ECE 729
Channel Coding — The Big Picture

In channel coding we can use either deterministic or random codes, and we can use either the average or maximal probability of error. Hence, we must introduce 4 notions of achievable rate.

Except in Section 5, X and Y are finite sets.

1. Achievability Using Deterministic Codes

A **deterministic encoder** is any function $f : \{1,\dots,N\} \to \mathsf{X}^n$. Observe that $f = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)$ is an element of

$$\underbrace{\mathsf{X}^n \times \cdots \times \mathsf{X}^n}_{N \text{ times}}.$$

There are $|\mathsf{X}|^{nN}$ possible encoders.

A **deterministic decoder** is any function $\varphi : \mathsf{Y}^n \to \{1,\dots,N\}$. If we enumerate the elements of Y^n as

$$\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{|\mathsf{Y}|^n},$$

then a decoder is specified by assigning each element to one of the numbers $1,\dots,N$. Hence, there are $N^{|\mathsf{Y}|^n}$ possible decoders.

1.1. Probabilistic Model

Let $f = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)$ be a given encoder. Let M be a $\{1,\dots,N\}$-valued random variable, and let $\boldsymbol{Y}$ be a Y^n-valued random variable with joint probabilities of the form

$$\mathsf{P}(M = i, \boldsymbol{Y} \in B) = \frac{1}{N} \cdot W_n(B|\boldsymbol{x}_i), \tag{1}$$

where $i \in \{1,\dots,N\}$ and $B \subset \mathsf{Y}^n$, and where $W_n(B|\boldsymbol{x})$ is shorthand for

$$\sum_{\boldsymbol{y}\in B} W_n(\boldsymbol{y}|\boldsymbol{x}),$$

and $W_n(\boldsymbol{y}|\boldsymbol{x})$ is a conditional probability mass function.

We can use the law of total probability and the law of substitution to write

$$\begin{aligned}
\mathsf{P}(\varphi(\boldsymbol{Y}) \neq M) &= \frac{1}{N}\sum_{i=1}^{N} \mathsf{P}(\varphi(\boldsymbol{Y}) \neq M | M = i) \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathsf{P}(\varphi(\boldsymbol{Y}) \neq i | M = i) \\
&= \frac{1}{N}\sum_{i=1}^{N} W_n(\{\boldsymbol{y} : \varphi(\boldsymbol{y}) \neq i\}|\boldsymbol{x}_i) \\
&=: e_n(f,\varphi).
\end{aligned} \tag{2}$$

1.2. Achievable-Rate Definitions

Definition D-A. A number $R \geq 0$ is achievable using **deterministic codes** under the **average** probability-of-error criterion if for every $\lambda > 0$, for every $\Delta R > 0$, for all sufficiently large n, there is a positive integer N with

$$\frac{\log N}{n} > R - \Delta R,$$

and there is a **deterministic code** (f,φ) with $f = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)$ such that

$$\frac{1}{N}\sum_{i=1}^{N} W_n(\{\boldsymbol{y} : \varphi(\boldsymbol{y}) \neq i\}|\boldsymbol{x}_i) < \lambda. \tag{3}$$

Of course, (3) is equivalent to writing $\mathsf{P}(\varphi(\boldsymbol{Y}) \neq M) < \lambda$ under the probabilistic model (1).

Definition D-M. A number $R \geq 0$ is achievable using **deterministic codes** under the **maximal** probability-of-error criterion if for every $\lambda > 0$, for every $\Delta R > 0$, for all sufficiently large n, there is a positive integer N with

$$\frac{\log N}{n} > R - \Delta R,$$

and there is a **deterministic code** (f,φ) with $f = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)$ such that

$$W_n(\{\boldsymbol{y} : \varphi(\boldsymbol{y}) \neq i\}|\boldsymbol{x}_i) < \lambda, \quad i = 1,\dots,N. \tag{4}$$

If (1) is replaced by $q(i)W_n(B|\boldsymbol{x}_i)$ for *any* probability mass function $q(i)$, then (4) implies $\mathsf{P}(\varphi(\boldsymbol{Y}) \neq M) < \lambda$.

It is obvious that if a rate satisfies Definition D-M, then it satisfies Definition D-A. In other words, if $\mathscr{C}_{\text{D-M}}$ denotes the set of achievable rates under Definition D-M, and if $\mathscr{C}_{\text{D-A}}$ denotes the set of achievable rates under Definition D-A, then

$$\mathscr{C}_{\text{D-M}} \subset \mathscr{C}_{\text{D-A}}.$$

The result that

$$\frac{1}{N}\sum_{i=1}^{N} \theta_i < \lambda$$

implies that at least $\lfloor N/2 \rfloor$ of the θ_i are less than 2λ shows, after decoder modification, that if a rate satisfies D-A, then it satisfies D-M. In other words,

$$\mathscr{C}_{\text{D-A}} \subset \mathscr{C}_{\text{D-M}}.$$

Hence,

$$\mathscr{C}_{\text{D-M}} = \mathscr{C}_{\text{D-A}}. \tag{5}$$

2. Achievability Using Random Codes

Let $\mathscr{U}_n^N$ denote the set of all encoder-decoder pairs (f,φ). The collection $\mathscr{U}_n^N$ is a finite set with $|\mathsf{X}|^{nN}N^{|\mathsf{Y}|^n}$ elements. If (F,Φ) is a $\mathscr{U}_n^N$-valued random variable, then (F,Φ) is a **random code**.

2.1. Probabilistic Model

Let (F,Φ) be a random code, let M be a $\{1,\ldots,N\}$-valued random variable, and let $\boldsymbol{Y}$ be a Y^n-valued random variable with conditional joint probabilities given by

$$\mathsf{P}(M=i,\boldsymbol{Y}\in B|F=f,\Phi=\varphi)=\frac{1}{N}W_n(B|\boldsymbol{x}_i), \tag{6}$$

where $f=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$. By the analysis leading to (2), we see that

$$\mathsf{P}(\Phi(\boldsymbol{Y})\neq M|F=f,\Phi=\varphi)=e_n(f,\varphi).$$

2.2. Achievable-Rate Definitions

Definition R-A. A number $R\geq 0$ is achievable using **random codes** under the **average** probability-of-error criterion if for every $\lambda>0$, for every $\Delta R>0$, for all sufficiently large n, there is a positive integer N with

$$\frac{\log N}{n}>R-\Delta R,$$

and there is a $\mathscr{U}_n^N$-valued **random code** (F,Φ) with $F=(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N)$ such that

$$\mathsf{E}\left[\frac{1}{N}\sum_{i=1}^{N}W_n(\{\boldsymbol{y}:\Phi(\boldsymbol{y})\neq i\}|\boldsymbol{X}_i)\right]<\lambda. \tag{7}$$

The left-hand side of (7) is just $\mathsf{E}[e_n(F,\Phi)]$.

Definition R-M. A number $R\geq 0$ is achievable using **random codes** under the **maximal** probability-of-error criterion if for every $\lambda>0$, for every $\Delta R>0$, for all sufficiently large n, there is a positive integer N with

$$\frac{\log N}{n}>R-\Delta R,$$

and there is a $\mathscr{U}_n^N$-valued **random code** (F,Φ) with $F=(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N)$ such that

$$\mathsf{E}\big[W_n(\{\boldsymbol{y}:\Phi(\boldsymbol{y})\neq i\}|\boldsymbol{X}_i)\big]<\lambda,\quad i=1,\ldots,N. \tag{8}$$

It is obvious that if a rate satisfies Definition R-M, then it satisfies Definition R-A. In other words, if $\mathscr{C}_{\text{R-M}}$ denotes the set of achievable rates under Definition R-M, and if $\mathscr{C}_{\text{R-A}}$ denotes the set of achievable rates under Definition R-A, then

$$\mathscr{C}_{\text{R-M}}\subset\mathscr{C}_{\text{R-A}}. \tag{9}$$

3. Connecting Random and Deterministic Codes

The **random coding argument** is that if $\mathsf{E}[e_n(F,\Phi)]<\lambda$, then there must be a realization (f,φ) with $e_n(f,\varphi)<\lambda$. In other words, if (7) holds, then there is a encoder-decoder realization such that (3) holds. Hence, if a rate satisfies R-A, then it satisfies D-A. Symbolically, we write

$$\mathscr{C}_{\text{R-A}}\subset\mathscr{C}_{\text{D-A}}. \tag{10}$$

If we combine (9) and (10), we obtain

$$\mathscr{C}_{\text{R-M}}\subset\mathscr{C}_{\text{R-A}}\subset\mathscr{C}_{\text{D-A}}. \tag{11}$$

4. Analysis of the DMC

For the DMC, $W_n=W^n$. Put

$$\mathscr{C}:=\Big\{R\geq 0:R\leq\sup_P I(P\times W)\Big\}.$$

For this set $\mathscr{C}$, we proved the forward result $\mathscr{C}\subset\mathscr{C}_{\text{R-M}}$,[1] and we proved the weak converse $\mathscr{C}_{\text{D-A}}\subset\mathscr{C}$. Combining these two results with (11) shows that all the sets in (11) are equal to $\mathscr{C}$. In fact, on account of (5), for the DMC,

$$\mathscr{C}_{\text{R-M}}=\mathscr{C}_{\text{R-A}}=\mathscr{C}_{\text{D-M}}=\mathscr{C}_{\text{D-A}}=\mathscr{C}.$$

5. Continuous Channels and Gaussian Channels

Let $\mathsf{X}=\mathsf{Y}=\mathbb{R}$. We now reinterpret $W_n(B|\boldsymbol{x})$ in the preceding sections as shorthand for

$$\int_B w_n(\boldsymbol{y}|\boldsymbol{x})\,d\boldsymbol{y},$$

where $w_n(\boldsymbol{y}|\boldsymbol{x})$ is a conditional density function.

The deterministic-code achievable-rate definitions are now easily extended adding the power constraint

$$\|\boldsymbol{x}_i\|^2\leq n\mathrm{P},\quad i=1,\ldots,N \tag{12}$$

for both the maximal and average probability-of-error definitions. Since both modified achievable-rate definitions use the same power constraint, it is easy to see that

$$\mathscr{C}_{\text{D-M}}(\mathrm{P})\subset\mathscr{C}_{\text{D-A}}(\mathrm{P}).$$

We extend the random-code achievable-rate definitions by adding the power constraint

$$\mathsf{E}\big[\|\boldsymbol{X}_i\|^2\big]\leq n\mathrm{P},\quad i=1,\ldots,N \tag{13}$$

for both the maximal and average probability-of-error definitions. We then have

$$\mathscr{C}_{\text{R-M}}(\mathrm{P})\subset\mathscr{C}_{\text{R-A}}(\mathrm{P}).$$

In the case of a memoryless channel ($w_n=w^n$), let

$$\mathscr{C}(\mathrm{P}):=\Big\{R\geq 0:R\leq\sup_{X:\mathsf{E}[X^2]\leq\mathrm{P}}I(X\wedge Y)\Big\},$$

where Y has conditional density $f_{Y|X}(y|x)=w(y|x)$. Since the random encoder we used implies

$$\mathsf{E}\big[\|\boldsymbol{X}_i\|^2\big]=\mathsf{E}\left[\sum_{k=1}^{n}X_{ik}^2\right]=\sum_{k=1}^{n}\mathsf{E}[X_{ik}^2]\leq n\mathrm{P},$$

we actually proved the forward result $\mathscr{C}(\mathrm{P})\subset\mathscr{C}_{\text{R-M}}(\mathrm{P})$. At this point we have established that $\mathscr{C}(\mathrm{P})\subset\mathscr{C}_{\text{R-M}}\subset\mathscr{C}_{\text{R-A}}$. Part of the extended Definition R-A is that (7) holds. We then used the random coding argument to get (3). *However, we do not*

[1] Recall that we used random codewords chosen i.i.d., and the decoder was taken to be a specific deterministic function of the random codebook.

know anything about the values of the $\|\boldsymbol{X}_i\|^2$ *or the* $\|\boldsymbol{x}_i\|^2$. We then extracted a subset of $\lfloor N/2 \rfloor - 1$ codewords $\boldsymbol{x}_i$ that satisfied (4) for 2λ *and* the power constraint $\|\boldsymbol{x}_i\|^2 \leq n\mathrm{P}$, and we modified the decoder to use fewer messages. Hence, we established that $\mathscr{C}(\mathrm{P}) \subset \mathscr{C}_{\text{R-M}}(\mathrm{P}) \subset \mathscr{C}_{\text{R-A}}(\mathrm{P}) \subset \mathscr{C}_{\text{D-M}}(\mathrm{P})$. As noted above, $\mathscr{C}_{\text{D-M}}(\mathrm{P}) \subset \mathscr{C}_{\text{D-A}}(\mathrm{P})$.

For the Gaussian channel we first showed that

$$\sup_{X:\mathsf{E}[X^2]\leq \mathrm{P}} I(X \wedge Y) = \frac{1}{2}\log\Big(1 + \frac{\mathrm{P}}{\mathscr{N}_0}\Big).$$

We were then able to prove the weak converse $\mathscr{C}_{\text{D-A}}(\mathrm{P}) \subset \mathscr{C}(\mathrm{P})$.

We have thus established that for the Gaussian channel under a power constraint ((12) for deterministic codes and (13) for random codes),

$$\mathscr{C}_{\text{R-M}}(\mathrm{P}) = \mathscr{C}_{\text{R-A}}(\mathrm{P}) = \mathscr{C}_{\text{D-M}}(\mathrm{P}) = \mathscr{C}_{\text{D-A}}(\mathrm{P}) = \mathscr{C}(\mathrm{P}). \qquad (14)$$

Remark. There are at least two alternatives to (13) that we could have considered. Since

$$\frac{1}{N}\sum_{i=1}^{N} \mathsf{E}\big[\|\boldsymbol{X}_i\|^2\big] \leq n\mathrm{P}$$

is implied by (13), it is easy to see that if (13) is replaced by this condition, then the analysis leading to (14) still holds. On the other hand, stronger techniques would be needed to analyze what happens if (13) were replaced by

$$\|\boldsymbol{X}_i\|^2 \leq n\mathrm{P} \quad \text{almost surely.}$$

Remark. In all of our work, we never used a random decoder except in so far as if $\varphi(\boldsymbol{y}) = \varphi_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N}(\boldsymbol{y})$ depends on the codebook, and if the codebook is random, then so is $\varphi(\boldsymbol{y}) = \varphi_{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N}(\boldsymbol{y})$. Hence, everything would go through unchanged if we restricted to decoders to be of this form.