Chapter 9

Channel Reliability Function

Po-Ning Chen
Department of Communications Engineering
National Chiao-Tung University
Hsin Chu, Taiwan 30050
Motivations

• \( R < C \) \( \Rightarrow \) \( P_e (n) \rightarrow 0 \).

For example,

– if it results in lower information transmission rate?

– if it is possible for one to consider to use the latter code even bits respectively, is it possible for one to consider to use the latter code even

– if \( C = 1 \) bit, and there are two optimal codes with rates 0.5 bits and 0.25

For example, \( p_e (w) \leq C > R \).
Definition 9.1 (channel reliability function)

Let $P_e(n, R)$ be the minimum probability error achievable for block codes of length $n$ with rate no less than $R$. Then the channel reliability function $E(R)$ is defined as the error exponent of $P_e(n, R)$, i.e.,

$$E(R) = \lim_{n \to \infty} -\frac{1}{n} \log P_e(n, R).$$

**Theorem 9.3 (random coding exponent)**

For DMC with generic transition probability $P_Y|X$, the random coding exponent is defined by

$$E_r(R) = \max_{0 \leq s \leq 1} \max \left\{ \sum_{x \in X} P(X=x) \frac{1}{1+s} \sum_{y \in Y} (\sum_{x \in X} P(Y=y|X=x) \log \frac{1}{1+s}) \right\}.$$

**Proof:** Similar to the proof of channel capacity, the code is randomly selected according to some input distribution $P_{\tilde{X}}$. The code is randomly selected according to some input distribution $P_{\tilde{X}}$. For sufficiently large $n$, the minimum probability error achievable for block codes of length $n$ with rate no less than $R$ is

$$P_e(n, R) \leq e^{-nE_r(R)}.$$
Step 1: Maximum likelihood decoder. Let \( \{c_1, \ldots, c_M\} \in \mathcal{X}^n \) denote the set of \( n \)-tuple block codewords selected, and let the decoding partition for symbol \( m \) (namely, the set of channel outputs that classify to \( m \)) be

\[
U_m = \{ y_n : P(Y_n | X_n)^{c_m} > P(Y_n | X_n)^{c_m'} \text{ for all } m' \neq m \}.
\]

Those channel outputs that are on the boundary, i.e.,

\[
\{ m' \neq m \} \text{ for all } m \text{ and } m', \quad (P(Y_n | X_n)^{c_m})_{uX|uX_d} < (P(Y_n | X_n)^{c_m'})_{uX|uX_d} : u \}
\]

be polynomially close to \( m \) (namely, the set of channel outputs that classify to \( m \)) be the set of \( n \)-tuple block codewords selected, and let the decoding partition for symbol \( m \) be

\[
\mathcal{C}_1, \ldots, \mathcal{C}_l \}
\]

denote the set of \( n \)-tuple block codewords selected. Let \( \mathcal{C}_1, \ldots, \mathcal{C}_l \}

\[
\text{Step 2: Property of indicator function for } s > 0. \text{ Let } \phi_m \text{ be the indicator function of } U_m. \text{ Then for all } s > 0. \text{ Let } \phi_m(\cdot) \text{ be the indicator function for } s > 0.
\]

\[
1 - \phi_m(y_n) \leq \left\{ \begin{array}{ll}
\frac{1}{1+s} & \text{for some } m \text{ and } m', \quad (P(Y_n | X_n)^{c_m})_{uX|uX_d} \leq (P(Y_n | X_n)^{c_m'})_{uX|uX_d} \\
& \text{for all } m' \neq m \end{array} \right.
\]

\[
\sum_{m' \neq m, 1 \leq m' \leq M_n} \left[ (P(Y_n | X_n)^{c_m})_{uX|uX_d} \right]^{u'Y_n \geq uY_n, u' \neq u} \geq (P(Y_n | X_n)^{u\theta})^{u \phi} - 1
\]
\[
\begin{align*}
\left\{ \left( \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \right\}_{s+1/1} & \sum_{u\notin \hat{u}} \left( \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \sum_{u\notin \hat{u}} = \\
\left\{ \left[ \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \right\}_{s+1/1} & \sum_{u\notin \hat{u}} \left( \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \sum_{u\notin \hat{u}} > \\
\left\{ \left[ (u\hat{f}) \frac{w}{u} - 1 \right] \left( \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \right\}_{u\notin \hat{u}} & \sum_{u\notin \hat{u}} \left( \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \sum_{u\notin \hat{u}} = \\
\left\{ \left( \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \right\}_{u\notin \hat{u}} & \sum_{u\notin \hat{u}} \left( \frac{w}{u} | u\hat{f} \right)^{uX|u\Lambda} \sum_{u\notin \hat{u}} > w|e \end{align*}
\]

Then denote the probability of error given codeword \( u \) transmitted. Then

\[ \text{Random-coding exponent} \]
\[ s([I]) \gtrless [s][I] \]

\[ \forall I, \exists s > 0 \text{ when } 1 \leq s \leq 1, \text{ and bounds on probability} \]

Step 4: Expectation of \( P_e \) |

\[ E[P_e | m] \leq E \left[ \sum_{y_n \in Y} P \frac{1}{1 + s} Y_n | X_n (y_n | c_m) \right] \]

\[ = \sum_{y_n \in Y} E \left[ \sum_{m' \neq m} P \frac{1}{1 + s} Y_n | X_n (y_n | c_m') \right] \]

where the latter step follows because \( \{c_m\} \) are independent random variables.

Step 5: Jensen's inequality for \( 0 < s \leq 1 \), and bounds on probability of error.

By Jensen's inequality, when \( 0 < s \leq 1 \),

\[ E[t_s] \leq (E[t]) s \]

**Random-coding exponent**
Therefore,

\[ \mathbb{E} \left[ \left( \mathbb{E} \left[ \frac{P_1}{1+s} \left| X_n \right| \left( \left| c_m \right| \left| \alpha \right| \right) \right] \right] \right] \leq \sum_{y_n \in Y} \mathbb{E} \left[ \left( \sum_{m' \neq m, 1 \leq m' \leq M_n} \frac{P_1}{1+s} \left| X_n \right| \left( \left| c_{m'} \right| \right) \right] \right] \]

Since the codewords are selected with identical distribution, the expectations

\[ \left( \left( \left( \mathbb{E} \left[ \frac{P_1}{1+s} \left| X_n \right| \left( \left| c_m \right| \right) \right] \right) \right) \right) ^ {u_N \geq \mu \geq \Gamma, \mu \neq \mu} \leq \sum_{y_n \in Y} \mathbb{E} \left[ \left( \sum_{m' \neq m, 1 \leq m' \leq M_n} \frac{P_1}{1+s} \left| X_n \right| \left( \left| c_{m'} \right| \right) \right] \right] \]

Therefore,

Random-coding exponent
\[
\left( (u^x|u^\tilde{f})_{(s+1)/1}^* d(u^x)^{u^x|u^\tilde{f}} \right) \bigotimes_{s \in W} u^W \geq \mathcal{E}
\]

Since the upper bound of \(E[P_e|m]\) is no longer dependent on \(m\), \(E[P_e]\) can be bounded by the same bound, namely,

\[
\left( (u^x|u^\tilde{f})_{(s+1)/1}^* d(u^x)^{u^x|u^\tilde{f}} \right) \bigotimes_{s \in W} u^W =
\]

\[
\left( \left( \left( u^x|u^\tilde{f} \right)_{(s+1)/1} d \right) \right) \bigotimes_{s \in W} u^W \geq
\]

\[
\left( \left( \left( u^x|u^\tilde{f} \right)_{(s+1)/1} d \right) \right) \bigotimes_{s \in W} \left( 1 - u^W \right)
\]

Hence, should be the same for each \(m\). Hence,
By using the fact that $P_{X^n}$ and $P_{Y^n|X^n}$ are product distributions with identical marginal, and taking logarithmic operation for both sides of the above inequality, we have desired result. Note that $\limsup_{n \to \infty} \left( \frac{1}{n} \log_2 M_n \right) = R$. 

$R = \sup_n W_n \log_2 (u/1) \left( \frac{u}{1} \right)^{\infty-n}$. 

By using the fact that $P_{X^n \mid Y^n}$ and $P_{X^n}$ are product distributions with identical marginal, and taking logarithmic operation for both sides of the above inequality, we have desired result. Note that $\limsup_{n \to \infty} \left( \frac{1}{n} \log_2 M_n \right) = R$. 

Random-coding exponent
Definition 9.4 (random-coding exponent)

The random coding exponent for DMC with generic distribution \( P_Y | X \) is defined by

\[
\mathcal{E}_r(R) \triangleq \max_{0 \leq s \leq 1} \left[ -sR + \mathcal{E}_0(s) \right],
\]

where \( \mathcal{E}_0(s) \) is defined as

\[
\mathcal{E}_0(s) \triangleq \max_{P_X} \left\{ -\log \sum_{y \in Y} \left( \sum_{x \in X} P_X(x) P_Y(y | x)^{1+s} \right)^{1+s} \right\}.
\]

The properties of \( \mathcal{E}_r(R) \) can be realized via the analysis of function \( \mathcal{E}_0(s) \) as follows:

1. \( \mathcal{E}_r(R) \) is non-increasing.
2. \( \mathcal{E}_r(R) \) is convex in \( R \).
3. \( \mathcal{E}_r(R) = 0 \) where \( \mathcal{C} \) is channel capacity, and \( \mathcal{E}_r(R - \delta) > 0 \) for all \( 0 < \delta < \mathcal{C} - \mathcal{C} \).
4. There exists \( R_c > 0 \) such that for \( 0 < R < R_c \), the slope of \( \mathcal{E}_r(R) \) is \(-1\).
The properties of random coding exponent

II:9-10

The properties of random coding exponent for \( \varepsilon = 0.2 \) is depicted in Figure 9.3.

Note that the input distribution achieving \((H)E\)' is uniform, i.e., \( 0 \leq \varepsilon \leq 1 \). The

where

\[
\max_{0 \leq s \leq 1} \left\{ -sR - \log \left( \frac{p(1 - \varepsilon)}{1 + s} + \frac{(1 - p)\varepsilon}{1 + s} \right) \right\} = (R)^E
\]

becomes

Example 9.6

For BSC with crossover probability \( \varepsilon \), the random coding exponent

For BSC channel with crossover probability \( \varepsilon \) and input distribution \( (d - 1, d) \):

\[
\begin{array}{c}
1 \\
\vdots \\
0
\end{array}
\begin{array}{c}
\varepsilon - 1 \\
\varepsilon \\
\varepsilon - 1
\end{array}
\begin{array}{c}
1 \\
\vdots \\
0
\end{array}
\begin{array}{c}
d - 1 \\
\vdots \\
d
\end{array}
\]

The properties of random coding exponent

II:9-10
The properties of random coding exponent II:9-11

Random coding exponent for BSC with crossover probability 0.2.
Expurgated exponent

II:9-12

Random coding exponent is derived based on random coding technique.

Since the codewords selection is unbiased, both the "good" codes and "bad" codes contribute the same when computing the expectation of error probability.

Therefore, if, to some extent, we can reduce the contribution of the "bad" codes, a better bound on channel reliability function may be found.

It can be expected that for $1 \leq m \leq \mathcal{W}$ and $\mathcal{W} \geq w \geq W$,

\[
\Pr_{\mathcal{H}}(e|w_{\mathcal{H}}) \leq (\mathcal{W})^{w|e}d \leq (\mathcal{W})^{w|e}d
\]

From a new codebook, a new codebook is formed, which is the error probability given codeword $m$ is transmitted.

After that, it chooses the first $\mathcal{W}$ codewords (whose error probability is smaller) to form a codebook $\mathcal{W}$,$\mathcal{H}$,$\mathcal{W}$,$\mathcal{H}$,$\mathcal{W}$,$\mathcal{H}$ and sorts these codewords in ascending order in terms of $\Pr_{\mathcal{H}}(e|w_{\mathcal{H}})$, which is the error probability given codeword $m$ is transmitted.

In stead of randomly selecting $\mathcal{W}$ codewords, expurgated approach first draws $2\mathcal{W}$ codewords to form a codebook $\mathcal{W}$,$\mathcal{H}$,$\mathcal{W}$,$\mathcal{H}$,$\mathcal{W}$,$\mathcal{H}$, and sorts these codewords in ascending order in terms of $\Pr_{\mathcal{H}}(e|w_{\mathcal{H}})$, which is the error probability given codeword $m$ is transmitted.

\[
\Pr_{\mathcal{H}}(e|w_{\mathcal{H}}) \leq \Pr_{\mathcal{H}}(e|w_{\mathcal{H}}) \leq \Pr_{\mathcal{H}}(e|w_{\mathcal{H}})
\]

Hence, a better codebook is obtained.

**Key 1:**

In stead of randomly selecting $\mathcal{W}$ codewords, a better bound on channel reliability function may be found.
Key 2: Expurged exponent

\[ \exists \, d > 0 \]

\[ \cdot \left[ \left( \nu \overline{\nu} \mathcal{C} \right)^{w|\alpha} \right] E \supseteq \left[ \left( \nu \overline{\nu} \mathcal{C} \right)^{w|\beta} \right]_{d/\beta} E \]

Note that by Lyapounov's inequality.

\[ \forall \, \rho > 0 \quad \left[ \rho \| X \| \right]_{\rho/\rho} E \geq \left[ \alpha \| X \| \right]_{\alpha/\alpha} E \]

Lyapunov's inequality.
Lemma 9.7 (existence of "good" code for expurgated exponent)

For a sequence of code size $M_n$ satisfying $\limsup_{n \to \infty} \left( \frac{1}{n} \log_2 M_n \right) = R$, there exists one block code with size $M_n$ and $P_e |_{m (\mathcal{C} \sim M_n)} \leq 2^{1/\rho E} \left[ P_e |_{m (\mathcal{C} \sim 2M_n)} \right]$.

(9.2.1)

Proof: Randomly draw $2^{M_n}$ codewords according to some input distribution $\tilde{P}_X^n$, and let the codebook be denoted by $\{c_1, c_2, \ldots, c_{2^{M_n}}\} = \mathcal{C}$.

Let $\phi(\cdot)$ be the indicator function of the set

\[ \{ \left[ \left( \frac{1}{n} \log_2 M_n \right) \mathcal{Z} \right]_{d/1} \geq t : \mathcal{X} \subset \mathcal{Y} \} \]

and let the codebook be denoted by $\mathcal{X}_d$. Lemma 9.7 (existence of "good" code for expurgated exponent)
By Markov's inequality, 

\[ E \left( \sum_{1 \leq m \leq 2^M_n} \phi(P_{\rho} | m (C \sim 2^M_n)) \right) \geq 2^{M_n} \]

Therefore, there exist at least one codebook such that

\[ u^W \geq \left( \binom{u^M_n}{\phi} + d \right) \phi \sum_{u^N_2 \geq u^1} u^W \]

By selecting \( M_n \) codewords from this codebook with \( \phi(P_{\rho} | m (C \sim 2^M_n)) = 1 \), a new codebook is formed, and it is obvious that \( (9.2.1) \) holds for this new codebook.
Expurgated exponent
\((\rho^u|_f)^uX|_u\Lambda D = (\omega^u|_u\tilde{f})^uX|_u\lambda D\rangle \quad \text{for some } n \quad \text{and} \quad m, \quad \text{i.e.},
\{m \neq \, \forall m | (\rho^u|_f)^uX|_u\Lambda D < (\omega^u|_u\tilde{f})^uX|_u\lambda D : m\} = \, \forall u\}

Step 1: Maximum Likelihood decoder. Let \(X \in \{\mathcal{C}_1, \ldots, \mathcal{C}_n\} \) denote the set of \(n\)-tuple block codewords selected, and let the decoding partition.

Proof: Randomly select \(2\) \(M_n\) codewords according to some input distribution \(\tilde{P}_X\).

Theorem 9.9 (Expurgated exponent) For DMC with generic transition pro-

\[E(x^e) \leq E(R)\]

\[
E(x^e) \leq E(R) \quad \text{with generic transition pro-}
\]

Definition 9.8 (Expurgated exponent) The expurgated exponent \((R(x^e)E)^{x^e}\) is defined by

\[E(x^e) \quad \text{with generic distribution} \quad P_X|_X \quad \text{is defined by} \quad \text{The expurgated exponent}
\]
Step 2: Property of indicator function for $s = 1$.

Let $\phi_m$ be the indicator function of $U_m$. Then for all $s > 0$, 

$$1 - \phi_m(y_n) \leq \begin{cases} 
\sum_{m' \neq m} \left[ P_{Y_n|X_n}(y_n|c_{m'}) P_{Y_n|X_n}(y_n|c_m) \right] \frac{1}{1+s} 
\end{cases}$$

(Note that this step is the same as random coding exponent, except only $s = 1$ is considered. By taking $s = 1$, we have)

$$1 - \phi_m(y_n) \leq \begin{cases} 
\sum_{m' \neq m} \left[ P_{Y_n|X_n}(y_n|c_{m'}) P_{Y_n|X_n}(y_n|c_m) \right] \frac{1}{2} 
\end{cases}$$

Step 3: Probability of error given codeword $c_m$ is transmitted.

Let $P_{e|m}(C \sim 2^M_n)$ denote the probability of error of $c_m$ given codeword $c_m$ is transmitted. Let

$$\left\{ \left[ \frac{(u|\Phi_u)_{Y_n|X_n|d}}{(u|\Phi_u)_{Y_n|X_n|d}} \right]_{w \neq w'} \right\} \geq (u|\Phi_u)^{w \phi} - 1$$

function of $\mu_u$. Then for all $s < 0$, let $\phi_m^w$ be the indicator function for $s = 1$. Let be the indicator will be arbitrarily assigned to either $m$ or $m'$.
Step 4: Standard inequality for $s' = 1$. It is known that for any $\exists \theta \neq \text{max} d_{ij}$

$$\left\{ \left( \sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \right)^{u \in \Theta} \right\}$$

$$\sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \geq$$

$$\left\{ \left( \sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \right)^{u \in \Theta} \right\}$$

$$\sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \geq$$

$$\sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \geq$$

$$\left\{ \left( \sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \right)^{u \in \Theta} \right\}$$

$$\sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \geq$$

$$\sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \geq$$

$$\left\{ \left( \sum_{u \in \Theta} (w \mid f)^{uX|uA} d^{(w \mid f)^{uX|uA}} \right)^{u \in \Theta} \right\}$$

Then

Expurgated exponent
Now since $\frac{\partial f}{\partial \rho} = \Delta d f_{\rho}/(\Delta d f_{\rho})$ (which implies $0 < \rho < 1$) we have the desired result.

$$\frac{\partial f_{\rho}}{\partial \rho} = \Delta d f_{\rho}/(\Delta d f_{\rho}) = (d f)_{\rho}$$
Step 6: Lemma 9.7.

From Lemma 9.7, there exists one codebook with size $M$.

\[ E_{\sum_{i=1}^{M_n} \left| \sum_{Y_n \in Y} \sqrt{P_{Y_n}} \right|} = \]

\[ \left\{ \left( \sum_{u \in \hat{u}} \frac{P_{X|uX} \sum_{u \in \hat{u}} P_{X|uX}}{\sum_{u \in \hat{u}} P_{X|uX}} \right)^{u_{W \geq 1}} \right\} = \]

\[ \left\{ \left( \sum_{u \in \hat{u}} \frac{P_{X|uX} \sum_{u \in \hat{u}} P_{X|uX}}{\sum_{u \in \hat{u}} P_{X|uX}} \right)^{u_{W \geq 1}} \right\} = \]

\[ \sum_{1 \leq m \leq 2M_n} \left( \sum_{Y_n \in Y} \sqrt{P_{Y_n}} \right)^{u_{W \geq 1}} \]
such that

\[
\left( u(x)^{uX} d(uX)^{uX} \sum \sum \right)_{s/1}^{(uW)} =
\left( (u \in u(x))^{uX} d(uX)^{uX} \sum \sum \right)_{s/1}^{(uW)} =
\left( (u \in u(x))^{uX} d(uX)^{uX} \sum \sum \right)_{d/1}^{(uW)} =
\left( (u \in u(x))^{uX} d(uX)^{uX} \sum \sum \right)_{d/1}^{(uW)} >
\left( (uW \in \Theta)_{d/1}^{uW} d \right)_{d/1}^{(uW)} >
\left( (uW \in \Theta)_{d/1}^{uW} d \right)_{d/1}^{(uW)} >
\]

such that

**Expected exponent**
By using the fact that $P \tilde{X}^n$ and $PY^\parallel |X^n$ are product distributions with identical marginal, and taking logarithmic operation for both sides of the above inequality, we have desired result.
Definition 9.10 (expurgated exponent)

The expurgated exponent for DMC with generic distribution \( P | X \) is defined by

\[
\frac{E_{\text{ex}}(R)}{R}\triangleq \max_{s \geq 1} \left\{ -s R + E_{\text{ex}}(s) \right\},
\]

where

\[
E_{\text{ex}}(s) \triangleq \max_{x \in X} \left( x \log \frac{\sum_{x' \in X} P_{x'}(x) P_{x'}(x')}{\sum_{y \in Y} \sqrt{P_{x'}(y)} P_{x'}(y)} \right).
\]

Lemma 9.11 (properties of \( E_{\text{ex}}(R) \))

1. \( E_{\text{ex}}(R) \) is non-increasing.
2. \( E_{\text{ex}}(R) \) is convex in \( R \). (Note that the first two properties imply that \( E_{\text{ex}}(R) \) is strictly decreasing.)
3. There exists \( R_{\text{cr}} \) such that for \( R > R_{\text{cr}} \), the slope of \( E_{\text{ex}}(R) \) is \( -\frac{1}{s} \).

The properties of expurgated exponent

The expurgated exponent with generic distribution \( P | X \) is defined by

\[
E_{\text{ex}}(R) \triangleq \max_{x \in X} \left( x \log \frac{\sum_{x' \in X} P_{x'}(x) P_{x'}(x')}{\sum_{y \in Y} \sqrt{P_{x'}(y)} P_{x'}(y)} \right).
\]
Example 9.12

For BSC with crossover probability $0.2$ (over the range of $(0, 0.192745)$), the expurgated exponent becomes

$$E_{\text{ex}}(R) = \max_{1 \geq p \geq 0} \max_{s \geq 1} \left\{ -sR - s \log \left[ p^2 + 2p(1-p) \left( \frac{2}{\sqrt{\varepsilon(1-\varepsilon)}} \right)^{1/s} \right] + \left( 1-p \right)^2 \right\}$$

where $(p, 1-p)$ is the input distribution. Note that the input distribution achieving the expurgated exponent, as well as random coding exponent, is uniform, i.e., $d = 1/2$. The expurgated exponent, as well as random coding exponent, are depicted in Figures 9.4 and 9.5.

\[ \text{Expurgated exponent (solid line) and random coding exponent (dashed line) for BSC with crossover probability } 0.2. \]
Expurgated exponent (solid line) and random coding exponent (dashed line) for BSC with crossover probability 0.2 (over the range of \((0, 0.006]\)).
Partitioning upper bound for channel reliability

Upper bounds:

• partitioning bound
• sphere-packing bound
• straight-line bound

Keys:

Hypothesis testing

Model: the receiver end can be modeled as:

\[ H_0 : c_m^u \neq \text{codeword transmitted} \]
\[ H_1 : c_m^u = \text{codeword transmitted} \]

Type II error, which can be computed using the theory of binary hypothesis

The channel decoding error given that codeword \( m \) is transmitted becomes the

output:

where \( m \) is the final decision made by receiver upon receipt of some channel

\[ \frac{H}{H} \]

Examples:

Hypothesis testing
Definition 9.13 (partitioning bound)

For a DMC with marginal $P_Y|X$, $E_p(R)_{\triangleq \max P_X\min \{P_{\tilde{Y}}|X : I(P_X, P_{\tilde{Y}}|X) \leq R\}}$.

Observation 9.14 (partitioning bound)

If $P_{\tilde{X}}$ and $P_{\tilde{Y}}|X$ are distributions that achieve $E_p(R)$, and $P_{\tilde{Y}}(y) = \sum_{x \in X} P_{\tilde{X}}(x) P_{\tilde{Y}|X}(y|x)$, then $E_p(R)_{\triangleq \max P_X\min \{P_{\hat{Y}}|X : D(P_{\hat{Y}}|X \parallel P_{\tilde{Y}}|P_X) \leq R\}}$.

In addition, the distribution $P_{\hat{Y}}|X$ that achieves $E_p(R)$ is a tilted distribution between $P_{\tilde{Y}}|X$ and $P_{\tilde{Y}}$, i.e., $E_p(R)_{\triangleq \max P_X D(P_{\lambda Y}|X \parallel P_{\tilde{Y}}|P_X)}$, where $P_{\lambda Y}|X$ is a tilted distribution between $P_{\tilde{Y}}|X$ and $P_{\tilde{Y}}$, and $\lambda$ is the solution of $D(P_{\lambda Y}|X \parallel P_{\tilde{Y}}|P_X) = R$.

For a DMC with marginal $P_X$, $E_p(Upper bound for channel reliability)$.
Theorem 9.15 (partitioning bound)

For a DMC with marginal $P_{Y|X}$, for any $\varepsilon > 0$ arbitrarily small, there exists an $\mathcal{C}$ with $\mathcal{C} \supseteq \mathcal{H} \supseteq \mathcal{H}$, such that

$$[(s)0\mathcal{E} - \mathcal{H}s]^{0<s}_{\text{max}} \supseteq \mathcal{H}\mathcal{D} \supseteq [(s)0\mathcal{E} - \mathcal{H}s]^{1>s>0}_{\text{max}}$$

Hence, the channel reliability $\mathcal{D}$ satisfies

$$\max \left\{ \mathcal{H} \Big[ \frac{-sR - \mathcal{H}0(s)}{1+s} \Big] \mathcal{H} \right\} = \max \left\{ \mathcal{H} \Big[ \frac{-sR - \mathcal{H}0(s)}{1+s} \Big] \mathcal{H} \right\}$$

$\mathcal{D}$ is the slope of $\mathcal{H}$ times $-1$.

Recall that the random coding exponent is

$$\max \left\{ s \geq 0 \left\{ \frac{-sR - \mathcal{H}0(s)}{1+s} \right\} \right\}$$

For optimizer $s^* \in (0, 1]$, the upper bound meets the lower bound.

Lemma 9.16

For a DMC with marginal $P_{Y|X}$, for any $\varepsilon > 0$ arbitrarily small, there exists an $\mathcal{C}$ with $\mathcal{C} \supseteq \mathcal{H} \supseteq \mathcal{H}$, such that

$$\mathcal{D} \supseteq (\varepsilon + \mathcal{H})\mathcal{E}$$
Partitioning exponent (thick line), random coding exponent (thin line) and expurgated exponent (thin line) for BSC with crossover probability 0.2.
A ball or sphere $\bullet$

- A ball or sphere centered at $a$ with radius $r$ is defined as

$$\{x \in (a, q) : \forall x \in q\}$$

- A ball or sphere $\bullet$

- Sphere-packing upper bound

- Sphere-packing upper bound

- Sphere-packing upper bound

- Sphere-packing upper bound
• To find the best codebook which yields minimum decoding error is one of the main research issues in communications.

• Roughly speaking, if two codewords are similar, they should be more vulnerable to noise or interference.

• Hence, a good codebook should be a set of codewords which look very different from others.

• In mathematics, such „codeword resemblance“ can be modeled as a distance function.

• Accordingly, a good codebook becomes a set of codewords whose minimum distance among codewords is largest. This is exactly the sphere-packing problem.
Example 9.17 (Hamming distance versus BSC)

For BSC, the source alphabet and output alphabet are both \{0, 1\}. The Hamming distance between two elements in \{0, 1\} is given by

\[ d_H(x, \hat{x}) = \begin{cases} 0, & \text{if } x = \hat{x} \\ 1, & \text{if } x \neq \hat{x} \end{cases} \]

For all \( m \), its extension definition to \( n \)-tuple is

\[ d_H(x^n, \hat{x}^n) = \sum_{i=1}^{n} d_H(x_i, \hat{x}_i) \]

It is known that the best decoding rule is the maximum likelihood ratio decoder,

\[ \phi(y^n) = m, \text{ if } P_{Y^n|X^n}(y^n|c_m) \geq P_{Y^n|X^n}(y^n|c_m') \text{ for all } m' \]

Therefore, the best decoding rule can be rewritten as:

\[ (w\mathcal{C}, w\widehat{\mathcal{C}})H_p \mathcal{C} \quad \text{if} \quad w = (w\hat{t})\phi \]

\[ \frac{2-1}{2} u(2-1) = \]

\[ (w\mathcal{C}, w\widehat{\mathcal{C}})H_p - u(2-1)(w\mathcal{C}, w\widehat{\mathcal{C}})H_p \mathcal{C} = (w\mathcal{C}, w\widehat{\mathcal{C}})uX|uX \mathcal{D} \]

For BSC with crossover probability \( \varepsilon \),

\[ P_{Y^n|X^n}(y^n|c_m) = \varepsilon d_H(y^n, c_m) (1 - \varepsilon)^{n-d_H(y^n, c_m)} \]

(9.4.2)

Therefore, the best decoding rule can be rewritten as:

\[ \phi(y^n) = m, \text{ if } d_H(y^n, c_m) \geq d_H(y^n, c_m') \text{ for all } m' \]
As a result, if two codewords are too close in Hamming distance, the number of bits of outputs that can be used to classify its origin will be less, and therefore, will result in a poor performance.

From the above example, two observations can be made.

- First, if the distance measure between codewords can be written as a function of the transition probability of channel, such as (9.4.2), one may regard the probability of decoding error with the distances between codewords.
- Secondly, the coding problem in some cases can be reduced to a sphere-packing problem. As a consequence, solution of the sphere-packing problem can be used to characterize the error probability of channels.
Theorem 9.18

Let $\mu_n(x, y)$ be the Bhattacharya distance between two elements in $X^n$. Denote by $d_{n,M}$ the largest minimum distance among $M$ selected codewords of length $n$. (Obviously, the largest minimum radius among $M$ disjoint spheres in a code space is half of $d_{n,M}$.) Then

$$
\limsup_{n \to \infty} - \frac{1}{n} \log P_e(n, R) \leq \limsup_{n \to \infty} \frac{1}{n} d_{n,M} = e^{nR}.
$$

Since according the above theorem, the largest minimum distance can be used to formulate an upper bound on channel reliability, this quantity becomes essential. We therefore introduce its general formula in the next subsection.

\[ \mu(x^n, y^n) = \log \sum_{y^n \in Y^n} \sqrt{P_{Y^n|X^n}(y^n|x^n)} \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n|X^n}(y^n|\hat{x}^n)} \]
The shaded area is $\mathcal{U}_c \ominus \mathcal{M}$. The shaded area is $\mathcal{A}_c \ominus \mathcal{M}_m \ominus \mathcal{M}_m'$. 

Relation of sphere-packing and coding
The largest minimum distance of block codes

- If the size of the code alphabet, \( q \), is an even power of a prime, satisfying \( q \geq 49 \), and the distance measure is the Hamming distance, a better lower bound than the Varshamov-Gilbert bound can be obtained through the construction of the Algebraic-Geometric code, of which the idea was first proposed by Goppa.
- Later, Zinoviev and Litsyn proved that a better lower bound than the Varshamov-Gilbert bound is actually possible for any \( q \geq 46 \).

History: Varshamov-Gilbert bound, McEliece bound, Elias bound, MeEliece bound

The largest minimum distance of block codes
• The distance spectrum formula is

\[ (W^u X)^u p \quad \min_{1 - W^u \geq w > 0} \max_{W^u \rightarrow p} \]

where \( W^u \) is the \( u \)-tuple code alphabet and \( p \) is the probability distribution.

\[ \cdot \left( \begin{array}{c}
\left( W^u X \right)^u p \quad \min_{1 - W^u \geq w > 0} \max_{W^u \rightarrow p}
\end{array} \right) \]

is the largest minimum distance is

\[ \min_{1 - W^u \geq w > 0} \max_{W^u \rightarrow p} \]

\[ \left( W^u X \right)^u p \]

The minimum distance is

\[ \min_{1 - W^u \geq w > 0} \max_{W^u \rightarrow p} \]

\[ \left( W^u X \right)^u p \]

where \( \nabla \) is the function that takes the minimum over the specified range.

\[ \{ 1 - W^u, \ldots, W^u, 1, 0 \} = W^u \]

A codebook with block length \( n \) and size \( M \) is represented by \( W \) and each entry \( c \) belongs to \( \nabla \).

\[ \min_{1 - W^u \geq w > 0} \max_{W^u \rightarrow p} \]

\[ \left( W^u X \right)^u p \]

denotes the distance.

\[ \nabla \]

The \( u \)-tuple code alphabet is denoted by \( X^u \).

Notations

Distance-spectrum formula

9.38
• Problem:
\[
\lim_{n \to \infty} n^{-1} d_{n,M} = e^{-nR} \quad \text{for fixed } R.
\]

• Key:
Random coding.

Theorem 9.19 (distance-spectrum formula)
\[
\sup_{X \Lambda} X(R) \geq \limsup_{n \to \infty} d_{n,M} \geq \sup_{X \Lambda} X(R + \delta) \quad (9.4.3)
\]
and
\[
\sup_{X \Lambda} X(R) \geq \liminf_{n \to \infty} d_{n,M} \geq \sup_{X \Lambda} X(R + \delta) \quad (9.4.4)
\]
for every \( \delta > 0 \), where
\[
\bar{\Lambda}_X(R) \triangleq \inf \left\{ a \in \mathbb{R} : \limsup_{n \to \infty} \left( \Pr \left\{ 1/n \mu_n(\hat{X}_n, X_n) > a \right\} \exp \{ nR \} \right) \right\}
\]
and
\[
\Lambda_X(R) \triangleq \inf \left\{ a \in \mathbb{R} : \liminf_{n \to \infty} \left( \Pr \left\{ 1/n \mu_n(\hat{X}_n, X_n) > a \right\} \exp \{ nR \} \right) \right\}
\]

Theorem 9.19 (distance-spectrum formula)

Key: Random coding.

For fixed \( \mathcal{H} \)

\[
y u = N \cdot u \quad \lim_{n \to \infty} \]

Problem:

Distance-spectrum formula
\[
\{ Y \mid \inf \sup \exists \{ u \mid \exists \}
\left( \left[ \left( u X \mid \hat{\omega} \right) u X \mid u \hat{\omega} \right] \underbar{z/1} D (u \mid \hat{\omega}) u X \mid u \hat{\omega} \right] \underbar{z/1} D \underbar{\sum} \underbar{\sum} \left( \left( u \mid \hat{\omega} \right) u X \mid u \hat{\omega} \right] \underbar{z/1} D \underbar{\sum} \underbar{\sum} \left( \left( u \mid \hat{\omega} \right) u X \mid u \hat{\omega} \right] \underbar{z/1} D \right)
\right)^{\infty \leftarrow u} \lim \sup : \alpha \begin{array}{c}
\left. \right|_{\sup \inf}^{\left. \right|_{\sup \inf}}
\end{array}
\} \left. \sup \inf \right|_{\sup \inf}^{\left. \sup \inf \right|_{\sup \inf}} \left( (Y) \right) \left( (Y) \right)
\]

Distance Spectral Formula
\[ \{ I = [z < Z] : z \text{ ess inf} \} \{ \nu \} = \limsup_{n \to \infty} \inf \{ \hat{X}^{(n)} : \hat{X}^{(n)} = X^{(n)} \} \]

\[ \{ \nu > (uX, uX)^{\infty} \} \{ \nu \} \limsup_{n \to \infty} \inf \{ \hat{X}^{(n)} : \hat{X}^{(n)} < \infty \} \]

where

\[ \nu_Y = \nu_Y \{
\begin{array}{ll}
\nu_Y > \nu, & \nu_Y > (X)^{\nu} \nu_Y \\
\nu_Y = \nu, & \nu_Y = (X)^{\nu} \nu_Y \\
\nu_Y < \nu, & \nu_Y < (X)^{\nu} \nu_Y
\end{array}
\]

Properties of distance-spectrum function
Properties of distance-spectrum function

General curve of $\Lambda_X(R)$.
Example 9.20

\[ X = \{ 0, 1 \} \]

\( \mu_n(\cdot, \cdot) \) is additive with marginal distance metric

\[ \mu(0, 0) = \mu(1, 1) = 0, \quad \mu(0, 1) = 1, \quad \mu(1, 0) = \infty \]

\( \bar{\Lambda}_X(R) = \sup_s \left\{ -sR - s \cdot \log_2 e^{-1/s} \right\} \]

\( \bar{R}_0(X) = -\log \Pr\{\hat{X} = X\} = \log 2 \)

\( \bar{D}_0(X) = \text{ess inf} \mu(\hat{X}, X) = 0 \)

\( \bar{R}_p(X) = -\log \Pr\{\mu(\hat{X}, X) < \infty\} = \log 4/3 \)

\( \bar{D}_p(X) = \mathbb{E}[\mu(\hat{X}, X)|\mu(\hat{X}, X) < \infty] = 1/3 \)

\( E = (0, 0)^n, \quad 0 = (1, 1)^n, \quad 0 = (0, 0)^n \)

\( \text{and is additive with marginal distance metric} \)

\( \{ I, 0 \} = X \)

Properties of distance-spectrum function
Properties of distance-spectrum function

\[ \left\{ \left[ \frac{4}{s - 1 - e + z} \right] \log s - \log - \right\}_{0<s} \]

Open questions: under what conditions is the Varshamov-Gilbert lower bound tight?
Lemma 9.21 (large deviation formulas for $\overline{\Lambda}X(R)$ and $\Lambda X(R)$)

\[ \overline{\Lambda}X(R) = \inf \left\{ a \in \mathbb{R} : \overline{\ell}X(a) < R \right\} \]

and

\[ \Lambda X(R) = \inf \left\{ a \in \mathbb{R} : \ell X(a) < R \right\} \]

where $\overline{\ell}X(a) \triangleq \limsup_{n \to \infty} -1/n \log \Pr \left\{ \frac{1}{n} \mu_n(\hat{X}_n, X_n) \leq a \right\}$ and $\ell X(a) \triangleq \liminf_{n \to \infty} -1/n \log \Pr \left\{ \frac{1}{n} \mu_n(\hat{X}_n, X_n) \leq a \right\}$.

\[
\left\{ v \geq (uX, uX)^{ur/l} \right\} \sim \mathcal{D} \mathcal{S} \{ u \overset{\infty}{\underset{l}{\to}} \inf \nabla \equiv (v)X \}
\]

and

\[
\left\{ v \geq (uX, uX)^{ur/l} \right\} \sim \mathcal{D} \mathcal{S} \{ u \overset{\infty}{\underset{l}{\to}} \inf \nabla \equiv (v)X \}
\]

where $\mathcal{D} \mathcal{S}$ denotes the sup- and the inf-large deviation spec,

\[
\{ \mathcal{H} > (v)X : \mathcal{H} \ni v \} \equiv (\mathcal{H})X \]

and

\[
\{ \mathcal{H} > (v)X : \mathcal{H} \ni v \} \equiv (\mathcal{H})X
\]

Lemma 9.21 (large deviation formulas for $\mathcal{H}X \nabla$ and $\mathcal{H}X\nabla$)

General Varshamov-Gilbert lower bound
\[
\sup_{X} \bar{\Lambda}(R) \geq \sup_{X} \bar{G}(R) \quad \text{and} \quad \sup_{X} \Lambda(R) \geq \sup_{X} G(R)
\]

where
\[
\bar{G}(R) \triangleq \sup_{s > 0} \left[ -sR - s \cdot \bar{\phi}(\theta) \right]
\]

(9.4.6)
\[
(s/I - \mathcal{X} \bar{\phi} \cdot s - \mathcal{H}s - \mathcal{Y})^{0<s} \sup_{X} \mathcal{G} = (\mathcal{H})^{X \mathcal{G}}
\]

(9.4.6)
\[
(s/I - \mathcal{X} \bar{\phi} \cdot s - \mathcal{H}s - \mathcal{Y})^{0<s} \sup_{X} \mathcal{G} = (\mathcal{H})^{X \mathcal{G}}
\]

where
\[
(\mathcal{H})^{X \mathcal{G}} \sup_{X} \mathcal{G} \geq (\mathcal{H})^{X \mathcal{V}} \sup_{X} \mathcal{V} \quad \text{and} \quad (\mathcal{H})^{X \mathcal{G}} \sup_{X} \mathcal{G} \geq (\mathcal{H})^{X \mathcal{V}} \sup_{X} \mathcal{V}
\]

Corollary 9.22 (Varshamov-Gilbert bound)

General Varshamov-Gilbert Lower Bound
It is conjectured that the channel reliability function is convex.

Therefore, any non-convex upper bound can be improved by making it convex.

This is the main idea of the straight line bound.

Therefore, any non-convex upper bound can be improved by making it convex.

It is conjectured that the channel reliability function is convex.

---

**Key:**

**Definition 9.23 (List Decoder)**

A list decoder decodes the outputs of a noisy channel by a list of candidates (possible inputs), and an error occurs only when the correct codeword transmitted is not in the list.

**Definition 9.24 (Maximal Error Probability for List Decoder)**

$P_{e,\text{max}}(n, M, L)$

$\equiv \min_{\mathcal{L} \text{ with } L \text{ candidates for decoder}} \max_{1 \leq m \leq M} \{ C \sim \text{with } \mathcal{L} \text{ for decoder} \}$

$P_e(m, n)$,

where $n$ is the blocklength and $M$ is the code size.

---

**Definition 9.25 (Average Error Probability for List Decoder)**

$P_e(n, M, L)$

$\equiv \min_{\mathcal{L} \text{ with } L \text{ candidates for decoder}} \frac{1}{M} \sum_{1 \leq m \leq M} \{ C \sim \text{with } \mathcal{L} \text{ for decoder} \}$

$P_e(m, n)$,

where $n$ is the blocklength and $M$ is the code size.
Lemma 9.26 (lower bound on average error)

For DMC,

\[ P_e(n, M, 1) \geq P_e(n_1, M, L, 1) \]

\[ P_{e, max}(n_2, L + 1, 1) \]

where \( n = n_1 + n_2 \).

Theorem 9.27 (straight-line bound)

\[ E(\lambda R_1 + (1 - \lambda) R_2) \leq \lambda E(R_1) + (1 - \lambda) E(R_2) \]

Proof: By applying the previous lemma with \( \lambda = \frac{n_1}{n} \), \( \log \left( \frac{M}{L} \right) \frac{1}{n_1} = R_1 \) and \( \log \frac{1}{u} = R_2 \), we can upper bound the sphere-packing exponent, and by partitioning exponent, and \( P_e(n, M, 1) \) by sphere-packing exponent, and

\[ \mathcal{E} \left( \lambda - 1 \right) + \mathcal{E} \left( \frac{1}{n_1} \right) \geq \mathcal{E} \left( \lambda - 1 \right) + \mathcal{E} \left( \frac{1}{n_1} \right) \]

Theorem 9.27 (straight-line bound)

\[ \mathcal{E} \left( R_1 \right) \leq \mathcal{E} \left( R_2 \right) \]

where \( \mathcal{E} \left( R_1 \right) = \mathcal{E} \left( R_2 \right) \)

\[ (1 + \lambda) \mathcal{E}(T W, 1) \geq (1, W, 1) \mathcal{E}(T W, 1) \]

Lemma 9.26 (lower bound on average error) For DMC, Straight-line bound