PAPER *Special Section on Information Theory and Its Applications*

Channel Capacity with Cost Constraint Allowing Cost Overrun**

Masaki HORI†*, *Nonmember and* **Mikihiko NISHIARA**†a), *Senior Member*

SUMMARY A channel coding problem with cost constraint for general channels is considered. Verdú and Han derived ε-capacity for general channels. Following the same lines of its proof, we can also derive ε-capacity with cost constraint. In this paper, we derive a formula for ε-capacity with cost constraint allowing overrun. In order to prove this theorem, a new variation of Feinstein's lemma is applied to select codewords satisfying cost constraint and codewords not satisfying cost constraint.

key words: *general channel, cost constraint, ε-achievability, Feinstein's lemma*

1. Introduction

Fundamental problems of channel coding for general channels were solved by Verdú and Han [1]. They derived formulas for channel capacity and ε-capacity. In addition to fundamental problems, there are some situations such that cost of codewords must be taken into account. Cost of a codeword is an abstract number representing such as the energy or the time spent to send it. As a mathematical model, cost is a real-valued function defined over the collection of channel inputs. A basic constraint is to limit the cost per symbol of every codeword to a certain constant [2]. Being more flexible with the cost constraint can potentially improve the efficiency of communication, and there are situations where it may not be practical to strictly adhere to the cost constraint. This is analogous to being tolerant of decoding errors in channel coding problems, which is referred to as ε-capacity. In this paper, we allow a certain probability of violation of cost constraint. This means that we have to spend much cost than the constraint for some codewords.

2. Formulation and Main Theorem

Through a mathematical formulation of the problem, we describe known results and some extensions we deal with.

Manuscript received February 25, 2023.
Manuscript revised June 26, 2023.
Manuscript publicized October 10, 2023.

†The authors are with Shinshu University, Nagano-shi, 380-8553 Japan.
*Presently, the author is with SCSK Corporation, Tokyo, 135-8110 Japan.

**This work was presented at the International Symposium on Information Theory and Its Applications 2022 (ISITA2022), Tsukuba, Japan, Oct. 17–19, 2022. This work was supported by JSPS KAKENHI Grant Number JP23K03851.

a) E-mail: mikihiko@shinshu-u.ac.jp

DOI: 10.1587/transfun.2023TAP0010

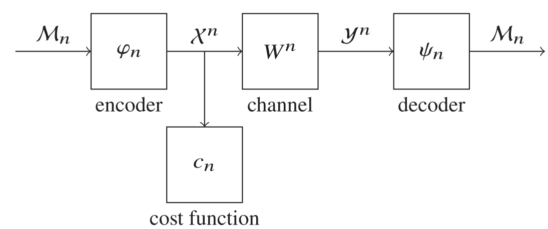


Fig. 1 Schematic diagram of our system.

2.1 Preliminary

Let $\mathcal{X}$ and $\mathcal{Y}$ be two abstract sets***. Consider general channel $\boldsymbol{W} \triangleq \{W^n\}_{n=1}^{\infty}$ with $\mathcal{X}$ and $\mathcal{Y}$ as input and output alphabet, respectively. This means that $W^n(\cdot|\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}^n$ is a distribution over $\mathcal{Y}^n$. We define a real-valued function c_n on the input alphabet $\mathcal{X}^n$ and call it the cost function. For $\boldsymbol{x} \in \mathcal{X}^n$, $c_n(\boldsymbol{x})$ is called the cost of $\boldsymbol{x}$.

We want to inform the destination of one of M_n messages through this channel. Let $\mathcal{M}_n \triangleq \{1, \ldots, M_n\}$ denote the set of messages. A selected message $m \in \mathcal{M}_n$ is encoded by the encoder $\varphi_n : \mathcal{M}_n \to \mathcal{X}^n$ into codeword $\varphi_n(m)$, which is fed into the channel. Observing the output of the channel, the decoder $\psi_n : \mathcal{Y}^n \to \mathcal{M}_n$ guesses the message selected at the encoder side. The guess of the decoder is not necessarily correct. The probability that the guess differs from the selected message is called the error probability and represented by

$$\varepsilon_n \triangleq 1 - \frac{1}{M_n} \sum_{m \in \mathcal{M}_n} W^n(\psi_n^{-1}(m)|\varphi_n(m)), \tag{1}$$

where $\psi_n^{-1}(m) \triangleq \{\boldsymbol{y} \in \mathcal{Y}^n | \psi_n(\boldsymbol{y}) = m\}$, which is called the decoding region for message m or codeword $\varphi_n(m)$. A pair (φ_n, ψ_n) of encoder φ_n and decoder ψ_n is called a code. Figure 1 depicts a schematic diagram of our system.

Let $\boldsymbol{X} \triangleq \{X^n\}_{n=1}^{\infty}$ and $\boldsymbol{Y} \triangleq \{Y^n\}_{n=1}^{\infty}$. Generally, given that $\boldsymbol{X}$ is fed into the channel, the output is denoted by $\boldsymbol{Y}(\boldsymbol{X})$. For each n, given that X^n is fed into the channel, the output is denoted by $Y^n(X^n)$. That is, we define

$$P_{Y^n(X^n)}(\boldsymbol{y}) \triangleq \sum_{\boldsymbol{x} \in \mathcal{X}^n} P_{X^n}(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}), \quad \boldsymbol{y} \in \mathcal{Y}^n. \tag{2}$$

***We describe the text as if these sets are countable. Use the Radon-Nikodym derivative and the integral in place of W^n/P_{Y^n} and the summation $\sum$, respectively, if necessary. For details, see [2], [3].

Every logarithm in this paper is the natural logarithm. Generally, the distribution of a random variable Z is denoted by P_Z.

2.2 Channel Capacity with Cost Constraint

We formulate a standard problem of channel coding with cost constraint for general channels.

Definition 1: For $\varepsilon \geq 0$, rate R is said to be (ε, Γ)-achievable if there exists a sequence of codes $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ satisfying

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon, \tag{3}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log M_n \geq R, \tag{4}$$

$$\frac{1}{n} c_n(\varphi_n(m)) \leq \Gamma \quad \text{for all } m \in \mathcal{M}_n \text{ and all } n. \tag{5}$$

Each inequality, namely (3), (4), and (5), represents the error constraint, the coding rate constraint, and the cost constraint, respectively. Especially, (5) requires that every codeword does not exceed the cost constraint.

Definition 2: We define (ε, Γ)-capacity as

$$C(\varepsilon, \Gamma) \triangleq \sup\{R | R \text{ is } (\varepsilon, \Gamma)\text{-achievable}\}. \tag{6}$$

In order to find a formula for (ε, Γ)-capacity, we need the following quantity.

Definition 3: For channel input $\boldsymbol{X}$ and output $\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{X})$, we define

$$\underline{I}_{\varepsilon}(\boldsymbol{X}; \boldsymbol{Y}) \triangleq \sup\left\{ R \,\middle|\, \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} < R \right\} \leq \varepsilon \right\}. \tag{7}$$

The following theorem is known [2].

Theorem 1: (ε, Γ)-capacity is given by

$$C(\varepsilon, \Gamma) = \sup_{\boldsymbol{X}: \Pr\{n^{-1} c_n(X^n) \leq \Gamma\} = 1} \underline{I}_{\varepsilon}(\boldsymbol{X}; \boldsymbol{Y}(\boldsymbol{X})), \tag{8}$$

where the supremum is taken with respect to all input processes $\boldsymbol{X}$ satisfying

$$\Pr\left\{ \frac{1}{n} c_n(X^n) \leq \Gamma \right\} = 1 \quad \text{for all } n. \tag{9}$$

2.3 Allowing Cost Overrun

Now, let U_{M_n} denote the random variable uniformly distributed over the set of natural integers not exceeding M_n. Note that (5) can be rewritten as

$$\Pr\left\{ \frac{1}{n} c_n(\varphi_n(U_{M_n})) \leq \Gamma \right\} = 1. \tag{10}$$

We generalize the definition of achievability at this viewpoint.

Definition 4: For $\beta \geq 0$, rate R is said to be $(\varepsilon, \beta, \Gamma)$-achievable if there exists a sequence of codes $\{(\varphi_n, \psi_n)\}_{n=1}^{\infty}$ satisfying (3), (4), and

$$\Pr\left\{ \frac{1}{n} c_n(\varphi_n(U_{M_n})) > \Gamma \right\} \leq \beta \quad \text{for all } n. \tag{11}$$

This constraint allows for the occurrence of cost overrun with a certain probability. For example, in the case of power constraint, it is generally necessary to adhere to the peak constraint, but some slight peak exceedance may be allowed by incorporating some form of protective circuitry. Inequality (11) mathematically represents such a situation.

Let us define a generalized channel capacity as follows.

Definition 5: $(\varepsilon, \beta, \Gamma)$-capacity is defined as

$$C(\varepsilon, \beta, \Gamma) \triangleq \sup\{R | R \text{ is } (\varepsilon, \beta, \Gamma)\text{-achievable}\}. \tag{12}$$

Then, we have the following theorem.

Theorem 2: $(\varepsilon, \beta, \Gamma)$-capacity is given by

$$C(\varepsilon, \beta, \Gamma) = \sup_{\boldsymbol{X}: \Pr\{n^{-1} c_n(X^n) > \Gamma\} \leq \beta} \underline{I}_{\varepsilon}(\boldsymbol{X}; \boldsymbol{Y}(\boldsymbol{X})). \tag{13}$$

If $\beta = 0$, of course, the situation is reduced to Theorem 1. The resulting formula for channel capacity is easy to suggest. However, we need new good codes achieving the capacity.

3. Code Construction Method

To prove Theorem 2, we prepare the following lemma, which is an extension of Feinstein's lemma [4]. This lemma is a key contribution of this paper.

Lemma 1: Suppose that a channel input X^n, a positive integer M_n, a real number $\gamma > 0$, and a subset $\mathcal{S}_n \subset \mathcal{X}^n$ are given. Let $Y^n = Y^n(X^n)$. Define

$$\mathcal{B}(\boldsymbol{x}) \triangleq \left\{ \boldsymbol{y} \in \mathcal{Y}^n \,\middle|\, \frac{1}{n} \log \frac{W^n(\boldsymbol{y}|\boldsymbol{x})}{P_{Y^n}(\boldsymbol{y})} \geq \frac{1}{n} \log M_n + \gamma \right\}. \tag{14}$$

Then, there exists a code (φ_n, ψ_n) such that

$$\varepsilon_n \leq \Pr\{Y^n \notin \mathcal{B}(X^n)\} + (1 + P_{X^n}(\mathcal{S}_n)) e^{-n\gamma} + \frac{1}{M_n}, \tag{15}$$

there are M_n codewords in total, and $\lceil P_{X^n}(\mathcal{S}_n) M_n \rceil$ codewords of those belong to $\mathcal{S}_n$, where $\lceil \cdot \rceil$ is the ceiling function.

If $\mathcal{S}_n$ is empty, this lemma is reduced to the original lemma. Subset $\mathcal{S}_n$ will be specified as a collection of sequences that satisfy the cost constraint when the lemma is utilized in later proofs. However, the lemma is considered to have generality and can be applied broadly because it is unrelated to the meaning of $\mathcal{S}_n$.

Proof: We choose codewords in two phases. In the first phase, we choose codewords among $\mathcal{S}_n$. And in the second phase, from the outside of $\mathcal{S}_n$.

Let us get started with the first phase. If $\Pr\{X^n \in \mathcal{S}_n\} = 0$, we directly proceed to the second phase because we have no codeword to choose in the first phase ($M'_n \triangleq 0$, see (36)). So, assume that $\Pr\{X^n \in \mathcal{S}_n\} > 0$. Define

$$\lambda' \triangleq \Pr\{Y^n \notin \mathcal{B}(X^n) | X^n \in \mathcal{S}_n\} + e^{-n\gamma}. \tag{16}$$

Note that this can be rewritten as

$$\Pr\{Y^n \in \mathcal{B}(X^n), X^n \in \mathcal{S}_n\} = (1 - \lambda' + e^{-n\gamma}) \Pr\{X^n \in \mathcal{S}_n\}. \tag{17}$$

As the first codeword $\varphi_n(1)$, we choose an $\boldsymbol{x} \in \mathcal{S}_n$ satisfying

$$W^n(\mathcal{B}(\boldsymbol{x})|\boldsymbol{x}) \geq 1 - \lambda', \tag{18}$$

and define the decoding region for $\varphi_n(1)$ as $\psi_n^{-1}(1) \triangleq \mathcal{B}(\varphi_n(1))$. Hereafter, we choose codewords one after another as follows. As the mth codeword $\varphi_n(m)$ for $m > 1$, we choose an $\boldsymbol{x} \in \mathcal{S}_n$ satisfying

$$W^n\left(\mathcal{B}(\boldsymbol{x}) \setminus \bigcup_{m'<m} \psi_n^{-1}(m') \middle| \boldsymbol{x}\right) \geq 1 - \lambda', \tag{19}$$

and define the decoding region for $\varphi_n(m)$ as

$$\psi_n^{-1}(m) \triangleq \mathcal{B}(\varphi_n(m)) \setminus \bigcup_{m'<m} \psi_n^{-1}(m'). \tag{20}$$

Assume that we have taken L' codewords and cannot choose any more codeword satisfying (19). Note that we have taken infinitely many codewords if $\lambda' \geq 1$. That is, $\lambda' < 1$ if L' is finite.

If L' is finite, we define

$$\mathcal{D}' \triangleq \bigcup_{m \leq L'} \psi_n^{-1}(m) \tag{21}$$

and evaluate L'. To do this, we will bound

$$\begin{aligned} \Pr\{Y^n \in \mathcal{B}(X^n),\ X^n \in \mathcal{S}_n\} &= \Pr\{Y^n \in \mathcal{B}(X^n) \cap \mathcal{D}',\ X^n \in \mathcal{S}_n\} \\ &\quad + \Pr\{Y^n \in \mathcal{B}(X^n) \setminus \mathcal{D}',\ X^n \in \mathcal{S}_n\} \end{aligned} \tag{22}$$

term by term. The first term is bounded by

$$\Pr\{Y^n \in \mathcal{B}(X^n) \cap \mathcal{D}',\ X^n \in \mathcal{S}_n\} \tag{23}$$
$$\leq \Pr\{Y^n \in \mathcal{D}'\} = \sum_{m \leq L'} \Pr\{Y^n \in \psi_n^{-1}(m)\} \tag{24}$$
$$\leq \sum_{m \leq L'} \Pr\{Y^n \in \mathcal{B}(\varphi_n(m))\} \tag{25}$$
$$= \sum_{m \leq L'} \sum_{\boldsymbol{y} \in \mathcal{B}(\varphi_n(m))} P_{Y^n}(\boldsymbol{y}) \tag{26}$$
$$\leq \sum_{m \leq L'} \sum_{\boldsymbol{y} \in \mathcal{B}(\varphi_n(m))} \frac{W^n(\boldsymbol{y}|\varphi_n(m))}{M_n} e^{-n\gamma} \tag{27}$$
$$\leq \frac{L'}{M_n} e^{-n\gamma}. \tag{28}$$

On (27), we applied the fact that

$$P_{Y^n}(\boldsymbol{y}) \leq \frac{W^n(\boldsymbol{y}|\boldsymbol{x})}{M_n} e^{-n\gamma} \quad \text{for } \boldsymbol{y} \in \mathcal{B}(\boldsymbol{x}). \tag{29}$$

The second term is bounded by

$$\Pr\{Y^n \in \mathcal{B}(X^n) \setminus \mathcal{D}',\ X^n \in \mathcal{S}_n\} \tag{30}$$
$$= \sum_{\boldsymbol{x} \in \mathcal{S}_n} P_{X^n}(\boldsymbol{x}) W^n(\mathcal{B}(\boldsymbol{x}) \setminus \mathcal{D}' | \boldsymbol{x}) \tag{31}$$
$$< \sum_{\boldsymbol{x} \in \mathcal{S}_n} P_{X^n}(\boldsymbol{x})(1 - \lambda') \tag{32}$$
$$= \Pr\{X^n \in \mathcal{S}_n\}(1 - \lambda'). \tag{33}$$

Therefore, (22) is bounded by

$$\Pr\{Y^n \in \mathcal{B}(X^n),\ X^n \in \mathcal{S}_n\} \leq \frac{L'}{M_n} e^{-n\gamma} + \Pr\{X^n \in \mathcal{S}_n\}(1 - \lambda'). \tag{34}$$

With (17), L' is evaluated as

$$L' \geq \Pr\{X^n \in \mathcal{S}_n\} M_n. \tag{35}$$

Hence, in any case, we have chosen at least L' codewords from $\mathcal{S}_n$. Here, let

$$M'_n \triangleq \lceil \Pr\{X^n \in \mathcal{S}_n\} M_n \rceil. \tag{36}$$

We formally employ $\varphi_n(1), \ldots, \varphi_n(M'_n)$ as codewords of our code. That is, for $m > M'_n$, codeword $\varphi_n(m)$ and decoding region $\psi_n^{-1}(m)$ become undefined.

Even if the cardinality of $\mathcal{S}_n$ is smaller than M'_n, it does not yield a contradiction. It just implies that a certain codeword is chosen twice or more. From (19), this situation takes place only if $\lambda' \geq 1$.

From here, we work on the second phase. If $\Pr\{X^n \in \mathcal{S}_n\} = 1$, we skip the second phase and proceed to the evaluation of the error probability ($M_n = M'_n$). So, assume that $\Pr\{X^n \notin \mathcal{S}_n\} > 0$. Define

$$\lambda \triangleq \Pr\{Y^n \notin \mathcal{B}(X^n) | X^n \notin \mathcal{S}_n\} + \frac{e^{-n\gamma}}{\Pr\{X^n \notin \mathcal{S}_n\}}. \tag{37}$$

Note that this can be rewritten as

$$\Pr\{Y^n \in \mathcal{B}(X^n), X^n \notin \mathcal{S}_n\} = (1 - \lambda) \Pr\{X^n \notin \mathcal{S}_n\} + e^{-n\gamma}. \tag{38}$$

As the mth codeword $\varphi_n(m)$ for $m > M'_n$, we choose an $\boldsymbol{x} \in \mathcal{X}^n \setminus \mathcal{S}_n$ satisfying

$$W^n\left(\mathcal{B}(\boldsymbol{x}) \setminus \bigcup_{m'<m} \psi_n^{-1}(m') \middle| \boldsymbol{x}\right) \geq 1 - \lambda \tag{39}$$

and define the decoding region for $\varphi_n(m)$ as

$$\psi_n^{-1}(m) \triangleq \mathcal{B}(\varphi_n(m)) \setminus \bigcup_{m'<m} \psi_n^{-1}(m'). \tag{40}$$

Assume that we have taken L codewords in total and cannot choose any more codeword satisfying (39). Similar to the first phase, L may be infinite.

If L is finite, we define

$$\mathcal{D} \triangleq \bigcup_{m\le L} \psi_n^{-1}(m) \tag{41}$$

and evaluate L. To do this, we can derive

$$\begin{aligned} &\Pr\{Y^n \in \mathcal{B}(X^n),\, X^n \notin \mathcal{S}_n\} \\ &\qquad \le \frac{L}{M_n} e^{-n\gamma} + \Pr\{X^n \notin \mathcal{S}_n\}(1-\lambda). \end{aligned} \tag{42}$$

With (38), L is evaluated as

$$M_n \le L. \tag{43}$$

Hence, in any case, we have chosen M_n or more codewords. We formally employ $\varphi_n(1),\ldots,\varphi_n(M_n)$ as the codewords of our code. Now, we completed our code.

Finally, we evaluate the error probability of our code. Before that, we prepare

$$\begin{aligned} &\lambda' - \lambda && (44)\\ &\le \lambda' - \Pr\{Y^n \notin \mathcal{B}(X^n) | X^n \notin \mathcal{S}_n\} - e^{-n\gamma} && (45)\\ &\le \Pr\{Y^n \notin \mathcal{B}(X^n) | X^n \in \mathcal{S}_n\} + e^{-n\gamma} - e^{-n\gamma} && (46)\\ &\le 1. && (47) \end{aligned}$$

With $0 \le \delta < 1$ such that $M'_n = \Pr\{X^n \in \mathcal{S}_n\}M_n + \delta$, the error probability is evaluated as

$$\begin{aligned} \varepsilon_n &= 1 - \frac{1}{M_n}\sum_{m\in\mathcal{M}_n} W^n(\psi_n^{-1}(m)|\varphi_n(m)) && (48)\\ &\le 1 - \frac{1}{M_n}\left(\sum_{m=1}^{M'_n}(1-\lambda') + \sum_{m=M'_n+1}^{M_n}(1-\lambda)\right) && (49)\\ &= \frac{M'_n\lambda' + (M_n - M'_n)\lambda}{M_n} && (50)\\ &= \frac{M'_n(\lambda'-\lambda) + M_n\lambda}{M_n} && (51)\\ &= \Pr\{X^n \in \mathcal{S}_n\}(\lambda'-\lambda) + \lambda + \frac{\delta}{M_n}(\lambda'-\lambda) && (52)\\ &\le \Pr\{X^n \in \mathcal{S}_n\}(\lambda'-\lambda) + \lambda + \frac{1}{M_n} && (53)\\ &= \Pr\{X^n \in \mathcal{S}_n\}\lambda' + \Pr\{X^n \notin \mathcal{S}_n\}\lambda + \frac{1}{M_n} && (54)\\ &= \Pr\{Y^n \notin \mathcal{B}(X^n)\} + (1+\Pr\{X^n \in \mathcal{S}_n\})e^{-n\gamma} + \frac{1}{M_n}. && (55) \end{aligned}$$

Note that the last evaluation is valid even if $\Pr\{X^n \in \mathcal{S}_n\} = 0$ or 1. □

4. Proof of Theorem 2

Here, we will see that Lemma 1 plays a substantial role in the construction of good codes.

Proof of Theorem 2: First, we describe the direct part of the proof. Consider any R satisfying

$$R < \sup_{\boldsymbol{X}:\Pr\{n^{-1}c_n(X^n)>\Gamma\}\le\beta} \underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}). \tag{56}$$

We will show that rate R is $(\varepsilon,\beta,\Gamma)$-achievable.

For some $\gamma > 0$, there exists an $\boldsymbol{X}$ such that

$$\Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma\right\} \le \beta \quad \text{for all } n, \tag{57}$$

$$\underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}) > R + \gamma. \tag{58}$$

We define

$$M_n \triangleq e^{nR}, \tag{59}$$

$$\mathcal{S}_n \triangleq \left\{\boldsymbol{x} \in \mathcal{X}^n \,\middle|\, \frac{1}{n}c_n(\boldsymbol{x}) \le \Gamma\right\} \tag{60}$$

and apply Lemma 1. Then, we obtain a code (φ_n,ψ_n) such that

$$\varepsilon_n \le \Pr\{Y^n \notin \mathcal{B}(X^n)\} + (1 + P_{X^n}(\mathcal{S}_n)\})e^{-n\gamma} + \frac{1}{M_n}, \tag{61}$$

there are M_n codewords, and $\lceil P_{X^n}(\mathcal{S}_n)M_n\rceil$ codewords of those belong to $\mathcal{S}_n$. From the definitions of $\mathcal{B}(\boldsymbol{x})$ and $\underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$, we have

$$\begin{aligned} &\limsup_{n\to\infty} \Pr\{Y^n \notin \mathcal{B}(X^n)\} && (62)\\ &= \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_{Y^n}(Y^n)} < R+\gamma\right\} && (63)\\ &\le \varepsilon. && (64) \end{aligned}$$

Then, the error probability is bounded as

$$\limsup_{n\to\infty} \varepsilon_n \le \varepsilon. \tag{65}$$

As for the coding rate, we can immediately verify

$$\liminf_{n\to\infty} \frac{1}{n}\log M_n \ge R. \tag{66}$$

Finally, as for the cost overrun, since $\lceil P_{X^n}(\mathcal{S}_n)M_n\rceil$ codewords belong to $\mathcal{S}_n$, we obtain

$$\begin{aligned} &\Pr\left\{\frac{1}{n}c_n(\varphi_n(U_{M_n})) > \Gamma\right\} && (67)\\ &= \frac{1}{M_n}\sum_{m\in\mathcal{M}_n} \mathbb{1}\{\varphi_n(m) \notin \mathcal{S}_n\} && (68)\\ &\le \frac{1}{M_n}(M_n - P_{X^n}(\mathcal{S}_n)M_n) && (69) \end{aligned}$$

$$= \Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma\right\} \tag{70}$$

$$\leq \beta, \tag{71}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. From (65), (66), and (71), we conclude that rate R is $(\varepsilon, \beta, \Gamma)$-achievable.

The converse part of the proof is done by following the same lines in the converse part of the proof of Theorem 1 [2, Theorem 3.6.6]. □

5. An Example

Here, we show an example of Theorem 2. Consider a binary noiseless channel. We divide input alphabet $\mathcal{X}^n$ into two subsets $\mathcal{X}^n_{(1)}$ and $\mathcal{X}^n_{(2)}$; $\mathcal{X}^n_{(1)}$ consists of all sequences that are binary representations of integers less than $2^{n/2}$, and $\mathcal{X}^n_{(2)}$ consists of the remainder. Assume that, it costs $C_n(\boldsymbol{x}) \triangleq n$ for $\boldsymbol{x} \in \mathcal{X}^n_{(1)}$ and $C_n(\boldsymbol{x}) \triangleq 2n$ for $\boldsymbol{x} \in \mathcal{X}^n_{(2)}$. Let cost constraint $\Gamma \triangleq 1$.

Since $W^n(Y^n|X^n) = 1$, Therem 2 yields that the channel capacity is given by

$$C(\varepsilon, \beta, \Gamma) = \sup_{\boldsymbol{X}:\Pr\{n^{-1}c_n(X^n)>\Gamma\}\leq\beta} \underline{H}_\varepsilon(\boldsymbol{X}), \tag{72}$$

where

$$\underline{H}_\varepsilon(\boldsymbol{X}) \triangleq \sup\left\{R \,\middle|\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{1}{P_{Y^n}(Y^n)} < R\right\} \leq \varepsilon\right\}. \tag{73}$$

If $\beta = 0$, that is, if we do not allow cost overrun, the channel capacity is $C(\varepsilon, \beta, \Gamma) = 1/2$ because $\underline{H}_\varepsilon(\boldsymbol{X})$ is maximized by the uniform distribution over $\mathcal{X}^n_{(1)}$ for $0 \leq \varepsilon < 1$.

On the other hand, if $\beta > 1 - \varepsilon$, the capacity becomes 1. To see this, let us construct a good code. Fix an $R < 1$. Let $M_n \triangleq 2^{nR}$ be the number of codewords. We choose βM_n distinct codewords from $\mathcal{X}^n_{(2)}$ and a codeword from $\mathcal{X}^n_{(1)}$, which is used $(1-\beta)M_n$ times. Note that $|\mathcal{X}^n_{(2)}| > \beta M_n$ for n large enough. The codewords are distributed uniformly. Then, the cost of codeword overruns Γ with probability β. Since the decoding error occurs only if a codeword from $\mathcal{X}^n_{(1)}$ is sent, the error probability does not exceed $1 - \beta < \varepsilon$. Hence, we can conclude that rate $R < 1$ is $(\varepsilon, \beta, \Gamma)$-achievable. This implies that the channel capacity is 1. Applying the output of the encoder, we can also verify $\underline{H}_\varepsilon(\boldsymbol{X}) = 1$.

6. Asymptotic Cost Constraint

In this section, let us consider an asymptotic cost constraint. We formulate asymptotic cost constraint as follows.

Definition 6: For input process $\boldsymbol{X}$, we define

$$\overline{c}_\beta(\boldsymbol{X}) \triangleq \inf\left\{\Gamma \,\middle|\, \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma\right\} \leq \beta\right\}. \tag{74}$$

Replacing (11) with an asymptotic constraint, we define a new achievability.

Definition 7: Rate R is said to be $(\varepsilon, \beta, \Gamma)$-achievable with asymptotic cost constraint if there exists a sequence of codes $\{(\varphi_n, \psi_n)\}_{n=1}^\infty$ satisfying (3), (4), and

$$\overline{c}_\beta(\{\varphi_n(U_{M_n})\}_{n=1}^\infty) \leq \Gamma. \tag{75}$$

Let $C'(\varepsilon, \beta, \Gamma)$ denote the supremum of $(\varepsilon, \beta, \Gamma)$-achievable rates with asymptotic cost constraint. Then, we have the following theorem.

Theorem 3:

$$C'(\varepsilon, \beta, \Gamma) = \sup_{\boldsymbol{X}:\overline{c}_\beta(\boldsymbol{X})\leq\Gamma} \underline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y}). \tag{76}$$

Proof: This theorem is proved by tracing the same lines of the proof of Theorem 2 with the diagonal line argument for the definition of $\mathcal{S}_n$ applying Lemma 1.

Consider any R satisfying

$$R < \sup_{\boldsymbol{X}:\overline{c}_\beta(\boldsymbol{X})\leq\Gamma} \underline{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y}). \tag{77}$$

Then, there exists an $\boldsymbol{X}$ satisfying

$$\overline{c}_\beta(\boldsymbol{X}) \leq \Gamma. \tag{78}$$

From the definition of $\overline{c}_\beta(\boldsymbol{X})$, for any natural number k, we have

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma + \frac{1}{k}\right\} \leq \beta, \tag{79}$$

which yields that

$$\Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma + \frac{1}{k}\right\} \leq \beta + \frac{1}{k} \tag{80}$$

for all n large enough. Conversely, for fixed n, there is the largest value of k satisfying (80). Let $\gamma_n \triangleq 1/k$ with such largest k (the diagonal line argument). Then, γ_n satisfies

$$\Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma + \gamma_n\right\} \leq \beta + \gamma_n \tag{81}$$

and vanishes as $n \to \infty$. Here, we define $M_n \triangleq e^{nR}$ and

$$\mathcal{S}_n \triangleq \left\{\boldsymbol{x} \in \mathcal{X}^n \,\middle|\, \frac{1}{n}c_n(\boldsymbol{x}) \leq \Gamma + \gamma_n\right\}, \tag{82}$$

and apply Lemma 1. Then, we obtain a code (φ_n, ψ_n) with M_n codewords in total. Out of these, $\lceil P_{X^n}(\mathcal{S}_n)M_n \rceil$ codewords belong to $\mathcal{S}_n$. Bounds for the error probability and the coding rate are derived in the same manner as Theorem 1. The probability of the cost overrun is bounded, for any $\delta > 0$, as

$$\Pr\left\{\frac{1}{n}c_n(\varphi_n(U_{M_n})) > \Gamma + \delta\right\} \tag{83}$$

$$\leq \Pr\left\{\frac{1}{n}c_n(\varphi_n(U_{M_n})) > \Gamma + \gamma_n\right\} \tag{84}$$

$$= \Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma + \gamma_n\right\} \tag{85}$$

$$\leq \beta + \gamma_n \tag{86}$$

for n large enough, which yields

$$\limsup_{n\to\infty} \Pr\left\{\frac{1}{n}c_n(\varphi_n(U_{M_n})) > \Gamma + \delta\right\} \leq \beta. \tag{87}$$

Since $\delta > 0$ is arbitrary, this implies that

$$\overline{c}_\beta(\{\varphi_n(U_{M_n})\}_{n=1}^{\infty}) \leq \Gamma. \tag{88}$$

Hence, R is $(\varepsilon, \beta, \Gamma)$-achievable with asymptotic cost constraint.

The converse part is omitted. □

7. A Stronger Criterion

In a recent work [5] on channel coding with cost constraint, a stronger criterion was introduced. Adapting to our situation, it is expressed by

$$\overline{c}'_\beta(\boldsymbol{X}) \triangleq \inf\left\{\Gamma \,\middle|\, \text{p-}\limsup_{n\to\infty} \mathbb{1}\left\{\frac{1}{n}c_n(X^n) > \Gamma\right\} \leq \beta\right\}, \tag{89}$$

where

$$\text{p-}\limsup_{n\to\infty} Z_n \triangleq \inf\left\{\theta \,\middle|\, \lim_{n\to\infty} \Pr\{Z_n > \theta\} = 0\right\} \tag{90}$$

for a sequence of real-valued random variables $\{Z_n\}_{n=1}^{\infty}$. Since we can observe that

$$\text{p-}\limsup_{n\to\infty} \mathbb{1}\left\{\frac{1}{n}c_n(X^n) > \Gamma\right\} \tag{91}$$

$$= \begin{cases} 0 & \text{if } \Pr\left\{\frac{1}{n}c_n(X^n) > \Gamma\right\} \to 0 \text{ as } n \to \infty, \\ 1 & \text{otherwise,} \end{cases} \tag{92}$$

criterion $\overline{c}'_\beta(\boldsymbol{X})$ for $0 \leq \beta < 1$ is equivalent to $\overline{c}_0(\boldsymbol{X})$ in our situation.

8. Conclusion

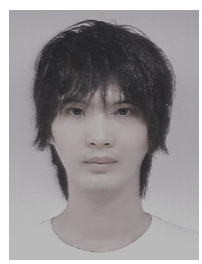
In this paper, we considered a channel coding problem with cost constraint allowing cost overrun for general channels. We formulated two types of constraints to cost overrun and derived channel capacities for each constraint. To show the two types of achievability, the new code construction method based on Feinstein's lemma played a substantial role. The authors are currently focused on proving the lemma using random coding argument.

Acknowledgments

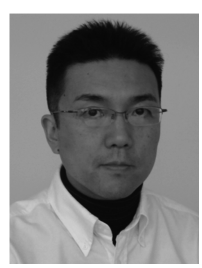
The authors thank anonymous reviewers for their helpful suggestions on this paper.

References

[1] S. Verdú and T.S. Han, "A general formula for channel capacity," IEEE Trans. Inf. Theory, vol.40, no.4, pp.1147–1157, July 1994.
[2] T.S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.
[3] S. Ihara, Information Theory for Continuous Systems, World Scientific, 1993.
[4] A. Feinstein, "A new basic theorem of information theory," Trans. IRE Prof. Group Inf. Theory, vol.IT-4, no.4, pp.2–22, Sept. 1954.
[5] M. Nishiara, "Channel coding with cost paid on delivery," IEICE Trans. Fundamentals, vol.E105-A, no.3, pp.345–352, March 2022.



Masaki Hori received the B. Eng. degree and the M. Eng. in electrical and computer engineering from Shinshu University in 2020 and 2022, respectively. He worked on the research for this paper while he was a student.



Mikihiko Nishiara received the B. Eng. degree in Communication Engineering from Shibaura Institute of Technology, Tokyo, in 1992, and the M. Eng. degree and the D. degree in information systems from the University of Electro-Communications, Tokyo, in 1998 and 2001, respectively. From 2001 to 2007, he was a research assistant of the Graduate School of Information Systems, the University of Electro-Communications. Since 2007, he has been an Associate Professor in the Faculty of Engineering, Shinshu University. His research interest is in information theory including Shannon theory, data compression, and theoretical analysis of communication systems. He is a member of the IEEE.