Capacity Achieving Codes for the AWGN Channel

Raj and Hesam

I. Basic Setup

We consider the setup shown in Fig. 1. A vector $\mathbf{x} \in \mathbb{F}_2^k$ is mapped into a codeword $\mathbf{y} \in \mathbb{F}_2^n$ by an

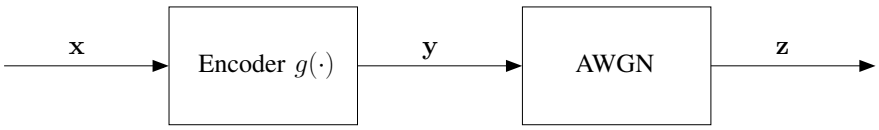

Fig. 1. Transmission over an AWGN

encoding function $g(\cdot)$. Since we will only consider linear codes, the encoder $g(\cdot)$ is a linear transformation from $\mathbb{F}_2^k$ to $\mathbb{F}_2^n$, with code generator matrix $\mathbf{G} \in \mathbb{F}_2^{k \times n}$. Hence

$$\mathbf{y}^\mathsf{T} = \mathbf{x}^\mathsf{T}\mathbf{G}.$$

The codeword $\mathbf{y}$ is sent over a (real) AWGN channel[1], whose output $\mathbf{z} \in \mathbb{R}^n$ may be written as

$$\mathbf{z} = \mathbf{y} + \mathbf{w},$$

where $\mathbf{w} \sim$ i.i.d. $\mathcal{N}\left(0, \frac{1}{\rho}\mathbf{I}\right)$. Hence ρ denotes the SNR of the AWGN channel.

Lemma 1: For the setup under consideration, $H(\mathbf{z}|\mathbf{x}) = H(\mathbf{z}|\mathbf{y})$, and hence $I(\mathbf{x};\mathbf{z}) = I(\mathbf{y};\mathbf{z})$.

[1]We assume standard BPSK modulation for transmission over the AWGN channel, i.e., $y_j \mapsto (-1)^{y_j}\ \forall\ j$. With a slight abuse of notation, we refer to both the binary codeword and the modulated symbols with the same notation $\mathbf{y}$; the one being referred to will be clear from the context.

Proof: We assume that the decoder is aware of the generator matrix $\mathbf{G}$. Intuitively, Lemma 1 is true since $\mathbf{y}$ is a deterministic function of $\mathbf{x}$. More rigorously, denote $N(\mathbf{G})$ as the cardinality of the left null-space of $\mathbf{G}$. We first notice that

$$p_{X|Z}(\mathbf{x}|\mathbf{z}) = \frac{1}{N(\mathbf{G})} p_{Y|Z}(\mathbf{G}^{\mathsf{T}}\mathbf{x}|\mathbf{z}) = \frac{1}{N(\mathbf{G})} p_{Y|Z}(\mathbf{y}|\mathbf{z}).$$

Further, we assume that the input is equiprobable, i.e., $p_X(\mathbf{x}) = \frac{1}{2^k} \ \forall \ \mathbf{x}$. Hence,

$$\begin{aligned} H(\mathbf{z}|\mathbf{x}) &= -\int\int p_{X,Z}(\mathbf{x},\mathbf{z}) \log p_{Z|X}(\mathbf{z}|\mathbf{x}) \ d\mathbf{z}d\mathbf{x} \\ &= -\int_{\mathbf{z}} \sum_{\mathbf{x}} p_Z(\mathbf{z}) p_{X|Z}(\mathbf{x}|\mathbf{z}) \log \frac{p_Z(\mathbf{z}) p_{X|Z}(\mathbf{x}|\mathbf{z})}{p_X(\mathbf{x})} \ d\mathbf{z} \\ &= -\int_{\mathbf{z}} \sum_{\mathbf{x}} p_Z(\mathbf{z}) \frac{p_{Y|Z}(\mathbf{G}^{\mathsf{T}}\mathbf{x}|\mathbf{z})}{N(\mathbf{G})} \log \frac{p_Z(\mathbf{z}) p_{Y|Z}(\mathbf{G}^{\mathsf{T}}\mathbf{x}|\mathbf{z})}{N(\mathbf{G})\frac{1}{2^k}} \ d\mathbf{z}. \end{aligned}$$

In the summation above, as $\mathbf{x}$ runs over $\mathbb{F}_2^k$, $\mathbf{y}$ runs over the code $\mathcal{C}$, with each value being taken $N(\mathbf{G})$ times. Therefore,

$$\begin{aligned} H(\mathbf{z}|\mathbf{x}) &= -\int_{\mathbf{z}} N(\mathbf{G}) \sum_{\mathbf{y}\in\mathcal{C}} p_Z(\mathbf{z}) \frac{p_{Y|Z}(\mathbf{y}|\mathbf{z})}{N(\mathbf{G})} \log \frac{p_Z(\mathbf{z}) p_{Y|Z}(\mathbf{y}|\mathbf{z})}{p_Y(\mathbf{y})} \ d\mathbf{z} \\ &= -\int_{\mathbf{z}} \sum_{\mathbf{y}\in\mathcal{C}} p_{Y|Z}(\mathbf{y},\mathbf{z}) \log p_{Z|Y}(\mathbf{z}|\mathbf{y}) \ d\mathbf{z} \\ &= H(\mathbf{z}|\mathbf{y}). \end{aligned}$$

■

In order to obtain lower bounds on $I(\mathbf{x};\mathbf{z})$, we will therefore analyse $I(\mathbf{y};\mathbf{z})$. Notice that

$$\begin{aligned} I(\mathbf{y};\mathbf{z}) &= H(\mathbf{z}) - H(\mathbf{z}|\mathbf{y}) \\ &= H(\mathbf{z}) - \frac{1}{2}\log\frac{(2\pi e)^n}{\rho^n}. \end{aligned} \tag{1}$$

II. Lower bound on the entropy

We now lower bound the entropy

$$H(\mathbf{z}) = -\int p(\mathbf{z}) \log p(\mathbf{z}) \ d\mathbf{z}.$$

Using Jensen's inequality,

$$\log\left(\mathbb{E}_{f_1}(f_2(x))\right) \geq \mathbb{E}_{f_1}\left(\log f_2(x)\right).$$

Set $f_1(\cdot)$ and $f_1(\cdot)$ to equal $p(\mathbf{z})$, we obtain

$$H(\mathbf{z}) \geq -\log\left[\int p^2(\mathbf{z}) \ d\mathbf{z}\right]. \tag{2}$$

Let the code $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{2^k}\}$. Notice that not all $\mathbf{c}_i$ are distinct, if $\mathbf{G}$ is not full-rank. The pdf of $\mathbf{z}$ is given by

$$\begin{aligned} p(\mathbf{z}) &= \frac{1}{|\mathcal{C}|}\sum_{i=1}^{|\mathcal{C}|}\mathcal{N}\left(\mathbf{z};\mathbf{c}_i,\frac{1}{\rho}\mathbf{I}\right) \\ &= \frac{1}{|\mathcal{C}|}\sum_{i=1}^{|\mathcal{C}|}\frac{\rho^{n/2}}{(\sqrt{2\pi})^n}e^{-\frac{\rho}{2}|\mathbf{z}-\mathbf{c}_i|^2}. \end{aligned}$$

$$\begin{aligned} \int p^2(\mathbf{z})\,d\mathbf{z} &= \frac{\rho^n}{|\mathcal{C}|^2(2\pi)^n}\int\sum_{i,j=1}^{|\mathcal{C}|}\exp\left\{-\frac{\rho}{2}\left[|\mathbf{z}-\mathbf{c}_i|^2+|\mathbf{z}-\mathbf{c}_j|^2\right]\right\}\,d\mathbf{z} \\ &= \frac{\rho^n}{|\mathcal{C}|^2(2\pi)^n}\int\sum_{i,j=1}^{|\mathcal{C}|}\exp\left\{-\frac{\rho}{2}\left[|\mathbf{z}|^2+|\mathbf{z}+\mathbf{c}_i-\mathbf{c}_j|^2\right]\right\}\,d\mathbf{z} \end{aligned} \quad (3)$$

We now make use of the following well-known result on the characteristic function of Hermitian quadratic form of complex Gaussian random variables (briefly, HQF-GRV).

Lemma 2: [1, Appendix 4] The characteristic function of the HQF-GRV $\Delta = \mathbf{z}^{\mathsf{H}}\mathbf{F}\mathbf{z}$, where $\mathbf{z} \sim \mathcal{CN}(\overline{\mathbf{z}}, \mathbf{R})$ is given by

$$\Phi_\Delta(s) = \mathbb{E}\left[\exp(-s\Delta)\right] = \frac{\exp(-s\overline{\mathbf{z}}^{\mathsf{H}}\mathbf{F}(\mathbf{I}+s\mathbf{R}\mathbf{F})^{-1}\overline{\mathbf{z}})}{\det(\mathbf{I}+s\mathbf{R}\mathbf{F})}.$$

Using the pdf of a complex GRV, we may rewrite Lemma 2 as

$$\begin{aligned} \mathbb{E}\left[\exp(-s\Delta)\right] &= \int\exp(-s\mathbf{z}^{\mathsf{H}}\mathbf{F}\mathbf{z})\frac{1}{\det(\pi\mathbf{R})}\exp\left\{-(\mathbf{z}-\overline{\mathbf{z}})^{\mathsf{H}}\mathbf{R}^{-1}(\mathbf{z}-\overline{\mathbf{z}})\right\}\,d\mathbf{z} \\ &= \mathbb{E}\left[\exp(-s\Delta)\right] = \frac{\exp(-s\overline{\mathbf{z}}^{\mathsf{H}}\mathbf{F}(\mathbf{I}+s\mathbf{R}\mathbf{F})^{-1}\overline{\mathbf{z}})}{\det(\mathbf{I}+s\mathbf{R}\mathbf{F})}. \end{aligned}$$

Using $\mathbf{F} = \mathbf{I}, s = \frac{\rho}{2}, \mathbf{R} = \frac{2}{\rho}\mathbf{I}, \overline{\mathbf{z}} = \mathbf{c}_j - \mathbf{c}_i$ in the above, we obtain

$$\frac{1}{\det\left(\frac{2\pi}{\rho}\mathbf{I}\right)}\int\exp\left\{-\frac{\rho}{2}|\mathbf{z}|^2\right\}\exp\left\{-\frac{\rho}{2}|\mathbf{z}-(\mathbf{c}_j-\mathbf{c}_i)|^2\right\}\,d\mathbf{z} = \frac{\exp\left[-\frac{\rho}{2}(\mathbf{c}_j-\mathbf{c}_i)^{\mathsf{H}}(2\mathbf{I})^{-1}(\mathbf{c}_j-\mathbf{c}_i)\right]}{\det(2\mathbf{I})}.$$

Using the above in (3), we obtain

$$\int p^2(\mathbf{z})\,d\mathbf{z} = \frac{1}{2^n|\mathcal{C}|^2}\sum_{i,j=1}^{|\mathcal{C}|}\exp\left\{-\frac{\rho}{4}|\mathbf{c}_j-\mathbf{c}_i|^2\right\}. \quad (4)$$

From (1), (2) and (4), we obtain

$$I(\mathbf{x};\mathbf{z}) \geq -\log\left[\left(\frac{\pi e}{2\rho}\right)^{n/2}\frac{1}{|\mathcal{C}|^2}\sum_{i,j=1}^{|\mathcal{C}|}\exp\left\{-\frac{\rho}{4}|\mathbf{c}_j-\mathbf{c}_i|^2\right\}\right].$$

Since BPSK modulation is used, $|\mathbf{c}_j - \mathbf{c}_i|^2 = 4d_H(\mathbf{c}_j, \mathbf{c}_i)$. Further, since we employ a linear code, we may further simplify the above to obtain

$$I(\mathbf{x};\mathbf{z}) \geq -\log\left[\left(\frac{\pi e}{2\rho}\right)^{n/2} \frac{1}{|\mathcal{C}|}\sum_{i=1}^{|\mathcal{C}|} \exp\left\{-\rho\ w_H(\mathbf{c}_i)\right\}\right],$$

where $w_H(\mathbf{x})$ denotes the Hamming weight of $\mathbf{x}$. If we define B_w to be the number of codewords with Hamming weight w, we may rewrite the above as:

$$I(\mathbf{x};\mathbf{z}) \geq -\log\left[\left(\frac{\pi e}{2\rho}\right)^{n/2} \frac{1}{|\mathcal{C}|}\sum_{w=0}^{n} B_w e^{-w\rho}\right] \tag{5}$$

III. Distance Spectrum of the Code

We consider the ensemble of low density generator matrix (LDGM) codes, where each column of $\mathbf{G}$ is a vector of weight d, drawn uniformly at random. To compute B_w for this ensemble, we first calculate the probability of a bit of the codeword being equal to 1. We denote this probability by P_1. Then, B_w would simply be equal to $\binom{n}{w}(P_1)^w(1-P_1)^{n-w}$, since the columns of the generator matrix $\mathbf{G}$ are drawn independent of each other.

In order to compute P_1, we first calculate the probability of having a bit of the codeword $\mathbf{y}$ equal to 1 if the data sequence $\mathbf{x}$ has weight j. We then multiply this value by the probability of a data word having weight j and sum up over all possible weights to obtain P_1. In other words, for $\mathbf{y}^\mathsf{T} = \mathbf{x}^\mathsf{T}\mathbf{G}$, we have:

$$P_1 = \sum_{j=1}^{n} Pr(y_1 = 1 | w_H(\mathbf{x}) = j) Pr(w_H(\mathbf{x}) = j) \tag{6}$$

where y_1 is the first bit of y (one can choose to work with any arbitrary component of $\mathbf{y}$ as it does not change the problem).

Assume we have an input with j number of ones. Without loss of generality, we assume that the first j bits are equal to 1. In order to have the first bit of $\mathbf{y}$ equal to 1 we must have to have odd number of ones in the first column of $\mathbf{G}$ in places where $\mathbf{x}$ is equal to 1. Therefore:

$$P_1^j \triangleq Pr(y_1 = 1 | w_H(\mathbf{x}) = j) = \sum_{i=1}^{\lceil j/2 \rceil} \binom{j}{2i-1} p^{2i-1}(1-p)^{j-2i+1}, \tag{7}$$

where $p = d/k$ is the probability of having a one in a particular coordinate of the first column of $\mathbf{G}$. Equation (7) reduces to:

$$P_1^j = \frac{1-(1-2p)^j}{2}. \tag{8}$$

Since the input $\mathbf{x}$ is chosen according to a uniform distribution, we can compute P_1 as follows:

$$\begin{aligned} P_1 &= \sum_{j=1}^{k} \frac{1-(1-2p)^j}{2} \frac{\binom{k}{j}}{2^k} \\ &= \frac{1}{2^{k+1}} \sum_{j=1}^{k} \binom{k}{j} (1-(1-2p)^j) \\ &= \frac{1}{2^{k+1}} \left[\sum_{j=1}^{k} \binom{k}{j} - \sum_{j=1}^{k} \binom{k}{j} (1-2p)^j \right] \\ &= \frac{1}{2^{k+1}} \left[2^k - (2-2p)^k \right] = \frac{1-(1-p)^k}{2}. \end{aligned} \tag{9}$$

As a result of equation (9), we can write B_w as:

$$\begin{aligned} \sum_{w=1}^{n} B_w e^{-w\rho} &= \sum_{w=1}^{n} \binom{n}{w} (P_1 e^{-\rho})^w (1-P_1)^{n-w} \\ &= (1-P_1+P_1 e^{-\rho})^n \\ &= \left[\frac{1+(1-p)^k}{2} + \frac{e^{-\rho}}{2} (1-(1-p)^k) \right]^n \\ &= \left[\frac{1+(1-d/k)^k}{2} + \frac{e^{-\rho}}{2} (1-(1-d/k)^k) \right]^n. \end{aligned} \tag{10}$$

As $k \to \infty$ we have that

$$\sum_{w=1}^{n} B_w e^{-w\rho} \to \left[\frac{1+e^{-d}}{2} + \frac{e^{-\rho}}{2} (1-e^{-d}) \right]^n. \tag{11}$$

We have from equations (5) and (11) that

$$I(\mathbf{x};\mathbf{z}) \geq -\log_2 \left[\left(\frac{\pi e}{2\rho} \right)^{n/2} \frac{1}{2^{Rn}} \left[\frac{1+e^{-d}}{2} + \frac{e^{-\rho}}{2} (1-e^{-d}) \right]^n \right], \tag{12}$$

where $R = k/n$ is the rate of the code. Notice that the right hand side above will diverge if

$$\left(\frac{\pi e}{2\rho} \right)^{1/2} \frac{1}{2^R} \left[\frac{1+e^{-d}}{2} + \frac{e^{-\rho}}{2} (1-e^{-d}) \right] > 1.$$

This threshold after which the bound diverges may be expressed in terms of the code rate R and the SNR ρ. In particular, the bound will diverge for all $\rho < \rho'(R,d)$, where the function $\rho'(R,d)$ is the solution of ρ obtained from equating the left and right hand sides of the above bound. In the region of convergence,

$$\frac{I(\mathbf{x};\mathbf{z})}{n} \geq -\log_2 \left[\sqrt{\frac{\pi e}{2\rho}} \frac{1}{2^R} \left(\frac{1+e^{-d}}{2} + \frac{e^{-\rho}}{2} (1-e^{-d}) \right) \right] \tag{13}$$

There is a problem with the above bound since it tends to ∞ with ρ

REFERENCES

[1] M. Schwartz, W. R. Bennett, S. Stein, *Communications Systems and Techniques*, McGraw Hill Book Company, 1966.