- A **scheme** is the (encoder, decoder) pair.
- An **encoder** can be thought of as a codebook $c_n = \{X^n(1), X^n(2), \ldots X^n(M)\}$ where $X^n(i)$ is the encoding of the i-th message.
- The **decoder** can be thought of as the mapping $\hat{J}(\cdot)$, which maps Y^n to a message in $\{1, \ldots, M\}$.
- The **rate** $= \frac{\log M}{n} = \frac{\log |c_n|}{n} = \frac{m}{n} \frac{\text{bits}}{\text{channel use}}$.
- $P_e = P(\hat{J} \neq J)$.

Sometimes we also have a transmission constraint:

$$\frac{1}{n}\sum_{i=1}^{n} \Lambda(X_i) \leq \alpha$$

E.g., $\Lambda(x) = x^2$ in wireless communication corresponds to a constraint on the average power of the transmission.

$$C = \text{ maximal rate of reliable communication}$$

$$C^{(I)} = \begin{cases} \max\limits_{P_x} I(X;Y) & \text{without a transmission constraint} \\ \max\limits_{P_x : \mathrm{E}\,\Lambda(X) \leq \alpha} I(X;Y) & \text{with a transmission constraint} \end{cases}$$

Theorem 1 (Channel Coding Theorem)**.** $C = C^{(I)}$

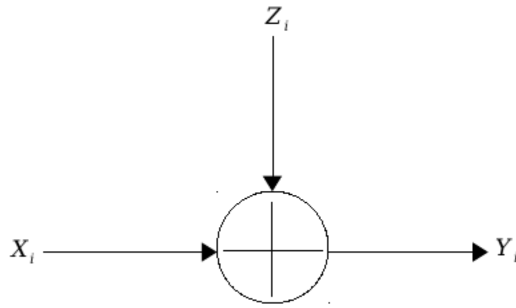

Figure 1: Additive White Gaussian Noise channel. Assume independence of X_i and Z_i.

3 Example III: AWGN channel

Recall that if $G \sim \mathcal{N}(0, \sigma^2)$, then

1. the differential entropy $h(G) = \frac{1}{2} \log 2\pi e \sigma^2$,
2. $h(X) \leq h(G)$, if X is any random variable with $E[X^2] \leq \sigma^2$.

The previous lecture we found that under transmission power constraint, the channel capacity of the AWGN channel is

$$C(P) = \max_{P_X : E[X^2] \leq P} I(X;Y).$$

We define an upper bound by the below calculation,

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(Y - X|X) \\ &= h(Y) - h(W|X) \\ &= h(Y) - h(W), \text{ by independence of W and X} \\ &\leq h(\mathcal{N}(0, P + \sigma^2)) - h(\mathcal{N}(0, \sigma^2)) \\ &= \frac{1}{2} \log 2\pi e(P + \sigma^2) - \frac{1}{2} \log 2\pi e \sigma^2 \\ &= \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2}\right), \end{aligned}$$

where the inequality follows from $Var(Y) = Var(X) + Var(W) \leq P + \sigma^2$ and the second equality holds because translation by a constant does not change the differential entropy, and X is a constant conditioned on X. Equality is achieved with $Y \sim \mathcal{N}(0, P + \sigma^2)$, which we find with $X \sim \mathcal{N}(0, p)$.

We now have our expression for the capacity of the AWGN channel:

$$C(p) = \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2}\right). \tag{1}$$

All that is left to prove is that any rate above the capacity is not achievable. Below is a geometric argument for why we might expect this to be the case. **Note:** $\frac{P}{\sigma^2}$ is commonly called the signal to noise ratio, or SNR.

3.1 A rough geometric interpretation

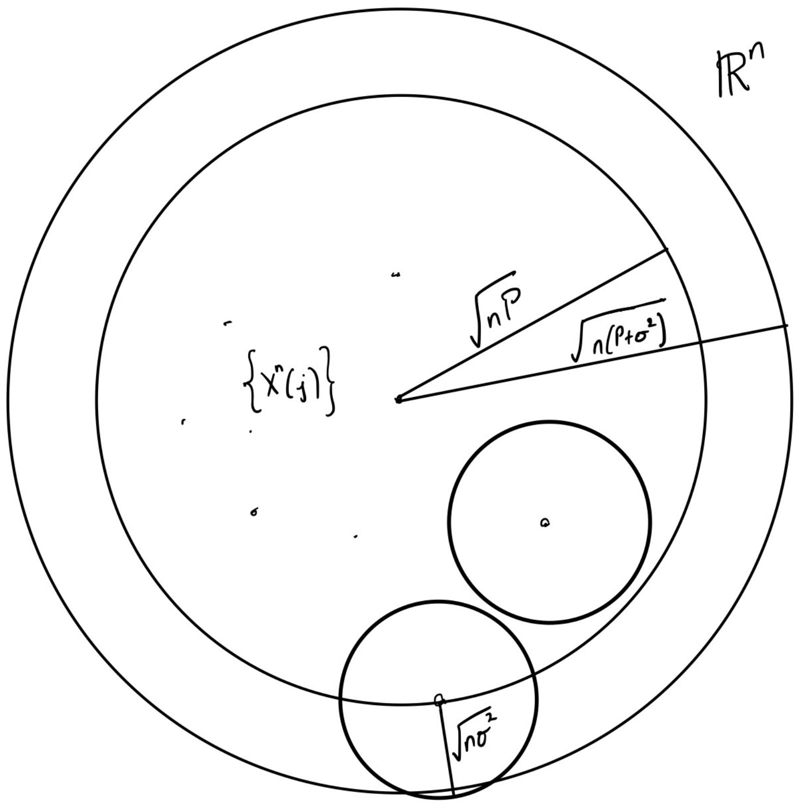

Figure 2: The space of message points lie within a sphere in $\mathbb{R}^n$, with radius set by the transmission power constraint.

Recall the power constraint over multiple uses of the channel:

$$\frac{1}{n}\sum_{i=1}^{n} X^2 \leq P \tag{2}$$

Or, equivalently:

$$\sqrt{\sum_{i=1}^{n} X^2} \leq \sqrt{nP} \tag{3}$$

If we represent the elements of our codebook as points in $\mathbb{R}^n$, we can interpret the LHS of (3) as the 2-norm of said points. We can think of the power constraint as being equivalent to constraining the points in the codebook to the sphere of radius $\sqrt{nP}$ centered at the origin.

Next, we consider the geometric interpretation of added noise:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} W_i^2 &\longrightarrow E[W_i^2] \text{ with high probability by L.L.N.}\\ &= \sigma^2\end{aligned}$$

It follows that for large n:

$$\sqrt{\sum_{i=1}^{n} W_i^2} \approx \sqrt{n\sigma^2} \tag{4}$$

The RHS of (4) can similarly can be interpreted as noting that addition of noise to the transmitted signal applies uncertainty in the form of a sphere of radius $\sqrt{n\sigma^2}$. Consider the expected channel output:

$$\begin{aligned} E[\sum_{i=1}^{n} Y_i^2] &= E[\sum_{i=1}^{n} X_i^2] + E[\sum_{i=1}^{n} W_i^2] \\ &\lesssim nP + n\sigma^2 \end{aligned} \tag{5}$$

Geometrically, we interpret (5) as saying that on average the channel outputs in such a power constrained scheme will lie within a sphere of radius $\sqrt{nP + n\sigma^2}$. In order to achieve reliable communication, we want the "noise balls" (given by spheres of radius $\sqrt{n\sigma^2}$ around each of the n-vectors $X^n(i)$) to be non-intersecting. This leads us to a constraint on the number of possible messages we can communicate, or the size of the codebook:

$$\begin{aligned} \text{number of messages} &\leq \frac{\text{volume of ball of radius } \sqrt{n(P+\sigma^2)}}{\text{volume of ball of radius } \sqrt{n\sigma^2}} \\ &= \frac{K_n(\sqrt{n(P+\sigma^2)})^n}{K_n(\sqrt{n\sigma^2})^n} \\ &= \left(\frac{P+\sigma^2}{\sigma^2}\right)^{\frac{n}{2}} \\ &= \left(1 + \frac{P}{\sigma^2}\right)^{\frac{n}{2}} \\ \Rightarrow rate = \frac{\log(\text{number of messages})}{n} &\leq \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right) \end{aligned} \tag{6}$$

Definition 2. *We refer to the n-vector $X^n(i)$ as the **message point** corresponding to entry i $(1 \leq i \leq M)$ in our codebook of size M.*

The Channel Coding Theorem tells us that the channel capacity is equal to the maxiumum achievable rate given in (6). Intuitively, we achieve this rate by optimally packing the noise balls of radius $\sqrt{n\sigma^2}$ into the sphere of radius $\sqrt{n(P+\sigma^2)}$ in a non-overlapping manner. The centers of the noise spheres are used as the **message points** $X^n(i)$ in our codebook. Note that although there are packing inefficiencies for low dimensions (i.e., for small n), these inefficiencies are overcome as n increases.

4 Joint AEP

We have the following setting:

$$
\begin{aligned}
X, Y &\text{ random variables on alphabets } \mathcal{X}, \mathcal{Y} \\
(X, Y) &\sim P_{X,Y} \\
X &\sim P_X \\
Y &\sim P_Y \\
(X_i, Y_i) &\text{ iid } \sim (X, Y) \\
p(x^n) &= \prod_{i=1}^{n} P_X(x_i) \\
p(y^n) &= \prod_{i=1}^{n} P_Y(y_i) \\
p(x^n, y^n) &= \prod_{i=1}^{n} P_{X,Y}(x_i, y_i)
\end{aligned}
$$

Definition 3. *The set of jointly ϵ-typical sequences is:*

$$
\begin{aligned}
A_\epsilon^{(n)}(X,Y) = \Big\{ (x^n, y^n) : &\left| -\frac{1}{n} \log p(x^n) - H(X) \right| \le \epsilon, \\
&\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| \le \epsilon, \\
&\left| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \right| \le \epsilon, \Big\}
\end{aligned}
$$

Theorem 4. *Joint AEP.*
Part A. *If (X^n, Y^n) formed by iid (X_i, Y_i):*

1. $P\Big((X^n, Y^n) \in A_\epsilon^{(n)}(X,Y)\Big) \xrightarrow{n \to \infty} 1$
2. $(1-\epsilon)2^{n(H(X,Y)-\epsilon)} \le \left| A_\epsilon^{(n)}(X,Y) \right| \le 2^{n(H(X,Y)+\epsilon)}$*, where the first inequality holds for sufficiently large n, and the second inequality holds for all n.*

Proof
We apply AEP, and convergence in probability on the three conditions of the jointly typical set. That is, there exists n_1, n_2, n_3 such that for all $n > n_1$, we have

$$
\Pr\left\{ \left| -\frac{1}{n} \log p(x^n) - H(X) \right| \ge \epsilon \right\} < \epsilon/3,
$$

and for all $n > n_2$, we have

$$
\Pr\left\{ \left| -\frac{1}{n} \log p(y^n) - H(Y) \right| \ge \epsilon \right\} < \epsilon/3,
$$

and for all $n > n_3$, we have

$$
\Pr\left\{ \left| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \right| \ge \epsilon \right\} < \epsilon/3.
$$

All three apply for n greater than the largest of n_1, n_2, n_3. Therefore the probability of the union the set of (x^n, y^n) satisfying these inequalities must be less than ϵ, and for n sufficiently large, the probability of the set $A_\epsilon^{(n)}$ is greater than $1 - \epsilon$.

Upper Bound:

$$\begin{aligned}
1 &= \sum p(x^n, y^n) \\
&\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}(X,Y)} p(x^n, y^n) \\
&\geq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}(X,Y)} 2^{-n(H(X,Y)+\epsilon)}, \text{ by definition of typicality} \\
&= 2^{-n(H(X,Y)+\epsilon)} \left| A_\epsilon^{(n)}(X,Y) \right| \\
&\Rightarrow \left| A_\epsilon^{(n)}(X,Y) \right| \leq 2^{n(H(X,Y)+\epsilon)}
\end{aligned}$$

Lower Bound:
By Part 1, $P\left((X^n, Y^n) \in A_\epsilon^{(n)}(X,Y)\right) \xrightarrow{n \to \infty} 1$.
Thus, for large n:

$$\begin{aligned}
1 - \epsilon &\leq P((X^n, Y^n) \in A_\epsilon^{(n)}(X,Y)) \\
&\leq \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} 2^{-n(H(X,Y)-\epsilon)} \\
&= 2^{-n(H(X,Y)-\epsilon)} \left| A_\epsilon^{(n)}(X,Y) \right| \\
&\Rightarrow \left| A_\epsilon^{(n)}(X,Y) \right| \geq (1-\epsilon) 2^{n(H(X,Y)-\epsilon)}
\end{aligned}$$

□

Part B. For $(\widetilde{X}^n, \widetilde{Y}^n)$ where $P_{\widetilde{X},\widetilde{Y}} = P_X \times P_Y$ (essentially you have sequences X^n, Y^n which are drawn from P_X and P_Y **independently**):

$$(1-\epsilon) 2^{-n(I(X;Y)+3\epsilon)} \leq P\left\{(\widetilde{X}^n, \widetilde{Y}^n) \in A_\epsilon^{(n)}(X,Y)\right\} \leq 2^{-n(I(X;Y)-3\epsilon)}$$