Chapter 6

Optimistic Shannon Coding Theorems for Arbitrary Single-User Systems

Po-Ning Chen

Department of Communications Engineering

National Chiao-Tung University

Hsin Chu, Taiwan 30050

Motivations

- The conventional definition of the minimum achievable fixed-length source coding rate $T(\boldsymbol{X})$ (or $T_0(\boldsymbol{X})$) for a source $\boldsymbol{X}$ (cf. Definition 3.2) requires the existence of reliable source codes for *all sufficiently large blocklengths*.
- Alternatively, if it is required that reliable codes exist for *infinitely many blocklengths*, a new, more *optimistic* definition of source coding rate (denoted by $\underline{T}(\boldsymbol{X})$) is obtained.
- Similarly, the *optimistic capacity* $\bar{C}$ is defined by requiring the existence of reliable channel codes for infinitely many blocklengths, as opposed to the definition of the conventional channel capacity C.

Why introducing Optimistic quantities?

- This concept of optimistic source coding rate and capacity has recently been investigated by Verdú *et.al* for *arbitrary* (not necessarily stationary, ergodic, information stable, etc.) sources and single-user channels.
- More specifically, they establish an additional *operational* characterization for the optimistic minimum achievable source coding rate ($\underline{T}(\boldsymbol{X})$ for source $\boldsymbol{X}$) by demonstrating that for a given channel, the classical statement of the source-channel separation theorem holds for every channel if $\underline{T}(\boldsymbol{X}) = T(\boldsymbol{X})$.

  (Classical statement of source-channel separation theorem) Given a source $\boldsymbol{X}$ with (conventional) source coding rate $T(\boldsymbol{X})$ and channel $\boldsymbol{W}$ with capacity C, then $\boldsymbol{X}$ can be reliably transmitted over $\boldsymbol{W}$ if $T(\boldsymbol{X}) < C$. Conversely, if $T(\boldsymbol{X}) > C$, then $\boldsymbol{X}$ cannot be reliably transmitted over $\boldsymbol{W}$.

  (Reliable transmission) There exits a sequence of joint source-channel codes such that the decoding error probability vanishes as the blocklength $n \to \infty$.
- In a dual fashion, they also show that for channels with $\bar{C} = C$, the classical separation theorem holds for every source.

Outlines

- In this chapter, we derive the general formula for $\underline{T}(\boldsymbol{X})$ and $\bar{C}$.
- The key to these results is the application of the generalized sup-information rate introduced in Chapter 2 to the existing proofs by Verdú and Han of the direct and converse parts of the conventional coding theorems.
- General expressions for the optimistic minimum ε-achievable source coding rate and the optimistic ε-capacity are also provided.

Optimistic source coding theorems

- Key: *two new bounds* due to Han on the error probability of a source code as a function of its size.

- Interestingly, these bounds constitute the natural counterparts of the upper bound provided by Feinstein's Lemma and the Verdú-Han lower bound to the error probability of a channel code.

- With this result, the formula for $\underline{T}(\boldsymbol{X})$ for information station sources reduces to

$$\underline{T}(\boldsymbol{X}) = \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

- This is in contrast to the expression for $T(\boldsymbol{X})$, which is known to be

$$T(\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} H(X^n).$$

- (Summary) The above result leads us to observe that for sources that are both stationary and information stable, the classical separation theorem is valid for every channel.

- Note that Vembu *et.al* demonstrate that for a given source $\boldsymbol{X}$, the separation theorem holds for every channel if its optimistic minimum achievable source coding rate ($\underline{T}(\boldsymbol{X})$) coincides with its conventional (or pessimistic) minimum achievable source coding rate ($T(\boldsymbol{X})$); i.e., if $\underline{T}(\boldsymbol{X}) = T(\boldsymbol{X})$.

Optimistic source coding theorems

Definition 6.1 An (n, M) fixed-length source code for X^n is a collection of M n-tuples $\mathscr{C}_n = \{c_1^n, \ldots, c_M^n\}$. The error probability of the code is

$$P_e(\mathscr{C}_n) \stackrel{\triangle}{=} Pr\left[X^n \notin \mathscr{C}_n\right].$$

Definition 6.2 (optimistic ε-achievable source coding rate) Fix $0 < \varepsilon < 1$. $R \geq 0$ is an optimistic ε-achievable rate if, for every $\gamma > 0$, there exists a sequence of (n, M_n) fixed-length source codes $\mathscr{C}_n$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R$$

and

$$\liminf_{n\to\infty} P_e(\mathscr{C}_n) \leq \varepsilon.$$

The infimum of all ε-achievable source coding rates for source $\boldsymbol{X}$ is denoted by $\underline{T}_\varepsilon(\boldsymbol{X})$. Also define $\underline{T}(\boldsymbol{X}) \stackrel{\triangle}{=} \sup_{0<\varepsilon<1} \underline{T}_\varepsilon(\boldsymbol{X}) = \lim_{\varepsilon\downarrow 0} \underline{T}_\varepsilon(\boldsymbol{X}) = \underline{T}_0(\boldsymbol{X})$ as the optimistic source coding rate.

Optimistic source coding theorems

Lemma 3.3 Fix a positive integer n. There exists an (n, M) source block code $\mathscr{C}_n$ for P_{X^n} such that its error probability satisfies

$$P_e(\mathscr{C}_n) \le Pr\left[\frac{1}{n}h_{X^n}(X^n) > \frac{1}{n}\log M\right].$$

Lemma 3.4 Every (n, M) source block code $\mathscr{C}_n$ for P_{X^n} satisfies

$$P_e(\mathscr{C}_n) \ge Pr\left[\frac{1}{n}h_{X^n}(X^n) > \frac{1}{n}\log M + \gamma\right] - \exp\{-n\gamma\},$$

for every $\gamma > 0$.

Optimistic source coding theorems

We can then use Lemmas 3.4 and 3.4 (in a similar fashion to the general source coding theorems) to prove the *general* optimistic (fixed-length) source coding theorems.

Theorem 6.3 (optimistic minimum ε-achievable source coding rate formula) For any source $\boldsymbol{X}$,

$$\underline{T}_\varepsilon(\boldsymbol{X}) \le \begin{cases} \lim_{\delta \uparrow (1-\varepsilon)} \underline{H}_\delta(\boldsymbol{X}), & \text{for } \varepsilon \in [0,1); \\ 0, & \text{for } \varepsilon = 1. \end{cases}$$

Information stable source

- It is already known for an information stable source $\boldsymbol{X}$,

$$\bar{T}(\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} H(X^n).$$

- We herein prove a parallel expression for $\underline{T}(\boldsymbol{X})$.

Definition 6.4 (information stable sources) A source $\boldsymbol{X}$ is said to be information stable if $H(X^n) > 0$ for n sufficiently large, and $h_{X^n}(X^n)/H(X^n)$ converges in probability to one as $n \to \infty$, i.e.,

$$\limsup_{n\to\infty} Pr\left[\left|\frac{h_{X^n}(X^n)}{H(X^n)} - 1\right| > \gamma\right] = 0 \qquad \forall\gamma > 0,$$

where $H(X^n) = E[h_{X^n}(X^n)]$ is the entropy of X^n.

Lemma 6.5 Every information source $\boldsymbol{X}$ satisfies

$$\underline{T}(\boldsymbol{X}) = \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

Information stable source

Observations:

- If the source $\boldsymbol{X}$ is *both* information stable *and* stationary, the above Lemma yields

$$T(\boldsymbol{X}) = \underline{T}(\boldsymbol{X}) = \lim_{n\to\infty} \frac{1}{n} H(X^n).$$

  This implies that given a stationary and information stable source $\boldsymbol{X}$, the classical separation theorem holds for every channel.

Optimistic channel coding theorems

- In this section, we state without proving the general expressions for the optimistic ε-capacity ($\bar{C}_\varepsilon$) and for the optimistic capacity ($\bar{C}$) of arbitrary single-user channels.

- **(Recall)** The general expressions of the conventional channel capacity

$$C = \sup_{\boldsymbol{X}} \underline{I}_0(\boldsymbol{X};\boldsymbol{Y}) = \sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}),$$

  and the conventional ε-capacity

$$\sup_{\boldsymbol{X}} \lim_{\delta \uparrow \varepsilon} \underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \le C_\varepsilon \le \sup_{\boldsymbol{X}} \underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}).$$

Definition 6.6 (optimistic ε-achievable rate) Fix $0 < \varepsilon < 1$. $R \ge 0$ is an optimistic ε-achievable rate if there exists a sequence of $\mathscr{C}_n = (n, M_n)$ channel block codes such that

$$\liminf_{n\to\infty} \frac{1}{n} \log M_n \ge R$$

and

$$\liminf_{n\to\infty} P_e(\mathscr{C}_n) \le \varepsilon.$$

Definition 6.7 (optimistic ε-capacity $\bar{C}_\varepsilon$) Fix $0 < \varepsilon < 1$. The supremum of optimistic ε-achievable rates is called the optimistic ε-capacity, $\bar{C}_\varepsilon$.

It is straightforward for the definition that C_ε is non-decreasing in ε, and $C_1 = \log|\mathcal{X}|$.

Definition 6.8 (optimistic capacity $\bar{C}$) The optimistic channel capacity $\bar{C}$ is defined as the supremum of the rates that are ε-achievable for all $\varepsilon \in [0, 1]$. It follows immediately from the definition that $C = \inf_{0\le\varepsilon\le 1} C_\varepsilon = \lim_{\varepsilon\downarrow 0} C_\varepsilon = C_0$ and that C is the supremum of all the rates R for which there exists a sequence of $\mathscr{C}_n = (n, M_n)$ channel block codes such that

$$\liminf_{n\to\infty} \frac{1}{n} \log M_n \ge R,$$

and

$$\liminf_{n\to\infty} P_e(\mathscr{C}_n) = 0.$$

Theorem 6.9 (optimistic ε-capacity formula) Fix $0 < \varepsilon < 1$. The optimistic ε-capacity $\bar{C}_\varepsilon$ satisfies

$$\sup_{\boldsymbol{X}} \lim_{\delta\uparrow\varepsilon} \bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \le \bar{C}_\varepsilon \le \sup_{\boldsymbol{X}} \bar{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}). \tag{6.2.1}$$

Note that actually $\bar{C}_\varepsilon = \sup_{\boldsymbol{X}} \bar{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$, except possibly at the points of discontinuities of $\sup_{\boldsymbol{X}} \bar{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$ (which are countable).

Theorem 6.10 (optimistic capacity formula) The optimistic capacity $\bar{C}$ satisfies

$$\bar{C} = \sup_{\boldsymbol{X}} \bar{I}_0(\boldsymbol{X};\boldsymbol{Y}).$$

Information stable channels

- It is known that the capacity of information stable channels is

$$C = \liminf_{n\to\infty} C_n,$$

where

$$C_n \stackrel{\triangle}{=} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

Definition 6.11 (Information stable channels) A channel $\boldsymbol{W}$ is said to be information stable if there exists an input process $\boldsymbol{X}$ such that $0 < C_n < \infty$ for n sufficiently large, and

$$\limsup_{n\to\infty} Pr\left[\left|\frac{i_{X^nW^n}(X^n; Y^n)}{nC_n} - 1\right| > \gamma\right] = 0 \qquad \forall \gamma > 0.$$

Lemma 6.12 Every information stable channel $\boldsymbol{W}$ satisfies

$$\bar{C} = \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

Information stable channels

Observations:

- It is known that for DMC,
$$\bar{C} = C.$$
- The same result holds for $modulo - q$ additive noise channels with stationary ergodic noise.
- However, in general, $\bar{C} \geq C$ since $\bar{I}_0(\boldsymbol{X};\boldsymbol{Y}) \geq \underline{I}(\boldsymbol{X};\boldsymbol{Y})$.

Examples

- We provide four examples to illustrate the computation of C and $\bar{C}$.
- The first two examples present information stable channels for which $\bar{C} > C$.
- The third example shows an information unstable channel for which $\bar{C} = C$.
- These examples indicate that information stability is neither necessary nor sufficient to ensure that $\bar{C} = C$ or thereby the validity of the classical source-channel separation theorem.
- The last example illustrates the situation where $0 < C < \bar{C} < C_{SC} < \log_2 |\mathcal{Y}|$, where C_{SC} is the channel strong capacity.

  Definition. The strong (or strong converse) capacity C_{SC} is defined as the infimum of the numbers R for which for all (n, M_n) codes with $(1/n)\log M_n \geq R$, $\liminf_{n\to\infty} P_e(\mathcal{C}_n) = 1$. This definition of C_{SC} implies that for any sequence of (n, M_n) codes with $\liminf_{n\to\infty}(1/n)\log M_n > C_{SC}$, $P_e(\mathcal{C}_n) > 1-\varepsilon$ for every $\varepsilon > 0$ and for n sufficiently large.
- It is shown that $C_{SC} = \lim_{\varepsilon\uparrow 1} \bar{C}_\varepsilon = \sup_{\boldsymbol{X}} \bar{I}(\boldsymbol{X};\boldsymbol{Y})$.
- We assume in this section that all logarithms are in base 2 so that C and $\bar{C}$ are measured in bits.

Example 6.13

- Consider a nonstationary channel $\boldsymbol{W}$ such that at odd time instances $n = 1, 3, \cdots$, W^n is the product of the transition distribution of a binary symmetric channel with crossover probability 1/8 (BSC(1/8)), and at even time instances $n = 2, 4, 6, \cdots$, W^n is the product of the distribution of a BSC(1/4).
- It can be easily verified that this channel is information stable.
- Since the channel is symmetric, a Bernoulli(1/2) input achieves

$$C_n = \sup_{X^n} (1/n) I(X^n; Y^n);$$

thus

$$C_n = \begin{cases} 1 - h_b(1/8), & \text{for } n \text{ odd;} \\ 1 - h_b(1/4), & \text{for } n \text{ even,} \end{cases}$$

where $h_b(a) \stackrel{\triangle}{=} -a \log_2 a - (1-a)\log_2(1-a)$ is the binary entropy function.

- Therefore, $C = \liminf_{n\to\infty} C_n = 1 - h_b(1/4)$ and $\bar{C} = \limsup_{n\to\infty} C_n = 1 - h_b(1/8) > C$.

Information stable channels

Example 6.14

- Let $\mathcal{N}$ be the set of all positive integers.
- Define the set $\mathcal{J}$ as

$$\begin{aligned}\mathcal{J} &\triangleq \{n \in \mathcal{N} : 2^{2i+1} \le n < 2^{2i+2}, i = 0, 1, 2, \ldots\} \\ &= \{2, 3, 8, 9, 10, 11, 12, 13, 14, 15, 32, 33, \cdots, 63, 128, 129, \cdots, 255, \cdots\}.\end{aligned}$$

- Consider the following nonstationary symmetric channel $\boldsymbol{W}$. At times $n \in \mathcal{J}$, W_n is a BSC(0), whereas at times $n \notin \mathcal{J}$, W_n is a BSC(1/2). Put $W^n = W_1 \times W_2 \times \cdots \times W_n$.
- Here again C_n is achieved by a Bernoulli(1/2) input $\hat{X}^n$. We then obtain

$$C_n = \frac{1}{n}\sum_{i=1}^{n} I(\hat{X}_i; Y_i) = \frac{1}{n}\left[J(n)\cdot(1) + (n - J(n))\cdot(0)\right] = \frac{J(n)}{n},$$

where $J(n) \triangleq |\mathcal{J} \cap \{1, 2, \cdots, n\}|$.

Information stable channels

- It can be shown that

$$\frac{J(n)}{n} = \begin{cases} 1 - \frac{2}{3} \times \frac{2^{\lfloor \log_2 n \rfloor}}{n} + \frac{1}{3n}, & \text{for } \lfloor \log_2 n \rfloor \text{ odd;} \\ \frac{2}{3} \times \frac{2^{\lfloor \log_2 n \rfloor}}{n} - \frac{2}{3n}, & \text{for } \lfloor \log_2 n \rfloor \text{ even.} \end{cases}$$

- Consequently,

$$C = \liminf_{n\to\infty} C_n = \frac{1}{3} \quad \text{and} \quad \bar{C} = \limsup_{n\to\infty} C_n = \frac{2}{3}.$$

Information unstable channels

Example 6.15 *The Polya-contagion channel:*

- Consider a discrete additive channel with binary input and output alphabet $\{0, 1\}$ described by
$$Y_i = X_i \oplus Z_i, \quad i = 1, 2, \cdots,$$
where X_i, Y_i and Z_i are respectively the i-th input, i-th output and i-th noise, and $\oplus$ represents modulo-2 addition.

- Suppose that the input process is independent of the noise process.

- Also assume that the noise sequence $\{Z_n\}_{n\geq 1}$ is drawn according to the Polya contagion urn scheme as follows:
  - an urn originally contains R red balls and B black balls with $R < B$;
  - the *noise* just make successive draws from the urn;
  - after each draw, it returns to the urn $1 + \Delta$ balls of the same color as was just drawn ($\Delta > 0$).
  - The noise sequence $\{Z_i\}$ corresponds to the outcomes of the draws from the Polya urn: $Z_i = 1$ if ith ball drawn is red and $Z_i = 0$, otherwise.

- Let $\rho \stackrel{\triangle}{=} R/(R + B)$ and $\delta \stackrel{\triangle}{=} \Delta/(R + B)$. It is shown that the noise process $\{Z_i\}$ is stationary and nonergodic; thus the channel is information unstable.

Information unstable channels

- We then obtain

$$1 - \bar{H}_{1-\varepsilon}(\boldsymbol{X}) \le C_\varepsilon \le 1 - \lim_{\delta \uparrow (1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}),$$

and

$$1 - \underline{H}_{1-\varepsilon}(\boldsymbol{X}) \le \bar{C}_\varepsilon \le 1 - \lim_{\delta \uparrow (1-\varepsilon)} \underline{H}_\delta(\boldsymbol{X}).$$

- It has been shown that

$$-(1/n)\log P_{X^n}(X^n) \to V \stackrel{\triangle}{=} h_b(U) \quad \text{in distribution,}$$

where U is beta-distributed with parameters $(\rho/\delta, (1-\rho)/\delta)$, and $h_b(\cdot)$ is the binary entropy function.

- Thus

$$\bar{H}_{1-\varepsilon}(\boldsymbol{X}) = \lim_{\delta \uparrow (1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) = \underline{H}_{1-\varepsilon}(\boldsymbol{X}) = \lim_{\delta \uparrow (1-\varepsilon)} \underline{H}_\delta(\boldsymbol{X}) = F_V^{-1}(1-\varepsilon),$$

where $F_V(a) \stackrel{\triangle}{=} Pr\{V \le a\}$ is the cdf of V, and $F_V^{-1}(\cdot)$ is its inverse.

- Consequently,

$$C_\varepsilon = \bar{C}_\varepsilon = 1 - F_V^{-1}(1-\varepsilon),$$

and

$$C = \bar{C} = \lim_{\varepsilon \downarrow 0} \left[1 - F_V^{-1}(1-\varepsilon)\right] = 0.$$

Information unstable channels

Example 6.16

- Let $\tilde{W}_1, \tilde{W}_2, \ldots$ consist of the channel in Example 6.14, and let $\hat{W}_1, \hat{W}_2, \ldots$ consist of the channel in Example 6.15. Define a new channel $\boldsymbol{W}$ as follows:

$$W_{2i} = \tilde{W}_i \quad \text{and} \quad W_{2i-1} = \hat{W}_i \quad \text{for } i = 1, 2, \cdots .$$

- As in the previous examples, the channel is symmetric, and a Bernoulli(1/2) input maximizes the inf/sup information rates.

Information unstable channels

- Therefore for a Bernoulli(1/2) input $\boldsymbol{X}$, we have

$$
\begin{aligned}
&Pr\left\{\frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)}\leq\theta\right\}\\
&=\begin{cases}
Pr\left\{\frac{1}{2i}\left[\log\frac{P_{\tilde{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)}+\log\frac{P_{\hat{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)}\right]\leq\theta\right\}, & \text{if } n=2i;\\
Pr\left\{\frac{1}{2i+1}\left[\log\frac{P_{\tilde{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)}+\log\frac{P_{\hat{W}^{i+1}}(Y^{i+1}|X^{i+1})}{P_{Y^{i+1}}(Y^{i+1})}\right]\leq\theta\right\}, & \text{if } n=2i+1;
\end{cases}\\
&=\begin{cases}
1-Pr\left\{-\frac{1}{i}\log P_{Z^i}(Z^i)<1-2\theta+\frac{J(i)}{i}\right\}, & \text{if } n=2i;\\
1-Pr\left\{-\frac{1}{i+1}\log P_{Z^{i+1}}(Z^{i+1})<1-\left(2-\frac{1}{i+1}\right)\theta+\frac{J(i)}{i+1}\right\}, & \text{if } n=2i+1.
\end{cases}
\end{aligned}
$$

- The fact that

$$-(1/i)\log[P_{Z^i}(Z^i)] \to V \stackrel{\triangle}{=} h_b(U) \quad \text{in distribution,}$$

where U is beta-distributed with parameters $(\rho/\delta, (1-\rho)/\delta)$, and the fact that

$$\liminf_{n\to\infty} \frac{1}{n} J(n) = \frac{1}{3} \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n} J(n) = \frac{2}{3}$$

imply that

$$\underline{i}(\theta) \stackrel{\triangle}{=} \liminf_{n\to\infty} Pr\left\{ \frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta \right\} = 1 - F_V\left(\frac{5}{3} - 2\theta\right),$$

and

$$\bar{i}(\theta) \stackrel{\triangle}{=} \limsup_{n\to\infty} Pr\left\{ \frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta \right\} = 1 - F_V\left(\frac{4}{3} - 2\theta\right).$$

- Consequently,

$$\bar{C}_\varepsilon = \frac{5}{6} - \frac{1}{2}F_V^{-1}(1-\varepsilon) \qquad \text{and} \qquad C_\varepsilon = \frac{2}{3} - \frac{1}{2}F_V^{-1}(1-\varepsilon).$$

- Thus

$$0 < C = \frac{1}{6} < \bar{C} = \frac{1}{3} < C_{SC} = \frac{5}{6} < \log_2|\mathcal{Y}| = 1.$$